

**The Hebrew University of Jerusalem  
Federmann School of Public Policy and Government**

**Analysing the Research-Policy ‘Gap’:  
The Phonics-Whole Language ‘Pendulum’**

Submitted in partial fulfillment of the requirements for the degree of Master of Public Policy

**Supervised by Professor David Levi-Faur**

Submitted by Ilana Pinshaw (ID: 328930193)  
April 2013

## Abstract

Which factors are likely to increase the use of research in public policy? The question is addressed in this thesis by examining the changes to beginner reading instruction in the United States in the 1960s and 1980s. It follows the swings that occurred between code-emphasis (phonics) and meaning-emphasis (whole language) approaches in these periods. While the question of reading instruction policy has been widely studied by reading researchers, the issue has not been systematically examined from the point of view of knowledge utilisation in public policy processes. The study asks why research supporting the phonics method of reading instruction succeeded or failed to penetrate reading instruction policy and practice. It uses four main hypotheses proposed by knowledge utilisation theorists: science push, demand pull, dissemination and interaction in order to understand the "research-policy gap". The findings suggest that no one model is able to predict all the changes that took place in the two periods under consideration. Nonetheless, the *Dissemination* model had the strongest explanatory power. Combining this model with the suppositions from the *Science Push* model and proposals put forward by knowledge utilisation theorists about processes of enlightenment, and the role of politics and tactics in utilisation, it is possible to decipher common processes in the utilisation of research findings between the two periods.

## **Acknowledgements**

I would like to express my deep gratitude to my supervisor, Professor David Levi-Faur, for his support and simple suggestions that many times clarified a very confused mind. Thank you for agreeing to take me on at a somewhat late stage, and bringing me to this finishing point. Without your guidance, it would never be complete.

Thank you to friends and family who very generously helped out along the way: Peggy Wahlhaus, Tanya Gruenewald and Miriam Shindler. Thank you to Dr. Anat Gophen, for the start of the journey. A final thank you to my parents, who made sure to remind me that my thesis was waiting in the event that I may have forgotten.

# Analysing the Research-Policy ‘Gap’: The Phonics-Whole Language ‘Pendulum’

## Table of Contents

<b>Abstract.....</b>	<b>2</b>
<b>List of tables and figures .....</b>	<b>5</b>
<b>Introduction.....</b>	<b>6</b>
Theoretical Background.....	8
Hypotheses.....	13
Outline of the thesis .....	14
<b>Chapter 1: The Great Debate .....</b>	<b>15</b>
<b>Chapter 2: Theory and Methodology.....</b>	<b>19</b>
Research Design.....	19
Utilisation or ‘Use’ .....	21
Independent Variables .....	23
Formulation of Reading Instruction Policy and Practice in U.S. Schools.....	26
<b>Chapter 3: Phonics Research during the 1960s and 1980s .....</b>	<b>34</b>
Research Quality .....	34
Research Relevance .....	47
Dissemination .....	57
Interaction .....	64
<b>Chapter 4: Reading Instruction Policy and Practice in the 1960s and 1980s.....</b>	<b>69</b>
1960s.....	69
1980s.....	75
<b>Chapter 5: Analysis.....</b>	<b>82</b>
<b>Chapter 6: Conclusions .....</b>	<b>94</b>
Implications for models and operational variables .....	94
Implications of findings for the definition of use .....	105
Strengths and weakness of the method .....	109
<b>Bibliography .....</b>	<b>114</b>

## List of tables and figures

Table 1: Comparison of Teacher-Training Textbooks .....	74
Table 2: Presence of the relevance variable in the 1960s and 1980s .....	97
Figure 1: The Science Push Model .....	9
Figure 2: The Demand Pull Model .....	9
Figure 3: The Dissemination Model .....	10
Figure 4: The Iteration Model .....	11
Figure 5: Status Groups and Institutions involved in the application of scientific research to elementary reading instruction .....	30
Figure 6: A modified <i>Science Push</i> Model .....	96
Figure 7: A modified <i>Dissemination</i> Model .....	102

## Introduction

Scholars and policymakers have long bemoaned the perceived dissonance between the findings of academic research and public policy, referred to as the ‘research-policy gap’. Criticisms that policymakers are not listening to researchers; that researchers are not producing relevant research; that time, money and precious resources are being wasted due to a lack of communication and congruence between policymakers, practitioners and researchers pepper popular and academic literature. These beliefs are particularly prominent in the education field. Critics attack education research as "stale, irrelevant, . . . of little use to policy makers" (Layzell, 1990 in Birnbaum, 2000) or for lacking scientific rigor or trustworthiness (Carnine, 2000; Slavin, 2002). Others criticise education policymakers for failing to implement evidence based best-practice (Denton, Vaughn, & Fletcher, 2003) or accuse them of only adopting policies that fit with existing ideologies or fashions (Slavin, 2002; Traub, 2002).

Numerous efforts have been made to narrow the “research-policy gap” with the goal of improving outcomes. In 1944, President Roosevelt asked Dr. Vannevar Bush, (then director of the United States Office of Scientific Research and Development), to consider how scientific research could be utilised in peacetime to improve “national health, [create] new enterprises, bring new jobs, and [better] the national standard of living” (President Roosevelt’s letter, para 2 in Bush, 1945). Nearly sixty years later, the “No Child Left Behind” legislation made federal funding in education contingent on the use of effective methods and instructional strategies that are based on scientific research. More recently, both teachers and researchers have been working to explore ways in which findings from neuroscience can inform curriculum or pedagogy (Geake, 2005; Geake & Cooper, 2003; Goswami, 2006; National Institute of Child Health and Human Development, 2000; OECD, 2007; Varma, McCandliss, & Schwartz, 2008; Willingham & Lloyd, 2007)

Scholars have also worked to bridge the gap, proposing ways for researchers to improve their research to make it more amenable to public use. The research- or knowledge-utilisation field developed in response to a desire among researchers throughout the social sciences to increase utilisation of their findings by decision makers in the public sphere with the aim of improving the efficacy and efficiency of policies in a variety of fields.

### ***The “Research-Policy Gap” and Reading Instruction***

The question of how beginner readers should be taught to read has been one of the most widely and publicly debated issues in education. Throughout the twentieth century it has suffered from ‘pendulum swings’ (Stahl, 1990) or successive ‘wars’ (Kim, 2008; Pearson, 2000) between different approaches to instruction. The debate between a code-emphasis or phonics approach, and the meaning-emphasis or whole language approach<sup>1</sup>, has for many come to exemplify the faults that perpetuate this gap. Carnine, for example, asks why, in the face of “overwhelming evidence” in favour of explicit teaching of phonemic awareness and phonics, the whole language approach dominated American classrooms during the 1990s (Carnine, 2000, p. 2).

Why indeed?

Understanding what prompted the repeated and bidirectional shifts between the competing approaches may provide insight into the policy process and the place of academic research within it. The current discussion on how best to translate discoveries in neuropsychology into educational curricula and instructional methods (see for example Ansari & Coch, 2006; Bruer, 2002; Geake & Cooper, 2003; Hall, 2005 as well as a dedicated new journal “Mind, Brain, and Education” from the International Mind, Brain, and Education Society), of which reading has been a particular focus, make the phonics-whole language case particularly relevant.

#### ***Aims***

To address this question, we turn to the knowledge utilisation field. A number of models of the relationship between research and policy or practice have been proposed, each seeking to explain the circumstances under which research is likely to influence policy. Each model proposes courses of action for scholars and decision makers to bridge the “gap” between the policy and research communities.

The phonics-whole language case, while being widely studied by reading researchers, has not been systematically examined by knowledge utilisation theorists. Notwithstanding the question of which approach should or should not have been used at any point in time, this

---

<sup>1</sup> Code-emphasis is a generalised term for instructional approaches that focus on teaching children to decode spelling-sound relationships, often referred to as phonics. Meaning-emphasis is the generalised term for instructional approaches that focus on reading for meaning and encourage sight recognition of words termed, in different periods, look-say, whole word or whole language approaches. These approaches will be discussed in further detail below.

thesis will seek to examine the reasons why research<sup>2</sup> supporting the phonics method of reading instruction succeeded or failed to penetrate reading instruction policy and practice during two periods: 1960s to early 1970s; and 1980s to early 1990s. To do so, I examine the similarities and differences in the determinants of utilisation of research knowledge supporting the phonics method of reading instruction during these two periods.

## **Theoretical Background**

Several theoretical models characterising the research-policy relationship have been formulated, each of which identify a particular aspect of the process as critical to determining research use (Davies, Nutley, & Walter, 2005; Landry, Amara, & Lamari, 2001a; Lavis, Ross, McLeod, & Gildiner, 2003; Wandersman, 2003; Weiss, 1979). The seven leading models are the *science push*, *demand pull*, *dissemination*, *interaction*, *political*, *tactical*, and *enlightenment* models. The first four models have been operationalised, and will serve as the basis for my hypotheses regarding the use of research in reading instruction policy and practice. The final three models are not predictive, and as such, constructing hypotheses is beyond the scope of this thesis. Nonetheless, they will be addressed in the conclusion in so far as they can explain variation not explained by the other models. For the present discussion, utilisation or “use” refers to cases in which research influences policy or practice. How use should be defined is another source of debate, which will be discussed in chapter 2.

### ***Science Push***

The *science push* model identifies research itself as the independent variable in the research-policy relationship. In this model, the success and pattern of the progress of knowledge from research to policy is determined by the quality of research conducted. The *science push* model presumes a linear progression from basic research to large scale implementation: policy developments originate in academia with research into the problem and its causes. This leads to the application of findings to a solution, and the development of an intervention. The intervention is trialled to assess its efficacy, in some cases replicated to assess success under different circumstances, and then applied on a large scale in the form of policy or practice protocols (Wandersman, 2003). Modelled on the paradigm of medical drugs and

---

<sup>2</sup> In general, the term “research” is used to refer to research conducted within academic institutions (“academic research”). What constitutes research is itself a point of contention within the reading field, and this will be discussed further in subsequent chapters.

procedures, products remain relatively unchanged as they move from academia to real-world implementation.

Since the research itself is the key factor in determining influence over policy, by improving the quality of the research (research design, validity and reliability)<sup>3</sup>, researchers can predictably increase utilisation (Landry et al., 2001a). For example, Carnine (2000) argues that education research is failing to impact policy because it relies too heavily on qualitative studies, on individual subjective judgements and on ideology. To become relevant to policymakers, he argues, education research must become “more like medicine”; it must utilise experimental designs, control groups and modern statistical research methods to rigorously assess programmes and practices (Carnine, 2000, p. 1).

Schematically, the model can be presented as follows:



Figure 1. The Science Push Model (adapted from Wandersman, 2003, p. 229)

### ***Demand Pull***

The *demand pull* model is similar to *science push* in that it too posits a linear progression from research to policy, though it differs in its prognosis of causation. Where the *science push* model argues that improving research quality will increase its use by policymakers, *demand pull* theorists contend that increasing the relevance of research to policymakers will increase use.

The outline of the model remains largely the same, but there is a meaningful difference as identifying problems shifts to precede research conduct:



Figure 2. The Demand Pull Model (based on the characterisation of the demand pull model by Weiss, 1979, and Landry et al., 2001)

---

<sup>3</sup> Landry et al. (2001a) details quality components further to include efficiency, compatibility, complexity, observability, trialability, validity, reliability, divisibility, applicability and radicalness

The other significant distinction in this model is between basic and applied research. ‘Applied’ or ‘use-driven’ research is defined as: investigations and innovations developed to address specific and practical problems identified by policymakers or practitioners (Furlong & Oancea, 2005). This type of research is distinguished from basic research, created by scholars to advance a theoretical knowledge base, which, according to *demand pull* theorists, is not capable of contributing to policy solutions or practices as it lacks relevance to real world problems (Keller, 1985 cited in Birnbaum, 2000). According to this model, research that users direct and which is designed to meet their needs will be able to bridge the perpetual gap between researchers and policymakers. When research is closely aligned with user needs, there is a higher level of research utilisation.

### ***Dissemination Model***

Like the *science push* and *demand pull* models, the *dissemination* model conceives the research-policy relationship linearly. Unlike these models, however, the *dissemination* model shifts the focus of the research-policy relationship from the research – its quality or relevance – to the communication of findings. Conceived visually, no real distinction is made between the first three boxes of the above models (though there is generally a preference for the *demand pull* model’s conception of the process). Instead, the focus shifts to the link between research and the policy world, hidden above in the arrows between the boxes and here upgraded to the critical second step.



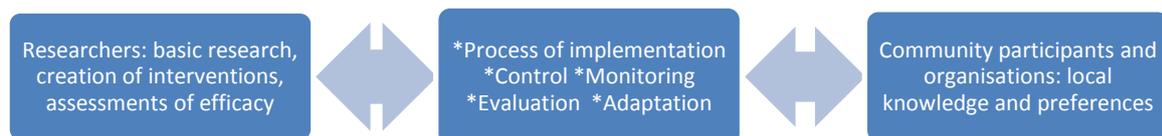
*Figure 3.* The Dissemination Model (adapted from Landry, Amara, & Lamari, 2001b, p. 399; Peterson, Rogers, Cunningham-Sabo, & Davis, 2007, p. 23)

According to this model, researchers must employ dissemination or communication strategies in order to ensure that their conclusions or recommendations will be read by policymakers (Denton et al., 2003; Landry et al., 2001a, 2001b; Peterson et al., 2007). Researchers can thus determine the degree to which their recommendations are adopted by increasing their investment in dissemination activities.

### ***Interaction Model***

The *interaction* model rejects the linear structure of the relationship between researchers and users described in the models above, proffering “a disorderly set of interconnections and back-and-forthness that defies neat diagrams” (Weiss, 1979, p. 428). The above models assume that researchers have the patent on knowledge of how to improve policy and practice. Theorists in the *interaction* model argue that other types of knowledge are also important. Users – policymakers, practitioners, communities – hold knowledge that is critical to the process of adapting policy recommendations to local contexts, which researchers lack (Wandersman, 2003). *Interaction* theorists argue further that researchers and policymakers differ in their goals, the resources available to them, their credibility, and the constraints of the contexts in which they work (Wandersman, 2003). Gaps are consequently due not only to knowledge disparities, but to differences in implementation ability between the two groups.

The *interaction* model advocates for active collaboration between researchers and policymakers in the creation, evaluation and implementation of policy alternatives, with shared ownership over results (Davies et al., 2005). According to the model, researchers should work to build the capacity of organisations to implement these policies. The model predicts that when there is a sustained and intense interaction between researchers and users, there is a higher likelihood that research findings will be utilised.



*Figure 4.* The Interaction Model (adapted from Wandersman, 2003, p. 238)

The contact between researchers and policymakers advocated for in this model is qualitatively different from other models, for example the communication activities in the *dissemination* model. In that case, contact is brief and unidirectional with knowledge passing uninterrupted from researcher to policymaker rather than the “co-production” advocated for in this model.

### ***Political and Tactical Models***

The *political* and *tactical* models defined by Weiss (1979) do not predict what types of research will be used, but rather *when* it will be used. The *political* model refers to cases in

which research is introduced by decision makers into an ongoing policy debate as “ammunition” to support an existing position. Desiring to add credibility to their position, politicians draw on existing research to provide seemingly objective support to their ideological position. Derided by most social scientists as an abuse of research, Weiss argues that as long as research findings are not distorted or misinterpreted, this is in fact an important and proper use of research. Moreover, “since the research finds ready-made partisans who will fight for its implementation, it stands a better chance of making a difference in the outcome” than even high-quality research that lacks existing advocates in the policy sphere (Weiss, 1979, p. 429).

Where the *political* model reduces research to the ideology of its findings, the *tactical* model removes the knowledge producing function of research entirely, diluting its *raison d’être* to simple existence. Research, in these cases, regardless of its content, becomes a tactic in bureaucratic politics – “proof of [government] responsiveness” or justification for a delay in decision making - “We are waiting until research is completed” (Weiss, 1979, p. 429). It can bear the burden of responsibility for unpopular decisions, or alternately grant credibility to agencies by virtue of their association with serious academics.

Both these models ignore the ‘intrinsic usefulness’ of research in favour of its ‘political usefulness’ – rather than demanding relevance to policy, high technical standards and convergence with other studies, the usefulness of research is determined by its timing, clarity of presentation, and political and financial feasibility (Tangri & Strasburg, 1979).

### ***Enlightenment Model***

The enlightenment model, proposed by Weiss (1979), reconceptualises the way in which research can impact policy. Rather than the findings of a specific study or group of studies informing a policy decision; over time the concepts and theoretical perspectives of research permeate the policy process, affecting how problems are conceived and considered. We cannot visually conceive this model as its paths are often untraceable, even by those directly involved. Decision makers may be unaware of the existence of the specific studies that led to the development of the concepts on which their decisions are based, having absorbed knowledge of concepts or findings indirectly through personal contacts, the media, colleagues, etc., who may even themselves be unaware of the original source or evidence.

This model does not predict when research use will take place, nor which research will be preferred. Nonetheless, it still envisions research as having broad potential to impact policy in meaningful and substantial ways such as determining which issues are on the agenda, informing the way in which policymakers think about issues, or introducing new language or new concepts that change the way issues are conceived (Weiss, 1979). The timeline for impact is extended substantially, as findings may take years to permeate the public decision making consciousness, beyond either the validity of findings or the lifetime of researchers (Birnbaum, 2000). Similarly, the locus of impact expands, with fields and industries not intended as targets, absorbing and utilising research in unexpected ways (Klein & Rosenberg, 1986).

The diffuse and indirect permeation of research into policy faces unique challenges. As researchers are not directly involved in the communication of findings, and there are no procedures to evaluate knowledge communicated, research communicated diffusely “dispenses invalid as well as valid generalizations. Many of the social science understandings that gain currency are partial, oversimplified, inadequate, or wrong... The indirect diffusion process is vulnerable to oversimplification and distortion, and it may come to resemble "endarkenment" as much as enlightenment.” (Weiss, 1979, p. 430).

## **Hypotheses**

On the basis of the first four models, the following hypotheses can be formulated:

H1: Research utilisation was determined by research quality (*Science Push* model)

H2: Research utilisation was determined by the relevance of research to user needs (*Demand Pull* model)

H3: Research utilisation was determined by the extent to which researchers disseminated findings to policymakers and practitioners (*Dissemination* model)

H4: Research utilisation was determined by the level of interaction between researchers and policymakers/practitioners (*Interaction* model)

These hypotheses will guide the analysis.

## **Outline of the thesis**

Chapter 1 provides an overview of “The Great Debate” including an introduction to the problems of literacy and reading instruction; the central differences in approaches to reading instruction; and a brief overview of key events in the debate between reading instruction approaches in the United States during the twentieth century.

Chapter 2 introduces the theoretical constructs that will guide the thesis. Methodology is discussed and the models described above are operationalised.

Chapter 3 examines four aspects of reading research during the two periods being studied corresponding to the four hypotheses – the quality of phonics research during the two periods, its relevance to users, efficacy of dissemination activities and the extent of interactions between researchers and policymakers. Drawn from a variety of historical sources, this section will attempt to establish the state of research during the two periods in order to formulate predictions based on the models for how research would be used by policymakers and practitioners in each period.

Chapter 4 explores the changes that took place in reading instruction policy and practice in each of the two periods, drawing in greater detail the picture of developments provided in chapter 1. This chapter will attempt to establish the extent of research “use” during each of the periods.

Chapter 5 explains research utilisation during the 1960s-70s and 1980s-90s by comparing the occurrence of the four aspects of research predicted by the models in the two periods with the extent of research use that resulted.

Chapter 6 draws broader conclusions from the findings about the models proposed, the research methodology used, and the way in which the dependent and independent variables should be measured.

## Chapter 1: The Great Debate

Literacy is one of the cornerstones of modern society. As such, methods of teaching children to read have remained high on the public agenda of many developed countries for over a century, since the advent of modern schooling (Hempenstall, 1997). During the colonial period and following independence, reading in the modern United States of America was taught by drilling children on the names of the letters, followed by their sounds and a few letter-sound correspondences (Venezky, 1987). As education in this period was generally associated with religious instruction, children would then proceed directly to reading, traditionally the bible.

The rise of developmental psychology in the early 20<sup>th</sup> century brought with it opposition to this ‘dry and meaningless’ approach to reading instruction (Kim, 2008). Already in the mid-19<sup>th</sup> century Horace Mann, then the Secretary of the Massachusetts State Board of Education, had condemned ‘phonic’ drills as “bloody, ghostly apparitions” and advocated for an approach to reading that taught children to recognise words as whole units (cited in Kim, 2008). This began the long and great debate between the “phonics” and “whole word” or “whole language” approaches to teaching children to read.

Learning to read is a complex process. It requires knowledge that written words encode spoken language; understanding that words are made up of letters (known in the literature as graphemes); and that these letters represent sounds, or phonemes (Juel, 1991). One of the challenges of learning to read, particular (though not exclusive) to English is the lack of consistency in the relationship between letters and sounds (‘grapheme-phoneme correspondence’). English has 26 letters, but around 45 phonemes, meaning that many letters represent more than one phoneme, while some phonemes are represented by more than one letter (Hempenstall, 1997). This produces inconsistent relationships between spelling and pronunciation where identical sounds are spelt differently (rough/ruff) or identical spellings that produce different sounds (rough/through).

This is the most obvious, though not the greatest challenge of learning to read. Another significant challenge is that while written language is segmented, spoken language is continuous<sup>4</sup>. It has been shown that in order to learn to read, children must understand that this continuous flow of language can be broken down into words, words into phonemes (so

---

<sup>4</sup> You may hear this when listening to a foreign language; it is impossible to say where one word ends and another begins. It can also be seen in ancient texts which prior to a certain period, lacked word divisions.

cat becomes /c/ /a/ /t/) and that these phonemes correspond to letters, or graphemes (Juel, 1991). Knowledge of this feature of spoken language has been termed “phonemic awareness”.

There are two main approaches to teaching children to read. These two approaches have been variously named, but were appropriately grouped by Chall (1967) into the “code-emphasis” approach and “meaning-emphasis” approach.

Code-emphasis approaches focus on teaching children to decipher the “alphabetic code” of written language. Children are traditionally taught certain letter-sound correspondences, then provided with words that utilise that correspondence; for example the sound /b/ with the letter [b], and then the words bat, baby, brother. The code-emphasis approach is most commonly referred to as “phonics”, “an approach to teaching reading which aims to sensitise children to the relationships of the spelling patterns of our written language to the sound patterns of our oral language” (Hempenstall, 1997, p. 404). There is no one phonics approach, but rather many different, often contradictory, approaches that differ on which correspondences should be taught to children and in what order (Hempenstall, 1997; Levin, 1967). The biggest divide within phonics is between the “explicit” or “synthetic” approach and the “implicit” or “analytic” approach. The former begins with the letters, associating each letter with a sound, and then blending sounds and letters to make words. The latter approach begins with whole words, teaching children to identify the sounds in the context of the word, rather than in isolation (National Institute of Child Health and Human Development, 2000). Both approaches presume that children must be taught letter-sound correspondences, because they are unlikely to discover them independently through reading.

Meaning-emphasis approaches on the other hand, argue that it is more important that children immediately understand the purpose of reading, that is, to gain meaning from a text. When beginning to read, children are taught to recognise words by sight, and immediately combine them into sentences and stories. This approach argues that children should be taught to read in the way that skilled readers are presumed to do so – by recognising words as whole units rather than as combinations of letters (Hempenstall, 1997). Meaning-emphasis approaches have variously been referred to as the “look-say”, “whole-word”, or “whole-language” approach. While there are differences between these methods - the whole-language approach for example involves a broader educational philosophy (see K. S. Goodman, 1989; Y. M.

Goodman, 1989) - they will be considered as one in this document as they are successors of one another in time, and all alternately sat in opposition to the phonics approach.

Meaning-emphasis approaches predominated much of the first half of the twentieth century. The mainstream approach during this period was to first present children with a large corpus of words to be sight read, after which some analytic phonics lessons would be introduced (Pearson, 2000). The emphasis was on meaning and comprehension, and analysis of the code sat alongside context and picture cues as clues that could be used to decipher new words (ibid.). Referred to at this time as the “look-say” or “whole-word” approach, this approach held such supremacy that a study in the mid-60s found that 98% of schools used practically identical basal reading series, unchanged since the 1930s, all utilising this approach (Barton & Wilder, 1964).

A few critical events occurred in 1955 and again in 1967 that sought – and eventually succeeded – to change the status quo. Consensus on beginning reading instruction in the 1950s was so strong that dissenters were seen as “the lunatic fringe” (Chall, 1967, p.295). In 1955 Rudolph Flesch published “Why Johnny Can’t Read”, a popular book condemning the “look say” approach to reading instruction and calling for a return to phonics (Flesch, 1955). The book was directed at parents, teachers and the general public, remained on the bestseller list for 30 weeks, and received widespread coverage in the popular press (Chall, 1967). The book sold millions of copies, and effectively created a public debate on the issue of reading instruction methodology.

Due to the uproar and the then-raging debate between supporters of the ‘look-say’ and ‘phonics’ approaches, the U.S. Board of Education, a federal government institution, decided to fund a study named “First Grade Reading Studies” that would compare a range of programmes in reading instruction and assess their efficacy (Bond & Dykstra, 1967; Bond, 1967). Released in 1967, it coincided with the release of another significant work that was to embody the phonics movement - Jeanne Chall’s “Learning to Read: The Great Debate” (Chall, 1967). The latter work synthesised existing research into reading processes, and reviewed practice in schools at the time. Both studies provided support for the phonics method of reading instruction.

Following these events, in the late 1960s and early 1970s, a substantial and noticeable shift commenced in how reading instruction was taught in American classrooms. Phonics became

the dominant method, and remained so until the early 1980s (Chall, 1983; Goodman, Shannon, Freeman, & Murphy, 1988; Pearson, 2000).

In the late 1970s and early 1980s, a movement arose to challenge the widespread acceptance of the phonics approach to reading instruction. Labelled the “whole language movement”, it advocated for a child-centred approach that again condemned the phonics approach as uninspiring for children (K. S. Goodman, 1967; Y. M. Goodman, 1989). Based in an approach that saw teachers as “co-learners” with students, it encouraged teachers to allow children to discover their own meaning in texts. As part of this philosophy, whole language theorists advocated for teaching children to read whole words by sight, based on the assumption that they would discover grapheme-phoneme correspondences themselves while learning to read and write (ibid). The whole language movement swept the nation, and by the late 1980s and early 1990s became the dominant method of reading instruction (Kim, 2008; Pearson, 2000).

The two changes that have been described here – the shift to a phonics approach in the late 60s and early 70s, and the move away from phonics in the late 80s and early 90s have been painted here in broad brushstrokes in order to provide context, but will be described in further detail in chapter four of the thesis. These two shifts will be the subject of the study, as we attempt to understand the changes in further detail and uncover their causes.

## Chapter 2: Theory and Methodology

This section will explore how the theoretical models described above will be operationalised. Study design and methodology will be discussed, and the dependent and independent variables that will be used to test the models will be defined.

### Research Design

Over the past decade, researchers in the field of knowledge utilisation have been testing the models described above in real-life policy scenarios. The dominant mode of measuring the use of research in policy and practice is through surveys in which either researchers or users are asked to reflect on their practice and identify factors that affect the use and non-use of research (Cornelissen, 2000; Rich, 1997). Types of surveys vary. For example, Landry et al., (2001a, 2001b) distributed questionnaires to a large number of faculty members in Canadian universities, asking them to (self-) report utilisation of their studies by practitioners, professionals and decision makers. Dimaggio and Ussem (1979) targeted the opposite end of the relationship, interviewing a select group of decision makers in arts organisations, asking them when they decided to use the results of a particular type of research (audience surveys) and why. Despite their differences in survey techniques and target groups, both attempt to assess the impact of particular studies on targeted groups of decision makers.

This approach is common in knowledge utilisation literature, but has been criticised for the limited types of research use generally considered by respondents (Rich, 1997). Rich (1997) questions whether the passage of a specific piece of information through an organisation *can* be documented: is it possible to track a piece of information as it passes from hand to hand, and can it be isolated from the effects of other sources of knowledge that inform decision making?

The changes in approach to the teaching of beginning reading that took place in the late 1960s and late 1980s are the cumulative effects of hundreds or even thousands of decisions by diverse groups of decision makers operating at different levels<sup>5</sup>. Nonetheless, the different decision makers moved in a common direction, and in many cases claimed to have a scientific basis to their decision making. Rather than address the impact of findings from a particular investigation, I seek to assess the impact of the field as a whole, as represented in major syntheses of research conducted in each of the two periods.

---

<sup>5</sup> This will be discussed further below in defining policy and practice in education

A different type of study was developed in response to the limitations of the above method, and can be applied in a complementary manner to this study. Termed ‘Social Framework Analysis’, it is in many ways the diametric opposite of the above approach (Cornelissen, 2000; Rich, 1997). Rather than trying to attempt to “identify how discrete bits of information influence specific behaviour, choices or actions”; this approach analyses the way in which concepts and terms that originate in academia infiltrate public decision making and affect how decisions are made and/or conceived (Cornelissen, 2000, p. 323). Since policymakers and practitioners will likely not be aware of the source of a concept or even its influence on them, and because the process may extend beyond the lifetime of the scholar who proposed the idea, evaluators use reasoning and judgement to assess the passage and development of concepts, sometimes supported by citation analysis (Rich, 1997). This approach is highly reliant on conjecture and as such it is easy to refute. It also limits the type of research that can be considered to works with highly innovative and challenging theoretical implications, such that their impact on frames of reference will be noticed.

