DOCTOR OF PHILOSOPHY

Using precision teaching strategies and tactics to increase essential skill fluency

Beverley, Michael

Award date:
2012

Awarding institution:
Bangor University

Prifysgol Bangor

Bangor University

Using precision teaching strategies and tactics to increase essential skill fluency

Michael Beverley

Thesis submitted to the School of Psychology, Bangor University, in partial fulfilment for the degree of Doctor of Philosophy

December 2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

First I would like to thank all of my supervisory committee: Dr Carl Hughes (First supervisor), Prof. Richard Hastings (Second supervisor), and Dr Paloma Marí-Beffa (Chair). All of them have contributed to my continued professional development whilst I have been working in the psychology department at Bangor University.

Particular thanks must go to my direct supervisory team; both respected and highly qualified academics in their own right. Carl has been an invaluable colleague, mentor and friend since I arrived here to work in 1996, and it was he who introduced me to the science of Precision Teaching and the wonders of Standard Celeration Charting—which I first used as an aid to give up smoking in January, 2003. Whilst I have known Richard for a shorter time, his advice, academic leadership, and insightful knowledge of innovative methods of data analysis have been invaluable; this liaison in turn has been instrumental in selecting some of the methods of data analysis used in this thesis. Thanks also go to Dr Tracey Lloyd for her collegial support during my time working in our department, and for her final proof reading of parts of this document. Final academic praise goes to Emily Tyler, also nearing completion of a PhD, who has been committed to the advancement of Precision Teaching in local schools and a most valued Teaching Assistant on the year three module—Evidence-based Behavioural Methods in Education.

Great appreciation is due to Kent Johnson and Kris Melroe. In 2010 I was privileged to be the recipient of a scholarship to study at Morningside Academy's Summer School Institute. My appreciation to them and the other staff at Morningside is immeasurable.

I would like to thank two practitioners of ABA (graduates of the Masters in ABA) who have been so influential in disseminating effective practices in Gwynedd; Bethan Mair Williams and Beverley Jones continue to work with vast enthusiasm to extend the network of educational establishments who adopt PT and use it to guide their teaching.

Thanks are also due to all the schools where the research took place, and to motivated teachers who work within those schools who have embraced PT methods and incorporated them into their everyday classroom practice. As this is a public document, I cannot name these people or the schools, but I am quite sure they will know who they are.

Additional thanks goes to all members of the *Standard Celeration Society* and its list serve. The members are a professional, friendly group of people who always offer both resources and advice to any person in order that information about PT can be disseminated to the wider public, so that the number of people using PT continues to grow.

Academic tributes complete, I now move to my family. My wife Angela and my son Matthew Stephen James Beverley both deserve thanks for their support over the duration of researching and writing this PhD. Whilst it is not easy working full time and completing a PhD, it is certainly not any easier for loved ones to have to endure some of the sacrifices in terms of my time, attention, and availability. Acknowledgments are also due to my brothers Steven and Peter Beverley.

Finally, obvious thanks go to my parents—James Beverley and Irmgard Luise Sofie Beverley (née Greve)—were it not for World War II, they would probably never have met, and the rest goes without saying.

Quotations

“A fo ben, bid bont” “He who would be a leader, must be a bridge.” The Mabinogion.

"You never make an error. You were doing what the laws of behavior are predicting you would do under those contingencies. An error is a faulty contingency; it is not something in you." - Julie Vargas 12/7/12. Invited Address: IPTC 2012

“Children are not retarded. Only their *behavior* in average environments is sometimes retarded. In fact, it is modern science's ability to design suitable environments for these children that is retarded.” (Lindsley, 1964, p. 62)

> If I were seriously ill and in desperate need of a physician, and if by some miracle I could secure either Hippocrates, the Father of Medicine, or a young doctor fresh from Johns Hopkins School of Medicine, with his equipment comprising the latest developments in the technologies and techniques of medicine, I should, of course, take the young doctor. On the other hand, if I were commissioned to find a teacher for a group of adolescent boys and if, by some miracle, I could secure either Socrates or the latest Ph.D. from Teachers College, with his equipment of the latest technologies and techniques of teaching, with all due respect to the College that employs me and to my students, I am fairly certain that I would jump at the chance to get Socrates (Bagley 1934, as cited in Snider, 2006, p. 174).

According to Malachi Constant “I was a victim of a series of accidents, as are we all.” (Vonnegut Jr., 1975, p. 137)

Summary

The area for this thesis was the application of Precision Teaching (PT) to the teaching of essential skills to various populations—university undergraduates, high school adolescents, and primary school children. Chapter 1 of this thesis begins with a brief overview describing problems that exist within the educational system, and how we might intervene to raise the performances of all learners by using PT's systems and strategies. Paper 1 was a small *n* study (conducted over 10 weeks) aimed at improving the rate of reading high frequency English words in a group (n = 5) of children struggling with reading. This was compared to a small comparison group that underwent Treatment as Usual (TAU; n = 2). Although the engagement time of the PT group was brief (M = 12 minutes per week), all the children showed significant gains in their accurate reading fluency. In Paper 2 we applied these same principles to teach statistics to undergraduates (Intervention, n = 24; TAU, n = 31). Results showed that the PT group performed significantly better at post-test in comparison to TAU. In addition the PT group performed better on their weekly module content tests. Paper 3 was conducted to increase maths performance, comparing the intervention group (n = 19) with a control group (n = 10). All the PT children performed significantly better on three of the four outcome measures of essential skill maths fluency (d = 1.25-1.67)—one of which measures had not been practiced providing some initial support for contingency adduction. In Paper 4 we used a simple flashcard intervention to increase recall of second-language vocabulary. The intervention spanned four weeks and took only 15 minutes per week for each child, yet the intervention group (n = 79) performed significantly better then the WLC group (n = 16) at post-test (d = 1.54). Both Papers 3 & 4 report data showing significant increase in the RCI measures and a low NNT. In all four papers Standard Celeration Charts were used to monitor learner's improvement.

Preface to Introductory Chapter

This thesis consists of four articles that have been published or are currently in submission to peer-reviewed journals—Chapters 2-5. In the context of my work in the university and its goals, I am employed to teach undergraduate and masters classes. However within the context of my PhD, my goal was to conduct research that could impact both teachers and learners within local educational services, including at the university. Although none of the research was funded, I aimed to provide educators with brief, effective interventions that are easy to implement, by training them in the use of Precision Teaching's (PT) systems and strategies. In addition I hoped that I could present results in a way that could be communicated to broader audiences. The settings for the research were therefore educational with various populations—University with undergraduates, high schools with young adolescents, and primary schools with younger children. It was necessary that the procedures put in place were not costly in terms of either physical resources or staff time required, but that they still could demonstrate pedagogical gains for children, in terms of added benefit, for those taking part. Chapter 1 of this thesis begins with a brief overview describing some problems that exist within the current educational system, and how we might intervene to raise the educational performances of all children by using the systems and strategies that PT has to offer. Then guiding principles and methodological components of PT are described, concluding with discussion of how PT can be effectively used within learning environments, as a measuring and monitoring tool to ensure that learning is not only taking place, but also growing at an appropriate rate to keep in line with any academic curriculum goals. In the conclusion of Chapter 1 an overview of the organisation of the chapters of the thesis are defined and a brief narrative of the studies in the following chapters is presented.

Chapter 1: Introduction

What are the Problems / Issues?

The provision of effective education for every child is of paramount importance, as a poor education can have devastating effects on future life choices and career prospects (Barrett et al., 1991; Baxter & Frederickson, 2005; U. S. Department of Education, 2002, 2004). A consistent relationship exists between poor basic skills and unemployment: people with poor basic skills are more likely to have left school at 16 with no qualifications; will have more difficulty retaining employment or achieving promotion; and children of parents with poor basic skills are more likely to struggle academically and therefore be less able to assist their children with academic matters (The Basic Skills Agency, 2001). Children therefore need to be provided from their early life with the best education possible through the best teaching methods available.