In order to somewhat overcome these limitations, I will be using elements of both approaches. Data used in this study will be obtained through primary texts and historical sources. Questions asked in survey research will be adapted to assess the research field on the one hand (represented by research syntheses), and the policy and practice domain on the other (represented by the outcomes of the collective decisions of educational policymakers and practitioners, rather than particular decisions). As far as it is possible, reasoning and judgement based on assessments from within the reading field will be used to analyse the way in which research has affected concepts within decision making.

The benefit of hindsight provides a unique perspective on change and the role of research. Looking at events that occurred 30 and 50 years ago enables a broader view of the ways in which research affected decision making, and enables us to extend the timeline of impact. It also enables us to look at policy on a national scale; rather than under what circumstances is research used by a particular school board, for example, under what circumstances is it adopted by school boards across the country? Similarly, it is able to go beyond the question of the impact of an individual study to assess the impact of a cohort of studies, their cumulative effects.

I apply a comparative method of analysis, assessing the similarities and differences in the determinants of utilisation of research knowledge supporting the phonics method of reading

instruction during the 1960s and the 1980s (Ragin, 1989). As Rich (1997) argues, it is difficult if not impossible to directly attribute decisions made to particular inputs of information (that is to connect cause to outcome). The comparative method attempts to bypass this problem by assuming that if cases of use have possible causes in common, that are absent from cases where use did not occur, then this is a likely cause of the event (Ragin, 1989). This method is reliant on variables developed as part of theoretical constructs, but is able to test those constructs through the congruence method (Faure, 1994; George & Bennett, 2005). Thus the models developed by knowledge utilisation theorists will be used to guide the comparative analysis.

Operationalising the models also requires that we operationalise the terms used and measured in the study.

### **Utilisation or ‘Use’**

In order to examine the use of research in policy, it is necessary that we define the term ‘use’. There is substantial variation in the way that this concept has been defined, resulting in significant disparities in the amount of research utilisation measured (Birnbaum, 2000; Rich, 1997). Knowledge utilisation scholars divide the perspectives on use into three categories: instrumental (product and process), conceptual, and symbolic (Landry et al., 2001a; Rich, 1997).

The ‘product’ perspective on use, the dominant conception of use when discussing the research-policy gap, refers to “cases where the knowledge of a single study induces users to make decisions that would not have been made otherwise”, i.e. the direct application of a research product by policymakers to create an intervention, understand a problem, or select from amongst competing policy alternatives (Landry et al., 2001a, p. 336). In order to meet this definition, findings must be understood correctly and implemented as advised by researchers, at least in part. Policymakers and practitioners are aware of the research on which they are basing their decision, and will usually justify their choice on the basis of the scientific evidence.

With the aim of expanding conceptions of the ways in which research is considered to influence policy and practice, scholars proposed a ‘process’ perspective of use. While the *product* perspective of use focuses on the contribution of scientific findings to decision outcomes – a programme, a policy, curriculum etc, the *process* perspective focuses on the

contribution of research to the decision making procedure. From this standpoint, the fact that certain policy alternatives are considered for implementation, even if they are not ultimately selected; changes to the selection or evaluation process, or the decision making process as a whole; are considered evidence of the use of research in policy and practice (Landry et al., 2001a).

Dissatisfied with what they considered this ‘instrumental’ approach to utilisation, critics proposed two new categories of use: *conceptual* and *symbolic*. *Conceptual* use refers to the ability of research to create new ideas, new theories, and new interpretations of facts without necessarily changing the actual content of the decisions (Landry et al., 2001b). *Conceptual* use goes beyond the “intrinsic” usefulness of studies to consider their “intellectual usefulness” – their contribution to society’s conceptual frameworks such as problem definition, understanding of how social institutions work or informing the dynamics of a problem (Tangri & Strasburg, 1979). Birnbaum (2000) argues that the common criticism that policymakers are not using research in decision making is misled by the instrumental view of utilisation, ignoring this “more meaningful”, and more common application of research.

*Symbolic* use refers to cases in which research is held up as evidence to support existing positions or legitimise held views (Landry et al., 2001b). Unlike the instrumental definition of use, use according to these theorists includes cases in which findings are misunderstood but nonetheless influential; cases in which dissemination is so diffuse and indirect that neither the researcher nor the practitioner is aware of the other’s role in implementation; or cases when decision making is influenced by non-contemporary data (Birnbaum, 2000; Davies et al., 2005; Rich, 1997; Weiss, 1982).

Hierarchies of use have been proposed that range from research being transmitted, received and read, to the findings being judged useful, to research being cited in decision making or influencing a decision, to actual implementation of recommendations (Landry et al., 2001a; Rich, 1997). This, however, is more appropriate for cases in which users can be asked directly about their decision making. It is difficult to assess these different levels retrospectively without interviewing decision makers and so only cases in which there is visible evidence of use can be considered. Nonetheless, it is perhaps relevant to consider the distinction between complete implementation, adaption of information, partial implementation, and steps towards implementation proposed by Larsen and Werner as progressively decreasing measurements of use (1981, cited in Rich, 1997). This allows for

cases in which it may be rational for users to reject, ignore or alter research findings due to other considerations of implementation.

### **Independent Variables**

Independent variables measure the different characteristics of research identified as critical by the models. Research here will largely refer to knowledge produced by academia. Rather than individual studies, we will be looking at the collective knowledge of the field during the designated time period. This will predominately be measured through assessment of works that synthesise past research.

### ***Research quality***

Research quality is the independent variable used to assess the *science push* model (H1). Research quality is primarily defined by traditional scientific criteria of validity, reliability and proper experimental design (Landry et al., 2001a). While most knowledge utilisation literature defines research quality in reference to the quality of a particular study, it is more relevant in this case to consider the quality of a research field at a particular time. When referring to the quality of the field, in addition to the standard criteria of validity and reliability, reading research theorists have assessed the degree to which studies are cumulative, comparative and consistent with one another (Barr, 1984; Bond, 1967; Chall, 1967; Levin, 1967); whether they collectively describe and support a broad definition of reading (K. S. Goodman, 1989); if they incorporate findings from other disciplines (K. S. Goodman, 1989; Levin, 1967); and assessments as to the conclusiveness of findings (Hempenstall, 1997; Kim, 2008). An answer to each in the affirmative is considered to increase utilisation.

The qualitative nature of this study enables the utilisation of multiple broad definitions. As I am not a reading research specialist, I will rely on the judgements of reading researchers to assess reading quality, noting in each case the definition of ‘quality’ that they use when making this judgement. The extent to which research continues to be cited by later researchers will also contribute to the judgement of the ‘quality’ of research at a particular time.

### ***Research relevance***

Research ‘relevance’, the variable used to represent the *demand pull* model (H2), refers to the extent to which studies are developed to address specific and practical problems. *Demand pull* theorists differ in whether they define relevance as being simply “directed towards a specific practical aim or objective” (OECD p.78), or whether they require that the problems be identified “on the ground, in the context of application” by policymakers or practitioners (Furlong & Oancea, 2005, p. 8). Both definitions will be considered here.

Measuring relevance according to this definition is complex. Landry et al., (Landry et al., 2001a, 2001b) identify four criteria related to the relevance of research to users<sup>6</sup>. Most researchers collect data to assess these criteria through direct questioning of one side of the chain; either asking researchers about their efforts to achieve relevance and their perception about how those efforts are received (e.g. Landry et al., 2001a, 2001b); or asking users about the relevance of research to their needs (e.g., Dimaggio & Useem, 1979). Having neither of these options open to this study, relevance will be assessed indirectly as outlined below in reference to Landry et al.’s criteria:

- 1) “Focus on advancement of knowledge” – this refers to the question of whether the work is basic research (focused on advancing scholarly knowledge) or applied (focused on advancing a public or private benefit), the latter equating with greater relevance. In reading research, research into the basic processes that underlie reading is considered ‘basic’ research, while studies that addresses instructional methods are considered applied (K. S. Goodman, 1989; Venezky, 1984). This will be assessed by examining the content of the studies.
- 2) “Focus on users’ needs” – this refers to the extent to which researchers consciously direct their research towards users’ needs as researchers perceive them. Here, rather than questioning researchers, testimony of the authors’ intentions from within their works will be used.
- 3) “Funding source” - the source of funding is used as an indication of whether or not the research is inward focused, where research that relies on internal university funding is less likely to be utilised than research that is externally funded, based on the assumption that “researchers who rely on external sources of funds are... more sensitive to the needs

---

<sup>6</sup> My division between *science push* and *demand pull* theories is not the same as Landry et al’s and thus some variables listed are used to assess the *science push* model rather than the *demand pull* model in those studies

of users located outside the academic milieu” (Landry et al., 2001a, p. 337)<sup>7</sup>. A different criterion that measures a similar concept is the source of the impetus for the research, in other words, who identifies and frames the problem. Research is considered to be more relevant when it is commissioned by users to address a problem they have discovered “in the context of application” and/or conducted “in-house” (Dimaggio & Useem, 1979; Furlong & Oancea, 2005; Landry et al., 2001a). The syntheses themselves as well as secondary sources reviewing research will provide data for this criterion.

- 4) “Perceptions of user context” – this includes whether practitioners and policymakers consider the research pertinent; whether it coincides with their needs as they define them; whether a specific audience is targeted; the credibility of researchers and timing of output. This will be assessed by examining historical sources for user context and comparing to research products.

### *Dissemination activities*

Dissemination activities, the variable used to represent the *dissemination* model (H3) are defined as activities through which researchers (or intermediaries) communicate findings to users. Landry et al., (2001a; 2001b) name two main aspects of dissemination:

- 1) Adaption of products to users, by making them readable, operational, implementable, and presented in an appealing format;
- 2) Dissemination efforts, which consists of meetings with users at various stages of the process. Into this latter category fall public and academic conferences (e.g. Hallinan, 1996) and, in the field of reading, textbooks and other curriculum materials (Denton et al., 2003).

Dissemination activities are distinguished from “indirect” or societal dissemination as per the *enlightenment* model in that they are purposeful – target audiences, regardless of their size, are selected and products are directed specifically to them. These activities are also distinguished from *interactive* ones in that communication is one-way, i.e. researchers communicate findings to users but there is no mechanism for feedback.

---

<sup>7</sup> Landry et al (2001a) present these as two variables, but essentially they are two ends of the same scale and can be considered one variable

### ***Interaction Activities***

Landry et al., (2001a) reduce the *interaction* hypothesis to one variable: the intensity of ‘linkage mechanisms’ between users and researchers. These informal contacts do not, in my opinion, adequately capture the *interaction* model. The important distinction between the *dissemination* and *interaction* models is that in dissemination activities, information flows one way: from researchers to users. In interaction activities, knowledge is “co-produced”, and efforts are made to bridge the resource and capacity gaps between the “two communities” (Landry et al., 2001a; Wandersman, 2003). Measurement of the *interaction* variable (H4) will thus involve assessments of the degree to which the knowledge of users is valued by researchers and integrated into their findings and recommendations; and whether scholars either take into account the capacity of policymakers and practitioners when proposing recommendations or make efforts to build the capacity of users to meet recommendations – such as accounting for the cost of implementing new programmes or providing relevant professional development for teachers (Chall, 1967; Denton et al., 2003).

## **Formulation of Reading Instruction Policy and Practice in U.S. Schools**

Beyond the general definitions of research and research use, it is necessary to examine the specific policy and practice structures of the case under investigation. Establishing the context in which reading policy and practice is formulated enables us to identify fields in which use may take place.

### ***Authority and Responsibility for Educational Policy***

Formal authority and responsibility for educational policy in the U.S. rests in the hands of the States by way of the 10<sup>th</sup> amendment, which states that any power not expressly granted to the federal government is held by the States or the people (Nelson & Weinbaum, 2006). As we will see, however, this formal authority had little meaning over most of the twentieth century, and de facto policy makers and practitioners had great influence. Assessment of the use of research in policy and practice in this field therefore necessitates that we examine the process by which curriculum has been formed in the U.S. over the twentieth century.

### ***Policy Products in Reading Instruction***

The most influential forces in the formation of educational curricula in American schools, as will be demonstrated below, are the basal reading programmes. Basal reading programmes referred to here as ‘basals’ or ‘basal readers’, effectively provide teachers with a complete

reading curriculum including instructional strategies, student workbooks, teachers' manuals, graded reading texts, and often other classroom aids, enabling them to touch on all aspects of instruction (Anderson, Hiebert, Scott, & Wilkinson, 1985).

The first popular basal reader series, the *McGuffey Eclectic Readers*, sold approximately 120 million copies from the 1830s until 1920 (Venezky, 1987). With around 4.3 million children aged 5 to 9 attending school in the continental US in 1990, and considering the fact that readers would have been kept from year to year and shared between children, the figure of 120 million copies indicates very wide, if not total, coverage (North, Willcox, & Gannett, 1904). Despite strong local control and almost no involvement of centralised authority, the use of common textbooks produced relatively uniform practice in reading instruction nationwide throughout the twentieth century (Venezky, 1987).

In the early 1960s, over 95 percent of all elementary schools in America used one of a number of remarkably similar basal series, and teachers across the country reported close to the same practices in reading instruction (Barton & Wilder, 1964). Teachers surveyed stated that reading series and their teacher's manuals had the greatest impact on their beliefs regarding the teaching of reading (Barton & Wilder, 1964). In her comprehensive study of reading instruction practice in this period, Chall concluded that "[m]ethods of teaching beginners how to read reach the classroom in the form of teaching materials... published commercially" (Chall, 1967, p. 297). Studies from the late 70s and early 80s estimate the proportion of instruction dominated by basal programme content and teacher's manuals as between 75 and 95 percent (Adams, 1990; Anderson et al., 1985; Dole, Rogers, & Osborn, 1987, p. 284; Keith, 1985; Shannon, 1982, p. 884). Mirroring Chall's early statement, almost twenty years later Anderson et al. report that basal programmes "strongly influence how reading is taught in American schools and what students read" (Anderson et al., 1985, p. 35).

We thus see strong evidence to support the statement that "the selection of a basal reader is tantamount to selecting the reading curriculum" (Farr, Tulley, & Powell, 1987, p. 268). The content of basal readers are commonly used as indices to make judgements about methods used in classrooms, and will continue to be utilised for this purpose in this study (Chall, 1996). Basal readers will thus be examined for their inclusion of research recommendations both as a critical *product* in reading instruction and due to their ability to indicate classroom practice.

Another important policy *product* that will be examined for use due to its importance in determining practice is that of teacher training. Barton and Wilder (1964), interviewing teachers on factors that influenced their practice found that teacher training was particularly important. 56% of teachers named “practice teaching connected with the teachers college” as important in influencing their beliefs regarding the teaching of reading, and 48% named “workshops, institutes and in-service courses” (Barton & Wilder, 1964, p. 379)<sup>8</sup>. Though this was lower than the 63% who named “reading series and teachers’ manuals” as influential, it nonetheless points to an important force in determining classroom practice. Teachers and administrators further qualified that without specialised training, teachers were insufficiently prepared to implement reforms and new approaches were likely to fail (Chall, 1967).

### **Policy Processes**

Basals have been shown above to be important *products* in beginner reading classrooms in the US during the relevant periods. Understanding how these reading series are produced and selected is crucial in determining the use of research in decision making *processes*.

While basal readers are produced by commercial publishing companies, the companies are only one set of players in the process of creating and selecting textbooks for first grade classrooms.

### ***Local Beginnings***

At the beginning of the twentieth century, local school boards held central responsibility and authority over education policy and practice (Howell, 2005; Miller, 2008). At the turn of the twentieth century over 80% of funding for schools came from local sources (Snyder & Hoffman, 1997). Though technically subject to state law, local school boards were largely autonomous in their control over curricula, hiring and firing of staff, and most other aspects of school governance (Howell, 2005). Textbooks were selected locally, often by district superintendents or local schools (Keith, 1985).

### ***Emerging involvement of the states and higher levels of government***

Beginning in the 1950s, as part of a nationwide response to the threat to American superiority posed by the successful USSR launch of Sputnik, the states began to assert their constitutional jurisdiction over educational policy, and state and federal governments began

---

<sup>8</sup> Multiple responses were allowed

to take on larger roles in funding education (Howell, 2005; Nelson & Weinbaum, 2006; Snyder & Hoffman, 1997). State-level authorities consolidated local school districts and shifted authority for matters such as teacher certification and curriculum from local communities to state-employed professionals and administrators (Howell, 2005). Of particular relevance for our field, many states began practicing what is known as state-level adoption of textbooks, whereby in order to reduce the cost of textbooks, they are bought in bulk by state agencies. In order to qualify for free or subsidised textbooks, schools must use one of the books selected by committees (Keith, 1985). By 1950, state-level adoption was common practice in 24 states ('adoption states'), with the remaining states ('non-adoption states') delegating responsibility for textbook selection to local, county or district school boards or textbook committees (Institute for Educational Development, 1969). Even in adoption states, however, local involvement was still important. In almost all adoption states (22), local "units" (boards of education, textbook selection committees, superintendents, chief educational officers, etc) were involved in the process of selecting materials in addition to state-level "units" (Institute for Educational Development, 1969).

Despite the ostensible increase in state control over textbook selection, individuals and groups outside the official decision making process continued to play important roles in determining textbook content and adoption. Barton and Wilder (1964) provided a detailed overview of the various "status groups" involved in the application of scientific research to elementary reading instruction, pictured in figure 5. Grouped by institution, these include government and foundations, universities, publishing companies, local teachers colleges, local school systems including teachers and local communities. Barton and Wilder note that some "status groups" were omitted, of particular note, state departments of education and professional associations of school administrators and teachers. Each of these groups will be examined (to varying extents depending on the availability of evidence) for their use of phonics research in determining reading policy and practice in the U.S. during the relevant periods.

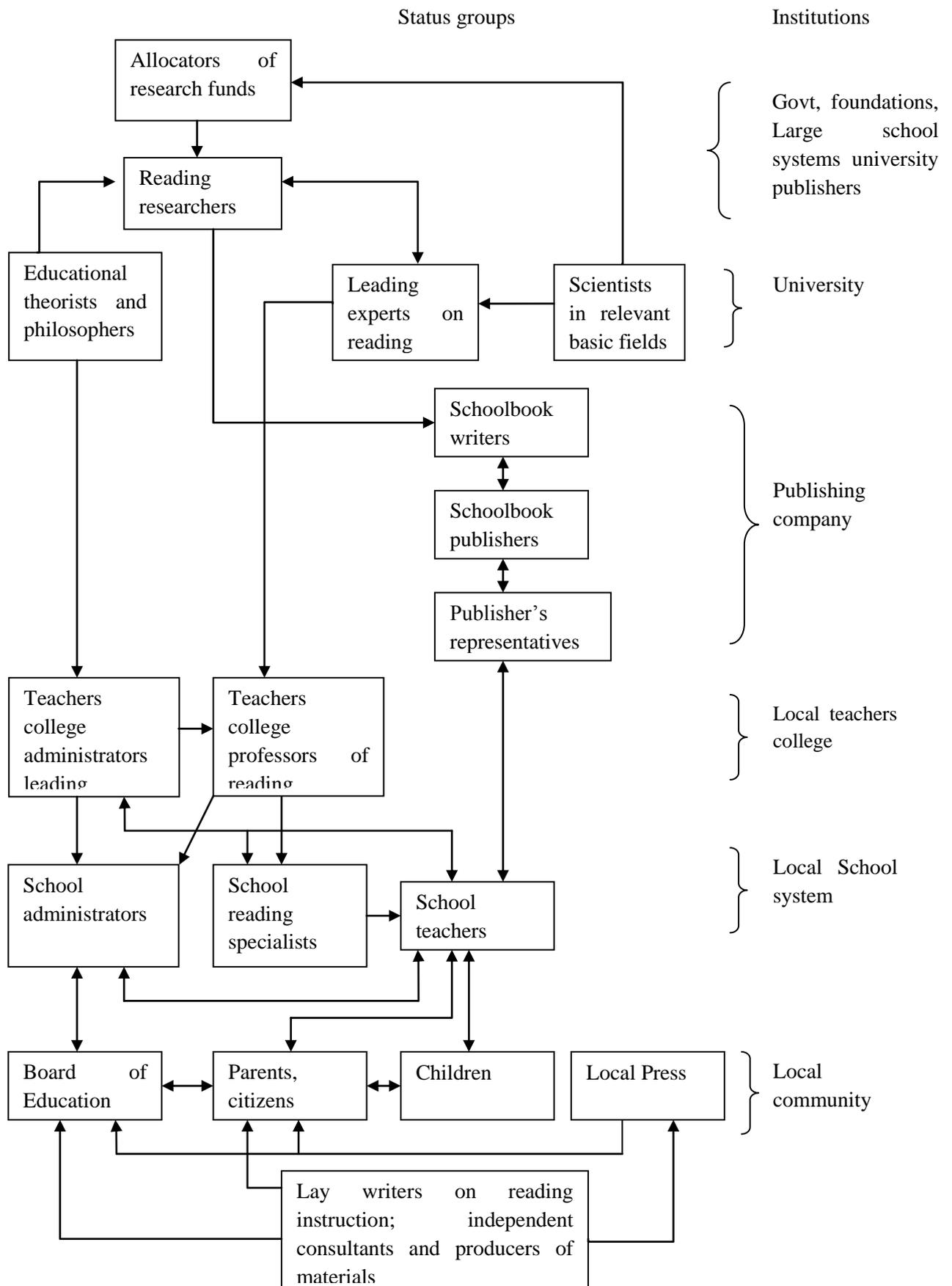


Figure 5. Status Groups and Institutions involved in the application of scientific research to elementary reading instruction (Barton and Wilder, 1964, p. 365)

Barton and Wilder's (1964) model was corroborated by studies of official textbook selection processes and their participants. Due to limited knowledge about products and methods among official participants in the process, and limited time to consider the available options, participants were often dependent on input from forces external to the official process, giving them substantial sway in the outcome of the selection process (Institute for Educational Development, 1969; Keith, 1985). External sources of information on educational products were varied, and included commercial, academic and professional groups. Salesmen and representatives from basal publishing companies; journals and periodicals; conferences and professional meetings; and 'conversations with other educational professionals, such as teachers or curriculum specialists' were all rated by more than 50% of decision makers as important for obtaining information about educational products (50.4; 50.9; 58.4 and 54.9 percent respectively; Institute for Educational Development, 1969, pp. 215–217). In addition to this indirect influence, company representatives were often also invited to meetings of textbook selection committees or of school district faculty to present (and sell) their materials (Chall, 1967; Farr et al., 1987). It should be noted that publishers did work to meet the demands of textbook selection committees, as best they could discern them (Institute for Educational Development, 1969).

In this initial survey of the 1960s, teachers were considered to be the most influential forces in the selection process, despite the fact that they were not official appointees in the process (Institute for Educational Development, 1969). Parents were also influential on textbook selection whether by applying pressure on textbook selectors (Chall, 1967) or through participation in selection committees (Farr et al., 1987). The presence of these issues on the public stage and in the media was thus influential in the selection process.

#### *Measuring 'use' in textbook selection processes – 1960s*

Unfortunately, there are no records I could access from the 1960s of the specific decisions made by local and state "units" involved in selecting textbooks. The diffusion of authority and financial restrictions limits the impact of any particular local- or state- level decision on the textbooks produced.

The collective decisions of elected officials across the country must therefore be inferred through the changes made to the basal readers created and adopted in this period. As explained, changes to basal readers resulted from basal publishers' perception of changes in the demand emanating from influential players in the selection processes and external to it.

Changes made to the content of teacher training, the public presence of related issues and funding for research are other areas in which the use of research in policy and practice can be assessed.

### *Changing dynamics of influence between the 1960s and 1980s*

Basal reading series did not lose importance between the two periods in question, however the selection processes changed in substantial enough ways as to add new factors to our analysis. Where once state involvement in setting curriculum was limited to “ensur[ing] a balanced program” (Shannon, 1991, p. 152), states and local authorities beginning in the 1970s began setting detailed curriculum requirements. Curricular frameworks were created by Boards of Education, outlining subject goals, pedagogical objectives, assessment methods etc, all of which guided textbook selection (Hoffman et al., 1994; Kerr, 1996; Kim, 2008)<sup>9</sup>.

Though other states also began developing similar criteria, the U.S. national market for educational textbooks is subject to what has been described as the “California” or “Texas effect” (Farr et al., 1987; Keith, 1985; Watt, 2009). Due to the substantial investment required to develop a basal programme, publishers could not afford to adjust their products to suit the desires of every textbook committee. Instead, they adjusted their products and timing to meet the adoption cycles and criteria of the largest players – California and Texas - and then marketed those products to a national market (Chall & Squire, 1991; Farr et al., 1987; Keith, 1985; Watt, 2009)<sup>10</sup>. Thus, despite the diffuse decision making apparatus of state and local control in the U.S., and the diversity of the population, remarkably similar practices could be found across the country and a variety of studies have repeatedly shown that basal series printed at the same time were virtually indistinguishable from one another (Barton & Wilder, 1964; Farr et al., 1987; K. S. Goodman et al., 1988; Hoffman et al., 1994; Pearson, 2000)<sup>11</sup>.

Beginning in the 1980s, these two states began to realise the importance of their large markets and began using their buying power to influence content (Keith, 1985). The

---

<sup>9</sup> Not all states and districts would necessarily have had comprehensive frameworks, but lists of criteria for textbook selection, often influenced by larger state boards of education, professional organisations or publishing companies, fulfilled a similar role (Farr, Tulley, & Powell, 1987).

<sup>10</sup> Occasionally adjustments will be made for other large state or district adoptions, but considering that the adoption processes of smaller states and districts are also subject to the California/Texas effect, these were generally minor adjustments to an existing product (Farr et al., 1987).

<sup>11</sup> Apparently they were also sometimes literally indistinguishable from one another. Farr et al (1987) report that textbook adoption committee members interviewed could not tell programmes apart even immediately after completing a multi-month review and selection of those same textbooks.

education boards set textbook selection criteria designed to “elicit a more tailored product from textbook publishers” (Keith, 1985, p. 32). Tailored, that is, to the curricular goals of that state. The importance of these curricular frameworks was not only in their content, but also in their timing. In the 1960s, publishing companies had to predict the interests of selection committees, and market their products at the appropriate time to increase the chance of adoption, while textbook selection committees were limited to what publishing houses had chosen to produce on the basis of their predictions (Chall, 1967; Institute for Educational Development, 1969). In the 1980s and 90s, Texas and California began to release the selection criteria over a year in advance of selection in order to provide time for publishing companies to produce products that met the new guidelines (Hoffman et al., 1994). Publishers, wanting to obtain the lucrative contracts these states offered, adapted textbook content to meet these criteria. For the reasons listed above, Texan and Californian frameworks become an important focus during the 1980s and 90s and have been selected for study for evidence of use during this period.

### ***1990s: Entrance of Legislation into Curricular Policy***

Legislation is traditionally a major focus of studies examining standing government policies. However, as we will see below, legislation relating to reading instruction was almost non-existent until the 1990s. Legislation that was relevant during the period being examined was restricted to predominately federal provisions for funding for research to inform curriculum (Nelson & Weinbaum, 2006). Setting curriculum and instructional methodology, throughout the period in question, was the domain of bureaucratic agencies – State Departments of Education, Boards of Education, Local School boards etc. - and legislators did not encroach on this domain. The end of the period under study is marked by the entrance, for the first time, of legislators into the setting of curricular priorities. As many as 18 states between 1994 and 1997 began mandating phonics as an instructional methodology for beginner reading (Kim, 2008). The initiation of this legislation highlights its prior absence, and provides further justification for the focus on basal reader content, teacher training and other influences in examining the use of research in policy products during this period.

## Chapter 3: Phonics Research during the 1960s and 1980s

This chapter examines the state of research during each of the two periods under examination – the 1960s to 70s and 1980s to 90s – which will serve as the basis for the analysis in chapter five. Data for this section is sourced from primary examination of major research syntheses from each period, combined with historical analysis of those syntheses. Each period is examined independently to establish the status of each of the four variables at that time: research quality, research relevance, dissemination and interaction. The content of the syntheses is briefly examined under research quality in order to establish whether findings appear in policy and practice, to be discussed in chapter four. Following this, a comparison of the two periods for each variable will be made that will serve as the basis for the analysis.

### Research Quality 1960s

In order to understand the research syntheses of the mid-1960s and the changes they produced, it is important to examine the state of research in the decade prior, both as a point of comparison and also because one of the syntheses – Chall's (1967) *Learning to Read: The Great Debate* – relied on these studies to draw conclusions.

Reflecting on the state of reading research in the mid-1960s, scholars found that while “abundant”, the research that had been collected to that point was “largely inconclusive, noncumulative and limited in size and scope” (Barton & Wilder, 1964, p. 389). The field suffered from “a dearth of synthesizers and theorists” working to draw broad conclusions from individual studies (Chall, 1967, p. 87). The result was that existing research on beginning reading “says nothing consistently...[and] if you select judiciously and avoid interpretations, you can make the research “prove” almost anything you want” (ibid. see also Bond, 1967; Hempenstall, 1997; Levin, 1967).

Research studies conducted between the 1930s and early 1960s were criticised for their poor experimental validity, lacking both internal and external validity (Bond, 1967; Chall, 1967; Venezky, 1984). Classroom experiments were poorly designed, usually comprising a direct comparison between two competing methods, with poor controls for external variables (Bond, 1967). Laboratory studies, usually once-off events of 30-60 minutes, were questioned for their ecological validity to classroom instruction, particularly as they usually used adults as subjects rather than children (Venezky, 1984). Most studies did not even attempt to address causation (Chall, 1967). Different groups of reading researchers conducted different

types of investigations, each published in their own journals and rarely were the groups' findings on similar questions integrated (ibid.).

Instructional methods under investigation (the “basal approach”, “phonetic method” etc) were poorly defined and often differed greatly between experiments. The absence of construct validity meant such that findings could not be generalised from a particular published programme to the larger approach (Chall, 1967; Robinson, 1967). This was not just a fault of comparative experiments; even among advocates of phonics, for example, there was little agreement on what constituted the “phonetic method” or how it should be taught (Spache, 1967). Assessments of reading ability varied greatly between experiments, and researchers established the superiority of one method over another for reading ability as a whole by presenting results in terms of significance on various subtests of standard instruments, without trying to explain the relation of the subtests to a reading construct (Chall, 1967). Studies constantly contradicted one another, and as they did not attempt to build on earlier research, there was no way to assess the reliability of findings (ibid).

The amount of funding for research and research positions has a significant impact on the quality of the field. Barton and Wilder (1964) conducted a survey of reading researchers in the early 1960s and found that only 1.6% received half or more of their income from research with the remainder teaching courses, teaching in schools, doing clinical practice or writing instructional materials for children as their main source of income. Because reading instruction was not yet considered an academic profession, and the majority of doctoral dissertations were conducted by teachers as part of professional development, less consideration was given to their contribution to furthering academic knowledge than their specific professional concerns (Chall, 1967). In addition, because the majority of their time was devoted to non-academic pursuits, reading researchers were cut off from research activity in the basic sciences, which prevented developments in other fields from being integrated into the pool of reading knowledge (Chall, 1967).

*Learning to Read: The Great Debate* (Chall, 1967) and The Cooperative Research Program in First-Grade Reading Instruction (known as First Grade Reading Studies; Bond & Dykstra, 1967) changed the standing of the research field by working to synthesise knowledge produced by research to draw conclusions based on “empirical scholarship rather than rhetoric” Pearson (2000, p. 8). Chall’s work, sponsored by the Carnegie corporation of New York, synthesised existing reading research evidence from 1910 to 1965. First Grade Reading

Studies (*FGRS*) was conducted as part of Project Literacy, a federally funded nationwide cooperative research venture. *FGRS* was the first major effort to coordinate multiple experimental studies into reading instruction methodology, comprising 27 studies conducted by leading researchers from a variety of fields.

Both projects worked to overcome many of the problems of comparison that had plagued the field to that point. Though their categories differed slightly, both Chall (1967) and Bond and Dykstra (1967) addressed the aforementioned problem of construct validity by reclassifying all instructional methods used in studies into five and six categories respectively. This enabled comparison between studies using the same approach, but was not without flaws. In both cases methods and materials were grouped “arbitrarily” (Bond & Dykstra, 1967, p. 45) on the basis of their common characteristics; and each method included a number of different programmes, allowing for a degree of difference in the implementation of instructional methods between studies. Bond and Dykstra were able to deal with the ‘reading ability’ construct more effectively, as every researcher participating in the study agreed to utilise the same measures of reading ability - the five sub-tests of the Stanford Achievement Test - while Chall was forced to repeat the method above, categorising tests into eight measures of reading ability.

In order to address problems of experimental validity and reliability of findings, Chall synthesised findings from multiple sources (30 experimental, 17 correlational and 10 clinical), triangulating findings while ignoring conclusions, in which, she argues, authors often disregarded their own findings. She also distinguished between the grade-level of subjects tested in order to examine possible developmental influences on reading acquisition. Nonetheless, she does acknowledge that doing so required that she was forced to relax her criteria for scientific rigor: “most [experimental] studies were unsatisfactory in some respects...Had I considered only studies that fulfilled all necessary experimental conditions, I would have been left with just a handful – if that many” (Chall, 1967, pp. 101–102).

In order to overcome the “piece-meal” approach of prior investigations, *First Grade Reading Studies* experimenters agreed upon certain criteria for all studies, including common research design and experimental controls, and shared measures to assess pre-reading capabilities and reading ability at the end of first grade. Classes were taught by the classroom teacher to ensure that they reflected the “real life of classrooms” (Ruddell in Schantz, 2002, p. 90). In addition, all researchers committed to collecting information about teacher, pupil, school, and

community characteristics – external variables that may be expected to influence reading acquisition. In addition to the individual analysis conducted by the researchers responsible for each experiment, all research results were made available to the project’s coordinating centre, which conducted its own comparative analysis of results. Despite these controls, Bond and Dykstra found that random assignment was not uniformly adhered to, as they sometimes recorded differences in the starting point of groups being compared within experiments. Analyses of covariance were used to remove some of this difference, however this relies on statistical, rather than experimental, control (see Bond & Dykstra, 1967, p. 101). More meaningful differences were found between classrooms, suggesting that non-method factors in the classrooms were playing a significant role. A wide range of group characteristics was present experiments were large, supporting the generalisability of results.

Both studies concluded that the available evidence did not indicate that any particular published programme was better than any another (despite this being the primary aim of most research to that point). Nonetheless, both studies did conclude that an initial code-emphasis (Chall’s term) or word study skills (Bond and Dykstra’s term) – direct phonics instruction – tended to produce higher reading scores than the look-say basal method. Bond and Dykstra (1967) found this already in the first grade. Chall (1967), whose analysis continued beyond the first grade, suggested a ‘look-say’ approach may produce faster gains in grade 1, but an initial emphasis on phonics produced better readers and spellers in the ‘long term’, that is by grade two and at least to grade three. Due to a lack of evidence, she could not conclude about the differences in results with the two kinds of approaches beyond the third grade. Chall and Bond and Dykstra concluded that the look-say basals and phonics programmes both produce reading failures, however Chall found that there was ‘*considerable* evidence’ that a meaning-emphasis approach at the beginning produces more serious failures (Chall, 1967, p. 176 emphasis in original). Both authors recommended that one approach should not be used exclusively; an initial code emphasis is important, but children benefited from variety, and elements of different programmes should be combined. At the end of her analysis, Chall reviewed the findings of *FGRS*, which had only just been partially released, and indicated that it appeared that there was convergence in their findings.

Breaking from the tendency to that point towards an “atheoretical” approach to reading theory (see Chall, 1967, 1996; Gaffney & Anderson, 2000; K. S. Goodman, 1969), Chall used the existing findings to propose a (at least partial) theory of reading acquisition, with hypotheses as to why differences were found in reading gains at different stages of

development and why certain children may benefit more from a particular approach than others.

Chall's (1967) work was very highly regarded in professional circles, influential in changing attitudes to phonics, and stimulated substantial follow-on research (Anderson et al., 1985; Hempenstall, 1997). Chall's work continues to be held up as a "classic" in the field and later experiments confirmed and expanded on her findings (Chall, 1996). Over 20 years later, Adams concludes that comparison studies "based on masses of data, gathered through formal experimental procedures...are point for point virtually identical to those at which Jeanne Chall arrived on the basis of her classroom observations and interpretive review of the literature" (Adams, 1990, p. 59).

Bond and Dystra's (1967) evidence was also well regarded, particularly for its success in bringing together leading researchers with diverse opinions, many of whom had vested interest as authors of basal series, to conduct a coordinated and balanced experiment that in the end did not favour any of them absolutely (Schantz, 2002).

### ***Limitations of 1960s reading research***

For all the gains made by the two major syntheses of the 1960s, they preserved many of the weaknesses of the research field of the period. Chall's analysis was based on studies conducted over the preceding 30 years which, as described, were methodologically weak. Some scholars disputed her ability to draw conclusions from such compromised data (e.g. Carbo, 1988). *FGRS* retained the standard research design of comparing method A to method B, limiting the type of conclusions that could be drawn. While all *FGRS* experiments utilised a basal reader programme as a control between the studies, no studies compared all the experimental programmes directly, so the substantial differences found between projects in achievement could not be explained (Bond & Dykstra, 1967). The natural classroom environment, while a positive attribute from the perspective of ecological validity, made control over the full range of variables difficult (Schantz, 2002). Some scholars disputed the validity of statistical methods used in *FGRS* though one later reanalysis confirmed findings (Barr, 1984).

The problem of inadequate basic research to support applied studies persisted in the cumulative works. Beyond the criticism that instructional methods given the same label were often not the same (Stauffer, 1967), whether instructional programmes were even able to

represent the principle that they espoused was questioned. The central principle of the phonics approach is that children are taught to identify the correspondence between letters and sounds, the “grapheme-phoneme relationship”. However these relationships were only fully mapped in the 1970s, and there was no consensus at this time on which pairings should be taught or in what order (Levin, 1967). In general, applied researchers working in the 1950s and 60s had to rely on basic research into language and cognition developed prior to the 1920s, as, due to an absence of funding, this type of research was negligible thereafter, until the passage of the Cooperative Research Act in 1954 (Venezky, 1984). This research was not “thorough or reliable by modern standards”, and was “too disconnected” to provide much of psychological interest, yet it remained the primary source of knowledge of perception at least until the 1960s, when experimental psychologists revived studies into the cognitive processes underlying reading (Venezky, 1984, p. 12)

Goodman (1967, 1969) proposed the most substantial criticism of the conclusions made by Chall (1967) and *FGRS*. He argued that by breaking reading up into separate skills, such as oral reading ability, researchers were testing that skill only, and not reading ability as a whole (Goodman, 1969). Without understanding how different processes interact with one another, and with language and thought, he argued, knowledge produced cannot contribute to a theory of reading able to “generate hypotheses and predict and explain reading behavior” (Goodman, 1969, p. 14). Chall recognised the theoretical flaws in her own work, presenting her findings as tentative: hypotheses that demand further testing rather than a comprehensive theory (Chall, 1967).

## **Research Quality 1980s**

The criticisms of reading research in the 1960s – that it failed to develop knowledge of the basic perceptual and cognitive processes involved in reading, the absence of a theory of reading and lack of a strong experimental corpus from which to draw from - drove the research of the following two decades.

The limited conclusions that could be drawn from *FGRS*, confirmed Chall’s (1967) argument that A vs B studies comparing different instructional methods had run their course (Pearson, 2000; Stahl & Miller, 1989). Even with better controls and coordination between studies, this type of studies were still unable to explain “what works for whom, how, why and under what

conditions”, and consequently, the contribution to collective knowledge and theory building was negligible (Pearson, 2004, p. 234).

A key development of reading research in the 1970s and 80s was that the focus shifted from testing which instructional method worked most effectively, to examining the cognitive processes of the child without reference to the teacher (Alexander & Fox, 2004). Driven by a flood of new money for basic research, cognitive psychologists and others entered the reading field, working to identify and understand the domain- and task- specific cognitive processes associated with reading acquisition and the order in which they are acquired (Alexander & Fox, 2004; Pearson, 2000). This new research enabled the development of a more thorough understanding of instructional practices and how they should be constructed than when they were studied directly in the earlier studies (Adams, 1990; Chall, 1996; Juel, 1991; Pearson, 2004).

The renewed interest in reading also resulted in exponential growth in the number and type of research studies in reading. While Chall (1967) was limited to 30 experimental, 17 correlational and 10 clinical studies considered strong enough to be included, syntheses of the 1980s benefited from a comparatively massive corpus of research – one estimate puts the number of articles and books published between 1966 and 1985 in the field of beginner reading at between four and six thousand (Grundin, 1985). Similarly, where Chall’s (1967) summary of existing research and Bond and Dykstra’s (1967) conclusions were almost entirely based on the results of classroom and laboratory experiments comparing the relative effectiveness of different instructional methodologies, this type of data represents only a small portion of reading research conducted in the 1970s and 1980s (Adams, 1990; Anderson et al., 1985).

Heeding past criticisms, and benefiting from a deeper understanding of learning, memory and perception, scholars began proposing theories of reading acquisition. The first comprehensive theory was proposed by Smith (1971, cited in Juel, 1991) based on Goodman’s (1967, 1969) suppositions. Experimental examinations of the hypotheses of this theory destabilised it, and an alternative, “staged” theory of reading acquisition was proposed in the early 80s, with two or three qualitatively different stages of development (Juel, 1991).

This model proposed that there are qualitative differences between the way beginner and experienced readers process words, with “stages” in development. The child first identifies words by attending to “random” features of the words or print environment (such as the ‘M’

of McDonalds, or the name of a toothpaste brand by its location on the tube), and cannot identify words or letters out of that set context. In the second stage, the child learns to decode the spelling-sound relationship, first “gluing to print” – abandoning prior pseudo-fluency and even meaning to focus on “sounding out” words, until they decipher the code and learn rule of correspondence between graphemes and phonemes. Eventually (in the third stage), they are able to automate this decoding process to recognise words immediately, acquiring fluency and freeing up attention resources to allow for more attention to meaning (Juel, 1991). A number of smaller experiments provided evidence for the existence of first stage, while a larger number of studies showed qualitative differences in errors produced by children with good spelling-sound knowledge and those without (Juel, 1991).

The significance of this staged theory of reading was that it provided clear support for the phonics method of instruction. Critically, it was able to explain several anomalies that had posed challenges to supporters of phonics in the past: why phonics is beneficial for beginner readers even if skilful readers appear to process words automatically; why a meaning-emphasis approach may provide initial gains but is disadvantageous in the long run; and why phonemic awareness in pre-readers was predictive of future reading ability (Adams, 1990; Juel, 1991).

Achievements in the theoretical realm had important implications for instruction. The most important of these was something of a resolution of how phonics should be taught. Phonics programmes had been criticised for the lack of consensus on which rules should be taught and in what order, and published programmes varied widely on which they selected (Adams, 1990). Critics claimed that grapheme-phoneme relationships were too complex and too numerous to be taught and that these relationships were not properly understood (Smith, 1971 cited in Kim, 2008; Levin, 1967).

Research conducted in the 1980s was able to resolve these questions as phonics was shown to be important not for the specific relationships that were taught but due to the ability of phonics instruction to foster theoretical concepts at the base of reading: phonemic awareness (the concept that the continuous flow of speech can be segmented into phonemes), alphabetic understanding (words are composed of letters) and cryptanalytic intent (print is encoded speech) (Juel, 1991; Kim, 2008; Pearson, 2000). Therefore, rather than teaching children to verbalise rules governing pronunciation, or years of instruction to cover every phoneme-grapheme relationship; it was found that if children were taught to functionally utilise the

most common phoneme-grapheme correspondences, they will develop the concept that these rules exist, and thus be able to infer other relations with sufficient opportunities to read both independently and orally with a teacher (Anderson et al., 1985; Juel, 1991).

The instructional implications for this latter clause were that lessons on a phonic rule should be paired with stories where a sizeable portion of the words use that rule (Adams, 1990; Anderson et al., 1985; Juel, 1991). It also implied that while teaching phonics rules is important at the beginning of instruction, as children progress it is more important that they are given sufficient opportunities to read and that they learn to enjoy reading, than that they are explicitly taught less common relationships (Juel, 1991). This effectively rejected the “false dichotomy between phonics and meaning”; the assumption that (though never advocated by researchers) knowing the alphabetic code and reading good literature were either/or choices (Kim, 2008, p. 42).

Better coordination with investigations in related fields, such as memory storage and retrieval, led to additional lessons for instruction. For example, the discovery that phonemes and letters function not only as codes to decipher words, but also as indices for storing and retrieving words, translated into recommendations for classroom tasks that cause children to attend to each letter in order, such as writing of whole words from dictation or in stories, rather than those that teach letters in isolation, such as worksheets where children fill letters into the blank spaces in words (Adams, 1990).

### ***Research Syntheses***

While thousands of articles and books were published during this period, two major studies, which synthesised the collective knowledge of this period for policymakers, will be the focus of the analysis: *Becoming a Nation of Readers* (Anderson et al., 1985) and *Beginning to Read: Thinking and Learning about Print* (Adams, 1990).

*Becoming a Nation of Readers* (Anderson et al., 1985), a report of the National Academy of Education’s Commission on Education and Public Policy, was commissioned by the National Institute of Education to collate and simplify the “critical mass” of research into reading acquisition and its application to reading instruction (Anderson et al., 1985, p. v). The report was produced by a panel of “nationally recognized reading experts” led by Dr. Richard C. Anderson, professor and director of the Centre for the Study of Reading at the University of Illinois (United States Department of Education, 1985, para. 4).

Five years later, at the behest of federal lawmakers, a new study was commissioned to translate *Becoming a Nation of Reader's (BNR)* general recommendation of phonics instruction into implementable policy (Adams et al., 1991). The resulting book, *Beginning to Read: Thinking and Learning about Print* (Adams, 1990), was arguably the most comprehensive summary of research into beginner reading published to that time (Pearson, 1990). To simplify findings for policymakers and practitioners, the long and complex book was compiled into a shorter summary document, prepared by Stahl, Osborn and Lehr in consultation with Adams and released together with the original (Stahl, Osborn, & Lehr, 1990).

Both studies were well regarded in the research community. These works are regularly cited as the leading summaries of available reading research (see for example Hempenstall, 1997; Kim, 2008; Pearson, 2004; Stahl, 1999). They were selected as two of the twenty-five most influential research syntheses in literacy (1967-1997) created from submissions from members of the Reading Hall of Fame, the only two syntheses from the 1980s synthesising research on all aspects of beginner reading to be included (Shanahan, 2000).

Adams was almost universally praised for the unprecedented thoroughness of her study, quality of her analysis, readability of her prose and the relevance of her work to instructional applications (see, for example, Allington, Chaney and Kapinus in Adams et al., 1991). Pearson termed Adams' (1990) book "the most complete review, within a single cover, of our expanding knowledge of...issues and research in early reading instruction" (Pearson, 1990, pp. vi-vii). Even Cullinan and Strickland, who disagreed with her conclusions, praised the book as a work of "impeccable" scholarship (Strickland & Cullinan, 1990, p. 433).

*BNR* was less of an achievement of academic analysis than *Beginning to Read*. Its importance lay in its success in making a large corpus of academic studies accessible to potential users in the policy and practice communities. The report was designed for a broad audience, to be readable like a textbook (Chall in Aaron, Chall, Durkin, Goodman, & Strickland, 1990). While the work was praised for its accessibility, it was also condemned for the alleged compromises to the research base that resulted (Aaron in Aaron et al., 1990). Condemnation was not universal; Chall defended the report, arguing that in the process of reconciling their vastly differing opinions, the members of panel managed to synthesise the available data "which too often seems to have conflicting and controversial findings" into a document that all could agree to sign their names (Chall, 1985, p. 123; Aaron et al., 1990) . Botel (1985)

concurrent, praising the high calibre of the “blue ribbon” authors and their success “elegantly” synthesising “a diverse and rich” body of scientific information, designating it an important work for policymakers and school leaders (Botel, 1985, pp. 260–262).

### *Conflict within the research field*

Reading researchers during this period faced a new challenge. While the debate of the 1960s was heated, it was nonetheless conducted in the same domain, and in a limited number of peer-reviewed journals. Goodman’s (1967, 1969) challenge to the reading concept effectively split the field in two, divided not only by their position on phonics, but also the type of research they conducted, the disciplines of their advocates, and their model of the reading acquisition process. This posed challenges to the decisiveness and consensus of the field.

Determining consensus in the research field, a factor of research quality, requires that we evaluate the strengths of the competing movements. Stanovich (1994) proposes three criteria for evaluating competing claims to knowledge:

1. Research must be peer reviewed as part of a process of establishing the credibility of findings and quality of research design.
2. Research should be replicated by independent researchers in order to verify findings.
3. There should be consensus in the relevant research community “whether or not there is a critical mass of studies that point towards a particular conclusion” (Stanovich, 1994, p. 287), established either through wide reading of the field or by frequency of citations. (Hempenstall, 1997).

On the first two points, whole language research faltered. Whole language theorists favoured the more qualitative, ethnographic and naturalistic methods due to their “harmony” with the whole language classroom (K. S. Goodman, 1989, p. 210). For awhile this research was accepted as the whole language movement led a shift in the reading research field away from quantitative analysis and towards these qualitative methods (Gaffney & Anderson, 2000). However, once the discourse of research shifted to an insistence on reliable and replicable research beginning in the mid-80s, the reliance of the whole language movement on naturalistic research alone, opposition to setting measurable standards for reading outcomes, and the fact that this research was generally not peer reviewed, replicated or independently validated led to the delegitimisation of the findings of these studies (Chall, 1996; Pearson, 2000, 2004). Whole language lacked the “mechanism for self-correction that is the foundation of the credibility of science” (American Physical Society Council, 1999).

Regarding the third point, consensus is difficult to define in this case.

The model of reading acquisition used by meaning-emphasis theorists which conceived reading acquisition as a natural ‘hardwired’ process which children would naturally acquire without direct instruction if exposed to enough text was prominent through the 1970s (Alexander & Fox, 2004; Pearson, 2000). Experiments designed to test these theories, however, disproved the major tenets (Adams, 1990; Juel, 1991).

A growing body of research showed that many children, particularly those with lower reading readiness, failed to infer spelling-sound relationships naturally without direct instruction (Adams, 1990; Hemenstall, 1997; Juel, 1991)<sup>12</sup>. Even for children with higher reading readiness, explicit phonics instruction was shown to speed acquisition of these principles (Juel, 1991). Further research found that, contrary to Goodman and Smith’s suppositions, poor rather than good readers relied on contextual cues for reading, which even amongst skilled readers were unreliable cues. (Adams, 1990; Chall, 1996; Kim, 2008). Indeed, it was found that what differentiated good from poor readers was knowledge of spelling-sound correspondences and automaticity of decoding (Adams, 1990; Juel, 1991).

On the other hand, “considerable evidence” from experimental, naturalistic, and case study research had been amassed which provided strong support for phonics principles and the benefits of direct instruction for acquisition of the alphabetic code (Juel, 1991, p. 774,778; see also Adams, 1990; Anderson et al., 1985; Chall, 1996). In addition to evidence cited above, longitudinal studies found long-term benefits for direct phonics instruction in terms of reading ability, particularly for children with “low reading readiness”, usually those from lower socio-economic backgrounds (Hemenstall, 1997; Juel, 1991). Scientific findings that phonics instruction provides a “better start in learning to read” for children, (Anderson et al., 1985, p. 37) were corroborated by results from the National Assessment of Educational Progress, which found that reading improved during the period of greater code emphasis in the 1970s, and that these gains continued long term, with higher reading ability among 17 year olds in 1988 (Chall, 1996). On the other hand, gains among beginner readers stagnated or declined during the 1980s with the reduction of phonics instruction in classrooms and basal readers (ibid.).

---

<sup>12</sup> Unlike spoken language, which children will naturally acquire if exposed to speech

Distrust in the veracity of results emerging from whole language supporters, alongside the weight of evidence in support of an initial code emphasis led to a loss of credibility within the research community for a meaning-emphasis only approach to reading instruction (Chall, 1996; Kim, 2008; Pearson, 2004; Stanovich, 1994). The mainstream consensus in the research community by the mid-80s was argued to be so firmly in favour of phonics that research shifted from whether phonics should be taught to the question of how best it should be taught (Anderson et al., 1985; Hempenstall, 1997; Juel, 1991).

Overall, according to the three criteria prescribed by Stanovich (1994), phonics was found to be of superior quality to the research produced by and used to support the whole language movement.

### **Comparing phonics research of the 1960s to that of the 1980s**

At the end of the 1960s, policymakers were presented with substantiated preliminary findings on reading instruction and reading acquisition that were both valid and reliable. These findings were still only partial, however, and based on a limited number of studies with methodological flaws. For the first time, theories of reading were proposed, at least partially supported by experimental evidence. Research into basic processes was still in its infancy, limiting the evidentiary basis on which applied researchers could base their proposals for instructional methods, though it did develop substantially following the passage of the Cooperative Research Act in 1954. Despite these limitations, the end of the 1960s marks a substantial rise in the quality of the reading research field, which had been relatively stagnant for over 30 years with the release of two high quality syntheses directed at practitioners and policymakers.

Like the research of the 1960s, the research of the 1980s concluded in favour of an early code-emphasis. Findings at this point were, however, firmer, more detailed and more conclusive (Chall, 1996). Where Chall's (1967) summary of existing research and Bond and Dykstra's (1967) conclusions were almost entirely based on the results of classroom and laboratory experiments comparing the relative effectiveness of different instructional methodologies, this type of data represents only a portion of the research summarised by Anderson et al. (1985) and Adams (1990). Research into cognitive, perceptive and other processes did not exist 20 years earlier provide readers with a far more detailed understanding of the reading process and how it is acquired. While in the 60s, Chall (1967) could only

hypothesise why an early code-emphasis may be beneficial for beginner readers, in the 80s major reading syntheses were able to conclude that an “early, focused and systematic emphasis on the code” was effective in assisting children to acquire the skills critical to fast, accurate word identification at the centre of skilled reading (Pearson, 2004, p. 245; Anderson et al., 1985; Hempenstall, 1997; Juel, 1991). Where Chall (1967) could only conclude an advantage for code-emphasis programmes, Adams (1990) can explain why they work, how they work, and how they can be better designed for greater effectiveness.

## **Research Relevance**

As mentioned in the introduction, Landry et al. (2001a, 2001b) identified four criteria to measure the relevance of research to users. Each of these criteria will be examined separately, according to the method delineated in the introduction, and comparison between the two time periods will be conducted per criterion.

### ***Focus on advancement of knowledge***

#### *1960s*

Applied research is defined as research directed at solving a real world problem faced by policymakers or practitioners (Furlong & Oancea, 2005). The research of the 1950s and 60s, which sought to determine which of the available reading programmes was most effective at teaching children to read, clearly fits this description. Basic studies were rare in this period. In fact a major criticism of the research of these decades was the weakness and paucity of studies into the basic processes of reading (Venezky, 1984). Thus the research of the 50s and early 60s can firmly be described as meeting the *demand pull* criterion of relevance due to its focus on public benefit. The 1960s syntheses retained this applied focus, though there was an attempt to make inroads towards advancing scholarly knowledge, i.e. understanding the reading process. The experimental studies included in *First Grade Research Studies* followed the model of the earlier studies, directed at assisting educators in their selection of a reading programme for the classroom, though they used improved methodologies to do so. *The Great Debate* had a slightly broader mandate; in addition to the experimental comparisons, Chall (1967) synthesised available research into the cognitive and perceptual processes involved in reading in an attempt to formulate a theory of reading. This latter type of research is considered ‘basic’ as it is conducted to “pursue the truth wherever it may lead” without certainty of its immediate real world implications (Bush, 1945, p. 12; Furlong & Oancea, 2005; Landry et al., 2001a). Due to the limited evidence available, however, this was only a

minor component of Chall's work and she retained the dominant focus of the period on the practical implications of findings for educational policymakers and practitioners.

Thus according to the 'focus on advancement of knowledge' definition of relevance, the research of the late 60s was slightly less relevant to users than the earlier works due to the minor shift away from the "purely" applied approach of the earlier works. Nonetheless, the works of this period retained the focus on applied knowledge and thus can be categorised as focusing on advancing a public benefit more than advancing scholarly knowledge.

### *1980s*

Increased funding for basic research, and consensus that the field had an inadequate understanding of the processes involved in reading and reading acquisition, pushed researchers of the 1970s and 80s toward a focus on advancing scholarly knowledge (Alexander & Fox, 2004; Pearson, 2000). Researchers all but stopped studying instructional methods directly, and instead sought to understand cognitive and learning processes, and develop theoretical models of reading and reading acquisition (Adams, 1990; Stahl & Miller, 1989). While this research was generally accompanied by explanations as to the implications of findings for instruction, the primary focus remained on basic processes and theoretical models— subjects that are far from the definition of applied research. *Becoming a Nation of Readers (BNR)* and *Beginning to Read*, on the other hand, aimed to translate the theoretical findings of this corpus of research into practical recommendations for the real world dilemmas of the classroom for implementation – a clear public benefit.

*BNR* (Anderson et al., 1985) was more applied in its nature, with the body of the text summarising research findings and directly applying them to the home and classroom; the specific studies on which findings were based relegated to the footnotes. *Beginning to Read* (Adams, 1990), on the other hand, contained detailed explanations of the findings, academic disputes and theoretical implications of research conducted, before venturing into explanations of their consequences for instruction. A much greater proportion of *Beginning to Read* is devoted to basic research than any of the other syntheses.

In sum, research conducted in the 70s and early 80s was, according to this definition, less relevant than that of the 1960s. Of the syntheses, *BNR* was more relevant, on par with the syntheses of the 1960s, while *Beginning to Read* was the most basic, and therefore least relevant, of the syntheses.

### *Focus on users' needs*

As this criterion requires an examination of the testimony of authors, only the syntheses will be examined here.

#### *1960s*

Both syntheses of the 1960s provide ample evidence of their authors' intention to meet users' needs. Chall introduces her work with the statement that the question of how to best teach children to read "is more than an academic question" (1967, p. 3). She goes beyond a broad claim that *The Great Debate* "is addressed to a wide audience" to delineate the various target groups and their differing needs:

"For authors, editors, and publishers of reading materials, I have attempted to provide evidence useful for making decisions on new reading programs. I have also included material for teachers and administrators who are faced with daily decisions on methods and materials, for their teachers in schools and colleges of education, and for parents..." (Chall, 1967, p. 3)

She repeatedly asks why research is not being applied to instruction, with the clear insinuation that it was conducted with this as its primary purpose.