The Right to Effective Education

According to some educationalists, children have the *right* to an effective education (Barrett et al., 1991) and parents have the right to expect the best education for their children. For this to happen methods of education must be based on sound, scientific evidence—not fads, frauds, and follies (Carnine, 2000; Kozloff, 2005). Furthermore, teaching practices must follow research proven methods that are based on evidence-based practice and validated and refined through practice-based evidence (Brusling, 2005; Carnine, 2000). Whatever the political debates about whether effective educational practices are regarded as civil rights, clearly there is little argument over whether educational practices should be *as effective* as possible given the current level of investment and resources within a society.

What is Effective Education?

According to Barrett et al. (1991) a good instructional system or teacher must fulfil a minimum of three criteria: "(1) It must be effective in helping students learn more rapidly then they would on their own; (2) what students learn must benefit both the individual and society as a whole; (3) it must employ positive rather than coercive or punitive methods.” (pp. 79-80).

This means that if a child who has been failing academically is to catch up with their age-appropriate grade level, they must learn at a faster rate than they had previously been learning (Johnson & Layng, 1994; Johnson & Street, 2004; Skinner, 1984). This is because by definition such children have been showing less than average or sub-optimal learning within the normal environment. Clearly then any failure to make faster gains in learning than the average child typically makes, will only result in them falling even further behind. Thus for these children it is vital that a method that can effectively monitor academic gains or losses is put in place. Arguably, if such a method had originally been in place then such academic losses might have been prevented.

Within education there are two broadly competing educational philosophies—traditional and progressive (Snider, 2006). Traditional education emphasises the need for students to progress through the curriculum in specific steps for each curriculum area, by achieving mastery of each preceding step, before progression onto the next; whilst, in contrast, progressive education promotes methods that are intended to foster children’s innate interest in learning; although this latter type of education relies on children progressing only as rapidly as they would learn on their own without any direct intervention (Fredrick, Deltz, Bryceland, & Hummel, 2002).

In order to develop and provide effective educational practices it is essential that education follow evidenced-based practice and practice-based evidence approaches. The former refers to the use of well-controlled scientific research in determining what has been shown to work. The latter refers to methods of monitoring practice that clearly demonstrates that practice is having the desired effects that were expected from the research evidence (i.e., in the case of education, is effective educational practice).

Viewed from this context it is clear that educational provision and approaches are too often based on current trends, rather than scientific evidence and these educational ‘fads’ cycle according to the prevailing approach to learning that may be currently in favour (Kozloff, 2005; Slavin, 2008). In many cases the approaches have not been adequately researched before they are rapidly adopted in educational settings throughout the country (Slavin, 1989). There have been effective teaching tools available for quite some time with accomplished research-proven track records (Bijou, 1970; Heward, 2005; Maloney, 1998; Moran & Malott, 2004), yet there appears to be a “knowledge to practice gap” (Heward, 2005, p. 317) that prevents their complete and wholesale adoption into mainstream educational practices (Lindsley, 1992b). This failure to adopt proven practices is analogous, for instance, to the delay in time between understanding the benefits of citrus fruits in the prevention and treatment of scurvy (1601), and its eventual implementation as a treatment, leading to it being finally eradicated in the mercantile marine—after a delay of 264 years (Barbash, 2011; Lamb, 2001; Mosteller, 1981; Skinner, 1984). Heward (2005) believes that there is a “more pressing (and possibly larger) gap we must close . . . between what research has discovered about effective instruction and what is practiced in the majority of classrooms.” (p. 317).

In her book, Snider describes how educational practices should follow the example from medicine in its timely transition toward becoming a mature science (Snider, 2006; Snider & Roehl, 2007). A view supported by Carnine (2000; Hargreaves, 1997) as he considers how education is at a similar juncture of development toward scientific status, as medicine was in the early part of the twentieth century before it rose to the challenge of becoming a mature profession.

Raising Educational Performance

There are a number of issues that need to be considered if you desire to raise the educational performance of young children. These issues revolve around remediation of existing failure, prevention of future failure, and (perhaps within the context of finite resources) the pursuit of excellence for all children. The first issue is how can we stop or reverse the failure that many children in our primary education system are experiencing. The second is how can we monitor learning effectively so that we can prevent significant failure before it occurs. And last, how can we implement effective educational practices that can produce excellent learning for all children?

Precision Teaching—Effective Measurement of Learning

Measurement of Performance in Education (Why it is Important)

Morningside Academy in Seattle, founded in 1980 by Dr Kent Johnson, is a school that incorporates the measurement of learning into their everyday practice. They guarantee a child's academic performance to improve by a minimum of one year's growth in their weakest academic skill, within six-months of tuition and fluency-building practise. As such it is an impressive example of the effectiveness that can be achieved to combat academic failure when performance is measured and the teaching adjusted accordingly to meet the requirements of individual learners (Johnson & Layng, 1994; Johnson & Street, 2004). Morningside Academy incorporates performance measures to determine each learner's academic improvement at different times throughout the academic year: daily (micro), weekly or monthly (meta), and once or twice yearly (macro). Data collected for each learner for every skill enable teachers to predict future growth and adjust instruction accordingly to ensure that they maintain the learner on the appropriate learning trajectory to meet targeted progress goals. If such interventions can remediate deficits then learners may be able to effectively move from their current position in this *distribution of deficit* and approach or even enter the normal distribution of their typically performing peers.

Most schools and educational systems already utilise Macro level systems of measurement (e.g., standardised yearly achievement tests), and these offer benefits of being able to locate children with respect to their peers on key areas of the curriculum. However, these types of measures have three important limitations: (1) they are expensive and time consuming to administer, (2) they cannot identify learning problems when they are first developing, and (3) they typically offer little guidance on how teachers can intervene with specific learning issues.

Conversely, Meta and Micro level assessments are designed to be quick to administer, specific in how they identify learning problems, and offer simple guidelines on what types of intervention are needed to correct particular failures in individual learning (Johnson & Street, 2004). Thus Macro level of assessment can be viewed more as testing devices used to ascertain children's current performance level, whilst both Meta and Micro assessments should be viewed more as taking precise, individual measurements to ensure that adequate learning is taking, and continues to take place (see Figure 1.1).

The main way that Meta and Micro level measures differ to Macro assessments is that the teacher can use these quick assessments to make instructional decisions. These levels of quick assessments should not be equated with testing children, but rather measuring learning in ways that can help teachers design individual programmes so that each child achieves their potential. The teaching process thus becomes reactive and individualised in the sense that teachers are far more able to identify learning problems in a timely fashion for each learner. These help the teacher alter learning strategies until adequate progress is being maintained, and in the case of learners who are falling behind their peers, aid them to make significant gains to catch up to the expected level.