Bond and Dykstra begin their study with the practical problem faced across the country that "every year hundreds of thousands of children begin the complex task of learning to read" (1967, p. 9). Experimental design also reflects their commitment to ensuring that findings are relevant to the context in which users work: studies were conducted in classrooms with real teachers in order to achieve the highest possible fidelity to real life situations. They expressly take into account factors such as class size, teacher competence and background of children that are critical in the real world, but malleable in the lab. Conclusions were expressed as recommendations for implementation in classrooms. Asked retrospectively about the impact of *FGRS*, researchers who participated in the studies responded in reference to its use, or lack thereof, in policy and practice, thereby highlighting this use of the studies as their primary concern (Schantz, 2002).

While the research of this period set higher standards for experiments than had been expected in the past, this did not detract from their commitment to producing research directly relevant to users involved in educational policy and practice.

1980s

Like Bond, Dysktra and Chall; Anderson et al. (1985) and Adams (1990) intended that their works be used by policymakers and practitioners to inform instruction and directed their works at meeting needs as they perceived them.

The *BNR* report was conducted by the National Academy of Education's Commission on Education and Public Policy, an institute that was expressly established to identify research that might inform educational policy (Anderson et al., 1985). The authors explicitly state that "[t]he purpose of this report is... to draw implications for reading instruction", and they acknowledge society's need for literate citizens as the drive for improving reading instruction (Anderson et al., 1985, p. 3). Perhaps the strongest support that they favoured users' needs is the statement that "the heavy trappings of scholarship are eschewed" in order to make the studies accessible to the widest possible audience (ibid., p. 4).

Adams' (1990) study was less focused specifically on users' needs as it instead intended to summarise and critique research evidence to uncover "what we collectively know about the knowledge and experience that are required for acquiring literacy" (Adams et al., 1991, p. 387; see also Adams, 1990; Pearson, 1990). While ultimately intended to inform educational policy, the book itself was academic and lengthy. For this reason, the summary was produced, selecting data that "teachers, school administrators, parents and other members of the interested public... would find enlightening, perhaps provocative, *but above all, useful.*" (Stahl et al., 1990, p. ii, emphasis added). Reviews of the work suggest that it was successful in this regard, (Chaney, Kapinus and Schwartz in Adams et al., 1991; all three reviewers held positions in local or state policy making entities).

### ***Funding sources***

1960s

During a period when limited funding was available from any source, Landry et al's (2001a, 2001b) criterion of internal vs external funding is not directly applicable to this period. Instead the impetus of the study will be used, i.e. whether the idea for the study originated with researchers or with users.

As mentioned, Barton and Wilder (1964) conducted a survey of reading researchers in the early 1960s and found that only 1.6% received half or more of their income from research with the remainder earning a living from practical pursuits such as teaching and textbook

writing. Similarly, Chall (1967) attests that the majority of research was conducted by teachers or administrators doing doctoral studies or action research. Thus whether the funding for a particular study conducted in the 1950s and 60s was directly funded by a university or subsidised by practical pursuits, it is likely that all researchers were, out of necessity, “outwardly focused”. Certainly, studies conducted as “action research” originate in the needs of users.

The major syntheses of our investigation were also funded externally, by the federal government and the Carnegie institute which were both engaged with policy, though not responsible for it. The impetus for the studies, however, did not originate there. Rather, it was researchers who identified the need for the studies. The idea for both studies originated in a conference of reading specialists held in 1959 at Syracuse University and were proposed to the relevant funders by reading scholars (Chall, 1967; Schantz, 2002).

Thus the measurement of “outward focus” that asks who instigated the research would classify the research of the 50s and early 60s as relevant because the research was driven primarily by educators (users) themselves, though it must be noted that the evidence for this period is indirect and inconclusive. The research syntheses of the late 60s, on the other hand, would be classified as less relevant as they were initiated based on a need identified by researchers.

### *1980s*

Both *BNR* and *Beginning to Read* were commissioned and funded by bodies of the federal government, qualifying their source of funding as external. This does not substantially differentiate them from the works of the 1960s. What does differentiate them is the impetus for the studies. Unlike during the 1960s where the impetus for synthesis was a research-derived need, in the 1980s policymakers were the ones calling for research syntheses to resolve fermenting debate (Anderson et al., 1985; *Human Services Reauthorization Act of 1986*, Title IX Sec.901 p.14). *Beginning to Read* was specifically commissioned by an act of congress. Like in the 1960s, the policymakers commissioning the 1980s studies lacked authority over educational policy, held by local and state decision makers. Nonetheless, they can be distinguished from the syntheses of the 1960s in that they were directed toward a need identified by the policy community, rather than one identified by reading scholars. By this stage, annual syntheses of research had made the findings of *BNR* fairly well known in the

research community, and so it was the user community who found it necessary to commission the study in order to inform policy (Koenke, 1986).

The evidence suggests that, according to this criterion, the syntheses of the 1980s were more “outwardly focused” and therefore relevant than those of the 1960s.

### ***Extent to which studies meet users’ context***

This criterion requires that we not only look at the characteristics of the research itself, but also the context of potential users and the suitability of the research to that context. This criterion has been subdivided into four aspects: user interest, research timing, researcher credibility and the specificity of the targeted group.

#### *1960s*

##### *User interest*

The *Sputnik* launch in 1957 directed public attention to education as decision makers and the general public sought to improve educational outcomes in order to compete with Russia (Nelson & Weinbaum, 2006; Schantz, 2002; Venezky, 1984). Flesch played on this sentiment to draw public attention to the field of reading, portraying the whole-word/look-say approach as a threat to democracy and likening the phonics approach to concepts of freedom and equality (Hempenstall, 1997). *Why Johnny Can’t Read* (Flesch, 1955) and the debate that followed challenged the hegemony of the look-say approach, forcing users to respond to public demand that reading instruction be improved and a phonics approach be considered (Chall, 1967; Hempenstall, 1997). Studies purporting to provide an answer to this question were thus valued by users and considered pertinent to the public agenda.

This receptive context did not exist when research into teaching strategies was being conducted in the 1930s, 40s and early 50s. Users appeared to be satisfied with the basal series produced 30 years earlier, and were not seeking an improved product (Barton & Wilder, 1964; Pearson, 2000). Research conducted before the mid-1950s faced a hostile user audience in the throes of the child-centred philosophy of education who ignored even well-grounded conclusions in support of phonics due to the past association of the phonics method with rote learning (Chall, 1967). Those same conclusions, when voiced by Chall (1967) in the new user context post-Sputnik, were then readily received (Chall, 1983).

### *Research timing*

The decision making for textbook selection occurred on average on a five-year cycle in most states and districts (Institute for Educational Development, 1969). The release of research, if it coincided with the right point on that cycle, could determine its relevance to users. However, as the timing of adoption varied across the thousands of school districts, and as no particular group of decision makers was more significant at this time than any other, it is not possible for me to determine when the release of research would be most pertinent to users in their staggered decision making cycles.

### *Researcher credibility*

The authors of both research syntheses would have been seen as credible by users. Chall at this time was a professor at Harvard University. She had served on the board of the International Reading Association from 1961 to 1964 and in 1967 founded the Harvard Reading Laboratory in 1967, so she was well established as a credible source both within academia and amongst reading professionals and practitioners.

The participating researchers of *FGRS* were selected via a competitive process through submission of proposals to the ‘Cooperative Research Council’ run by respected leaders from the field. (Bond, 1967; Schantz, 2002). Friendship between Francis Keppel, then Secretary of Education, and Donald Durrell, the “architect” of *FGRS* enhanced the credibility of the project (Schantz, 2002, p. 78). Bond already held a professorship at the University of Minnesota, and many of the participants including Bond and Chall were considered by fellow academics to be “giants in the reading field” (Vilscek in Schantz, 2002, p. 84)

### *Specificity of targeted group*

Both Bond and Dykstra (1967) and Chall (1967) directed their works at the multitude of players involved in setting educational policy and practice nation-wide, rather than at a specific group of decision makers in a particular location. They were by no means adapted to meet the unique needs of one decision making body or another.

Thus overall, the syntheses of the 1960s were well matched with the user context, though there were aspects in which this match was not complete: the works were not directed at specific users, and large groups of users disagreed with the findings of the studies. It thus depends on which definition of relevance to context is used, and will require comparison to the 1980s time period.

1980s

*User interest*

The *Becoming a Nation of Readers* report was part of a wider drive to improve educational outcomes in response to "the widespread public perception that something is seriously remiss in our educational system" (Gardner, 1983, Introduction, para. 2). A report on the state of education in the U.S., *A Nation at Risk* (1983), set a particular tone for public debate about education that called for 'back-to basics' school reform and greater accountability (Finley, 2000; Nelson & Weinbaum, 2006). These values favoured a phonics emphasis, setting the stage for the favourable reception of *Becoming a Nation of Readers*. Some even argued that the reception was too favourable, with policymakers ignoring the subtlety of recommendations to latch on to the recommendation that they favoured – phonics instruction (Shannon in Adams et al., 1991; K. S. Goodman, 1989).

*Beginning to Read* fell during a time of even greater heightening of standards and requirements, and the U.S. Congress was particularly favourable to a report that supported the inclusion of phonics in beginner reading instruction (Finley, 2000; Hargreaves & Goodson, 2006). At the same time, the book actually contradicted the specific purpose set out by the commissioner of the study. Federal legislators had specifically requested a table of beginning reading instructional programmes and methods indicating their correspondence to the recommendations of *BNR* and the quality of the phonics instruction (*Human Services Reauthorization Act of 1986*, Title IX Sec.901 p.14).

The Department of Education ignored the request for a table of programmes, modifying the questions passed on to Adams to more basic ones regarding the place of phonics in beginning reading instruction. Adams praised the Department for this revision, arguing that the original request oversimplified the issues and that this was an obstacle to producing meaningful solutions for educators (Adams in Adams et al., 1991). The degree to which Adams' highly praised (1990) book really met the needs of users is thus questionable. On the one hand, Adams fulfilled the mandate given her by the Department of Education, and worked to build the capacity of decision makers to make informed decisions about instruction based on a thorough understanding of the research. On the other, Congress, and presumably textbook selection committees at all levels, had requested a list of textbooks that met recommendations as a guide for selecting textbooks for use in schools, and the book by no means fulfilled this request (McCulloch, accessed 20 June 2012). Overall Adams met the interest of users, but probably offered greater complexity than was desired.

An additional complexity is the diversity of users' interests at this time. While the 'excellence' movement was the dominant position, prominent in federal and state politics, it did come in conflict with an increasingly influential grassroots movement towards greater teacher independence. Teachers were being encouraged to value their own knowledge as professionals, and reject the model of teachers as bureaucrats implementing decisions determined outside the classroom by higher authorities (K. S. Goodman, 1989; Pearson, 1989). Teacher support groups pushing for collegial professional development sprung up around the country (Y. M. Goodman, 1989) and teachers began demanding to take part in decision making for their classrooms (Hoffman et al., 1994). This approach was more hostile to a federally-funded report that offered "universal" solutions and which valued academic knowledge over professional knowledge (see, for example Wilkerson, 1988) though it did nod to professional knowledge by including one teacher on the committee (Anderson et al, 1985).

#### *Research timing*

While it was the federal government that commissioned the reports, it was the states who were important in terms of scheduling. As described above, at this time, the state educational bodies of Texas and California were the most influential in terms of curriculum. California's educational framework was set to be released in 1987 for 1989 adoption; Texas adopted new textbooks for the 1986/7 school year and then again in 1992/3 (Hoffman et al., 1994). Chall and Squire found that the development of "a total reading program" can take up to five years, though textbook publishers commonly released revisions of existing programmes every three years (Chall & Squire, 1991, p. 124). Hoffman et al. (1994) argue that five of the largest basal series publishers released new programmes in 1993 in response to Texas' 1990 Proclamation. Thus it is likely that research published three years before the expected selection dates can be incorporated into the design of new programmes. *Becoming a Nation of Readers*, published in 1985, would thus have likely been too late to influence textbooks published for the 1986/7 Texan cycle, but would have been sufficiently early to meet the Californian 1989 deadline and the later Texan adoption. *Beginning to Read* was released in 1990, the same year as the Texas Proclamation, and thus was well timed for Texas' 1993 adoption cycle.

#### *Researcher credibility*

Again the authors of both research syntheses would have been seen as credible by users. The Commission responsible for *BNR* was termed "blue ribbon" by the U.S. Department of

Education (United States Department of Education, 1985, para. 4), and was specifically established by the National Academy of Education to identify research to inform educational policy (Anderson et al., 1985, foreword). Similarly, the Center for Reading, which oversaw and released Adams's work, was established by the Department of Education, so it would certainly have been considered credible, at least within the federal government. Adams, being a single researcher, may have had less credibility than *BNR*'s panel of experts, but her endorsement from Pearson, the director of the Center and founding editor of the canonical *Handbook of Reading Research*, would have held considerable weight.

A factor that may have produced a difference in credibility between the two periods was the method of synthesis selected. While Chall (1967) and Bond and Dykstra (1967) used quantitative methods of synthesis, both *BNR* and *Learning to Read* utilised qualitative methods. *Becoming a Nation of Readers* relied on the collective judgement of the experts on the panel to decide which studies to include and what conclusions to draw from the available evidence (Anderson et al., 1985). Adams relied on her own judgement based on an "exhaustive" search of all available research into basic processes and instructional practices in early reading and other topics related to beginner reading (Adams, 1990; Pearson, 1990). The qualitative method enables broader and more detailed conclusions to be drawn from a wider variety of sources, methods and disciplines than the quantitative method, which restricts analysis to studies with experimental designs and meaningful effect sizes. This method also has drawbacks. Quantitative analyses are transparent in their selection of studies and method of analysis, while qualitative reviews are "more dependent on researchers' judgment and insight", leaving their conclusions more open to challenges (Shanahan, 2000, p. 210).

Thus, while Adams argued that she worked to reconcile data from educational and psychological journals with that from sociolinguistic and other perspectives not usually considered by psychologists (Adams, 1990; Adams in Adams et al., 1991), she was still criticised by Yetta Goodman, a leader of the whole language movement, who for excluding differing opinions and critiques of cited studies (Goodman in Adams et al., 1991). Similarly, the *BNR* panel was accused of selectively including only a small sample of the thousands of available articles, heavily biased towards articles authored by panellists and their colleagues (Grundin, 1985). Undoubtedly the dispute is, at least in part, a result of the different research philosophies of the two schools of thought. The lack of explicit selection criteria nonetheless contributes to the uncertainty.

### *Specificity of targeted group*

Like in the 1960s, the syntheses were targeted nationally, and not at specific decision makers responsible for setting curricula. A difference here was that *BNR* adapted recommendations for each group of users, offering different recommendations according to their role (Anderson et al., 1985). They were still described on a national level (teachers/schools/students were not differentiated by geographical area or demography) but there was greater differentiation between users.

All in all, the release of the syntheses in both periods coincided with user concern with improving educational outcomes and interest in a more ‘scientific’ approach, which suited the phonics recommendations. Research timing cannot be assessed for the 1960s, but in the 1980s, the studies were well-timed to meet the adoption cycles of the late 80s and early 90s. Authors of all syntheses were highly credible, and none of the studies targeted a particular group of users. Adams’ work was probably the least relevant of the four according to this criterion, due to the violation of the requested criteria. *BNR* was more relevant due to the strong federal support for phonics and the division of chapters by user context. Overall, the periods were largely similar in their relevance according to the ‘user context’ criterion.

## **Dissemination**

### *Adaption of products to users*

One criterion by which dissemination is measured is the effort that researchers invest to make reports more amenable to readers. Works that are easier to understand, in a format that is accessible to users, and have specific and implementable recommendations are considered to be of higher dissemination value.

### *1960s*

#### *Format/Accessibility*

One of the criticisms of research produced before the 1960s was that it was published exclusively in academic peer-reviewed journals, and was inaccessible to users (Venezky, 1984). The publication of *The Great Debate* in book form using non-academic language, and the summary of *FGRS* in a practitioner journal, suggests preliminary steps towards making research more accessible to users. Previously, in order to be acquainted with the collective findings of the reading research field, users would have had to themselves read and reconcile often contradictory developments published simultaneously in multiple journals. The unique

appeal of *The Great Debate* was that it condensed the debate in one place, analysed the synthesised findings and offered explanations for apparent contradictions, providing users for the first time with easy access to the accumulation of research.

On the other hand, these original documents were quite bulky – *The Great Debate* was a full length book while *FGRS* was written in the style of academic articles, with extended statistical analysis. Working to minimise these impediments, Chall (1967) wrote short prose summaries at the beginning of each chapter, and Bond and Dykstra (1967) cleared the final conclusions of technical language and references to effect sizes, writing clear recommendations with practical implications. A year after the initial publication Dykstra published a summary of the follow-up to *FGRS* in *The Reading Teacher* (Dykstra, 1968). Pearson (1992) described the item as “an excellent model of how to report research for a broad audience”, noting in particular that it succeeded in clarifying the conclusions of the original report that had been regularly misinterpreted (Pearson, 1992, p. 382). Chall’s book was described as “thorough, disciplined, and readable” (Adams, 1990, p. 39) and it became “an educational best seller” within its first year, speaking to its general popularity (Chall, 1983).

### *Applicability*

Educational psychologists involved in reading research of the 1950s and 60s were criticised for failing to interact regularly with reading professionals, concern themselves with the practical side of reading or have adequate understanding of “the complexities of curricular design and classroom practice” that would allow translation of research results to the classroom (Venezky, 1984, p. 27). Chall and the participants in *FGRS* spent considerable time working with teachers, reading specialists and policymakers, and as a result demonstrated a far greater understanding of the complexities of the classroom. Indeed, eight of the ten participants in *FGRS* interviewed by Schantz (2002) had experience teaching in a school classroom before their transition to research, though only one of them had taught in the elementary years when reading is taught (grades 1-3). The syntheses made a number of practical recommendations for instruction, but they were limited to rather general prescriptions such as an early code-emphasis, utilisation of multiple complementary methods of instruction, and raising expectations of children’s vocabulary and acquisition rate (Bond & Dykstra, 1967; Chall, 1967). The field could not at this time describe in detail how instruction should look or point to a preferred programme. Most recommendations were suppositions

rather than definite findings. Both authors emphasised how little was known, and how much still needed to be explored.

According to both aspects of adapting products to users, the syntheses of the 1960s improved upon the research of the previous decade(s), though they by no means fulfilled all the requirements of this criterion.

### *1980s*

#### *Format/Accessibility*

Both syntheses of the 1980s made concerted efforts to make the developments and findings of two decades of research accessible to a user audience. Anderson et al. (1985) did this by sidelining controversies, excluding statistical analysis, and focusing on research findings rather than the research process (Botel, 1985; Grundin, 1985; Koenke, 1986). Adams' (1990) long, dense and 'jargon riddled' book was condensed into a shorter and simpler summary, similar to *BNR* in concept (Adams in Adams et al., 1991). A number of academics and even educators expressed concern that these simplifications would be misconstrued – and indeed they were (Shannon, Goodman, Kapinus and others in Adams et al., 1991). Nonetheless, these compromises did make the documents more “useful to busy policymakers... [and] helpful in policy and planning discussions” (Kapinus in Adams et al., 1991, p. 379). Thus both the original *BNR* report, and the *Beginning to Read Summary* successfully adapted research products to make them accessible and readable for users. Compromises to research integrity that may have resulted from minimising controversies in the field do not detract from the higher level of dissemination.

#### *Applicability*

Like *The Great Debate* and *FGRS*; *BNR* and *Beginning to Read* were “not intended to be a substitute for curriculum” (Adams in Adams et al., 1991, p. 389). Rather, the works were intended to build reading professionals' understanding of how reading is acquired and developed, providing concrete recommendations that could be integrated into existing strategies. Additionally, the selection of authors indicates effort to ensure that knowledge of the classroom would be considered in writing the reports. A teacher was included on the panel of *BNR* and professional educators were employed as advisors (Anderson et al., 1985). According to Pearson (1990), Adams had long been involved in ventures to translate complex research issues into operational classroom procedures.

Both syntheses contained highly applicable recommendations for teachers, parents, school administrators, teaching colleges, textbook authors and other reading professionals. Parents, for example, were advised to read aloud to their children while asking simple questions about the texts (Anderson et al., 1985). The developments of the field enabled the authors to outline far more specific recommendations for instruction than had been possible twenty years earlier. For example, Anderson et al. (1985) recommended that phonics be taught by blending sounds rather than separating them, and Adams advised that text selections should utilise phonics relations taught. These were far more specific recommendations than ‘a code-emphasis beginning’, which had been Chall’s (1967) conclusion.

The improvements that were made in terms of applicability of recommendations are what particularly distinguished the syntheses of the 1980s/90s from those of the 1960s.

### *Dissemination Efforts*

It is difficult to assess the full scope of dissemination activities during an historical period as activities such as informal meetings between users and researchers and even conference proceedings do not necessarily leave a public record. The assessment below is thus based on the available evidence. Whether this accurately represents the full scope of dissemination activities is unfortunately a limitation of this study.

#### *1960s*

##### *Developing educational products*

Adapting research reports is only one step prescribed by dissemination theorists. Researchers should further “adapt their products so as to facilitate their appropriation by users” (Landry et al. 2001a, p. 338). In the context of high level of dependence by teachers on published materials, this means working to integrate findings into basal reading series, teachers’ manuals and training courses.

Barton and Wilder list the practice of researchers working with publishing houses to produce a commercial basal programme for the classroom as one of the primary routes by which research impacts policy (Barton & Wilder, 1964). This was not a step taken by any of the authors of our syntheses during this period. The individual researcher-developed non-basal reading programmes tested in *FGRS* did not succeed in penetrating the market, so changes that took place after this time cannot be attributed to the packaging of those programmes (Pearson, 2000; Schantz, 2002). The recommendations produced by Chall (1967) and Bond

and Dykstra (1967) from their respective syntheses, were therefore accompanied by less investment in dissemination than earlier (partisan) researchers operating in the 1950s and 60s.

Chall did, however, consult for a number of children's television programmes, introducing her findings to the general public through programmes such as Sesame Street and The Electric Company (Woo, 1999). She also served in her professional career as "a member of numerous scholarly organizations, editorial boards, policymaking committees, and state and national commissions... a consultant for children's encyclopedias, an educational comic book, educational software, and educational television", though the timing of each of these positions was unavailable ("Jeanne Chall Reading Lab", accessed 18 May, 2002).

### *Using professional platforms*

Aside from products developed specifically by researchers for users, dissemination efforts include participation by researchers in professional fora, which disseminate findings to a wider user audience. *The Reading Teacher*, a journal for education professionals, dedicated three whole volumes to the *First Grade Reading Studies* and their second grade follow up (May 1966, October 1966 and May 1967). Another twenty-five articles on the subject of beginner reading were published in the 1968, 1970, 1972 and 1974 issues (Pearson, 1992). Articles by Chall, Bond, Dysktra and other participants in *FGRS* appeared in the above editions.

Conferences are another common platform for researchers to disseminate their findings, lecturing directly to various professionals and allowing at least some small degree of feedback or questioning. While it is difficult to assess the full number of conferences held addressing this topic during this period as records of such events are scarce, we do know that *The Great Debate* was the subject of two full day professional conferences and symposia in the year following its release (Chall, 1983) and the three authors of the syntheses were featured in the International Reading Association's annual convention in April of 1968 ("Preview Program," 1968).

While the International Reading Association (IRA) could not by any means be described as being dedicated to supporting a code-emphasis approach to beginner reading, the prominence of these authors in the IRA – Chall was a member of the board 1961-1964, and Bond's presentation to the convention was listed under the banner of a "distinguished leaders addresses" - suggests that it was a forum through which findings would have been readily, if critically, disseminated ("Preview Program," 1968, p. 20).

Thus, while the research synthesis authors did not translate their findings into a published reading programme (indeed, by their own admission, this would not have been entirely possible), substantial efforts were made to disseminate findings to a wide practitioner audience through existing professional platforms, and to the wider public through television programmes and Chall's 'best selling' book.

*1980s*

#### *Dissemination of products*

While neither Adams nor Anderson et al. converted their recommendations into classroom textbooks (indeed Adams explicitly avoided a request to do so), extensive efforts were made to disseminate the reports to users to educate them on findings. The *Becoming a Nation of Readers* report was published publicly by the U.S. Department of Education and made freely available at minimal cost (\$4.50; Anderson et al., 1985). The Department also encouraged secondary distribution: The National Council of Teachers of English was asked to distribute *BNR*, which they did after some debate (though they expressed their opposition to recommendations in a comprehensive rebuttal to the report; Davidson, 1988).

These efforts seem to have been successful; The Center for the Study of Reading reports 300,000 copies of *BNR* in circulation<sup>13</sup> (Center for the Study of Reading, accessed 4 June 2012).

#### *Using professional platforms*

Whereas the 1960s syntheses were able to rely on structured dissemination frameworks such as International Reading Association conferences and professional journals like *The Reading Teacher* to communicate findings to users; during the 1980s, the whole language movement dominated these structures, and limiting the dissemination of products and ideas supporting a phonics approach.

An examination of article titles in *The Reading Teacher* in the year following the publication of *Becoming a Nation of Readers* found only one article relating to the report – two brief reviews, one supportive, one critical (Botel, 1985; Grundin, 1985). Pearson (1992) analysed 20 years of articles (in even numbered journals) published in *The Reading Teacher*, a journal run by the International Reading Association aimed at practitioners. He found that only four

---

<sup>13</sup> This is a current figure but it is fair to presume that the bulk were released in the first few years. Perhaps signifying that you cannot satisfy everyone, the Riggs Institute, a non-profit literacy agency that firmly supports phonics argues that "the government appropriated little funding to distribute this report to teachers and textbook selection committees who need it" (McCulloch, para. 20, accessed 20 June 2012).

of the 288 articles published between 1984 and 1991 focused exclusively on phonics and word identification - 1.4% of the articles published in that period. In comparison, 61 articles in this period were included on 'integrated language arts' (21%), which Pearson associates with the whole language approach (Pearson, 1992). Pearson contends that the absence of articles reporting on the rather substantial developments that were being made by reading researchers outside the whole language school suggests that practitioners and meaning-emphasis advocates were getting advice on the role of word identification and phonics in reading not from respected scholars, but from basal teachers' manuals and professional textbooks (Pearson, 1992). Considering that basal readers of this period were criticised by code-emphasis scholars for not accurately reflecting research of the period (Adams, 1990; Anderson et al., 1985; Juel, 1991), this is a fairly damning portrayal of the dissemination of phonics research in this period.

Professional conferences and teacher support groups were another major means of dissemination. Assessing the presence of research supporting a code-emphasis within the IRA and its various channels is difficult, however they do seem to have been dominated by whole language advocates. Dr Lynn Gordon, currently the head of the phonics special interest group at the IRA reported that she formed the group "out of a frustration with the obvious whole language ideological slant of IRA during the 1980s and 1990s" (personal communication, May 22, 2012). The group, which aims to share phonics resources and researchers among reading professionals was formed in 2002. The only other "phonics support group" in evidence, The National Right to Read organisation, was established in 1993 (National Right to Read Foundation, accessed May 30, 2012). By comparison, the Language Experience Special Interest Group, which included whole language and other meaning-emphasis approach supporters, was established by Stauffer in 1970 (Guzzetti, 2002). Local support groups disseminating whole language research and resources were so plentiful that they were united under a 'whole language Umbrella' in 1990 (Gilles, 1996).

Across the country, teacher support groups created by teachers and other school staff were used to disseminate whole language principles, discuss concepts and determine how these findings could be applied to practice (Battista, 1990; K. S. Goodman, 1989; Y. M. Goodman, 1989; Pearson, 1989). Unlike the informal and personal contacts usually used to convey knowledge from one teacher to another described by Chall (1967), this type of dissemination was organised and extensive. The groups, each with dozens or even hundreds of members, spread like an "epidemic, wildfire, manna from heaven", eventually shifting from "a

collection of guerrilla sorties” to “the conventional wisdom” (Pearson, 2004, p. 230; 220). Meetings grew to be larger than regional conferences of the International Reading Association (Pearson, 2004).

Dissemination efforts were not only targeted at teachers. Lobbyists targeted state legislatures and education departments convincing them to change textbook adoption policies and – in what some argued was a departure from the whole language philosophy – reshape the nature of the basal readers to meet whole language principles (Kim, 2008; Pearson, 2004).

The dissemination efforts associated with *BNR* and *Beginning to Read* seem to be primarily focused at the federal level, though there are suggestions of communication with the professional bodies of the states (e.g. Boards of Education; Adams et al., 1991). Direct dissemination to teachers, lower level educational bureaucrats and state legislatures seems to have been the preserve of the whole language movement.

The whole language movement’s hold on the major professional platforms for reading educators through the 1980s and 90s prevented phonics advocates from using these platforms to disseminate findings to users as they had during the 1960s. As a result, despite independent efforts to disseminate findings directly to users, their reach was far more limited than it had been in the earlier period.

## **Interaction**

This variable divides into two criteria: regard for teachers’ knowledge, representing co-production of knowledge; and accounting for the capacity of users, which covers attempts to build the capacity of users to meet recommendations, and adaption of recommendations to fit with users’ existing capacity.

### ***Regard for teachers’ knowledge***

#### *1960s*

Both syntheses provide evidence that they valued the knowledge of classroom teachers and what it could contribute to reading scholarship. Chall (1967) visited more than three hundred classrooms in the course of her study to talk with teachers, supervisors and principals about the teaching methods they were using. This information, and the difficulties they faced implementing new programmes was integrated into *The Great Debate*.

*FGRS* used classroom teachers rather than researchers to convey (teach) experimental and control materials, making them active participants in the production of research results (Schantz, 2002).

Both *The Great Debate* and *FGRS* acknowledged the importance of teachers in instruction, noting that this factor can exceed methodology in its influence on reading acquisition amongst children (Bond & Dykstra, 1967; Chall, 1967; Schantz, 2002). What makes good teachers successful, or if their abilities can be taught to other teachers was, however, not discussed. While *FGRS* did use classroom teachers in implementation of the experiment, their unique contributions to instruction were not recorded. Therefore, while this factor was acknowledged, it was not sufficiently investigated and as such could not inform the collective knowledge of the time. This is a lower form of interaction than the co-production of knowledge. Both syntheses did highlight this area as one demanding further investigation.

#### *1980s*

The two syntheses of the 1980s have different records regarding the value of users' knowledge. The Commission on Reading that produced the *BNR* report included a first grade teacher and an ex officio member from the Department of Education in addition to eight professors from different fields on the panel. Consultants included school teachers, principals and reading specialists; district reading and curriculum coordinators; and editors from basal publishing companies as well as other professors and researchers. This is a far greater inclusion of users' knowledge and step towards co-production of information than either of the 1960s syntheses. Nonetheless, the body of the report drew its findings from research. Users' knowledge may have provided context and detail, but did not go as far towards equality of status for researchers and users in the production of knowledge as did the whole language movement. Adams' (1990) work suggests an even lower level of interaction, as all listed consultants hailed from academia.

As mentioned above, phonics had at this time strong competition for the attention of policymakers and practitioners in the form of the whole language movement. The competition between the two schools of thoughts was never sharper than in the efforts that the latter made in reaching out to teachers. Above I have outlined the efforts that were made to disseminate whole language philosophy to teachers and other educators. This, however, does not adequately convey the substantial change in approach that the whole language movement fostered. Whole language blurred the boundaries of dissemination and interaction,

using teacher support groups and other means not only to spread ideas, but to encourage each group and individual to take responsibility for developing their own knowledge. Goodman described it thus:

“...these empowered teachers are generating [energy and innovations that] are producing a new base of practical knowledge for teachers and researchers to draw upon. A good deal of this development is passed from teacher to teacher in personal contacts, in teacher support groups, and in local conferences. Teachers are not only sharing their classroom innovations, they are collaborating with researchers and conducting their own research as they teach. “(K. S. Goodman, 1989, p. 215).

Teachers responded positively to the idea of having control over decision making for their classrooms and, as mentioned, the movement spread like wildfire. The high level of interaction described by Goodman (K.S., 1989) above was not found alongside any of the major phonics works. The major works of the period supporting a phonics approach were traditional works of academic research, drawing conclusions from painstakingly acquired evidence, but failing to fully integrate professional knowledge as it did not meet scientific criteria for reliability and validity.

In summary, while Chall (1967) and Bond and Dykstra (1967) did nominally include teachers in their processes of knowledge production, that knowledge was considered separately from knowledge acquired through the scientific method. Their input was limited to considerations of how recommendations could be adapted to the classroom, excluding the contribution that it could make to understanding how children learn to read. *BNR* did include users on the panel writing the report, though examination of the report content suggests that their input played a similar role to that of the earlier syntheses. The critical difference affecting the phonics school between the 1960s and 1980s was again the comparison to the whole language movement in the 1980s. In the 1960s, it was still assumed that instructional decisions should be made outside the classroom, by appointed officials or delegated to basal publishers. In the 1980s, the phonics school was going against the burgeoning movement of teacher independence led by the whole language movement, and thus the phonics school's failure to meet *interaction* criteria for facilitating use was heightened through comparison with the competing whole language movement.

### *Accounting for the capacity of users*

As mentioned, there is often a capacity gap between researchers and practitioners, with practitioners unable to implement recommendations due to knowledge or resource deficiencies. Researchers who attempt to either account for this differential, or bridge it by building capacity, rate higher on the *interaction* scale.

#### *1960s*

*FGRS* is, in many ways, a textbook example of how to build practitioner's capacity to implement research recommendations. Each of the studies trained teachers to implement the programmes; implementation was supervised both by the project managers and by the coordinating centre; and teachers were given feedback to improve implementation according to the criteria. This is true, however, only for the few dozen teachers who participated in the study, and no efforts were made to provide the same level of training to teachers outside the study such that it could affect the level of change we are examining on a national scale. Even looking only at these cases, there is reason to suggest that teacher training was insufficient to overcome the realities of the educational system of the time. One project manager reported that the teachers in her study were "required to go back to teaching the basal reading program" after the study finished (Vilscek in Schantz, 2002, p. 99). Other participants similarly expressed that their involvement was insufficient to overcome the power of the basal publishing industry. Indeed, it could be argued that the recommendation of both syntheses to get rid of the basal series did not take into account schools' inability to independently create reading programmes or train teachers to teach to recommendations.

Teacher training forums provided another platform for capacity building. Chall, Bond and Dykstra were all involved in teacher training. At the time of the publication of her book, Chall taught at the Harvard Graduate School of Education. In 1966, she founded the Harvard Reading Laboratory "for the purposes of training teachers and reading specialists, conducting research in reading, and serving the local community" ("Jeanne Chall Reading Lab," accessed 18 May 2002, para. 1). Guy Bond and Robert Dykstra were both professors of education at the University of Minnesota in the late 60s. It is fair to assume that in teaching future teachers, each author disseminated the results of their studies to users.

The participation of reading researchers in teacher training was not new for the reading field. According to Barton and Wilder (1964), researchers in the early 60s were spending as much as 80% of their time communicating research, via teacher training colleges, and regular contact with teachers, reading specialists, principals and parents, so it is even possible that this involvement represents a step down from the earlier period.

### 1980s

The idea behind the federally commissioned reports of the 80s and 90s was to build users' capacity to understand the concepts and processes involved in reading, so as to make better informed decisions about teaching children to read. In this way, the works attempted in a concerted manner to close the information gap that exists between researchers and practitioners during implementation, meeting a need that whole language, perhaps, failed to address. The whole language movement encouraged increased teacher independence in lesson planning, yet did not supply the required professional development to compensate for the recommended reduced reliance on published materials (Pearson, 2004). Less skilled teachers combined new materials with adaptations of the old in an uninformed and frequently unsuccessful manner (McGill-Frazen, 2000). Policymakers in California conflated whole language with whole-class instruction, argued by some to be the reason for the dismal results of that state in the National Assessment of Educational Progress in 1992 (Pearson, 2004).

Like the authors of the 1960s syntheses, many of the authors of the 1980s syntheses worked in teacher education though, due to an increase in funding for research, they were able to devote a much higher percentage of their time to research, at the expense of teacher training.

It appears that in neither period did researchers truly account for the capacity of users as advocated in the *interaction* model. Through its focus on practical and easily implementable recommendations for users, rather than detailed understanding of how those recommendations were drawn, *BNR* and the *Beginning to Read Summary* may have better accounted for the capacity of users than the earlier syntheses. *FGRS* provides the best example of building capacity of users, by working with teachers to implement the study, however this work was not replicated on a larger scale. Overall, these differences are relatively minor, and it can largely be concluded that in neither period did researchers supporting a phonics approach properly account for the capacity of users.

## Chapter 4: Reading Instruction Policy and Practice in the 1960s and 1980s

This chapter will examine the state of reading instruction policy and practice in the 1960s and 1980s. Specifically, the changes that took place in reading instruction policy and practice, and the degree of their alignment with research findings will be examined in order to assess research utilisation. In keeping with the definition of policy “use”, assessment of changes to reading instruction policy and practice in accordance with research will be assessed according to three categories: *product* use, *process* use, and *conceptual* use.

### 1960s

#### *Product Use*

The first use that will be examined is that of inclusion of research findings in basal series. As mentioned in chapter two, published reading programmes were highly important for the role that they played determining classroom content. This issue falls under the *product* definition of use, focusing on the end product of the reading textbooks rather than the process by which they were formed.

Meaning-emphasis basals using the “Look-Say” method dominated American classrooms from 1930 until the late 1960s and remained almost unchanged throughout the era (Barton & Wilder, 1964; Chall & Squire, 1991; Chall, 1967; Pearson, 2000). Though as many as 18 publishers flourished in this era, the various basal series were, for all intents and purposes, virtually identical (Barton & Wilder, 1964). Classroom practice was also practically identical across the country, with over 95 percent of all elementary schools in America using the series to direct the majority of instruction (Barton & Wilder, 1964; Chall & Squire, 1991). Reading instruction was thus guided by meaning-emphasis principles: children were first taught to recognise a corpus of words by sight, thereafter some analytic phonics was introduced as an auxiliary cue for identifying words (Pearson, 2000; Schantz, 2002).

The changes to basal readers and reading instruction practice in the decade following the publication of *The Great Debate* and *FGRS* were, in certain ways, revolutionary and stand in marked contrast to the stagnation of the first half of the century. Popp (1975) compared beginning reading programs published after 1967 with those published in the 50s and early 60s (cited in Chall, 1983; Chall & Squire, 1991). She found a significantly stronger code

emphasis in those published after *The Great Debate*. Phonics instruction in the later books was more systematic, more letter-sound correspondences were taught; and phonics instruction began earlier, right from the beginning of formal reading instruction (Popp, 1975 in Chall, 1983). Chall's (1967) recommendation that readers be made more difficult, increasing the number of words taught and reducing repetition, was also followed in the basal readers, though only in the first grade, and not throughout the series as had been recommended (Chall, 1983; Pearson, 2000). Other phonics-oriented teaching materials appeared during this period – phonics workbooks, decoding games, kits; even *Sesame Street* began to teach phonics to viewers (Chall, 1983). Meaning-emphasis programmes did not disappear during this period, but even these programmes increased the amount of phonics instruction, the number of decodable words and the number of new words in first-grade books (ibid.).

While Flesch (1979, 1981) argued that no changes were made to the basal readers in the 25 years following *Why Johnny Can't Read* (cited in Chall, 1983), Popp's (1975) findings are considered more credible and are widely cited by later scholars (e.g. Adams, 1990; Chall, 1983; Pearson, 2000). Most authors attribute this change to Chall's (1967) synthesis, *FGRS*, and other research conducted at the time (Chall & Squire, 1991; Chall, 1983; Pearson, 2000).

An additional case of research use in educational *products* after 1967 was the change made to standardised reading achievement tests, which reflected “almost universal acceptance of decoding as a major objective for the primary grades” (Chall, 1983, p. 4). The 1971 edition of the tests added a subtest on “word analysis” to the traditional “word meaning” and “paragraph reading” components of the test (Chall, 1983). This test forced teachers to teach decoding skills in order to prepare pupils for the test<sup>14</sup>.

Despite the vast changes to reading instruction practice in the early 70s, not all of the syntheses' recommendations were adopted. According to Pearson (2000), while Chall's (1967) recommendation that phonics be implemented early and systematically was adopted, and stories diversified, the recommendation that this be carried through until at least the fourth grade was not, and neither was the recommendation that grade levels be re-evaluated so as to increase the challenge at each grade level. Chall's (1967) favouring of synthetic

---

<sup>14</sup> Pearson (2000), while acknowledging the change and Chall's role in it, considers the rise of the “mastery learning movement” and the development of “criterion-referenced tests” during this period co-contributors to the change (2000, p10).

phonics was only taken up by some publishers, and then only partially; most basal publishers simply moved the analytic phonics lessons from the grade 2-4 books of the 1960s into the grade 1 basals of the 1970s (Pearson, 2000). That said, Chall's recommendation for explicit phonics instruction was by no means categorical; she acknowledged that at that time there was insufficient evidence available to decide convincingly in favour of a particular phonics method (Chall, 1967, 1983).

The inconsistency in changes made poses problems for linear models that define use as unmodified adoption of researchers' recommendations. Even though the changes were radical, they were only partial, and recommendations were not implemented precisely nor in full. The continuing dominance of the basal readers in American classrooms poses an even greater problem in this regard.

The basal reading series were not the only available option for teachers seeking a published reading programme to apply to their classrooms. *FGRS*, actually a synthesis of 27 studies, had tested a number of published reading programmes against the traditional basals and found each more effective than the basal-control (Bond & Dykstra, 1967; Dykstra, 1968). These "pretenders to the early reading throne" were, however, not adopted to replace the discredited basal series (Pearson, 2000, p. 9). Rather, it was those very same basal publishers who had received a "critical sacking" from Chall and "empirical thrashing" from *FGRS* were the first to adopt their recommendations, and they became the dominant product of the 1970s (Pearson, 2000, p. 9; Chall, 1983). Goodman et al argue that overturned the pretence that basals reflected the pinnacle of scientific discovery by showing them to be "a product of convention not science" yet use of the basal series actually increased (or at least remained at extraordinary high levels), and teachers continued to "follow the manuals literally" on the presumption that this represented best practice (Goodman et al., 1988, p. 24). The failure of *FGRS* to damage the basal readers' hegemony in fact led some of the participants in the study to conclude that it failed to impact practice in the short or long term (Schantz, 2002).

### ***Process Use***

Utilising the *process* perspective brings up additional cases of research use, ones in which research effected the decision making process regardless of the final outcome of that process. From this perspective, the idea that instructional methodology should be an issue of consideration for teachers and schools, and the acceptance that phonics-emphasis

programmes presented a viable alternative to the standard curriculum, both qualify as evidence of the use of research in policy.

The changes to the basal readers do not only provide evidence for changes to an end *product* – produced, it must be noted, by commercial publishers. They also provide evidence of the changes that were taking place among various authorities empowered to select textbooks, a policy *process*. When these processes were studied in the late 1960s, it was found that while states regulated the process by which textbooks were selected, they rarely regulated the criteria by which they could be chosen, beyond limits on cost (Institute for Educational Development, 1969). The idea that the methodology used to teach children to read should be a criterion for selection of textbooks was therefore a new one. Though the processes at this time have not been studied directly, public demand must have reached a critical mass nationwide in order to stimulate this scope of change among the major publishers. The need for a critical mass of interest can be seen by comparison with the prior decade when individual schools interested in adopting a phonics programme found that no commercial one fulfilling their criteria existed (Chall, 1967). Thus the widespread changes to basal reading programmes described above could only be possible if widespread changes had been made to selection criteria by decision makers across the country.

The second *process* use of research in reading instruction can be found in the textbooks used to train teachers in educational colleges and universities. While the textbooks could be considered a policy *product*, they have been included under *process* due to their primary role of informing teachers' decision making regarding their teaching methods. In this case, they also provide evidence of the publisher's *process* of determining which method should govern the textbook.

Comparing those published between 1972 and 1978 to those published between 1955 and 1965, Chall (1983) found a considerable shift following the publication of *The Great Debate* and the *FGRS* findings. Whereas before 1965, 82% of methods textbooks had favoured a meaning emphasis, 18% a combined approach and none a code-emphasis approach; by 1972-78 teacher training textbooks were fairly evenly split between the three methods, with a slight advantage for a combined approach (see table 1. below).

Classification of texts	1955-1965		1972-1978	
	Number	%	Number	%
Meaning-emphasis	9	82%	5	31%
Combined Approach	2	18%	6	38%
Code Emphasis	0	0	5	31%
	Total = 11			

*Table 1.* Comparison of Teacher-Training Textbooks: Changes in Methodology from 1955-1965 Period to 1972-1978 Period (Chall, 1983, p. 33 of introduction).

Even those textbooks that retained a meaning-emphasis, or utilised a combined approach, contained marks of the impact of phonics research. Meaning emphasis textbooks increased the proportion of the text devoted to phonics instruction from 2.4% to 3.6% (though, interestingly, combined textbooks reduced the same proportion from 7.5% to 6.0%). Of particular importance for the *process* perspective is the finding that almost all textbooks supporting either a meaning emphasis or combined approach referred to *FGRS* and/or *The Great Debate* when justifying their methodology selection (Chall, 1983).

Most textbooks correctly cited Bond and Dykstra's (1967) finding that no single approach was best, though some drew from this the conclusion that "the" combined approach was therefore superior (Chall, 1983). This was contrary to Bond and Dykstra's (1967) actual conclusions (clarified further by Dykstra, 1968) that while combinations (plural) of programmes were often superior to single approaches, the different programmes were not effective in all situations and thus no single "combined approach" could be recommended.

Four of the six 'combined' approach textbooks stated that the studies showed that the teacher is the most crucial factor, however in their final report, Bond and Dykstra concluded that "teacher experience and efficiency ratings are only slightly related to pupil success" (1967, p. 119)<sup>15</sup>. Chall's (1967) findings were often only cited in order to dismiss them, and occasionally the main conclusions were ignored while the study was used to "document a minor fact" (Chall, 1983, p. 36). Interestingly, while only one of the five code-emphasis books referred to *FGRS* and only two referred to *The Great Debate*, five of the six combined

<sup>15</sup> Interestingly five out of seven researchers involved in the studies interviewed by Schantz (2002) also remembered this conclusion incorrectly.

textbooks referred to each of the studies, and four of the meaning emphasis books referred to *FGRS*, three to *The Great Debate*.

The inconsistency with which Chall, Bond and Dykstra's findings were understood again brings into question the use of research in policy. While theorists such as Rich (1997) accept citations as evidence of use regardless of whether they misinterpret findings or are selective in their references as research use; linear theorists such as Carnine (2000) would reject this as a legitimate form of use. Indeed, it may even be more appropriate to consider these instances of use as *symbolic*, whereby research is used as a 'symbol' from which to derive legitimacy while its content is more or less ignored.

Nonetheless, the mere presence of these citations, regardless of their content, speaks to another form of *process* use. Research here has altered the policy decision making process by introducing the idea that reading instruction curricula should be evidence-based, forcing teacher trainers to justify their decisions against the leading research of the day – *FGRS* and *The Great Debate*. Regardless of whether selections were indeed made in line with Chall and Bond's recommendations, the fact that research becomes a part of the decision making process is, according to the *process* perspective of use, an example of the utilisation of research in policy formation.

This use of research in decision making can also be seen among policymakers allocating research funding; according to Chall, in the late 60s and early 70s most grant proposals for research cited *The Great Debate* in order to establish legitimacy; it also informed the rationale and planning of research implemented by the U.S. Office of Education (Chall, 1983).

### ***Concept Use***

If we allow for a *conceptual* definition of use, we can see broad use of research in the policy and public realm. While the creation of “the Great Debate” over reading instruction methodology can in all likelihood be attributed to Rudolf Flesch (1955); Chall's *The Great Debate* and *FGRS* succeeded in changing the nature of that debate. From the virtual shouting match of the 50s and 60s, with both sides tending to extremes, in the 1970s debaters were more willing to listen to one another and accept more moderate views (Chall, 1983). The nature of the debate shifted; with the “strong consensus for a single-process, meaning-emphasis approach to early reading... broken”, users, publishers and researchers began

discussing which type of phonics to teach<sup>16</sup> and how reading should be conceived (Chall, 1983, p. 3). Chall's (1967) terminology of "code-emphasis" and "meaning-emphasis" became widely accepted to the point where the terms were cited without attribution (Chall, 1983). This influence even reached the professional community, as represented by the professional journal *The Reading Teacher*, where references to phonics foci such as letters, syllables, word and subword units, as well as the term 'phonics' became commonplace in the late 1960s, a shift Gaffney and Anderson (2000) attribute to Chall's book. Furthermore, the high standard set by Chall (1967) and *FGRS* necessitated a shift in the modus operandi for contributors to *The Reading Teacher*; from opinion-driven position papers to research reports (Stahl & Fisher, 1992). Among parents, the teaching of letters and sounds was legitimised through the popularity of *Sesame Street* and *The Electric Company*, themselves products of the research findings (Chall, 1983).

With all that, however, the major *conceptual* shifts took place prior to the major research studies of the period. In fact, the funding of *FGRS* and *The Great Debate* can themselves be seen as evidence of the *conceptual* use of research. Chall contended that "the consensus on beginning reading instruction in the 1950s and early 1960s was so strong that only a variety of attacks together could challenge it... everyone implicitly assumed that most of what was important to know about teaching the young child to read had already been discovered" (Chall, 1967, p.295). Thus the creation of the debate – a creation of a question, a problem that needed solving, was a *conceptual* shift. The decisions of the Carnegie Foundation and the Cooperative Research Branch of the United States Office of Education to 'resolve' the debate, even the idea that research could provide an answer beyond the knowledge of individual teachers, are all products of that shift.

## 1980s

The availability of evidence necessitates that different policies and practices be considered in the 1980s to the 1960s. Nonetheless, the basic categories of *product/process/concept* enable comparison across time periods.

---

<sup>16</sup> i.e. direct/synthetic or indirect/analytic

### ***Product Use***

In the 1980s, basal readers moved to the centre of the debate over beginner reading. As competing research schools sought to set the instructional agenda, each conducted their own analysis of the basal reader content and the changes that took place. While the conclusions varied, guided by their particular agenda, close examination of their analyses shows that apparently contradictory findings can be reconciled.

As supporters of the whole language approach, Goodman et al. decried the “considerable focus on phonics in all the basals” available in the mid-1980s, arguing that there was a focus on decoding skills and word identification at the expense of comprehension (Goodman et al., 1988, p. 66). Chall (1996), a phonics supporter, argued on the other hand that there was a significant decline in the amount of phonics taught in the 1980s as compared to the basals of the 70s, as comprehension and word meanings captured more instruction time. Both Pearson (2000) and Anderson et al. (1985), more even-handed in their support for either approach, report that during this period, all major published programmes included substantial material for teaching phonics from grade one. Pearson (2000) notes, however, that beginning in the early 80s there was greater emphasis on comprehension at the expense of skills. An explanation that fits all this evidence is that in the early 1980s, while phonics instruction did maintain the prominence it gained in the 1970s, its grip on curriculum waned, and more meaning-emphasis content was integrated into instruction. Chall and other phonics advocates lament this decline, as it was in opposition to increasingly strong research evidence supporting a code-emphasis beginning. For Goodman et al. and others, on the other hand, this decline was insufficient; the continued presence of phonics skills instruction in early lessons contrary to their view of how instruction should look.

Beyond the question of phonics vs no-phonics, the matter of *how* phonics was taught is relevant to the assessment of research use in instructional policy and practice. Despite the rather emphatic findings of the 1980s in favour of systematic phonics, implicit or analytic phonics continued to be the dominant approach of the basal readers (Anderson et al., 1985; Goodman et al., 1988)<sup>17</sup>. Similarly, despite recommendations that reading selections be guided by the phonics skill being taught, phonics skills were often taught in isolation in basal

---

<sup>17</sup> Though Goodman et al. do not actually distinguish between explicit and implicit phonics, their analysis clearly points to this conclusion. For example in a workbook exercise used to illustrate typical basal method, children were asked to underline which word in the pair has the same vowel sound as ‘bed’: break/bend, sheep/shell, an intrinsic task.

readers of this period, with little or no connection to the reading selections that followed (Adams, 1990; Anderson et al., 1985; Hoffman et al., 1994). Thus the basals of the 1980s, while they initially appear to reflect the research consensus on phonics, were not in line with recommendations emerging from the “scientific” literature of the period. Considering that the phonics emphasis was a holdover from an earlier period, and the change at this time was away from phonics, this further degrades the influence of phonics research on the basal readers.

The decline of phonics in the basals of the early 80s paled in significance to the seismic shift that took place in the late 80s and early 90s. The most notable change, and relevant for our analysis, was the demotion of decoding skills from “front and centre” in the 70s and 80s to “appendix-like status” in the early 90s (Pearson, 2000, p. 20). Phonics skills were removed from student workbooks, though some phonics remained in the teacher’s manuals, to be taught as part of the group of skills that children needed for reading (ibid). Where phonics had been the centre of skill development in the basal reading programmes adapted in 1986/87, those submitted for approval in 1993-4 focused more on comprehension and study skills, at the expense of decoding skills (Hoffman et al., 1994)<sup>18</sup>. Phonics, when it was taught, was taught analytically when it was encountered circumstantially in literature, contrary to the aforementioned recommendations. Even then, it was of marginal importance, the general focus on individual interpretation of texts, identification of repetition and rhythm and discussion of story elements (Hoffman et al., 1994; Pearson, 2000). The model of publishers writing or adapting texts to conform with lessons or to control vocabulary for ease of decoding was entirely rejected, with almost all stories taken directly from published literature (Hoffman et al., 1994). The resulting difficulty of the texts acted as an obstacle to acquisition of the code and were contrary to the findings of phonics research (Pearson, 2000).

While in the 1960s, basal reading programmes were examined as a proxy for the individual decisions of textbook selection committees, in the 1980s these decisions can be examined directly. As mentioned in chapter two, during this period state educational frameworks began

---

<sup>18</sup> Hoffman et al. (1994) analysed the basal reading programmes submitted to the Texas State Board of Education for the 1993-4 school year, comparing them to those adopted in 1986-7. Though this analysis is specifically of the programmes submitted for adoption in Texas, it is possible to generalise from this to nationwide practice. Hoffman et al. note that the changes seen in Texas in 1993-4 were expansions of earlier changes begun by most publishers a few years earlier in California. They note further that the texts were created with the intention of being marketed nationally in the following years, a reading that is supported by their finding that in most cases there were no notable differences between competing basal programmes published at the same time. (Hoffman et al., 1994).

to dictate curricular content. Analysis will focus on the reading frameworks of California and Texas as the most influential in the creation and selection of basal readers and curriculum across the country.

In 1987 and 1990, the Californian and Texan State Boards of Education respectively released new reading frameworks that both reflected and created a revolutionary shift in the definition of reading and the conception of how it should be taught. Both frameworks called for an approach to reading that centred on children's independent construction of knowledge and the use of genuine literature, sidelining decoding strategies (Hoffman et al., 1994; Pearson, 2000). The insistence on genuine literature rejected any form of vocabulary control in texts, producing the changes to the readers described above. Texas' reading curriculum did include provision for "the use of basic phonics", though it was only one of "several sources of information [used] simultaneously" (Proclamation of the State Board of Education Advertising for Bids on Textbooks: Proclamation 68, 1990, cited in Hoffman et al., 1994). The Californian framework was more overtly critical of phonics, but both de-emphasised skill instruction and phonics skills (Kim, 2008). As can be seen above, basal publishers responded to the Texan and Californian guidelines with a "remarkably different product" to any that had ever appeared on the basal market, cementing their influence (Pearson, 2000, p. 20). The strength of support for this meaning-emphasis approach reflected the dominance of the whole language approach among state-level decision makers, and the marginalisation of phonics.

#### *Instruction and relation to Basals*

Another product use that can be examined in this period is that of teachers' classroom practice. Basal programmes continued to be highly influential in determining how reading was taught in elementary schools in the 1980s. Various reports from the late 70s and early 80s estimate the proportion of instruction dominated by basal programme content and teacher's manuals as between 75 and 95 percent (see Adams, 1990; Anderson et al., 1985, p. 35; Chall, 1996; Dole et al., 1987, p. 284; Shannon, 1982, p. 884). A survey of teaching practice among fourth grade teachers conducted by the National Assessment of Educational Progress in 1992 found that 83% of teachers reported having a moderate or heavy whole language emphasis in their reading classrooms (Stahl, 1999). California, where state push for meaning emphasis had been more comprehensive, recorded significantly higher reliance on meaning-emphasis approaches (Kim, 2008).

Nonetheless, their influence was not total; classroom observations suggest that teachers largely followed the basal programmes and the teachers' manuals, but also borrowed from other, supposedly competing, methods, selected on the basis of their experience or what "makes sense" (Anderson et al., 1985; Durkin, 1984; Pearson, 1989, 2004; Stahl, 1999). Thus, phonics did not disappear entirely from classrooms. Only 33% of teachers nationally reported little or no reliance on phonics, 52% in California. This is of course self-reporting. Whole language advocates have argued that teachers claiming to have a whole language classroom were not necessarily keeping to the tenets of the movement (Pearson, 2004). The same could be said for the phonics that was taught. Pearson reports that "considerable resistance" to whole language persisted around the country, however this seems to be a factor of teachers preserving old practices rather than them integrating new findings (Pearson, 2000, p. 23). Indeed the tendency of teachers to use eclectic teaching practices, particularly by integrating new ideas with old experience, has been repeatedly noted (e.g. Chall, 1967; Pearson, 2004).

Clearly there are limits to the extent to which policy can prescribe practice. But that is not to say that they are entirely disconnected. The variation described above suggests that while phonics retained a hold in pockets among some teachers, overall it lost most of its influence.

### *Process Use*

The only area in which I have found evidence of use according to a *process* definition in this period is in the decision of federal policymakers relating to research funding. The decision was a policy *product* however gives clues as to *process*. Where *FGRS* received its grant from an open pool of funding for educational research, both *BNR* and *Beginning to Read* were specifically commissioned by federal policymakers to examine the issue of reading instruction methodology for beginning readers, the former by the Secretary of Education, the latter by Congress (Adams et al., 1991; Anderson et al., 1985). This was an unprecedented level involvement in programming by the federal government, in a domain that was clearly states' authority, and signifies the importance that the debate had acquired. The demand that curriculum be evidence-based further indicates the degree to which research had penetrated policy, resulting in a higher estimation of research and expectation of what it can contribute to policy and practice and validation of the phonics schools' approach to education research.

Research also managed to affect the *product* that resulted from this *process*. The call that lead to *Beginning to Read* specifically requested research into phonics and how it should be

taught. Considering that the previous years had been dominated by an interest in research on comprehension, the focus on beginning instruction in general and phonics specifically was a significant change (Pearson, 1990).