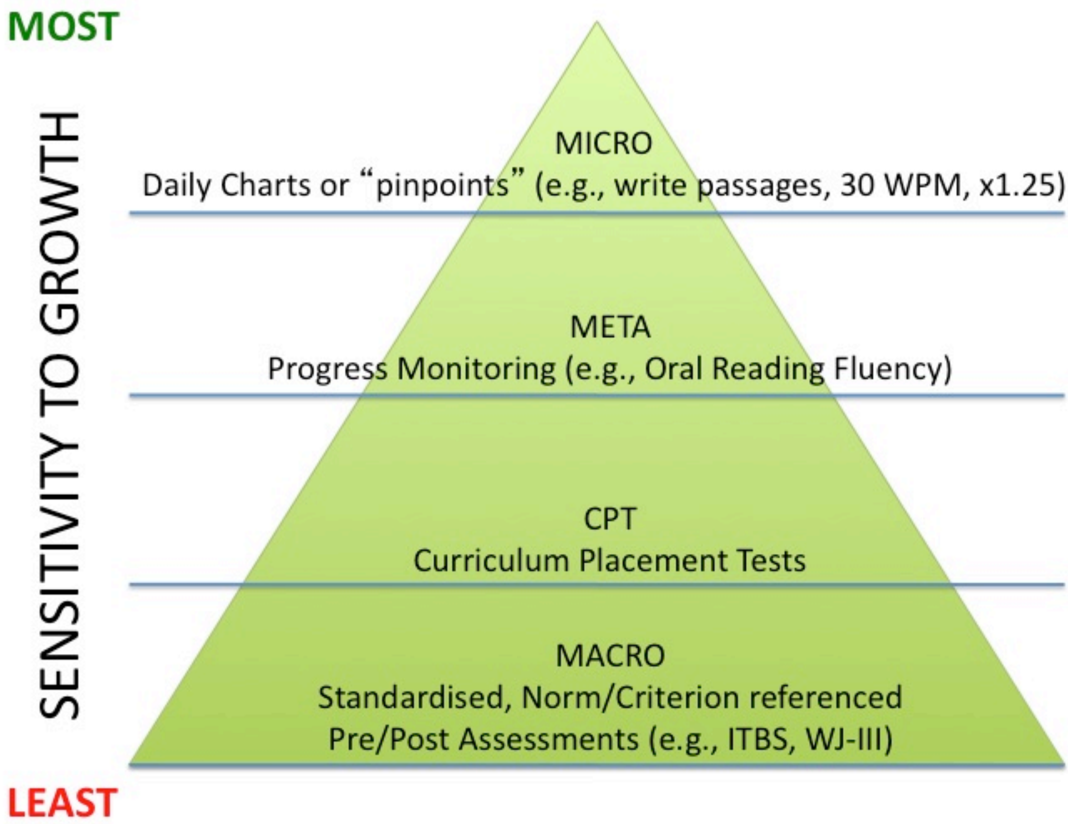


Figure 1.1. Using a multi-level system of assessments to inform instructional decisions and determine programme effectiveness. After Johnson, K. R., & Street, E. M. (2004). The Morningside model of generative instruction: What it means to leave no child behind. Concord, MA: Cambridge Centre for Behavioral Studies.

Curriculum Based Measurement (CBM), Response to Intervention (RtI), and Precision Teaching (PT) Measurement Systems

There are three main methods that allow measurements to guide teaching within the classroom, and through their use the teacher is facilitated to react to the data inductively: CBM, RtI, and PT. Each of these methods allow a scientific approach to be adopted in the classroom and also each is of course an example of practice-based evidence.

These three procedures also mesh nicely with the whole philosophy of the RtI model (see Figure 1.2), which is ". . . designed to allow for early and effective responses to children's learning and behavioral difficulties, provide children with a level of instructional intensity matched to their level of need and then provide a data-based method for evaluating the effectiveness of instructional approaches" (Fox, Carta, Strain, Dunlap, & Hemmeter, 2009, p. 1).

Because each child is an individual, each child's educational needs differ; these approaches allow those with the least need to progress as normal, but also enable children who do require additional support to receive that support in a planned, focussed and timely manner, and the effect of such support can be measured effectively and adjusted accordingly as and when required. Therefore the implications of increasing the more widespread use of procedures such as RtI, CBM, and PT could significantly impact educational achievement across all essential skills areas.

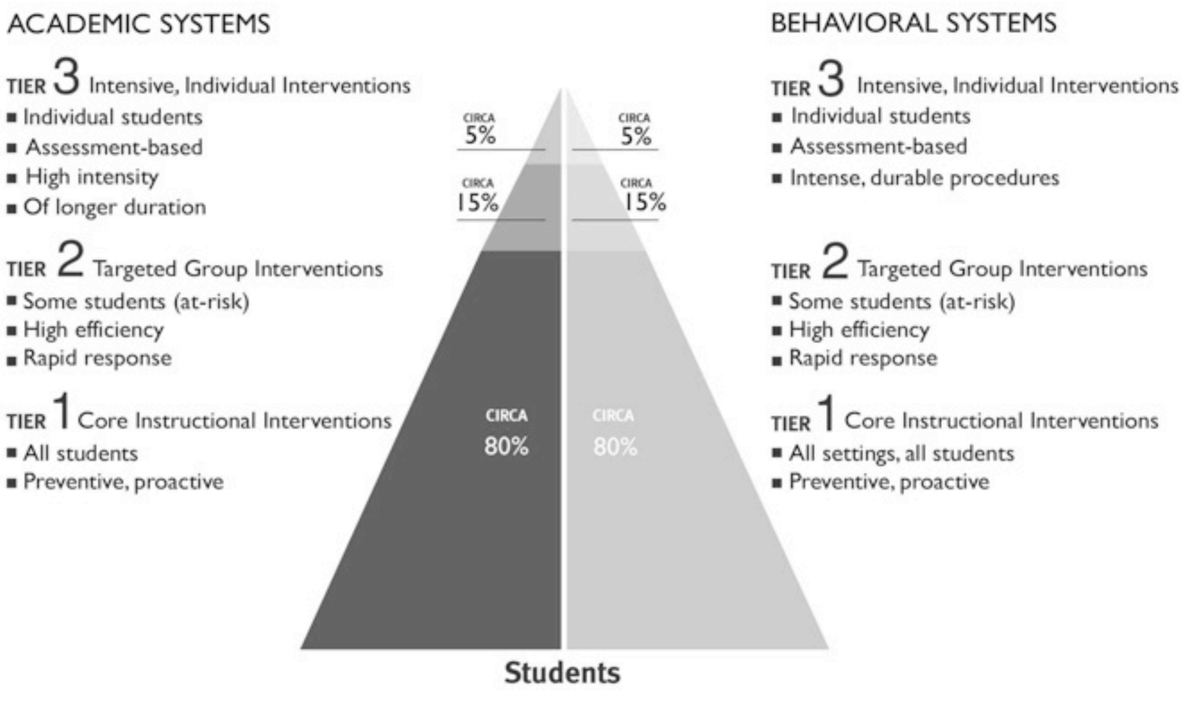


Figure 1.2. Response to Intervention three-tier pyramid. After Fox, L., Carta, J., Strain, P., Dunlap, G., & Hemmeter, M. L. (2009). Response to Intervention and the pyramid model.

Within the RtI model, all pupils are initially targeted for the core primary interventions with progress being measured approximately once every three months. Should data indicate that a more intensive level of intervention be required, individual children receive additional support with more strategic secondary interventions (measured every two weeks) or an even more intensive individualised tertiary intervention, with progress being measured once per week (National Centre for Response to Intervention, 2010). If even more intensive support is required then the daily measurements of a system like Precision Teaching would benefit learners even more.

Precision Teaching What is it?

One evidence-based educational method that has had almost five decades of success is Precision Teaching (PT). PT was founded and coached by Ogden Richardson Lindsley (Lindsley, 1995a). Lindsley had been one of B. F. Skinner's PhD students and many of the fundamental principles of PT stem from the free operant research originally carried out in Skinner's laboratories (Lindsley, 2002). Free operant in this context refers to "students [that] are free to respond at their own pace without having restraints placed on them by the limits of the materials or the instructional procedures of the teachers" (Lindsley, 1995a, p. 10). Additionally PT uses frequency of response as its measure of effectiveness: Skinner reported that his greatest two contributions and legacy to science were his use of frequency (or rate) as a measure of performance and the cumulative response recorder (Evans, 1967; as cited in Lindsley, 2010, p. 23).

PT can be given an arbitrary starting date of 1964; this was when Lindsley published his seminal paper (Eshleman, 1990; Lindsley, 1964). PT has been effectively applied with many different skills and with many different populations; ranging from children with special needs (Gryiec, Grandy, & McLaughlin, 2004; McDowell & Keenan, 2001a; Zambolin, Fabrizio, & Isley, 2004); children in mainstream schools to college students (Johnston & Pennypacker, 1971; Vieitez, 2003) and the elderly. Even though PT has been shown to be effective across different time periods, settings and curriculum, there is still resistance to adopting the approach (Kubina & Yurich, 2012).