### *Concept Use*

The *conceptual* definition ascribes greater influence to the meaning-emphasis approach, though additional examples of use of phonics research are also found. While Pearson (2004) suggests a conventional wisdom supporting a code-emphasis approach in the mid-1980s, whole language was clearly on the ascent and by other measures appears to have dominated professional concepts of reading at this time.

References to the term phonics in the *Reading Teacher*, a practitioner journal, dropped during the 1980s, along with associated terms such as letters, syllables, word and subword units, even as they rose in *Reading Research Quarterly*, the peer-reviewed journal by the same publisher. At the same time, the percentage of articles referring to whole texts, comprehension and ‘whole language’ rose, reflecting “the lure” of ideas proposed by Goodman, Smith and other meaning-emphasis advocates (Gaffney & Anderson, 2000, p. 60). This is particularly noteworthy considering that “a buildup of references to a topic in research journals [generally] precedes a buildup of references to this topic in practitioner journals”; the disconnect in the case of phonics apparently the only exception in this period (Gaffney & Anderson, 2000, p. 72). This *conceptual* shift also affected the type of research being conducted. The 1980s and early 1990s saw an increasing numbers of researchers were doing qualitative and naturalistic research, and teacher-as-researcher became the “rallying cry in educational scholarship” (Gaffney & Anderson, 2000, p. 58), a concept that whole language scholars led. At the same time, use of experimental and quantitative methods – the type of methods often being used to support phonics claims - waned.

In the late 80s, whole language was termed an “epidemic” and “a fact of life in literacy curriculum” (Pearson, 1989, p. 230); by the early 1990s, “the conventional wisdom, the standard against which all else was referenced” (Pearson, 2004, p. 219).

There were nonetheless hints of the “endarkenment” process described by Weiss (1979) as a risk in *conceptual* use; both phonics and whole language advocates complained of popular misinterpretation of their recommendations. In the phonics case, advocates were accused of forcing a choice between learning decoding skills and reading for meaning, where in reality

they were considered as complementary (Adams and Shannon in Adams et al., 1991; Koenke, 1986).

The *conceptual* attitude to use also produces a greater number of policy products. The traditional view of policy products assumes that upon publication, research findings will be used to shape policy decisions. Birnbaum (2000) recommends extending the timeline of influence, to uncover more instances of use. Looking beyond our period, we find that between 1994 and 1997, eighteen states introduced bills that mandated phonics instruction in the early grades or at minimum, teacher training on the topic; by 1999, similar bills were passed or pending in 36 states (Kim, 2008; Pearson, 2000). California, once spearheading the push for whole language, replaced its 1987 literature-emphasis educational framework in 1996 with one that, though purporting to be balanced, favoured explicit instruction of decoding skills and phonics methods (Stahl, 1999). This move was undoubtedly triggered by the National Assessment of Educational Progress (NAEP) results of 1992, which saw a drop in reading outcomes, particularly in California (Kim, 2008). Others have attributed it to the rise in demand for a “scientific” approach to education, driven by the right-wing (Shannon in Adams et al., 1991; Stahl, 1999) . Nonetheless, the swiftness of the changes suggests that phonics had been preserved as an option, one with the evidentiary support to withstand public scrutiny now directed at state policymakers’ decisions regarding reading instruction. The phonics argument that the whole language approach was contrary to research findings led it to take the blame for the poor NAEP results and California’s failures in particular.

## Chapter 5: Analysis

This thesis sought to examine the reasons why research supporting the phonics method of reading instruction succeeded or failed to penetrate reading instruction policy and practice during two periods: 1960s to early 1970s; and 1980s to early 1990s. The thesis examined the similarities and differences in the determinants of utilisation of research knowledge supporting the phonics method of reading instruction during these two periods, and the similarities and differences in utilisation in policy and practice in the same periods. Based on the evidence discussed above, it is now possible to examine the predictive power of each of the hypotheses that form our theoretical framework, that is, the *Science Push*, *Demand Pull*, *Dissemination* and *Interaction* models. The non-predictive *Enlightenment*, *Political* and *Tactical* models will also be examined.

### **H1: Research utilization was determined by research quality (*Science Push* model)**

The *Science Push* model predicts that utilisation will be largely determined by the quality of scientific evidence (Landry et al., 2001a; Wandersman, 2003). As predicted by the model, the improvement in research quality produced by the syntheses of the 1960s as compared with the preceding decades was followed by broad increases in utilisation. This is evident in changes to basal readers, teacher training methods and standardised testing; changes in the terminology used by researchers and raised standards for experiments; the change in the nature of the debate; the fact that instructional method was considered as a selection criterion; the legitimacy of phonics amongst parents; and the idea that curriculum should be evidenced based (Chall & Squire, 1991; Chall, 1983; Gaffney & Anderson, 2000; Pearson, 2000; Stahl & Fisher, 1992).

The failure of the individual programmes in *FGRS* to take hold despite their higher quality can be explained by the lack of cumulative evidence in support of individual programmes (Bond & Dykstra, 1967). Individual studies supported particular programmes but contradicted one another and as such did not meet the “conclusiveness” aspect of the quality definition. Similarly, the prevalence of analytic rather than synthetic phonics in the basals of the 1970s (Pearson, 2000), contrary to Chall’s recommendation, can be argued to have failed because her recommendation was equivocal and the evidence base on this question inconclusive (Chall, 1967).

While the *Science Push* model is able to explain the noticeable shift that took place in the 1960s, as well as some changes that did not take place at this time, it fails to explain the changes of the 1980s. The research of the 1980s was of substantially higher quality compared to that of the 1960s, yet its publication was followed by a decrease in use by teachers, educational boards, and basal publishers (Chall, 1996). This period also saw a reduction in the use of phonics concepts in practitioner journals (Gaffney & Anderson, 2000). Phonics that was taught was not taught as recommended (Adams, 1990; Anderson et al., 1985; Juel, 1991). While in the 60s the prevalence of analytic phonics could be explained by faults in the literature, by the 1980s evidence for this recommendation was substantial and well established (Adams, 1990; Juel, 1991). Furthermore, as phonics was removed from reading instruction, meaning-emphasis methods took its place, gaining dominance in almost all areas of policy and practice (Anderson et al., 1985; Chall, 1996; Hoffman et al., 1994; Kim, 2008; Pearson, 2000), despite code-emphasis research being of far higher quality according to the *Science Push* definition (Adams, 1990; Chall, 1996; Pearson, 2004). These findings contradict the predictions of the *Science Push* model.

The *Science Push* model also fails to explain the continued dominance of basal reading programmes in classrooms in the 1970s after the emphatic condemnation of their efficacy by Chall, Bond and Dykstra (Pearson, 2000). The creation of “the great debate” challenging the hegemony of meaning-emphasis approaches in the mid-50s is also a poor fit for the model, considering the poor quality of research during this period (Chall, 1967).

It should be noted that the failure of phonics research to influence educational policy and practice in the 1980s was not total. While the whole language approach dominated this period among practitioners and local and state government, the phonics approach dominated decision making at the federal level, influencing federal research funding and legislation and the demand for evidence-based curriculum (Adams and Shannon in Adams et al., 1991).

Furthermore, the explosion of phonics bills in the mid-late 1990s following the dire results of the National Assessment of Educational Progress (Kim, 2008), suggests that the research of the 1980s had not gone unnoticed, but rather required an additional catalyst to prompt use. Arguably, its quality had enabled it to be placed on stand-by, attracting policymakers in search of a method with “evidence-based” success. While this does not fit the traditional *Science Push* hypothesis, it does provide support for the idea that improving the quality of research will ultimately increase its use in policy and practice, even if that use is delayed.

Overall, the explanatory power of the *Science Push* model is relatively weak and the model fails comparative analysis. Nonetheless, the evidence does seem to suggest that improving research quality will eventually increase its utilisation in policy and practice. The relationship is not linear or immediate as suggested by the model, and its use appears to demand an external catalyst. Poorer quality research is also able to successfully infiltrate educational policy and practice in direct contradiction with the model. Nonetheless, over the long term, the quality of research appears to have influenced its use by policymakers and practitioners.

## **H2: Research utilization was determined by the relevance of research to user needs (*Demand Pull* model)**

Unlike the research quality variable which was fairly consistent within each time period (i.e. the 1980s improved on all parameters of quality compared to the 1960s), the degree to which phonics-school researchers met *Demand Pull* criteria changes depending on the definition of ‘relevance’ used. I will therefore consider each aspect separately before considering the variable as a whole.

One definition was the degree to which research was directed toward advancing a public benefit as opposed to advancing scientific knowledge, with ‘basic’ research deemed less relevant than ‘applied’ research (Landry et al., 2001a, 2001b). By this definition, the research studies on which the syntheses of the 1960s were based were far more relevant than those of the 1980s; the research of the 1960s focused on instructional methodology at the near exclusion of research into basic processes (Venezky, 1984). In comparison, the research of the 1980s synthesised in *BNR* and *Beginning to Read* was directed at developing and testing a theoretical model of reading acquisition and understanding the cognitive processes involved in reading (Adams, 1990; Alexander & Fox, 2004; Juel, 1991; Pearson, 2000; Stahl & Miller, 1989). The lower use in the 1980s is therefore consistent with the relevance hypothesis. On the other hand, the syntheses of the 1980s (*BNR* and the *Beginning to Read* summary) were exclusively designed to advance a public benefit, unlike the 1960s syntheses which also intended to advance knowledge (Anderson et al., 1985; Bond & Dykstra, 1967; Chall, 1967; Stahl et al., 1990).

A second prediction, that external funding for studies will increase use, fails to explain any of the changes that took place. All syntheses were externally funded, however the 1980s syntheses were commissioned independently by federal policymakers while the idea for both

*FGRS* and *The Great Debate* emerged from a conference of researchers (Adams, 1990; Anderson et al., 1985; Chall, 1967; *Human Services Reauthorization Act of 1986*, 1986; Pearson, 1990; Schantz, 2002).

The suitability of the works to the user context varied between groups of users. Among teachers, interest in the findings of research regarding instructional methodology was greater in the 1960s due to public and parental demand for improved instruction (Chall, 1967; Hempenstall, 1997; Nelson & Weinbaum, 2006; Venezky, 1984). In the 1980s, when a movement for greater teacher independence in decision making prevailed, interest was lower and the credibility of the authors was damaged by teachers' rejection of academic authority and questions that were raised about study selection in the syntheses (K. S. Goodman, 1989; Pearson, 1989).

For federal policymakers, research outcomes were considered more pertinent in the 1980s than in the 1960s, as in the 80s educational outcomes were being used as political platforms by federal politicians, while in the 1960s federal policymakers were focused on school resources and had not yet engaged with curriculum (Shannon in Adams et al., 1991; Finley, 2000; Howell, 2005; Nelson & Weinbaum, 2006). Interest among state policymakers should have been greater in the 1980s, once states began developing educational frameworks to guide curricula, though the demand that such policies be evidence-based only emerged in the 1990s and 2000s (Farr et al., 1987; Keith, 1985; *No Child Left Behind Act of 2001*, 2002; Pearson, 2000, 2004).

Among basal publishers, interest in the findings would have been higher in the earlier period due to their need to predict demands; once guidelines were being set by states their interest in research likely waned (Chall & Squire, 1991; Farr et al., 1987; Hempenstall, 1997; Hoffman et al., 1994; Institute for Educational Development, 1969; Keith, 1985; Watt, 2009).

This criterion of the 'relevance' variable was more accurate in its predictions of use, and interest among users ran fairly parallel to the changes that occurred. As described above, greater use was found among teachers and basal publishers in the 60s and federal policymakers in the 80s. States began to utilise phonics research only once demands that educational policy be evidence based emerged towards the twenty-first century (Kim, 2008; Pearson, 2004). It should be noted however that determinations of user interest rest on rather broad suppositions, limiting the strength of the findings.

A fourth prediction, that use would be determined by the specificity of the user group targeted, favours the 1980s for higher use due to efforts made in the syntheses to adapt recommendations to each target audience, with for example different recommendations for teachers and parents (Anderson et al., 1985). On the other hand, these are still broad groups, and specific decision makers were not targeted, so it is possible that neither work meets the specificity criterion.

On two definitions of the variable there was insufficient difference between the 1960s and 1980s to warrant comparison: focus on user needs (all four research syntheses were produced with the express purpose of informing policy and practice and determined to meet users' needs as the authors perceived them. At the same time, all had questionable success in achieving this goal from a user perspective); and credibility (high credibility among all synthesis authors, though credibility of source selection was questioned for the 1980s syntheses) (Adams, 1990; Adams et al., 1991; Anderson et al., 1985; Bond & Dykstra, 1967; Chall, 1967; Schantz, 2002; Stahl et al., 1990).

Overall, it appears that the definition of relevance in the *Demand Pull* model is insufficiently defined to draw firm conclusions. Very little distinguishes the two periods in terms of most definitions of relevance, yet the two periods produced very different patterns of use. One aspect of relevance – user context – was consistent with utilisation amongst different users in both periods.

### **H3: Research utilisation was determined by the extent to which researchers disseminated findings to policymakers and practitioners (*Dissemination* model)**

The *Dissemination* model predicts that the degree to which researchers adapt their products to users, and make efforts to disseminate those products to a user audience, will determine the extent of research use in policy and practice (Denton et al., 2003; Landry et al., 2001a, 2001b; Peterson et al., 2007). The 1960s syntheses produced (and reproduced) research in a format far more suitable to users than in the past, synthesising multiple studies in one location with the practical implications of the findings clearly outlined in non-academic language (Adams, 1990; Bond & Dykstra, 1967; Chall, 1967; Dykstra, 1968; Pearson, 1992; Venezky, 1984). Research findings were disseminated in practitioner journals, through professional conferences and the International Reading Association, and on children's TV shows (Chall, 1983; "Jeanne Chall Reading Lab," accessed May 18 2012., "Preview Program," 1968;

Pearson, 1992; Woo, 1999). As predicted by the model, this resulted in higher utilisation by users, which can be seen not only in the broad use outlined above, but also in the popularity of the works (Chall, 1983; Pearson, 1992).

The aspect of the *Dissemination* model that focuses on the adaptation of products falters in comparative analysis. In both periods, the syntheses were designed to be accessible to users and all authors made efforts to produce works that would be amenable to broad groups of readers (Adams, 1990; Anderson et al., 1985; Bond, 1967; Chall, 1967). The 1980s syntheses were, however, slightly more successful in this regard (*BNR* and the summary of *Beginning to Read*) in that they were written as reports using non-academic language, limited to around 150 pages and divested of statistical analysis and technical language that non-academic users may find an impediment to understanding (Anderson et al., 1985; Stahl et al., 1990). They were also able to form far more detailed and practical recommendations for users than in the earlier period (Adams, 1990; Anderson et al., 1985; Chall, 1996; Pearson, 1990). Nevertheless, utilisation was lower.

The aspect of the variable that focuses on dissemination activities is able to explain the divergent fates of phonics research in the two periods. In both periods, efforts were made to distribute the research directly to users. Secondary dissemination through professional networks, however, was far more limited in the 1980s. Traditional communication networks used by Chall, Bond and Dykstra, such as the International Reading Association conferences and *Reading Teacher* journal, were inaccessible to phonics advocates of the 1980s because they were controlled at that time by whole language advocates (Gordon, in correspondence 22/5/12; Pearson, 2004). The power of these traditional structures was also weakened by the explosion of teacher support groups disseminating whole language principles from teacher to teacher (Battista, 1990; Gilles, 1996; K. S. Goodman, 1989; Y. M. Goodman, 1989; Pearson, 1989, 1992). These groups, under the whole language umbrella, came to rival the International Reading Association in size and influence (Pearson, 2004). Whole language advocates were also able to influence state policymakers through direct lobbying (Kim, 2008). The presence of effective and widespread dissemination in both positive cases of use (phonics 1960s and whole language 1980s) and its absence in both negative cases of use (phonics pre-1967 and phonics 1980s) suggests that it is the primary determinant of use in the discussed cases.

The case of the 1980s and the impact of competition on dissemination highlights a factor not expressly noted by the *Dissemination* model, that - in this case at least - dissemination is heavily reliant on secondary dissemination and the use of professional networks. Unlike quality, or even relevance, where multiple works can attain high levels of the variable, the dominance of one approach over a dissemination network is at the exclusion of other approaches.

The support for a phonics approach at the federal level may also affirm this hypothesis, as *BNR* and its findings were primarily disseminated federally, by the US Department of Education and the Center for the Study of Reading (Adams and Shannon in Adams et al., 1991; Anderson et al., 1985; Center for the Study of Reading, accessed 4 June 2012.). On the other hand, these advocates may have had pre-existing agendas that favoured this approach (Shannon in Adams et al., 1991), their support a product of ideology rather than research influence, fitting the *Political* model better than *Dissemination*.

The *Dissemination* model is also able to explain the creation of the debate and the provision of funding for the 1960s syntheses, as they were preceded by the wide circulation of albeit questionable research in *Why Johnny Can't Read* and in the popular media (Chall, 1967; Kim, 2008). However, more than a decade and a half passed between *Why Johnny Can't Read* and the changes to the basals and classroom practice, suggesting that dissemination is necessary, but insufficient, to trigger use in all areas. Similarly, despite the high level of dissemination of whole language, it lost its status in the 1990s following the results of the NAEP (Kim, 2008; Pearson, 2004), suggesting that dissemination alone does not determine utilisation. Nonetheless, overall the *dissemination* model appears to have the greatest explanatory power of the models.

#### **H4: Research utilization was determined by the level of interaction between researchers and policymakers/practitioners (*Interaction* model)**

The *Interaction* model predicts that the extent to which users' capacity is accounted for in developing knowledge, and the degree to which knowledge is co-produced by users and researchers, determines research utilisation (Davies et al., 2005; Wandersman, 2003; Weiss, 1979). The supporters of this model can be divided into two camps: those that focus on the capacity of users believe that the research policy gap lies in the differences between the "two communities" of researchers and users; while those that advocate co-production of

knowledge focus on the necessity for sustained and intense interaction between researchers and users (Landry et al., 2001a; Wandersman, 2003).

The evidence on the first factor was mixed. Green argued that research could be more influential if scholars focused on developing good quality processes that practitioners could use to design and implement effective programmes rather than packaged products (cited in Wandersman, 2003). Adams' refusal to recommend particular curricula, preferring to provide policymakers and practitioners with knowledge on which they could base decisions, seems to be more or less in line with this concept (Adams in Adams et al., 1991). Chall's *The Great Debate* (1967) could be read similarly. Therefore, the model would predict that the syntheses of the 1960s would be more successful than the packaged programmes that preceded them; and that there would be greater use of phonics research in the 1980s than in the 1960s as recommended processes became more detailed and were adapted to the division of labour and the capacity of each group of users (Adams, 1990; Anderson et al., 1985). The results did not follow this prediction.

On the other hand, it could be argued that the information that Adams, Chall and others provided was not suited to the capacity of users. It is questionable whether teachers dependent on basal readers for classroom structure, curriculum and teacher training would have had the ability or time to convert the research recommendations into classroom worksheets and exercises (Barton & Wilder, 1964; Venezky, 1987). If this is the case, however, it is not able to explain the differing fates of phonics research in the 1960s and 1980s.

The strongest evidence that bridging capacity gaps is insufficient to determine use was the finding that teachers who participated in *FGRS* were not permitted to perpetuate the new methods after the end of the study, even when the new programme produced better results than the existing basal system (Schantz, 2002). Overall it appears that differences in capacity alone cannot explain the different patterns of use of phonics research during the two time periods.

The second definition of *Interaction* also faced challenges. The level of co-production of knowledge was also higher in *BNR* than in *The Great Debate* or *FGRS*, due to the inclusion of a teacher on the composing panel and policymakers as consultants (Anderson et al., 1985), yet this factor did not appear to influence utilisation. One could contend that in neither period

did phonics researchers demonstrate a high level of interaction with users as recommended by the model. This is particularly clear when compared to the methods employed by the whole language movement. In close alignment with the recommendations of the *Interaction* model, whole language was co-produced by teachers and researchers working closely together, and knowledge produced by teachers working alone was also highly regarded (K. S. Goodman, 1989). Teachers were encouraged through support groups to take responsibility for developing their own knowledge, and the idea of universal applicability of solutions was rejected in favour of localised decision making (K. S. Goodman, 1989; Y. M. Goodman, 1989; Pearson, 1989). On the other hand, because dissemination and interaction processes were conducted simultaneously often through the same fora, it is difficult to separate out the effects of the different variables.

Combining the two aspects of the *Interaction* model could also explain the downfall of the whole language movement. While the linkage mechanisms with users were strong, ultimately it failed to properly account for the capacity of teachers, assuming that they would be able to adopt the principles of the movement and reduce reliance on published materials without compensatory training (McGill-Frazen, 2000; Pearson, 2004). This assumption was erroneous and led to incorrect application of whole language principles, which ultimately felled the movement. It could thus be argued that while interaction may not be critical to the utilisation of research, it can have powerful enhancing effects if conducted well.

Overall, while the *Interaction* model is able to explain the rise and fall of the whole language movement, it is unable to explain the success of phonics research in the 60s in the face of its failure in the 80s.

### **Non-predictive models – Enlightenment, Political and Tactical models**

#### ***Enlightenment model***

The *Enlightenment* model predicts that research will gradually and indirectly influence policy as it permeates society over time and changes relevant conceptions. The lack of specificity in this hypothesis prevents it from predicting where and when utilization will take place based on the presence or absence of different variables. Nonetheless, it is able to explain long term patterns of change.

The effects of *Enlightenment* on policy making can be seen most clearly in the spate of phonics bills passed in the 1990s. No revolutionary new information had been released at this time that can claim responsibility for this shift. Clearly, the crisis of the 1992 NAEP results created a user context that was receptive to the results of research. However, as we saw above, identifying a time in which users are receptive to research does not provide an explanation for which research they will be receptive to. The *Enlightenment* model provides an explanation where the *Demand Pull* model cannot. Kim notes that “despite the difficulty of drawing firm causal links between instruction and achievement” it was the Californian language arts framework and the whole language movement that policymakers blamed for the poor reading results in the state (Kim, 2008, p. 99). Appropriating Weiss’ (1979, 1982) characterisation of how policymakers use research can provide an explanation for this. Policymakers chose to reject the specific findings of phonics research through the 1980s and early 1990s, yet the “concepts and theoretical perspectives” of the research permeated the policy process, affecting how the issues were considered (Weiss, 1979, p. 429). Thus when a crisis arose that suggested that the current system was failing children, latent ideas - the general idea that instructional methodology may determine a child’s reading abilities and the specific idea that the evidence supporting a meaning-emphasis approach was flimsy and poorly substantiated - suddenly gained greater weight, emerged from their dormancy and guided decision making, even when strict causality could not be proven.

The fact that this occurred in this particular case is particularly significant in light of the fact that dissemination of phonics research in the 1980s was relatively poor. Even weak dissemination, it appears, is capable of influencing policymaking as the minority of readers aware of the research pass on generalised findings<sup>19</sup>, and over time (a long time in the case of poor dissemination) ideas penetrate the public consciousness.

Additional evidence for the model was the fact that in all cases, it took at least a decade for a movement to go from conception to implementation (the first group of phonics research accumulated during the 1950s-60s and was implemented in the 70s; whole language accumulated during the 1970s-80s and was implemented 80s-90s; the second phase of phonics research accumulated 1970s-80s and was adopted in the 1990s). It does appear that Chall and Bond and Dykstra’s syntheses are responsible for prompting the changes to policy and practice that were witnessed in the 1970s. However, considering that Chall’s synthesis

---

<sup>19</sup> This links to the political model below

was a re-examination of existing research which had been ignored in the past, and based on Chall's own testimony that research was ignored because it was assumed that teaching methods could not be improved (Chall, 1967); it is fair to assume that the preceding decades of percolating ideas prepared policymakers to be receptive to those syntheses.

A final important lesson from the *Enlightenment* model is its resistance to control. As seen above and in the *Dissemination* model, secondary dissemination is highly important and this is often the main means by which research findings reach users. However, as Weiss (1979, 1982) predicted, this process leaves findings open to distortion and the conclusions that reach users are not always the same as those drawn by researchers. This was seen repeatedly in our cases: the idea that teaching phonics was to the exclusion of reading for meaning; the isolation of phonics lessons from text assignments; the conflation of whole language with a whole classroom approach; and the fact that it was the basal readers that absorbed phonics recommendations in the 1970s even though Chall, Bond and Dykstra had condemned them as a teaching tool (Adams, 1990; Anderson et al., 1985; Juel, 1991; Pearson, 2000, 2004). Nonetheless, while “endarkenment” did occur, over time the overgeneralisations were minimised and practice came to resemble research findings (Kim, 2008).

### ***Political and Tactical models***

Two other models predict that research will be used by policymakers when it is politically expedient to do so. These models cannot predict use on the basis of research form or content, as it will be determined by changing political interests. Though this is perhaps not the type of use desired by researchers, cases played an important role in the creation and dissemination of research.

The *Political* model is often viewed with a great deal of cynicism, as research is used uncritically to justify an existing position. For example, during the reading wars of the 50s and early 60s, research was frequently held up to justify the superiority of a particular reading programme even though there were few trustworthy findings on which to base policy decisions (Barton & Wilder, 1964; Bond, 1967; Levin, 1967; Spache, 1967). Similarly, phonics research was repeatedly used – in teacher training textbooks, by the US Department of Education, in basal readers - to justify existing opinions by taking findings out of context and ignoring qualifications (Shanahan in Adams et al., 1991; Anderson et al., 1985; Chall, 1983). Yet as Weiss (1979) predicts, these readymade advocates played an important role in

heightening public knowledge of the studies, kickstarting *Enlightenment* processes for research with poor direct dissemination.

The research field similarly benefitted from *Tactical* use of research by policymakers. While using research to provide proof of government responsiveness, policymakers also provided funding for major research studies (Nelson & Weinbaum, 2006). In the 1960s and 1980s when poor educational standards were seen as a threat to national security, the federal government provided funding for research which could claim to raise standards (Ansari & Coch, 2006). The studies could not possibly be utilised by those who funded them, as educational policy was controlled by the states. Nonetheless, the tactic of providing funding for research as evidence of federal politicians' interest in the public good benefited researchers and politicians alike.

As mentioned, the 1992 NAEP results created a crisis to which policymakers were forced to respond. The rush to “evidence based” policies (to which phonics research could lay claim) in the 90s and 2000s when criticism was high can also be seen through the prism of the *Tactical* model as research provided the guise of certainty or (at least) rational decision making for threatened policymakers operating in an ambiguous environment (Weiss, 1982).

*Political* and *Tactical* uses of research cannot be predicted on the basis of the research itself. Rather, these uses are dependent on changing political circumstances. While these may not be the types of research use that most scholars seek, they nonetheless are important as they further dissemination of findings, research is granted the mantle of relevancy, and these circumstances can lead to increased funding for academic research.

## Chapter 6: Conclusions

This paper sought to examine the use of research in policy and practice relating to the phonics – whole language ‘pendulum’ of reading instruction in the US in the 1960s and 1980s. Seven models of utilisation were examined – *Science Push*, *Demand Pull*, *Dissemination*, *Interaction*, *Enlightenment*, *Political* and *Tactical* with independent variables drawn from the knowledge utilisation literature. The cases were examined from a historical perspective and rather than investigate individual studies or specific decision makers, the impact of entire fields of research on nationwide policy and practice was assessed. The analysis utilised comparative analysis based on an examination of historical documents from the reading research field. Attempts were made to diversify sources, utilising supporters of phonics and of whole language as well as more independent views.

Comparative analysis of the two periods led to the conclusion that the *Dissemination* model best explains the changes that took place, as well as those that failed to occur. More successful efforts by researchers to disseminate their findings, through professional platforms in particular, led to greater utilisation of research by practitioners and policymakers in the short-to-medium term. Failure to disseminate findings reduced cases of utilisation. Nonetheless, other models were also able to explain the nature of the changes. Higher research quality (*Science Push*) was shown consistently to affect use in the long term, though this use may be significantly delayed if dissemination is insufficient, or if the research does not fit with user interest. Over periods of 10-20 years, processes described in the *Enlightenment* model enabled the integration of research findings into practice and policy as user context and interest shifted. The general implication of the findings for researchers is that investment of resources in the dissemination of findings, particularly to professional networks, is highly important if use by policymakers and practitioners is desired.

### Implications for models and operational variables

#### *Science Push*

In our cases it was found that “research quality” was best assessed by determining the agreement and conclusiveness of findings within a research field. This factor was shown to be more important in determining use than was the “scientific” quality (reliability/validity) of individual studies. This demands a reconceptualisation of “research” in the knowledge utilisation literature from a single study to an accumulation of studies synthesised to represent

the culmination of the field. Convergent findings from multiple studies are needed before a discovery can be deemed “conclusive”. This can take years as studies frequently contradict one another and disputes must be resolved. Yet, without a level of consensus it is difficult to expect that policymakers will be able to draw recommendations from research studies.

Research quality, even when defined as consensus in a field, while important, was not sufficient to determine use. Furthermore, while the variable was strong, the model was not. The linear progression of research to use posited by the model was shown to be overly simplistic. This oversimplification has led to its rejection by knowledge utilisation theorists even as it is championed by advocates in education (Carnine, 2000; Landry et al., 2001a; Weiss, 1979). Nonetheless, if we retain the variable, but refine the model by adding dissemination efforts after pilot studies are designed and conducted (direct and secondary dissemination, dissemination by policy advocates) and lengthen the arrows to allow for more time to pass, then the model holds in the two cases under study.

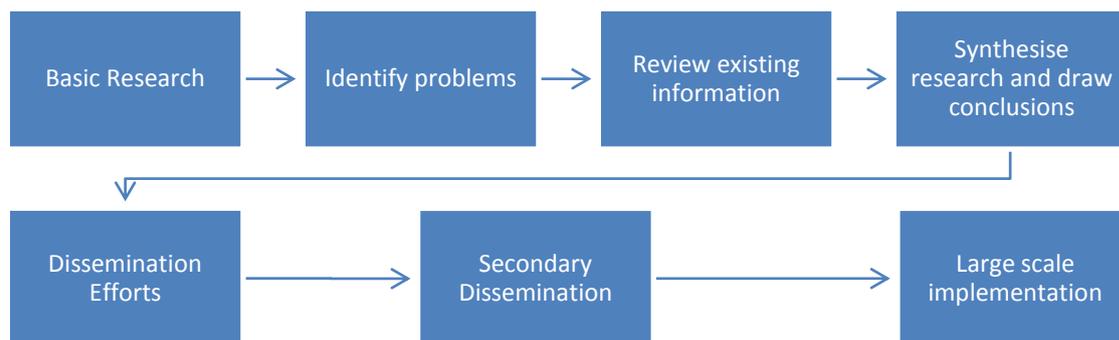


Figure 6. A modified *Science Push Model*

One reason for the problems of the model was the overly literal imitation of research progression from “the medical model” (e.g. Carnine, 2000). Even in the field of health, while research may occasionally provide “magic bullets” in the form of a drug or vaccine mass produced by a pharmaceutical company; more commonly successful “evidence-based treatment strategies” also require a trained practitioner to implement the treatment appropriately and effectively (Wandersman, 2003, p. 234). Convincing those practitioners to utilise the approved strategies, and training them to do so, requires additional steps not included in the original model.

One case in which the small studies to large scale replication model fit rather well was that of the basal reading programmes. Like prescription drugs, the programmes were designed to be

duplicated and distributed in the form of textbooks, worksheets and other materials. Individual differences and the possibility of error was minimised as much as possible by providing teachers with scripts for instruction to accompany the readers (Venezky, 1987). It is interesting to note that in both cases it is a commercial interest – pharmaceutical companies, basal publishers – that fit the original model, while policymakers and practitioners cannot be so easily manipulated.

It is important to note that the *Science Push* model failed the comparative test. It does not explain the success of the whole language movement or its later demise<sup>20</sup>. This suggests that *Science Push* is not the primary determinant of research use. More cases are needed to test its veracity.

### ***Demand Pull***

A key problem of the *Demand Pull* model is that a measurable variable to test the model has not been defined consistently by knowledge utilisation scholars. Different definitions have been proposed, but they produce divergent results. As we see in table 2, studies frequently met one definition but not another, and the different definitions did not appear to have similar patterns of variance.

		1960s	1980s
Advancing a public benefit (applied research)		✓	✗
Focused on user needs		✓	✓
Projects externally funded/externally initiated		✗	✓
User context	User interest	✓	✓ federal ✗ teachers
	Timing	N/A	✓ ('89-'93 adoptions)
	Credibility	✓	✓ authors ✗ study selection
	Specificity of target audience	✗	✗

*Table 2.* Presence of the relevance variable in the 1960s and 1980s according to various definitions of relevance.

<sup>20</sup> This will be discussed further in the discussion of the *Dissemination* model

The definition of the variable with the greatest explanatory power was user interest (user context). In both periods it was found that phonics research was more likely to be utilised when users were interested in what research had to say, usually driven by external demands for improved policy/practice. The unfavourable consequence of this reality was described by Weiss to discourage expectations raised by the *Science Push* model:

“unless a social condition has been consensually defined as a pressing social problem, and unless the condition has become fully politicized and debated, and the parameters of potential action agreed upon, there is little likelihood that policy-making bodies will be receptive to the results of social science research” (Weiss, 1979, p. 427).

But the reverse seems also to be true. Once an issue has been consensually defined as a pressing social problem, politicised and debated, and parameters for action set, there is a high likelihood that research will be used in decision making. If we only view policy making in the short term, then this scenario is certainly as rare as Weiss portrays. An historical view however, reveals that it does occur with some regularity, though the times that it occurs on a particular issue may be spaced decades apart. Oftentimes, this confluence of events occurs after a sudden feeling of crisis amongst the public. Classic crises from our cases – Sputnik followed by *Why Johnny Can't Read*; the portrayal of the education system in *A Nation at Risk*; the 1992 NAEP results – sparked public demand for research to provide better solutions (Chall, 1967; Flesch, 1955; Gardner, 1983; Kim, 2008; Nelson & Weinbaum, 2006). In such times of crisis, driven by forces independent of research and policy, the public turns to research for answers. This finding is not dissimilar to the idea of a policy window proposed by Kingdon (1995).

These crises created a “user context” that was receptive to researcher input. This definition of context is perhaps narrower than the one proposed by the *Demand Pull* model. Here context translates to circumstances when users are pressured to make a decision and feel the need for external authority to make or justify that decision (similar to the desire for rationality described above). The fact that research timing did not appear to affect utilisation seems to confirm Weiss (1982) and Birnbaum’s (2000) argument that policymakers are often influenced by yesterday’s ideas. Research need not fit the policy cycle exactly; good relevant research can be used at any time, and will only be used when the context is right for users. To be able to predict when this will occur, or respond when it does, is unnecessary and is often detrimental to research quality (Birnbaum, 2000; Chall, 1967).

If we accept that users will be receptive to research only at certain times, the question still remains: which research will they be most receptive to?

The traditional basic/applied definition, which proposes that applied research is more likely to be used than basic research, did not explain patterns of use. Based on the evidence, this can perhaps be explained by the overly binary distinction between basic and applied research. The principle behind the basic/applied dichotomy is that users will likely not be able to draw practical conclusions from even very well established and consensual knowledge building findings. It is reasonable to conclude that users cannot be expected to understand and create interventions on the basis of highly technical and narrow academic articles. However, the idea that in order to answer “real problems” researchers must conduct research that examines those questions directly, rather than try to develop a knowledge base, produced poorer quality research studies (Chall, 1967). Rather, researchers were able to draw more detailed and concrete recommendations for instructions on the basis of research conducted into the cognitive and perceptual processes involved in reading (basic research) than they had been able to when instruction was studied directly (applied) (Adams, 1990; Alexander & Fox, 2004; Chall, 1996; Juel, 1991; Pearson, 2004). The reading research field seemed to support Bush and Birnbaum’s calls for the preservation of independent scholarship in order to maintain high scientific standards and the knowledge base on which policy making is based (Birnbaum, 2000; Bush, 1945). That is, to address real problems, you need a strong knowledge base, and if you have a strong knowledge base, you are likely to be able to address real problems.

At the same time, studies were not conducted purely to develop a general knowledge base, but rather aimed to test a theory of reading acquisition directly relevant to reading instruction (Juel, 1991). Thus perhaps a better distinction of research types would be that proposed by Stokes: pure basic; pure applied; and Pasteur’s quadrant of “use-inspired basic research” – research that is inspired both by a “quest for fundamental understanding” and by “considerations of use” (cited in Furlong & Oancea, 2005, p. 7). Though this factor was not strong enough to override other variables in determining use, it does appear that research will be most relevant to users when it fits into this latter category.

The “importance of funding” criterion of relevance was also not able to explain findings. I was unable to use the source of funding for research to measure interest in the outside world as the periods did not differ significantly on this measure. For this reason, a different measure

for the same concept was used: whether the problem that prompted the studies was identified by users or by researchers. Yet again the differences were unable to explain utilisation results. Nonetheless, while the question of who initially identified the problem or instigated the research was not consequential, whether users were *also* convinced that the problem existed was important. During a period when there was a general feeling that everything there was to know had already been discovered, research was unable to penetrate public decision making. Once the public was convinced that there was a question to be answered, they became interested in hearing what researchers had to say (Chall, 1967). This factor resembles the user interest variable mentioned above, and should perhaps be considered as an additional measure for that variable.

The other aspects of the relevance variable: timing, user credibility, specificity of target group and focus on user needs, were not able to contribute to our understanding of why research was or was not used in the cases examined. This may be a result of a critical flaw in the *Demand Pull* model. The *Demand Pull* models assumes that a specified group of authorised decision makers can be identified who are responsible for the policy or practice that researchers wish to influence. Yet, it was found that forces outside of that group were frequently the most influential in decisions made (Institute for Educational Development, 1969). Furthermore, discrete groups of policymakers with the same responsibilities – individual school boards for example – tended to move in unison. While these groups of decision makers may be technically independent, the fact that we were able to see sweeping changes across the nation during periods of change suggests that their independence of thought is limited.

The *Demand Pull* model was able to connect heightened interest in research findings to increased utilisation, however aspects of the variable that attempted to predict what research would be considered “relevant” at such times were unable to explain the differing results of phonics research in the two periods.

### ***Dissemination***

The *Dissemination* variable proposed had two main components. The format of the research – its language, location, design, etc. – does not seem to have been of particular importance (though it is possible that all the syntheses met a minimum threshold of accessibility above which difference was not important). The second element – dissemination efforts – was shown to be of great importance in determining use. The dissemination activities with the

greatest differentiating effects were those over which researchers did not have full control: dissemination through professional platforms<sup>21</sup>. Findings that were disseminated via popular and professional mediums such as TV programmes, teachers' networks and publications, mass media, etc., had substantially higher utilisation in policy than findings which were only disseminated as research publications. Because users were already engaged regularly with these platforms, they did not need to make a special effort to locate relevant research. Most likely, communication of research through these fora also granted legitimacy to the research due to its association with a source considered credible. While the authors of *BNR* may have been highly regarded in the research community, and by those commissioning the study, when it was rejected by the National Council of Teachers, an organisation with which teachers had a longer association, its credibility amongst teachers was most likely damaged. The role played by secondary dissemination could also explain why the qualifications to the findings were often disregarded even when the main findings were accepted; the passage of information through multiple players functioned like a game of broken telephone where details are easily lost. This finding also explains why format was not found to be important – most users are not actually reading the research but rather receiving it second hand.

While the model was correct in its prediction that increasing dissemination of findings will increase use, it did not account for an important modifying variable in this equation – the influence of competition. As mentioned, dissemination through popular and professional platforms was shown to be particularly important due to the ready access to users and conferred legitimacy. However, the number of these platforms is limited, and the time that can be dedicated to research questions even more so. Moreover, it appears that in both periods, professional organisations in the field of reading instruction seemed to “take a side” in the debate. Researchers' access to users via these platforms was therefore oftentimes determined by the congruence between their findings and the position taken by the supervising organisation.

---

<sup>21</sup> An interesting finding is that dissemination was often conducted by individuals or groups not affiliated in any way with the research. Basal publishers integrated principles of the phonics books into textbooks in the 1970s and of whole language in the 1980s. Both movements explicitly opposed the reader system (K. S. Goodman, Shannon, Freeman, & Murphy, 1988; Pearson, 2000, 2004) yet this system became an important disseminator of (some of) both groups' findings. For researchers, the problem of dissemination by groups with whom they have no affiliation is that many fallacies are adopted along with some findings. It would be interesting to pursue the effects of attempts to work together with such interests.

An additional adjustment to the model recommended on the basis of the evidence is the addition of an evaluation block by users between receipt and acceptance of research. On the basis of their own knowledge, experience, logic and interests and with the addition of feedback from the results of implementation over time, users evaluate information before determining whether it should inform their decision making (on this matter, see Rich, 1997). Evidence of this was seen in the lack of implementation after the publication of *Why Johnny Can't Read* despite its wide distribution. While the book may have created debate in the 1950s and early 1960s, it was apparently insufficient to warrant large scale change to policy or practice. Similarly, while whole language dominated teacher practice in the 1980s, many teachers, though aware of the movement, rejected its findings and continued with a phonics approach (Pearson, 2000). In circles where whole language was widely accepted, feedback that suggested that the approach was producing poor results led to the re-evaluation of previously accepted concepts and rejection of the approach (Kim, 2008). Injection of an evaluation by users box can also explain why a variety of users in the 1970s chose to adopt some recommendations made by Chall while rejecting others even when they were aware of both.

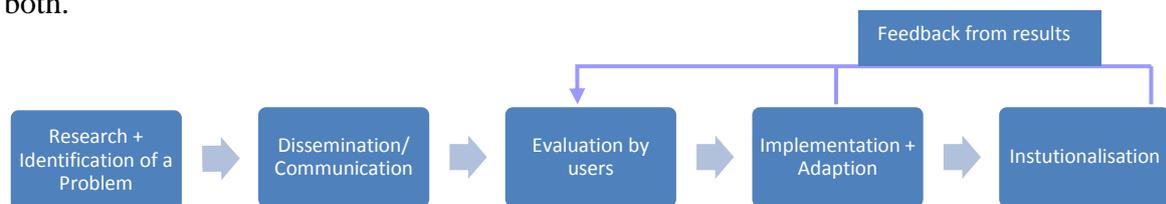


Figure 7. A modified *Dissemination* Model

### ***Interaction***

It is difficult to draw general conclusions about the *Interaction* model based on the cases of phonics research as little relevant evidence was found in either period. Nonetheless, some limited conclusions about the model can be drawn.

The hypothesised capacity gap between the two communities clearly existed and formed a barrier to implementation, yet it could not alone explain differing levels of utilisation. Often, when users lacked the capacity to incorporate research knowledge into their practice (policy *products*), the knowledge was instead incorporated into decision making *processes*. For example, phonics and whole language theorists were unable to overcome teachers' reliance on basal readers. Nonetheless, principles such as the importance of teaching code or of

reading meaning were incorporated into the process of selecting curricula. The problem for researchers was that this led in many cases to the misapplication of recommendations in the final policy products. Thus, while the capacity gap is not a barrier to use, it raises the likelihood of incorrect or insufficient application of research findings.

Linkage mechanisms between researchers and users appeared to have greater potential to impact utilisation. Despite failing the tests of agreement and difference for our cases, it does appear that “sustained and intense interaction” between users and researchers has the potential to substantially raise utilisation. Token inclusion of users in the process of creating knowledge (e.g. one teacher on the *BNR* panel) did not appear to satisfy the criteria of co-production. On the other hand, the popularity of the whole language movement amongst teachers is testament to the power of true partnership. Nonetheless, there is something to be said for the fact that this knowledge was of poorer quality than that produced by researchers. Overall, it appears that while the absence of interaction will not prevent use, the presence of interaction can raise increase the likelihood that research is used, and used as recommended.

Previously it was mentioned that there are two camps within the *Interaction* model – the “two communities” and the “sustained and intense interaction” camps. Combining these two elements may have greater explanatory power than either alone. Teachers and other users were not necessarily able to maintain high scientific standards in the production of knowledge. Researchers expecting that users will convert their findings to policy products must recognise this limitation and make efforts to build this aspect of user capacity if they wish to increase utilisation. At the same time, the inclusion of users to provide context and knowledge of implementation is highly important, as is the “buy-in” that results from being included as equals in knowledge production. There is a component of ideology in the balance between these two sometimes conflicting elements, and researchers (and users) may differ in their preferences. Examination of additional cases with strong *Interaction* elements is necessary to clarify this variable further.

### ***Enlightenment***

The *Enlightenment* model was not initially considered as a hypothesis because it does not make specific predictions about when use will take place, rather seeks to explain changes that take place after the fact. Nonetheless, it was correct in its supposition that research would over time permeate policy making and that research could change concepts that guide decision making.

As predicted, the diffuse nature of this process leads to many misconceptions, and invalid generalisations were dispensed alongside valid ones (Weiss, 1979). The pattern by which this occurred was consistent. General principles derived from research findings would be dispersed, but during circulation, details and qualifications would be omitted, similar to the process of secondary dissemination discussed above. As “enlightenment” is even more diffuse of a process, greater misunderstandings are to be expected. Nevertheless, over the course of decades many of the invalid lessons were weeded out and better quality research prevailed.

The idea that it is almost impossible to affect the utilisation of research as *Enlightenment* processes cannot be controlled (Birnbaum, 2000), appears to be overly pessimistic. As mentioned, *dissemination* was a powerful factor in the short term and *interaction* can also be used to raise the accuracy of findings communicated. Furthermore, higher *quality* research seems to prevail in the long term, as feedback loops report on the results of poor research. The *enlightenment* process seems to be occurring continuously, gradually displacing short term changes as concepts are adjusted. For proponents of neglected research, and advocates of use of research in policy, this may be heartening, if frustrating in its slowness.

### ***Tactical and Political***

The *Tactical* and *Political* models expanded our consideration of the function of research. Beyond its value in creating knowledge - the primary function of research from a scientist’s perspective - research, independent of its content, is also able to provide the appearance of rational choice, critical in modern thought (Weiss, 1982). In periods of uncertainty, scientifically sound research was able to provide assurance to users – and through them to the public - that they were making the “right” choice.

The clear lesson of the *tactical* and *political* models was that there are benefits to these types of research use, even if they distort the original purpose of the research. These types of use are important to enhance the status and funding for research. At a time when the relevance of research to policy is being questioned, threatening government funding for research, this type of use is particularly important (Davies et al., 2005; Kaestle, 1993). It is also a way that dissemination of research can be furthered, through well-placed advocates.

### Summary of findings from the knowledge utilisation models

Based on the findings, it appears that in these cases *dissemination* was the most important variable in determining utilisation. More effective *dissemination* strategies border on *interaction*, engaging directly with users primarily through professional platforms. *Dissemination* did not function alone, as acceptance of findings was dependent on evaluation by users. While partially dependent on the individual experience of the user, the criteria by which users evaluated research was variable, and could sometimes be manipulated. Higher quality, accessible and more relevant research is preferred by users, but will only be considered at times when the research meets their interests. Through *enlightenment* processes, research is able to shape concepts and ideas that change user context and increase user receptiveness to research, which in turn increases use. Alternatively, *interaction* with users can increase buy-in and credibility which again makes users more receptive to findings. Over time, results of implementation feed back to users, potentially altering their evaluation of accepted research. As better quality research tends to produce better results, high quality neglected research can prevail in the long term<sup>22</sup>.

Political crises can also effect changes to user context and their receptiveness to research. In our cases, strong public interest in evidence-based policies led policymakers and practitioners to seek out and incorporate high quality research into their processes and products. In the 1960s, the media furore around the phonics-whole word debate alongside the competition with Russia drove publishers and teachers to seek out proven solutions. Similarly, in the 1990s, panic over the NAEP results led state policymakers, by that time the dominant influence, to existing phonics research. The research was used when it was deemed politically expedient and tactically correct to do so (*Political/Tactical*). Yet, as above, there was also a need for users to know of the existence of the research (*Dissemination*), and to have already considered it as a viable alternative (*Enlightenment*). Higher quality research was more likely to stand the test of time (*Science Push*).

Overall, in the short term, use is most likely to occur when dissemination is high and user context is receptive/user interest is high. In the long term, however, high quality research can permeate decision making and alter user context, leading to changes to policy products and processes.

---

<sup>22</sup> It should be admitted that this could also be explained by the tendency toward “pendulum swings” between approaches (Slavin cited in S. A. Stahl, 1990). A longer timeline would be needed to assess this hypothesis.

## **Implications of findings for the definition of use**

There was high correlation between the different definitions of use – higher *product* use was associated with higher *process* and higher *conceptual* use. Nonetheless, there were substantial differences between the three perspectives so as to justify them being considered separately.

The main criticisms that have been launched at the *product* perspective are that it is too narrow and that *product* type uses rarely occur (Birnbaum, 2000). The results concur strongly with the first criticism. A purely *product* perspective would have neglected many important cases of use such as the rising importance of instructional methodology in decision making processes, increased funding for research, and changing conceptions of reading instruction. Furthermore, a pure application of the *product* definition – that findings must be understood correctly and implemented as advised by researchers – would have eliminated virtually all the product uses that were found. Nonetheless, instrumental uses of research in policy products were not as rare as had been supposed. Broad changes to the criteria for selection set by textbook selection committees followed by sweeping changes to basal readers and classroom practice were found in both periods. The *product* definition should not be applied to the exclusion of other types of use, and it should be expected that application will almost only even be partial; however it is an important aspect of use to which researchers can aspire.

The *process* perspective – another instrumental type of use – has suffered from similar criticisms to those launched at the *product* perspective. It is an important addition to the *product* perspective especially in cases such as this one where commercial interests (basal publishers) and practitioners (teachers and principals) dominate construction of policy products. In these cases, formal policymakers play a greater role in determining processes than they do products. Use of research in policy processes could be seen among more diverse groups of policymakers than could research use in policy products. Research use was evidenced in the design of teacher training, in federal funding for research and in textbook selection. It is likely that research use was evident in processes that I was unable to examine due to the choice of methodology.

The *process* perspective was able to shorten the timeline for research influence. Changes to basal readers were witnessed only in the 1970s. Calls first by the public and then by selection committees for textbooks to use evidence-based instructional methodology pre-dated these changes by as much as a decade and a half.

The *process* perspective also provided unique insights as to why research may be considered and rejected by decision makers. The assumption for many scholars has been that rejection indicates a weakness in the research. Examination of the decision making process through citations of research in teacher training textbooks reveals diverse reasons why a particular user may accept or reject findings. As mentioned, most teacher training textbooks published between 1972 and 1978 cited *The Great Debate*, *FGRS*, or both, when justifying the instructional methodology chosen (Chall, 1983). Citations of *The Great Debate* and *FGRS* were more frequent in meaning-emphasis textbooks than in code-emphasis ones, suggesting that research was being used to defend a decision more often than it was being used to make one. Misinterpretation of findings suggests failures in transmission or cognition of the research. Citations of minor aspects of the research to justify choices while larger recommendations are ignored suggests deliberate selectivity in accepting the findings, likely in accordance with pre-existing positions. A *product* perspective would simply have noted the shift in textbook emphasis (from 82% meaning-emphasis and 0% code-emphasis in books published 1955-1965 to 31% and 31% respectively in books published 1972-1978) and perhaps questioned why the shift was not greater. A *process* perspective reveals that research was being acknowledged and used in the process of creating almost all of the textbooks, even if the final product did not reflect this consideration. The *process* perspective is thus able to provide insight into the research-policy gap that the *product* perspective overlooks. Users often choose to reject research conclusions for (from their perspective) rational reasons. Those reasons may or may not be legitimate from an outside perspective, but they are not based in ignorance of the research as has been supposed.

It could be argued that this is *symbolic* use of research as it involves research being held up to legitimise held views. It is an important and interesting finding that research is used symbolically both by supporters of the research (such as code-emphasis textbooks in the 1970s and the Secretary of Education in the 1980s) and by its opponents (meaning-emphasis textbooks in the 1970s and arguably the National Council of Teachers in the 1980s). *Symbolic* use should perhaps be considered as a subset of *process* use as *symbolic* use involves the apparent inclusion of research in the decision making process without any actual changes to the content of the decision (*product*) or the way the issue is conceived (*concept*).

Whether *symbolic* use is considered legitimate probably depends on which side of the debate you stand, though *symbolic* use may be important for all researchers. Even if policymakers do

not intend to listen carefully to the findings of research (or intend to, but fail to do so), calls for policy to be evidence-based translate into a higher profile for research and a larger pool of funding. Partisan federal support for phonics in the 1980s provided funding for *Beginning to Read*, one of the most impressive and thorough research studies of the period.

The *concept* perspective provided in some way the most meaningful cases of research use, as research began to influence the way that issues were perceived and discussed. The creation of public debates and changes to the nature of those debates sparked public interest that led to an increase in demand for research findings. Changes were made to the terminology used to discuss reading acquisition in practitioner journals, and changes to the type of research that was respected by those journals. These changes reflect alterations in the way that reading was conceived by practitioners, a necessary precursor to transforming attitudes to the issue. The “intellectual usefulness” of the studies – their contribution to the dynamics of problems – was indeed meaningful (Birnbaum, 2000) as these changes preceded substantial changes to policy processes and products. While a *product* perspective sees only the failure of phonics research in the 1980s, a *concept* perspective sees its success in creating new interpretations of facts so that for example when reading levels were seen to have dropped in 1992, whole language was immediately blamed.

This type of use is important within a context in which debates were firmly partisan and it often appeared that research was only being used *symbolically* to reinforce an existing position. For researchers frustrated by partisan use or misuse of their findings, the *conceptual* perspective is an important balm. While today findings may be manipulated, used or ignored to fulfil an agenda, the ideas contained therein will gradually permeate thinking and in the long term be legitimately considered during decision making.

A problem that was found with *process*, *symbolic*, and *conceptual* uses was that they leave less clear record than *product* uses. Decision makers do not always record the process of their decision making, so a research method reliant on historical sources is naturally limited in its ability to assess the use of research in policy processes. Direct questioning of decision makers can provide greater illumination in this regard, though this method has its own limitations aforementioned.

It is often difficult to identify when concepts have changed, or to attribute those changes to a particular source, particularly as individuals are often unaware of the ideas, theories and

interpretations of facts that govern their thought processes. Citation analysis, the traditional method, would reveal only a sliver of these changes. It is easy to understand in this context why *product* uses are the preferred object of study and goal for researchers. It is likely that in this study many cases of *process*, *symbolic*, and *conceptual* uses have been missed.

Examination of the types and timing of uses identified by each of the different perspectives provides a new profile for how change to policy occurs. It has been assumed by many that if changes are made to the way researchers act – better quality, more relevant research; greater efforts to disseminate findings; greater interaction between researchers and users - then recommendations will be adopted. But these results seem to suggest that product changes can only occur when processes and concepts support that product. *Conceptual*, *process* and *product* uses may therefore be viewed as a progressive hierarchy of change, or as stepping blocks of utilisation. First concepts must change so that users are willing (and able) to consider the ideas proposed by research. Then these concepts are integrated into decision making processes, their status raised so that a new factor is considered as part of decision making. Only once these two changes have taken place, changes which may take years, is the scene set for product changes to take place. If concepts and processes are already in alignment with the findings, then adoption is more likely to occur and change to products can be faster. The changes that will take place in such a situation will also be less revolutionary because they will only be within the confines of existing concepts and processes.

This proposal is perhaps a new formation of what has been termed “user context”. As we saw, the most important factor in user context was user interest. User interest is engaged when users see an issue as a problem and believe that research can offer solutions. For a non-problem to be considered a problem to which research can provide solutions requires a change in the concepts surrounding that issue. The difference between the traditional conception of user context and this proposal is that traditionally the assumption is that user context is independently determined and researchers should adapt and direct their products to meet that context. Here, the proposal argues that user context is dynamic and receptive to change. Researchers can change user context to one that understands the world “better” and hence is receptive to what scholars have to say. Changing user context certainly takes longer and is more difficult, but it is also more revolutionary. It also neutralises the concern that researchers or research quality will be compromised by trying to fit with changing user interests.

## Strengths and weakness of the method

The methods selected for the collection of data and analysis of findings have substantial flaws. Nonetheless, they were able to provide a unique perspective to the issues that could not have been accomplished through traditional knowledge utilisation methodology.

### *Strengths*

There are a few major problems with knowledge utilisation's traditional empirical methods of inquiry that rely on surveys of decision makers or researchers or citation analysis. The first is that decision makers are oftentimes "unaware of when and how they use research" (Weiss, 1982, p. 623). Decision makers integrate knowledge acquired from academic sources into their general store of knowledge; asking users to report their use of research would lead to severe underreporting of cases (ibid.). In the case of *conceptual* use, users may not even be aware that their thoughts are being shaped by new concepts.

The second problem is that users often do not receive research ideas directly from the source. Secondary *dissemination* and *enlightenment* enables users to absorb research knowledge via colleagues, the media, or other non-academic sources, as they likely will not be aware of the source of these ideas (Weiss, 1982). Researchers, if they had specifically targeted a group of decision makers in town X, would have been completely unaware of the effects they had had on decision makers in towns Y and Z. Decision makers in town Z would not be able to cite the influences on their decision, even if they replicated research recommendations precisely. Again, asking users when they use research would be misleading.

The follow-on effects of *dissemination* and *enlightenment* were not limited to new geographical locations; "authorised decision makers" represent only a fraction of the individuals and groups influencing reading instruction policy and practice who were absorbing research findings. Research that originally aimed to assist schools in the process of selecting a commercial reading programme for classrooms ended up influencing teacher training, federal funding, practitioner journals, local and state educational boards, and others. Searching for use only among targeted users would ignore many utilisation cases.

A third problem is that "few decisions in government are made by a single decision maker or even a small group of decision makers, and almost no decisions of sufficient scope to qualify for the category of policy" (Weiss, 1982, p. 622). For scholars seeking to understand policy

formation, therefore, studying individual decisions is irrelevant. This was clear in our cases. Practice was so close to uniform across the country that it could be considered almost as one policy even though it was actually comprised of thousands of separate decisions (Barton & Wilder, 1964; Farr et al., 1987). This was often because policymakers were influenced by decisions made by their peers in the neighbouring town (or neighbouring state) (Farr et al., 1987). Focusing on only one set of decision makers would have neglected the influence of hundreds or thousands of others.

A fourth problem exists on the other side of the equation: the characterisation of research. If “research” is defined as the consensus of a field rather than single studies, then utilisation is dependent on a research field rather than on individual researchers, and actions undertaken by a single researcher cannot be meaningful. Studies which are designed to assess the effects of one particular researcher or one particular study on policymakers and practitioners will therefore naturally find very few cases of utilisation.