PT is not, as its name might imply, a method of teaching; it can be effectively applied to any curriculum area and at any instructional level. PT is "basing educational decisions on changes in continuous self-monitored performance frequencies displayed on 'standard celeration charts.'" (Lindsley, 1992a, p. 51). It has been further described as "a system of tactics and strategies for the self monitoring of learning" (Lindsley, 1997b, p. 537). PT is fundamentally a method of measuring learning that can help teachers make timely decisions about the effectiveness of teaching for each child, and help teachers ensure that *every child in a class* maintains successful learning.

PT can be best viewed as a system that provides an effective learning navigation tool that guides both the teacher and the learner along the most effective, direct path toward a skill frequency aim in the shortest possible time for each learner (Hughes, Beverley, & Whitehead, 2007). The value of precision teaching lies in identifying a subject area in which the child is failing to progress, followed by a daily session of teaching, fluency building, monitoring and evaluating progress in order to optimise learning (Lindsley, 1992).

It is argued that children start to experience problems in learning when they are moved onto skills that are more advanced without having acquired the prerequisite fluency performance on the basic skills. For example, a child who is not fluent at *multiplication tables* would likely have trouble when asked to complete a more advanced task (e.g., long division sums) that requires them to use multiplication skills. Similarly, if a child is not fluent at saying the sounds of alphabet letters, they will likely experience problems when they come to read words that contain those letters.

The Four Guiding Principles of PT

Precision Teaching has four guiding principles: (1) Focus on Directly Observable Behaviour, (2) Frequency as a Measure of Performance, (3) The Standard Celeration Chart, and (4) The Learner Knows Best (White, 2000).

Focus on directly observable behaviour. PT focuses on directly observable behaviour that can be accurately counted and recorded (Neal, 1981). To define and operationalise what constitutes observable behaviour Lindsley (1991, 1997a) devised the *dead person's test*. If a dead person can exhibit the same behaviour then it is not a valid behaviour to count. Additionally, behaviour is not considered and counted in isolation but takes in the dimension of time spent behaving by counting specific movement cycles. Precision teachers must ensure that when defining behaviour the behaviour is (1) observable and therefore countable, (2) you are counting movement itself, and not the absence of movement (e.g., sitting still or not swearing), and (3) ensuring that it is a movement that you are counting rather than a *label* (Alper & White, 1971; White, 1986, 2000).

Frequency as a measure of performance. Lindsley discovered through his research that frequency was between 10 to 100 times more sensitive to detect changes in patterns of behaviour than percent correct (Lindsley, 1995a). Traditional measures of academic performance are usually taken using percent correct as the measure, but this measure does not inform sufficiently about performance change, because it leaves out the most important information—that of time taken to complete the activity (Eshleman, 1992; Lindsley, 1995a). Additionally it ignores the fact that corrects and errors can differ in frequency independently from each other, and as such are not mutually exclusive (Binder, 1996, 2001). Within a PT framework frequency of correct and incorrect scores in academic tasks can provide additional information that affords insight into each pupil's proficiency at a particular subject and goes beyond what can be discerned from percentage correct data (Kubina & Morrison, 2000).

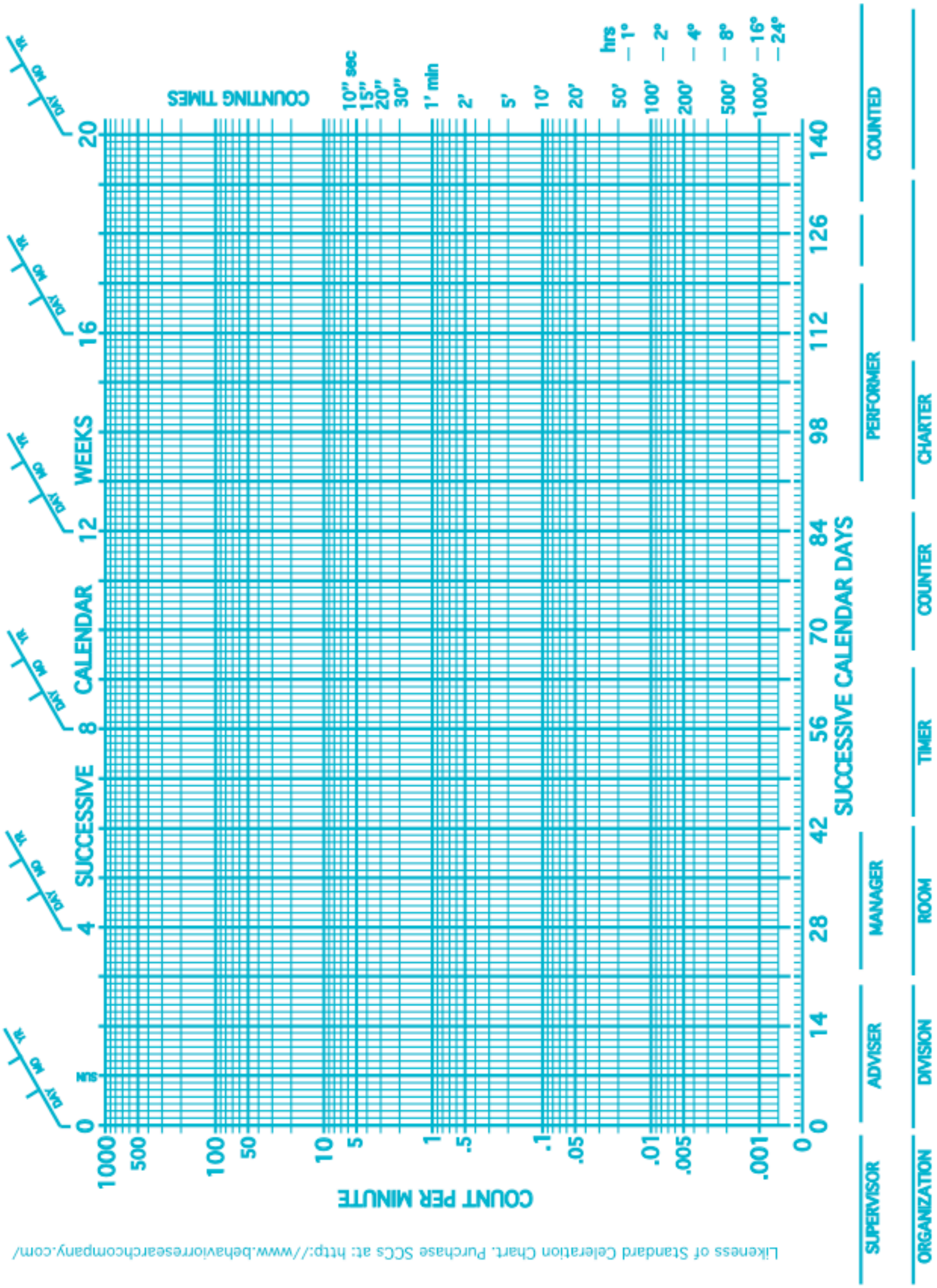


Figure 1.3. Likeness of the daily Standard Celeration Chart.

Standard Celeration Chart. PT involves the use of SCC (see Figure 1.3) to display performance data obtained from timed probes and are used to obtain a ‘snap-shot’ of the child’s performance on that skill. These charts possess a calendar scale along the x-axis to accommodate 140 successive days and a multiply-divide scale on the y-axis; according to precision teachers changes in behavioural frequencies are best represented both graphically and mathematically in multiply and divide proportional changes. As a learning chart the SCC has several advantages over standard display methods typically used: it offers the potential to record the full range of frequencies of human behaviour from 1 per day to 1,000 per minute—which in turn allows the discovery of functional relationships between two or more behaviours; it allows performance data to be monitored over an entire semester on one single visual display; the log scale allows the measurement of celeration (rate of learning over time); it can be used as an effective ‘diagnostic tool’ to accurately predict future performance, guide instructions, and measure the effects of interventions introduced to attempt to increase the rate of learning (Lindsley, 1995a; Neal, 1981; White & Neeley, 2003). The SCC has been described in detail in a number of previous publications (Calkin, 2003, 2005; Graf & Lindsley, 2002; Pennypacker, Gutierrez Jr., & Lindsley, 2003; White & Neeley, 2003), and in addition– within the present thesis in papers 1 and 2.