The strength of the method used is that by tracking the passage of an idea over a broad time frame and geographical reach, we are able to track research knowledge even when participants are not conscious of its passage. An historical view enables us to follow research knowledge even when it changes hands and form and is converted say from “intellectual knowledge” to “practical knowledge”, “small talk” or “unwanted knowledge” (Machlup cited in Rich, 1997, p. 14). One user, reading a research study, may initially consider the knowledge intellectually, after which they convert the information to practical knowledge that can be applied in their practice. When, in that form, it is passed on to a colleague, it is no longer conceived as a research product though the influence of research is critical. Studying citations or asking users would not reveal this passage of research. As outsiders witnessing changes, we are able to hypothesise the dissipated effects of research concepts. Through an historical perspective, we are able to witness changes to concepts that may not be visible to those immersed in them. By refusing to limit ourselves to a particular target audience, we are able to see the multiplicity of ways in which policy is formed. If “[t]he diffuse process of research use that we are calling “enlightenment” is highly compatible with the diffuse processes of policy making” (Weiss, 1982, p. 635), then the diffuse perspective provided by historical analysis is compatible with both of these, capable of measuring changes that cannot be seen from up close.

Another advantage of the methodology is that it enables us to witness the long term effects of research. As was seen, research can have both short term and long term effects, and different processes govern these two routes to influence which can only be seen from the purview of history.

The analysis methodology is also uncommon in knowledge utilisation literature. One problem with analysis that uses the input/output model is that often the passage of information cannot be traced or isolated from other inputs (Rich, 1997). Comparative analysis attempts to bypass this problem by not trying to follow the routes of influence, but rather assuming that if similar cases have only one causal variable in common, then that variable is probably responsible for producing the result under examination (Ragin, 1989). The comparative method, though flawed, enables us to connect changes to the variables that preceded them with an estimation of causation without having to map the passage of information, which in cases of *enlightenment* at least, is an impossible task.

### ***Weaknesses***

A significant limitation of this research was the fact that this thesis did not have the capacity to study the processes by which individual groups of policymakers made their decisions. This meant that to a large extent, *process* uses were neglected in the study. Because of this, and due to a method similar to social framework analysis, the methodology “relies heavily on reason and judgment” (Rich, 1997, p. 21). Assumptions as to reasoning were made on the basis of changes witnessed. Efforts have been made to base all interpretations in literature but reason and judgement were necessarily major components of both data collection and analysis.

As far as was possible, efforts were made to use both methods of agreement and of disagreement. This necessitated consideration of the two cases under study as multiple cases based on different instances of use (Faure, 1994), and on occasion to utilise pre-60s phonics and 1980s whole language as complementary cases, even though those cases were not examined to the same extent.

Also important, and a necessary feature of comparative approach (Faure, 1994), is that the thesis has only considered hypotheses related to research, derived from knowledge utilisation theory. As has already been noted in discussions of user context, factors entirely independent of the academic world can influence policy changes. Because the changes were quite specific

and did cite research findings, it is reasonable to assume that research played a significant causative role and that characteristics related to it may have been significant. Nonetheless, it is possible that a third or fourth factor external to both policy and research caused the changes and the relationship between research and policy was one of correlation rather than causation. More likely, again as was raised in discussions of user context, there was an interaction between research and other issues not under examination. Even among the variables being examined, the comparative method is hampered by multiple causality (Ragin, 1989). Multiple causality has been hypothesised, but this necessitated analysis outside strict application of the method.

**To sum up**, this thesis began with the claims that there is a substantial gap between research and policy, caused perhaps by policymakers not listening or researchers producing irrelevant research. This examination has shown that neither of these claims are entirely true. Policymakers were not listening at all times, but at important crossroads, research certainly got a hearing. Research takes time to accumulate and produce applicable recommendations, but continuous efforts were made to make research relevant to policy and practice. The gap was not as wide as it sometimes appears.

With that, it may be possible to close the gap further. In most cases, individual studies cannot hope to influence policy because alone they do not provide the scope or certainty necessary to form a justifiable basis for policy. Researchers working together, however, can make efforts to close the gap, by directing dissemination efforts to professional platforms or other platforms where policymakers and practitioners are already engaged. Retaining a central commitment to high scientific quality means that the recommendations that they are making will produce better results that will hold out in implementation. Greater efforts should also be made to work with implementers to build their knowledge and capacity to implement recommendations as recommended. Changes to policy require a conceptual environment and processes that support those policies and these must be changed before products can follow. Researchers must respect the fact that at times it may be rational for policymakers to reject research due to other considerations.

An important lesson is that many research findings take decades to be absorbed by the public and policymakers. I believe that it is the tendency to view research use in the short term only that has led to the negative conclusion that studies rarely leave “any discernible mark on the direction or substance of policy” (Weiss, 1979, p. 428) and the view of the research-policy

gap. It is true that individual studies will rarely have the immediate type of impact that has been hoped by many. However, collectively and over time, research can markedly alter both the direction and the substance of policy.

The implication for scholars seeking to examine how research knowledge has been utilised by practitioners and policymakers is that a broader view is required. Considering the limitations of this method as well as the more common survey and citation methods, it may be most effective to combine approaches and triangulate findings (Hanney, Gonzalez-Block, Buxton, & Kogan, 2003). These two case studies are too narrow to have real theoretical implications for the proposed models. Nonetheless, adaptations to the models have been proposed that may benefit from further testing.

The field of knowledge utilisation has primarily focused on public health. It may be time for education scholars to engage with this field to develop greater understanding of how knowledge is utilised by the diverse practitioners and policymakers engaged in education.

## Bibliography

- Aaron, I. E., Chall, J. S., Durkin, D., Goodman, K., & Strickland, D. S. (1990). The past, present, and future of literacy education: Comments from a panel of distinguished educators, Part II. *Reading Teacher*, 43(6), 370–380. doi:Article
- Adams, M. J. (1990). *Beginning to Read: Thinking and Learning about Print*. Massachusetts: The MIT Press.
- Adams, M. J., Allington, R. L., Chaney, J. H., Goodman, Y. M., Kapinus, B. A., McGee, L. M., ... Williams, J. P. (1991). Beginning to Read: A Critique by Literacy Professionals and a Response by Marilyn Jager Adams. *The Reading Teacher*, 44(6), 370–395.
- Alexander, P. A., & Fox, E. (2004). A Historical Perspective on Reading Research and Practice. *Theoretical Models and Processes of Reading*, 5, 33–68.
- American Physical Society council. (1999). What is Science. *American Physical Society*. Retrieved March 11, 2013, from [http://www.aps.org/policy/statements/99\\_6.cfm](http://www.aps.org/policy/statements/99_6.cfm)
- Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a Nation of Readers: The Report of the Commission on Reading*. (p. 155). IL: University of Illinois. Retrieved from [http://eric.ed.gov/ERICWebPortal/search/detailmini.jsp?\\_nfpb=true&\\_ERICExtSearch\\_SearchValue\\_0=ED253865&ERICExtSearch\\_SearchType\\_0=no&accno=ED253865](http://eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED253865&ERICExtSearch_SearchType_0=no&accno=ED253865)
- Ansari, D., & Coch, D. (2006). Bridges over troubled waters: education and cognitive neuroscience. *Trends in Cognitive Sciences*, 10(4), 146–151. doi:10.1016/j.tics.2006.02.007
- Barr, R. (1984). Beginning Reading Instruction: From Debate to Reformation. In P. D. Pearson (Ed.), *Handbook of Reading Research* (Vol. I, pp. 545–581). New York: Longman Inc.
- Barton, A. H., & Wilder, D. E. (1964). Research and practice in the teaching of reading: a progress report. In M. B. Miles (Ed.), *Innovation in Education*. New York: Bureau of Publications, Teachers College Columbia University.
- Battista, C. (1990, April 1). Teacher Ignites Lifetime Love for Words. *New York Times*. Retrieved June 4, 2012, from <http://www.nytimes.com/1990/04/01/nyregion/teacher-ignites-lifetime-love-for-words.html?pagewanted=all&src=pm>
- Birnbaum, R. (2000). Policy Scholars Are from Venus; Policy Makers Are from Mars. *The Review of Higher Education*, 23(2), 119–132.
- Bond, G. L. (1967). First-grade reading studies: an overview. In J. L. Frost (Ed.), *Issues and Innovations in the Teaching of Reading* (pp. 335–343). Glenview III: Scott, Foresman.
- Bond, G. L., & Dykstra, R. (1967). The Cooperative Research Program in First-Grade Reading Instruction. *Reading Research Quarterly*, 2(4), 5–142.
- Botel, M. (1985). Book Commentary; Becoming a Nation of Readers: A Review. *The Reading Teacher*, 39(3), 260–262.

Bruer, J. T. (2002). Avoiding the pediatrician's error: how neuroscientists can help educators (and themselves). *Nature Neuroscience*.

Bush, V. (1945). *Science the Endless Frontier: A Report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development*. Washington, DC: U.S. Government Printing Office. Retrieved from <http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm#ch3.3>

Carbo, M. (1988). Debunking the Great Phonics Myth. *The Phi Delta Kappan*, 70(3), 226–240.

Carnine, D. (2000, April). Why Education Experts Resist Effective Practices (And What It Would Take To Make Education More Like Medicine). Thomas B. Fordham Foundation, Washington, D.C. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED442804>

Center for the Study of Reading. (n.d.). Resources for Educators. *Center for the Study of Reading*. Retrieved June 4, 2012, from <http://csr.ed.uiuc.edu/pubs/resources.html>

Chall, J. S. (1967). *Learning to Read: The Great Debate*. New York: McGraw-Hill.

Chall, J. S. (1983). *Learning to Read: The Great Debate* (Second.). New York: McGraw-Hill.

Chall, J. S. (1985). Afterword. In *Becoming a Nation of Readers: The Report of the Commission on Reading* (pp. 123–125). IL: University of Illinois. Retrieved from [http://eric.ed.gov/ERICWebPortal/search/detailmini.jsp?\\_nfpb=true&\\_ERICExtSearch\\_SearchValue\\_0=ED253865&ERICExtSearch\\_SearchType\\_0=no&accno=ED253865](http://eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED253865&ERICExtSearch_SearchType_0=no&accno=ED253865)

Chall, J. S. (1996). *Learning to Read: The Great Debate* (Third.). Fort Worth: Harcourt Brace & Company.

Chall, J. S., & Squire, J. R. (1991). Beginning Reading Instruction: From Debate to Reformation. In R. Barr, M. R. Kamil, Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research* (Vol. II, pp. 120 – 146). New York: Longman Inc.

Cornelissen, J. P. (2000). Toward an understanding of the use of academic theories in public relations practice. *Public Relations Review*, 26(3), 315–326. doi:10.1016/S0363-8111(00)00050-3

Davidson, J. L. (Ed.). (1988). *Counterpoint and Beyond: A Response to "Becoming a Nation of Readers."* Urbana, Illinois: National Council of Teachers of English. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED298432>

Davies, H., Nutley, S., & Walter, I. (2005). *Approaches to assessing the non-academic impact of social science research: Report of the ESRC symposium on assessing the non-academic impact of research 12th/13th May 2005*. Research Unit for Research Utilisation, School of Management, University of St Andrews. Retrieved from <http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Forums/attach.aspx?a=64>

Denton, C. A., Vaughn, S., & Fletcher, J. M. (2003). Bringing Research-Based Practice in Reading Intervention to Scale. *Learning Disabilities Research and Practice*, 18(3), 201–211. doi:10.1111/1540-5826.00075

- Dimaggio, P., & Useem, M. (1979). Decentralized Applied Research: Factors Affecting the Use of Audience Research by Arts Organizations. *The Journal of Applied Behavioral Science*, 15(1), 79–94. doi:10.1177/002188637901500107
- Dole, J. A., Rogers, T., & Osborn, J. (1987). Improving the Selection of Basal Reading Programs: A Report of the Textbook Adoption Guidelines Project. *The Elementary School Journal*, 87(3), 283–298.
- Durkin, D. (1984). Is There a Match between What Elementary Teachers Do and What Basal Reader Manuals Recommend? *Reading Teacher*, 37(8), 734–44.
- Dykstra, R. (1968). The Effectiveness of Code- and Meaning-Emphasis Beginning Reading Programs. *The Reading Teacher*, 22(1), 17–23.
- Farr, R., Tulley, M. A., & Powell, D. (1987). The Evaluation and Selection of Basal Readers. *The Elementary School Journal*, 87(3), 267–281.
- Faure, A. M. (1994). Some Methodological Problems in Comparative Politics. *Journal of Theoretical Politics*, 6(3), 307–322. doi:10.1177/0951692894006003003
- Finley, S. J. (2000). *Instructional Coherence: The Changing Role of the Teacher*. Austin, Texas: Southwest Educational Development Laboratory. Retrieved from <http://www.sedl.org/pubs/teaching99/changingrole.pdf>
- Flesch, R. F. (1955). *Why Johnny can't read - and what you can do about it*. New York: Harper.
- Furlong, J., & Oancea, A. (2005). Assessing Quality in Applied and Practice-Based Educational Research: A Framework for Discussion. Oxford University Department of Education Studies. Retrieved from [http://www.esrc.ac.uk/ESRCInfoCentre/Images/assessing\\_quality\\_shortreport\\_tcm6-8232.pdf](http://www.esrc.ac.uk/ESRCInfoCentre/Images/assessing_quality_shortreport_tcm6-8232.pdf)
- Gaffney, J. S., & Anderson, R. C. (2000). Trends in Reading Research in the United States: Changing Intellect. In M. R. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (Vol. III, pp. 53–77). Mahway, NJ: Lawrence Erlbaum Associates.
- Gardner, D. P. (1983). *A nation at risk: The imperative for educational reform* (p. 48). Washington, DC: The National Commission on Excellence in Education. Retrieved from <http://dx.doi.org/10.1002/pam.4050030216>
- Geake, J. (2005). Educational neuroscience and neuroscientific education: in search of a mutual middle-way. Article. Retrieved July 7, 2010, from <http://arrrts.gtcni.org.uk/gtcni/handle/2428/49038>
- Geake, John, & Cooper, P. (2003). Cognitive Neuroscience: implications for education? *Westminster Studies in Education*, 26(1), 7. doi:10.1080/0140672030260102
- George, A. L., & Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*. Cambridge, Massachusetts: MIT Press.
- Gilles, C. (1996). The Future of Whole Language. *International Journal of Progressive Education*, 2(2). Retrieved from <http://inased.org/ijpev2n2/gilles2.htm>

- Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, 6(4), 126–135. doi:10.1080/19388076709556976
- Goodman, K. S. (1969). Analysis of Oral Reading Miscues: Applied Psycholinguistics. *Reading Research Quarterly*, 5(1), 9–30.
- Goodman, K. S. (1989). Whole-Language Research: Foundations and Development. *The Elementary School Journal*, 90(2), 207–221.
- Goodman, K. S., Shannon, P., Freeman, Y. S., & Murphy, S. (1988). *Report Card on Basal Readers*. New York: Richard C. Owen Publishers, Inc.
- Goodman, Y. M. (1989). Roots of the Whole-Language Movement. *The Elementary School Journal*, 90(2), 113–127.
- Goswami, U. (2006). Neuroscience and education: from research to practice? *Nat Rev Neurosci*, 7(5), 406–413. doi:10.1038/nrn1907
- Grundin, H. U. (1985). Book Commentary; A Commission of Selective Readers: A Critique of Becoming a Nation of Readers. *The Reading Teacher*, 39(3), 262–266.
- Guzzetti, B. J. (2002). *Literacy in America: An Encyclopedia of History, Theory and Practice*. ABC-CLIO.
- Hall, J. (2005). *Neuroscience and education : a review of the contribution of brain science to teaching and learning* (No. SCRE research report;121). Scottish Council for Research in Education. Retrieved from <http://www.scre.ac.uk/cat/1860030904.html>
- Hallinan, M. T. (1996). Bridging the Gap between Research and Practice. *Sociology of Education*, 69, 131–134.
- Hanney, S. R., Gonzalez-Block, M. A., Buxton, M. J., & Kogan, M. (2003). The utilisation of health research in policy-making: concepts, examples and methods of assessment. *Health Research Policy and Systems*, 1(1), 2. doi:10.1186/1478-4505-1-2
- Hargreaves, A., & Goodson, I. (2006). Educational Change Over Time? The Sustainability and Nonsustainability of Three Decades of Secondary School Change and Continuity. *Educational Administration Quarterly*, 42(1), 3–41. doi:10.1177/0013161X05277975
- Hempenstall, K. (1997). The whole language-phonics controversy: A historical perspective. *Educational Psychology*, 17(4), 399–418. doi:10.1080/19404150509546797
- Hoffman, J. V., McCarthy, S. J., Abbott, J., Christian, C., Corman, L., Curry, C., ... Stahle, D. (1994). So What's New in the New Basals? a Focus on First Grade. *Journal of Literacy Research*, 26(1), 47–73. doi:10.1080/10862969409547836
- Howell, W. G. (2005). *Besieged: school boards and the future of education politics*. Brookings Institution Press.
- Human Services Reauthorization Act of 1986. , Pub. L. No. 99-815 § 901 (1986). Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED277451>

Institute for Educational Development. (1969). Selection of Educational Materials in the United States Public Schools. Institute for Educational Development. Retrieved from <http://www.eric.ed.gov/PDFS/ED044030.pdf>

Jeanne Chall Reading Lab. (n.d.). *Harvard Graduate School of Education*. Retrieved May 18, 2012, from <http://www.gse.harvard.edu/academics/masters/langlit/chall/>

Juel, J. S. (1991). Beginning Reading. In R. Barr, M. R. Kamil, Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research* (Vol. II, pp. 759–788). New York: Longman Inc.

Kaestle, C. F. (1993). The Awful Reputation of Education Research. *Educational Researcher*, 22(1), 23–31.

Keith, S. (1985). Choosing textbooks: A study of instructional materials selection processes for public education. *Publishing Research Quarterly*, 1(2), 24–37. doi:10.1007/BF02683573

Kerr, J. (1996). State Textbook Selections Signal Rejection of “Whole Language” Reading Approach. *Associated Press/Daily News*.

Kim, J. S. (2008). Research and the Reading Wars. In F. M. Hess (Ed.), *When Research Matters: How Scholarship Influences Education Policy* (pp. 89–111). Harvard Education Press.

Kingdon, J. W. (1995). *Agendas, Alternatives, and Public Policies* (2nd ed.). New York: Harper Collins College Publishers.

Klein, S. J., & Rosenberg, N. (1986). An Overview of Innovation. In R. Landau & N. Rosenberg (Eds.), *The Positive sum strategy: harnessing technology for economic growth* (pp. 276–306). National Academies Press.

Koenke, K. (1986). ERIC/RCS: Is BaNoR a Banner? *Journal of Reading*, 29(4), 362–365.

Landry, R., Amara, N., & Lamari, M. (2001a). Utilization of social science research knowledge in Canada. *Research Policy*, 30(2), 333–349. doi:10.1016/S0048-7333(00)00081-0

Landry, R., Amara, N., & Lamari, M. (2001b). Climbing the Ladder of Research Utilization. *Science Communication*, 22(4), 396–422. doi:10.1177/1075547001022004003

Lavis, J., Ross, S., McLeod, C., & Gildiner, A. (2003). Measuring the impact of health research. *J Health Serv Res Policy*, 8(3), 165–170. doi:10.1258/135581903322029520

Levin, H. (1967). Reading Research: What, Why, and for Whom? In J. L. Frost (Ed.), *Issues and Innovations in the Teaching of Reading* (pp. 315–327). Glenview III: Scott, Foresman.

McCulloch, M. (n.d.). Phonics is Phonics - or is it? *The Riggs Institute*. Retrieved June 20, 2012, from <http://www.riggsinst.org/GreatDebate.aspx>

McGill-Frazen, A. (2000). Policy and instruction: What is the relationship? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (Vol. III, pp. 889–908). Mahway, NJ: Lawrence Erlbaum Associates.

Miller, M. (2008). First, Kill All the School Boards. *The Atlantic*, (January/February 2008). Retrieved from <http://www.theatlantic.com/magazine/archive/2008/01/first-kill-all-the-school-boards/6579/>

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching Children to Read: An Evidence-Based Assessment of the Scientific Literature on Reading and Its Implications for Reading Instruction*. Washington, DC: U.S. Government Printing Office. Retrieved from <http://www.nichd.nih.gov/publications/nrp/report.cfm>

National Right to Read Foundation. (n.d.). About Us. *National Right to Read Foundation*. Retrieved May 30, 2012, from <http://www.nrrf.org/aboutus.html>

Nelson, A., & Weinbaum, E. (2006, January). Federal Education Policy and the States, 1945-2009: A Brief Synopsis. New York State Archives, Albany. Retrieved from [http://www.archives.nysed.gov/edpolicy/altformats/ed\\_background\\_overview\\_essay.pdf](http://www.archives.nysed.gov/edpolicy/altformats/ed_background_overview_essay.pdf)

No Child Left Behind Act of 2001. , 115 STAT. 1425 (2002). Retrieved from <http://www2.ed.gov/policy/elsec/leg/esea02/index.html>

North, S. N. D., Willcox, W. F., & Gannett, H. (1904). *Abstract of the twelfth census of the United States, 1900* (Third.). Washington: U.S. Government Printing Office.

OECD. (2007). Understanding the Brain: The Birth of a Learning Science. Organisation for Economic Co-operation and Development. Retrieved from [http://www.oecd.org/document/60/0,3343,en\\_2649\\_35845581\\_38811388\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/60/0,3343,en_2649_35845581_38811388_1_1_1_1,00.html)

Pearson, P. D. (1989). Reading the Whole-Language Movement. *The Elementary School Journal*, 90(2), 231–241.

Pearson, P. D. (1990). Foreword. In *Beginning to Read: Thinking and Learning about Print* (pp. v–viii). Massachusetts: The MIT Press.

Pearson, P. D. (1992). “RT” Remembrance: The Second 20 Years. *The Reading Teacher*, 45(5), 378–385.

Pearson, P. D. (2000). Reading in the twentieth century. Michigan State University/CIERA. Retrieved from <http://www.ciera.org/library/archive/2001-08/0108pdp.pdf>

Pearson, P. D. (2004). The Reading Wars. *Educational Policy*, 216 –252. doi:10.1177/0895904803260041

Peterson, J. C., Rogers, E. M., Cunningham-Sabo, L., & Davis, S. M. (2007). A framework for research utilization applied to seven case studies. *American Journal of Preventive Medicine*, 33(1 Suppl), S21–34. doi:10.1016/j.amepre.2007.03.009

Preview Program: Thirteenth Annual Convention of the International Reading Association: Boston, Massachusetts, April 24-27, 1968. (1968). *The Reading Teacher*, 21(4), 1–32.

Ragin, C. C. (1989). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press.

Rich, R. F. (1997). Measuring knowledge utilization: Processes and outcomes. *Knowledge and Policy*, 10(3), 11–24. doi:10.1007/BF02912504

Robinson, H. M. (1967). Assessing the experimental evidence for various beginning reading plans. In J. L. Frost (Ed.), *Issues and Innovations in the Teaching of Reading* (pp. 327–334). Glenview III: Scott, Foresman.

Schantz, P., J. (2002). *The first grade studies in retrospect* (Doctorate). University of Texas, Austin. Retrieved from <https://repositories1.lib.utexas.edu/bitstream/handle/2152/902/schantzp022.pdf?sequence=2>

Shanahan, T. (2000). Research synthesis: Making sense of the accumulation of knowledge in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (Vol. III, pp. 209–226). Lawrence Erlbaum Associates.

Shannon, J. S. (1991). Politics, Policy and Reading Research. In R. Barr, M. R. Kamil, Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research* (Vol. II, pp. 147–168). New York: Longman Inc.

Shannon, P. (1982). Some Subjective Reasons for Teachers' Reliance on Commercial Reading Materials. *The Reading Teacher*, 35(8), 884–889.

Slavin, R. E. (2002). Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational Researcher*, 31(7), 15–21.

Snyder, T. D., & Hoffman, C. M. (1997). *Digest of Education Statistics*. DIANE Publishing. Retrieved from [http://books.google.co.il/books?id=qOg4x5mEkfwC&pg=PA108&lpg=PA108&dq=Statistics+of+Public+Elementary+and+Secondary+Day+Schools,+1955%E2%80%931984%E2%80%93&source=bl&ots=sd9QHKRynO&sig=m5p3vrC4X0T4J6pM-ZWtb5PkTZA&hl=en&ei=yiMpTrjYI8iVOs\\_T6MMK&sa=X&oi=book\\_result&ct=result&resnum=5&ved=0CDQQ6AEwBA#v=onepage&q=Statistics%20of%20Public%20Elementary%20and%20Secondary%20Day%20Schools%201955%E2%80%931984%E2%80%93&f=false](http://books.google.co.il/books?id=qOg4x5mEkfwC&pg=PA108&lpg=PA108&dq=Statistics+of+Public+Elementary+and+Secondary+Day+Schools,+1955%E2%80%931984%E2%80%93&source=bl&ots=sd9QHKRynO&sig=m5p3vrC4X0T4J6pM-ZWtb5PkTZA&hl=en&ei=yiMpTrjYI8iVOs_T6MMK&sa=X&oi=book_result&ct=result&resnum=5&ved=0CDQQ6AEwBA#v=onepage&q=Statistics%20of%20Public%20Elementary%20and%20Secondary%20Day%20Schools%201955%E2%80%931984%E2%80%93&f=false)

Spache, G. D. (1967). Limitations of the phonetic approach to developmental and remedial reading. In J. L. Frost (Ed.), *Issues and Innovations in the Teaching of Reading* (pp. 194–198). Glenview III: Scott, Foresman.

Stahl, N. A., & Fisher, P. J. L. (1992). “RT” Remembrance: The First 20 Years. *The Reading Teacher*, 45(5), 370–377.

Stahl, S. A. (1990). Riding the Pendulum: A Rejoinder to Schickedanz and McGee and Lomax. *Review of Educational Research*, 60(1), 141–151.

Stahl, S. A. (1999). Why Innovations Come and Go (and Mostly Go): The Case of Whole Language. *Educational Researcher*, 13 –22. doi:10.3102/0013189X028008013

Stahl, S. A., & Miller, P. D. (1989). Whole Language and Language Experience Approaches for Beginning Reading: A Quantitative Research Synthesis. *Review of Educational Research*, 59(1), 87–116.

Stahl, S. A., Osborn, J., & Lehr, F. (1990). "Beginning To Read: Thinking and Learning about Print" by Marilyn Jager Adams. A Summary. University of Illinois at Urbana-Champaign. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED315740>

Stanovich, K. E. (1994). Romance and Reality. *Reading Teacher*, 47(4), 280–91.

Stauffer, R. G. (1967). The Verdict: Speculative Controversy. In J. L. Frost (Ed.), *Issues and Innovations in the Teaching of Reading* (pp. 343–345). Glenview III: Scott, Foresman.

Strickland, D., & Cullinan, B. (1990). Afterword. In *Beginning to Read: Thinking and Learning about Print* (pp. 425–434). Massachusetts: The MIT Press.

Tangri, S. S., & Strasburg, G. L. (1979). Can Research on Women Be More Effective in Shaping Policy? *Psychology of Women Quarterly*, 3(4), 321–43.

Traub, J. (2002, November 10). Does It Work? *New York Times, Education Life*. Retrieved from <http://www.arthuracademy.org/newyorktimes.html>

United States Department of Education. (1985, May 1). Becoming a Nation of Readers [Press release]. United States Department of Education. Retrieved from [http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?\\_nfpb=true&\\_ERICExtSearch\\_SearchValue\\_0=ED253865&ERICExtSearch\\_SearchType\\_0=no&accno=ED253865](http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED253865&ERICExtSearch_SearchType_0=no&accno=ED253865)

Varma, S., McCandliss, B. D., & Schwartz, D. L. (2008). Scientific and Pragmatic Challenges for Bridging Education and Neuroscience. *Educational Researcher*, 37(3), 140–152. doi:10.3102/0013189X08317687

Venezky, R. L. (1984). The History of Reading Research. In P. D. Pearson (Ed.), *Handbook of Reading Research* (Vol. I, pp. 3–38). New York: Longman Inc.

Venezky, R. L. (1987). A History of the American Reading Textbook. *The Elementary School Journal*, 87(3), 247–265.

Wandersman, A. (2003). Community Science: Bridging the Gap between Science and Practice with Community-Centered Models. *American Journal of Community Psychology*, 31(3), 227–242.

Watt, M. G. (2009). Research on the Textbook Selection Process in the United States of America. *IARTEM e-journal*, 2(1), 1–22.

Weiss, C. H. (1979). The Many Meanings of Research Utilization. *Public Administration Review*, 39(5), 426–431.

Weiss, C. H. (1982). Policy Research in the Context of Diffuse Decision Making. *The Journal of Higher Education*, 53(6), 619–639.

Wilkerson, B., C. (1988). A Principal's Perspective. In *Counterpoint and Beyond: A Response to "Becoming a Nation of Readers."* Urbana, Illinois: National Council of Teachers of English. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED298432>

Willingham, D. T., & Lloyd, J. W. (2007). How Educational Theories Can Use Neuroscientific Data. *Mind, Brain, and Education*, 1(3), 140–149.

Woo, E. (1999, December 2). Jeanne Chall; Longtime Phonics Advocate. *Los Angeles Times*. Retrieved from <http://articles.latimes.com/1999/dec/02/local/me-39690>