Learner knows best. This is one of the most important of the guiding principles, as it emphasises that the learner is at the centre of the learning process. And that the learner’s data when viewed on the SCC will guide the teacher in making effective decisions for that individual in a timely manner (Lindsley, 1972, 1995a). As White (1986) succinctly points out:

> Essentially, in order to be responsive to the pupil's needs the teacher must be a student of the pupil's behavior, carefully analyzing how that behavior changes from day to

day and adjusting the instructional plan as necessary to facilitate continued learning. Precision Teaching offers a set of procedures designed to assist in that process (p. 1).

Key Methodological Characteristics of PT

Some key methodological characteristics of PT are: component/composite analysis, fluency training, time probes, tailoring practice materials to the progress of individual children based on learning pictures and the use of a standardised graphical display (referred to as the Standard Celeration Chart (SCC; Chiesa & Robertson, 2000). These are detailed within the following chapters.

PT measurements are taken by using short timed probes to take a 'snap-shot' of the child's learning. This kind of information gives the potential for rapid gains in academic performance, as decisions can be immediately be made, based on the data, on how best to improve learning for the individual child, and more importantly, rapidly be able to tell from future data whether the changes implemented had the planned effect on learning.

What does a PT approach look like?

Emphasis on Fluency and Outcomes

A fluent performance is one that can be demonstrated with both accuracy and speed (Binder, 1996) and appears to be an effortless, almost automatic process. Fluency has long been accepted as a required prerequisite for more advanced performances in the martial arts, music, and sports (Bloom, 1986), but has not been so readily adopted as a prerequisite of proficient academic performance (Lindsley, 1992b). Yet this is exactly the type of performance required to enable learners to acquire a skill to a true mastery level of competent performance, which in turn affords them the ability to progress on to the next level, knowing that they have truly mastered the preceding one and therefore have a firm foundation on which to build. There was a time that all children rehearsed their multiplication tables until they could recall them by rote. This type of learning has ceased in most schools; criticisms about *rote learning* and *drill and kill* were the reasons for this withdrawal from the practise.

Practicing until fluent is essential; rote learning or *drill and kill* would only be unacceptable if the child is left to continue to practice important skills in which they have already reached a level of mastery. This situation represents an aspect of ineffective education because time is spent on mastered tasks that would be better spent on learning other novel skills. The extent to which this is the case in modern educational practice is not known because there are no definitive guidelines on what level of skill is considered mastery within basic skill domains. *Fluency Based Instruction* (FBI) in contrast, conducts brief practice timing periods that allow peak performance, and it builds fluent performance in composite skills before application to more complex curricula items. The main issue here is that it is a case of throwing the baby out with the bathwater—the practice was good, but the endless drill with no AIMS or real concrete information on whether fluency had been achieved was the factor that was rightly criticised—PT measurement systems retain the benefits of traditional practice but negate the negative aspects that drained motivation when skills were practiced beyond the point of mastery.

Achieving fluency has been demonstrated to be a significant contributory factor towards positive educational outcomes (Binder, 2003), and slower performance may be indicative of a child presenting difficulties solving mathematic problems (Geary & Brown, 1991; Jordan & Montani, 1997). The general importance of automatic retrieval of basic information, and regular practice, has been supported by previous research (Bloom, 1986; Sweeney, Sweeney, & Malanga, 2001; U. S. Department of Education, 2009; Willingham, 2009) with many different populations (Beverley, Hughes, & Hastings, 2009; Binder, Haughton, & Bateman, 2002; Binder, Haughton, & Van Eyk, 1995; Bloom, 1986; Hughes et al., 2007; Johnson & Layng, 1992, 1994, 1996; Kubina, Commons, & Heckard, 2009; Kubina & Morrison, 2000; Kubina, Morrison, & Lee, 2002).

Johnson and Layng (1994, 1996) purport that only when performance becomes automatic, effortless, and can be performed at an appropriate speed, do the benefits or rather outcomes of fluent performance emerge. These benefits are captured in the acronym RESA (Retention, Endurance, Stability, Application; see for example Binder, 1996) that describe the key outcomes of fluent performances, whatever the skill domain. Learners must be able to retain skills over time without practice, be able to carry out the skill for extended or functional time periods without displaying fatigue, be able to maintain fluent performance in the presence of distraction, and be able to apply the learnt skills to novel or 'real-world' situations (Haughton, 1972; Kubina, 2002; McDade, Brown, & Goggins, 1983; Mercer, Mercer, & Evans, 2000). These benefits (RESA) are very clearly based on earlier research that supports a four-level instructional hierarchy of acquisition, fluency building, generalization, and application or adaption (see Figure 1.4, Haring & Eaton, 1978). It is also vital that students are fluent in *composite* skills of a *component* task to ensure that learners do not become *cumulative dysfluent*—a recognition that a learner will be hampered from performing component (Binder, 1996; Lindsley, 1995b) tasks that rely on a previous fluent performance of prerequisite composite tasks. This may be likened to the relationship between procedures and concepts: where the building up of procedural fluency may contribute to greater understanding of or access to more advanced concepts (Briars & Siegler, 1984; Cowan et al., 2011; Gelman & Meck, 1983; Greeno, Riley, & Gelman, 1984; Rittle-Johnson & Siegler, 1998).

Level	Emphasis	Strategies
Acquisition	Accuracy of response	1. Demonstration 2. Models 3. Cues 4. Routine drill
Fluency or Proficiency	Speed	1. Repeated novel drills 2. Reinforcement
Generalization	Novel stimulus	1. Discrimination training 2. Differentiation training
Adaption	Adapted response	1. Problem solving 2. Simulations

Figure 1.4. Learning Hierarchy. After Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hansen, C. L. (1978). *The fourth R: Research in the classroom*. Columbus, Ohio: Charles E. Merrill.

Steps to PT

In order to implement an effective PT intervention, you need to

Pinpoint or describe the behavioural objective. In this are included the correct and incorrect performances, the *AIM range* for expected frequency, and the *learning channel* (how the task is presented and how the learner responds).

Count and teach – using any of the evidence based procedures that demonstrate effective education.

Measure and decide. If learning is happening, then we continue; if learning is not happening, we make a change and monitor the effectiveness of that change over the subsequent days.

Following this simple three-step cycle allows you to ascertain whether the teaching methods employed are producing the intended academic gains for each individual learner

(Fuchs, 2004; Johnson & Street, 2004; Lindsley, 1995a, 1997a; Potts, Eshleman, & Cooper, 1993; Raybould, 1984; Raybould & Solity, 1988; Roberts & Norwich, 2010). If these gains in learning are not forthcoming, different teaching approaches or changes in instruction can be introduced and their effect on the learner's progress monitored.

Methods and Research

Experimental Methodologies

When conducting research, a variety of designs and methods of data analysis are available from which to select—all of which have their own strengths and weaknesses associated with them. For this thesis I have attempted to integrate both group and single-case designs; and in addition use some innovative measures to convey the improvement of the individuals whose data are usually ignored within the process of group analysis. I chose to amalgamate the two approaches to better communicate the findings and make them relevant to the broadest audience possible. In doing so I wanted to present my findings to readers who might not normally benefit from accessing behaviourally based research, and who may be more used to viewing data with elements of group design analysis.

Behaviour Analysis has a long history of criticising the use of the statistics and group design research for a number of conceptual and practical reasons (Johnston & Pennypacker, 2009; Sidman, 1960; Skinner, 1938); first group statistics work to remove sources of variability in behavioural data mathematically. In removing this variability they are essentially removing individual differences—an aspect that is important for a behavioural analysis that clearly defines behaviour as a function of the interaction between an individual and their environment (Johnston & Pennypacker, 2009); and secondly, on a purely practical level, a focus on individuals is essential in any approach that is going to be broadly applicable to people who are biologically and experientially unique, especially if practice is to be manipulated with respect to an individual in order to be educationally beneficial. Despite these criticism however, there are a number of benefits of examining data at a group level.

First, it allows the evaluation of interventions at a large scale group level, which in turn allows the possibility of completing Randomised Control Trials (RCT), which are the accepted gold standard of research for those outside of behavioural psychology and those that generally make decisions on practice policy. Second, it allows such findings to be generalised to the wider population in contrast to a single subject design, which cannot by its very nature be generalised in the same way. Nevertheless, the importance of being able to communicate findings from research to a wide audience is clear, and it is also clear that the large majority of people seem to currently understand this type of research approach more than single case designs. Yet, both methodologies have their benefits (Field, 2009; Gravetter & Wallnau, 2007).

PT research studies typically follow a Single Subject methodology in their design and analysis (Horner et al., 2005; Kazdin, 2010; Kennedy, 2005; White, 2002). In the majority of the papers presented in this thesis, the aim was to evaluate the effectiveness of PT fluency-based interventions at both the group level and the level of the individual child. Many group studies typically use only tests of significance to statistically support improvements when comparing outcome measures between two or more groups. However, the reliance solely on tests of statistical significance to evaluate the effects of teaching interventions has, in addition to the issues raised above, at least two drawbacks: (1) statistical significance levels are greatly affected by the study's sample size (i.e., nothing to do with the intervention, Gravetter & Wallnau, 2007); and (2) although they provide the researcher with group effect sizes, they convey no meaningful information about the magnitude of the change achieved for *each individual* in the intervention (Baguley, 2012; Field, 2009). To address the first limitation, researchers typically employ effect size measures, which are not overly affected by the sample size, and can describe the magnitude of the observed effect (Cumming, 2012); however group research design can never successfully address the second limitation, because as the mean is the focus of the analysis they do not pay attention to any individual data from the dataset. Additionally, not all effect sizes that are used are equally meaningful or clear. Even though standardised mean differences are commonly used, their interpretation is not intuitive to all consumers of research.

Measures

In order to provide further information than the typical group statistical analysis, I have adopted two measures of change that have been commonly used in different areas of health and educational research, and integrated them into my methodology to provide a measure of effectiveness of the intervention at the individual level (Reliable Change Index, RCI) and the relative potency of the intervention in comparison to existing treatments (Numbers Needed to Treat, NNT).

The Reliable Change Index was first used in clinical settings (Jacobson & Truax, 1991), but recently has been effectively employed to evaluate the outcomes of educational interventions with children with autism (Eldevik et al., 2010; Eldevik, Hastings, Jahr, & Hughes, 2012; Remington et al., 2007). The RCI takes into account the reliability of the measures over time and the typical variation of scores within the population. So for the educational studies I conducted in Chapters 4 and 5, the RCI was utilised to provide an index of how much improvement an individual is required to make, before this can be concluded to be statistically, and therefore educationally significant for that learner. The NNT is an effect size (again that originates from a medical model) that allows us to determine the number of people who would have to receive the intervention (be treated) in order to have an impact on one additional person (Kraemer et al., 2003) than would have occurred with *treatment as usual*. Effectively, when someone receives an intervention one of three things can happen:

they can deteriorate, remain unchanged, or improve as a result of the treatment. The NNT measure when applied appropriately can give a real objective sense of the effectiveness of the intervention, which can be simply communicated to people. These two measures are described in further detail in paper 3 (p. 84-85).

Structure of Thesis and Background to the Included Papers

The following four papers, which form the body of my thesis, have either been published or are in submission for publication to peer reviewed academic journals. Whilst most of the papers were conducted in school settings, I also wanted to explore the use of these same tactics and strategies in statistic classes with undergraduates within my department. All papers sought to focus on building component skill fluency with the goal of extending participant's fluency at composite tasks that depend on the fluent performance of these prerequisite component skills.

Paper 1 was a single case design study with a small sample size, the purpose of which was to explore the effects of a brief (10-week) intervention targeted at improving the accurate reading fluency of children who were demonstrating significant reading problems. The study was conducted using two types of instructional stimuli: SAFMEDS (Say All Fast Minute Everyday Shuffle) cards and randomised word sheets. Children's progress was recorded on data sheets and plotted on SCCs. Although the amount of time that children from the PT group spent engaged in the intervention was only 12 minutes per week (range 8-16 minutes), all the children showed significant gains in their accurate reading fluency.

As I am employed as a lecturer within the university, I wanted to apply these principles and procedures to my own teaching and try and impact a problem that a significant number of psychology undergraduate students report—poor knowledge of statistics and basic statistical concepts. Paper 2 describes a study in which we investigated the effectiveness of the use of SAFMEDS and charting of individual progress on SCC would have on the performance of psychology undergraduates studying on their research methods and statistics module. Outcomes of the intervention group were measured on two dimensions; performance on a pre- and post-test of randomly selected questions from the core module textbook, and their performance in comparison with other students on their weekly statistic exam. Because of my familiarity with both single-case and group designs, I also wanted to combine these two approaches. For this I extended the analysis of pre- and post-test data to look not only for significance between group means, but also to tease out the magnitude of improvement for each individual who was a part of that group mean.

In July, 2010 I was fortunate to receive a scholarship for a place on the Summer School Institute (SSI) to study at Morningside Academy and to gain practice in using their methods. This trip was invaluable for both my personal development within my PhD and for the development of the third year module that I teach at Bangor University (Evidence-based Behavioural Methods in Education). At Morningside I saw and experienced first-hand many of the techniques that I had only previously learned about from textbooks and journal articles. The trip allowed me to deepen my understanding of a complex and fascinating subject area and as a result of this, in partnership with my mentor and PhD supervisor Dr. J Carl. Hughes, we delivered our first free teacher-training event on Thursday 2nd September, 2010 in North Wales. The training event focussed on introducing these methods to teachers currently working within local schools, so they would able to use them in their future teaching. Although we had provided such training at conferences and as part of university degree programmes, this was our first event aimed at providing such training for teachers. Since then I have continued to provide training and support to many mainstream and special educational schools in the area, helping them to incorporate the use of PT into their normal school teaching, and working toward the eventual goal of collaborating with schools to run after school Learning Centres (associated with the university) to help children who are experiencing difficulty with their essential skill fluency. The eventual aim is to have students who study at Bangor gain experience working with children as part of their curriculum. This will be beneficial in two ways: (1) the children will receive the benefits of individualised tuition and fluency building practice in skills they are not yet fluent in, and (2) the undergraduates will acquire invaluable practical experience working in a real world setting with struggling learners.

Paper 3 was aimed at increasing essential numeracy skills in primary school children. Children worked on component elements of maths to improve their fluency, but additionally the research sought to discover whether this increased component fluency would impact on their performance of a composite task that they had not practiced during the intervention. Results from this study add support for this.

The last paper in this dissertation, Paper 4, focused on learning vocabulary in second language learning classes in a large mainstream secondary school. The focus of Paper 4 differed to that of Paper 3 in that we aimed to introducing the procedures within a model that trained teaching staff to implement the interventions themselves, with less researcher involvement. Inevitably there were trade off problems with the reduction in possible methodological rigour that can exist when conducting real-life research; however there is also a benefit of programme sustainability, in the increased likelihood for teachers to incorporate the strategies into their everyday classroom activities in the future. In the last two research studies (Papers 3 & 4) I also aimed to further extend the sophistication of the analysis and the ease with which I could communicate results with a broader audience, by incorporating two new measures of individual change and efficacy: the RCI and NNT.

Chapter 2: Using precision teaching to increase the fluency of word reading with problem readers[1]

[1] This chapter has been published as: Hughes, J. C., Beverley, M., & Whitehead, J. (2007). Using precision teaching to increase the fluency of word reading with problem readers. *European Journal of Behavior Analysis, 8*(2), 221-238.

Abstract

This study describes the effectiveness of a brief intervention aimed at increasing the rate (frequency) of reading common English words for five pupils who were experiencing problems with reading. The intervention employed frequency-building procedures aimed at increasing the rate at which the children could read words, and precision teaching (PT) to monitor the effectiveness of the intervention. The children in the treatment as usual (TAU) condition received the standard reading support during the same period. All of the pupils' were tested at pre- and post-intervention on the frequency at which they could read words accurately, and all were given a standardised reading test. Following intervention the outcome measures for fluency were taken: maintenance, endurance, application, and stability. All five of the PT children increased their rate of accurate word reading, and two made significant gains on the standardised tests of reading; however, the TAU children did not improve on any of the measures. The study provides additional evidence supporting the effectiveness of PT and frequency-building procedures, and that successful intervention need not require expensive or time-consuming resources.

The ability to read is perhaps the most important academic skill. Reading is the primary medium for the transfer of knowledge in all curriculum areas (Cawley, Miller, & Carr, 1990), and as such is a prerequisite skill required for much of what children learn in schools (Barber, 1997). A child without at least an average grade-level reading ability will find it difficult to progress adequately in core curricula and is already experiencing, or is facing, academic failure. With this failure the child will almost certainly lack a number of the key repertoires necessary to enter many employment domains in a modern information driven society (see e.g., Doyle, 1983). According to Salvia and Hughes (1990) the understanding of written text determines a student's success or failure in all post primary academic subjects. Indeed, Stuart (2006) asserts that many children in the UK are not reading satisfactorily by the end of primary education and are thus unprepared to effectively engage with the secondary school curriculum.

In addition to the cost to individuals, a survey conducted for the Basic Skills Agency by Gallup (ALBSU, 1993) estimated that inadequate basic skills cost UK industry in excess of £4.8 billion a year. Furthermore, the National Assembly for Wales estimated the annual cost to the Welsh economy of poor basic skills was £588 million (The Basic Skills Agency & The National Assembly for Wales, 2001). A survey conducted by the Basic Skills Agency (1988, as cited in The Basic Skills Agency, 2001) found that 780,000 people in Wales had literacy and numeracy problems. Furthermore, by the age of 11, 22% of Welsh children had not reached the required level for English, and by the age of 14 this had increased to 38%.

The Basic Skills Agency has made a commitment to improve the standards of basic skills across England and Wales (The Basic Skills Agency, 2001). Although reading is a complex task, in a simplified sense it can be described as the ability to decode or identify words and comprehend what has been read (King-Sears, 1994). Some of the key skills in reading include, *decoding* (breaking words down into sound elements or phonemes and then blending these sounds into whole words) and *word recognition* (fluent identification of whole words); without such skills, understanding of text will be difficult. One approach to learning that has had over four decades of success in improving basic skills is *precision teaching* (PT) (Beck & Clement, 1991; Johnson & Layng, 1994; Johnson & Street, 2004; Kubina & Starlin, 2003).

PT builds frequency in basic skills and helps teachers ensure that *every child in a class* is learning successfully. PT is a general approach that can determine whether an instructional method is achieving its aims. It is not, as the name might imply, a method of teaching; however, it does encourage certain practices that are not commonly used within UK classrooms, such as short practice sprints, daily measurement of performance, and data driven pedagogical decisions. It could be more accurately described as *precision measurement* because it is primarily a sensitive, student-centred learning measurement tool (Boyce, 2003). The value of PT lies in identifying a specific skill area (e.g., reading numbers) in which the child is either failing to progress or is performing below a level considered to be competent; this is then followed by a daily session of teaching, frequency building, monitoring and evaluating progress, and adapting to this evidence in order to optimise learning (Lindsley, 1992a).

There are four foundation principles of PT: (1) The child knows best–in the sense that the child's response to a task or learning approach is the best indicator of whether the teaching method is working with that child; (2) Focus on observable behaviours–this is a practical rule so that accurate counts can be taken to monitor whether a child is improving in response to the current teaching method; (3) Use frequency measures to monitor performance–PT focuses on rate or frequency of responding which can only be measured using the number (or count) of correct and incorrect responses within a given timeframe; and (4) Use a standardised graphical display, the Standard Celeration Chart (SCC), to record performance measures and make instructional decisions. These principles have been described in detail in a number of previous publications (e.g., Calkin, 2003; West, Young, & Spooner, 1990; White, 1986; White & Neely, 2004). Some of the key methodological characteristics we discuss here are: (1) element/compound analysis, (2) frequency-building, and (3) measuring learning or performance.

(1) Element/compound analysis refers to conducting an analysis of each compound (or complex) task in terms of what pre-skills, elements, or components, are required to carry out that task (Barrett, 1979; Binder, 1996; Haughton, 1972, 1980; Kubina & Morrison, 2000; McDowell & Keenan, 2001b); in the current paper we will use the terms *elements* and *compounds* (Kubina & Morrison). Precision teachers state that one of the main reasons a person begins to experience problems in learning a task is because he or she is dysfluent at some of the prerequisite elements of that task. For example, if a child is dysfluent at decoding some of the basic sounds of English words, he or she will likely experience problems when reading novel words that contain those elements. According to Binder (1996) acquiring fluency in prerequisite key skill elements can improve the future learning of new skills that are dependent on these prerequisites.

(2) Frequency-building is a method used to develop both *accuracy* and *appropriate speed* in key elements (Binder, 1990, 1991). Being fluent at a skill has been described as “automatic”, “effortless”, “smooth” and “second nature” (Kubina & Morrison, 2000, p. 89). Accuracy at appropriate speed is vital in assessing performance levels because it is a significant indicator of expertise and therefore can help us discriminate between performers who have, or have not, mastered a skill (Binder, 2003; Chiesa & Robertson, 2000). For example, two children might score equally on an exercise using a percentage correct measurement system, but one of them might have taken thirty minutes to complete the exercise and the other only five; it could be said that the latter child is more accomplished;

however, when using percentage correct we cannot distinguish between levels of mastery beyond 100% accuracy. Despite this limitation percentage correct remains the prevailing unit of measurement in most educational settings. Frequency is the measurement system used in PT and refers to count in time, or the number of responses during a specified timeframe; in PT this is often 1 minute. This aspect of PT has its origins directly from one of Skinner's main contributions to the scientific study of behaviour: adopting *rate* (a synonym of frequency) as the basic datum of the science of behaviour.

Within the PT framework the objective is to move the learner toward skill mastery on specified curriculum sequences. When the learner reaches a fluent performance on a skill, particular *learning outcomes* are expected, such as, retention and application of skills and knowledge even in the absence of instruction. This list was later extended so that the learning outcomes of fluent performance included *retention*, *endurance*, *stability,* and *application*—captured in the acronym RESA (Fabrizio, 2004; Fabrizio & Moors, 2003; Johnson & Layng, 1992)**.** According to Binder (1996) *retention* refers to "the relation between behavior frequencies at two points in time, between which the individual has had no opportunity to emit the behavior" (p. 164). The term *endurance* specifies how well a person can perform over prolonged periods of time (Binder et al., 1995). Without endurance, pupils are unable to concentrate for extended periods, are increasingly likely to make errors, and are likely to perform negative emotional behaviours (Binder, 1996). *Application* has been described "as the convergent relationship of elements and compounds" (Kubina & Morrison, 2000, p. 92). And finally, *stability* refers to maintenance of performance in the presence of distractors (Lindsley, 1995a).

(3) Measuring learning or performance. PT is fundamentally a method of measuring learning that can help teachers make timely decisions about the effectiveness of teaching for each child. Practice sprints or time probes are often used and offer an effective way of

measuring a pupil's frequency or rate of performance on a particular skill. The child's performance might be measured for one minute, or alternatively, the time probes can be adapted depending on the complexity of the task. The count is nevertheless converted (either by division or multiplication) to give a count per minute (frequency or rate) score. The main point about these practice sprints is that they are a very efficient way of ensuring massed practice opportunities (aimed at building fluency) and allowing objective decisions about whether current teaching methods are improving learning (measuring student performance). As mentioned, PT involves the use of SCC to display performance data obtained from timed probes. The SCC has been described in detail in a number of previous publications (Calkin, 2003, 2005; Graf & Lindsley, 2002; Pennypacker et al., 2003; White & Neely, 2004).

Informal testing often occurs implicitly within the day-to-day practice of teaching in many schools; however, the main way that PT differs from this is that the measurement probes are very short and precise, the teacher (or child) take data daily, and, importantly, use these data to make instructional decisions at least on a weekly basis. Over successive days of practice, the data (corrects and errors) plotted on SCC produce *learning pictures*. When used effectively these learning pictures can indicate, for each child, whether a task is too difficult, too easy, or just right; whether the child requires further instruction, further practice, or should move on to more complex tasks; and, most importantly, whether the teaching methods being adopted are having the desired effects. In short PT provides teachers with all the information they should require to make effective decisions about learning.

In this way the PT approach helps the teaching process become more effective in three ways: it becomes highly *reactive*–in the sense that teachers can identify learning problems as they happen for every child; it becomes *objective*–in the sense that the teacher knows precisely how every child is performing on the particular skills they are practicing; and it becomes highly *individualised*–in the sense that very quickly each child is working only on skills upon which he or she require practice. These also help defend PT from accusations that it is similar to “drill-and-kill” and other out-dated educational practices.

Thus optimal learning can be achieved through short frequent measurements that inform teachers about the learning progress of their students and guide them to alter learning strategies effectively until adequate progress is maintained. We describe the approach as *navigated learning*, because the teacher uses the information from these measurements to navigate the child through the learning sequence in the fastest way possible, in a similar manner to the way a captain of a ship uses a compass, or a coach helps an athlete improve performance.

Much research has confirmed the effectiveness of PT to improve pupils’ academic performance. An example of this is research conducted by Chiesa & Robertson (2000). Five children underwent a twelve-week programme drawing on the principles of PT. Prior to the programme the five children were the lowest performers in the class on long division. They worked on elements of long division, practiced to build frequency, and progressed at individual rates through the programme; during these periods the rest of the class had a normal mathematics lesson. Following the programme the five PT children out-performed all but one of their peers on a test of long division despite having spent no time working on long division during the 12-weeks.

The Morningside Model of Generative Instruction (Johnson & Layng, 1992; Johnson & Street, 2004) is another example of the effectiveness of PT. This is a programme implemented at the Morningside Academy, Seattle, and uses PT on a daily basis to ensure children are learning rapidly and becoming fluent at key skills. Many pupils who attend the academy are typically labelled as having ‘learning disabilities’ or ‘attention deficit disorder’ when they arrive. Morningside report that over a 10-year span, pupils have consistently advanced during one academic year anywhere from one to almost four grade levels in reading, language, arts, and maths. This is especially impressive considering these achievements come from students who prior to attending Morningside were progressing at less than one year per academic year. In the UK, the *National Literacy Strategy* has recommended similar optimistic aims. They state that children who are behind in reading should be progressing at twice the rate of their peers to ensure there is a narrowing of the educational gap; if this kind of performance improvement is not achieved by the children, then the school should provide more effective teaching (DfES, 2003).

The current study explored the effects of a 10-week programme designed to increase the accurate word reading frequency with five pupils whose reading age was significantly below their average peers. The words targeted for this brief intervention were high frequency words that make up over 80% of children's literature (Maloney, Brearly, & Preece, 2001). These selected words, once learnt, will potentially enable pupils to engage with a broader range of reading materials. Very brief teaching interventions will allow teachers to assess whether PT techniques can be successfully implemented within a mainstream setting with children who have significant reading problems.

Method

Participants

Seven children (11-12 years old) participated; all were from a secondary comprehensive school in North Wales. The children were randomly assigned to either the PT intervention (n = 5); or treatment as usual (n = 2), who received the school's standard reading support during the allocated time. The ages of the PT children ranged from 11 years 6 months to 12 years 8 months (Mean: 12 years 2 months). The ages of the children who did not receive PT procedures (here known as the TAU children) ranged from 11 years 7 months to 11 years 11 months (Mean: 11 years 9 months). All seven children were selected because they were having significant problems reading and were receiving remedial support that took place in the school's Supportive Studies classroom.

Prior to beginning the study, consent was obtained from the School Board of Governors, the head teacher, and the child's parents. Ethics approval was obtained from the School of Psychology Ethics Committee, Bangor University.

Setting

All seven children had been assigned a teaching assistant (TA) for the period—a 20-minute session, 3 to 4 times each week. The children attended the Supportive Studies classroom to participate in this study in place of attending the scheduled school assembly time.

Stimuli and Apparatus

Digital timers were used to time practice sprints (e.g., 1-min, 30-sec timing periods). Error words were practiced using white-boards and non-permanent marker pens.

Other materials included standard celeration charts (SCC) and data recording sheets. The SCC were used to record each child's best scores on a session-by-session basis. One data sheet was used to record all of the data for each session the child attended: date; name of the child; the skill being practiced; the aim (i.e., the frequency goal the child was aiming for); the score obtained for each time probe (i.e., the number of correct responses and learning opportunities); length of each time probe (e.g., 1-min, 30-sec); and any comments, including the words read incorrectly within each practice session.

Instructional and Testing Materials

There were 2 types of instructional stimuli: word sheets and SAFMEDS (Say All Fast a Minute Every Day Shuffle) cards.

Word sheets. There were two word sheets: words 1-220 and 221-420. The first 1-220 words were obtained from the Dolch list (Dolch, 1948).The remaining 200 words were taken from the most common vocabulary words in English (Maloney et al., 2001), except for those already in the first 220 Dolch list. Each sheet contained a randomised list in Times font, size 14, double spaced, and were printed double-sided on A4 paper, with 110 words on each side. The randomised sheets were produced in Microsoft Excel© (see Appendix A).

SAFMEDS. Initially we produced 420 SAFMEDS that were 9 x 4 centimetres in size and each had a common English word printed on one side. The words were the same as those used for the word lists previously described.

Each child had a folder that contained all the materials and sheets they required for their sessions.

Design and Measures

The study used a single subject design were each individual child receiving the PT intervention was monitored on each practice attempt throughout the study. In addition, a number of measures of reading were taken prior to and following the intervention for the PT children in order to assess whether the reading practice would have effects on general reading ability as measured on the school's standard scales. The TAU children were also tested at the same points to monitor effects of the standard reading support over the same period of time.

In addition to being used to gain pre- and post-test scores, the information from these tests conducted prior to the intervention was used to select the children who participated. The two standardised tests used to select the children were the Group Reading Test II (GRT II) (NFER-Nelson, 2000) and the Vocabulary Scale of the Middle Years Information System (MidYIS) year 7 (CEM, 2007); both tests are standardised measures and are commonly used in schools throughout the UK. The GRT II is administered every 6 months; it measures both decoding and reading comprehension (see Table 2.1). The participants had been tested two months before the PT programme began and were retested one month after the PT programme ended.

For the purposes of this study we used the Vocabulary Scale section of the MidYIS test, which measures pupils' reading and vocabulary skills; the participants completed the year 7 version of this test to correspond with their year of study. The national average score for pupils of the same age is 102.3 and equates to a grade B on the Vocabulary Scale. The scores for all 7 children on both the GRT II and the Vocabulary Scale of the MidYIS year 7 test reflected reading ages considerably below their chronological age.

Table 2.1

The Chronological Ages (Years:Months), Reading Ages measured using Group Reading Test II (GRTII), and the scores and grades on the Vocabulary Scale of MidYIS Year 7 tests for the PT (P) and Treatment as usual (TAU) children.

PT (P) TAU	Gender	Age (Y:M)	GRTII Reading Age (Y:M)	MidYIS (Year 7)	MidYIS Grade
P1	Male	11:04	7:03	77	D
P2	Male	11:05	6:09	89	D
P3	Male	12:05	6:10	80	D
P4	Female	12:06	<6:05*	83	D
P5	Female	12:03	<6:05*	67	D
TAU1	Male	11:09	<6:05*	79	D
TAU2	Male	11:05	6:07	75	D

*The test is unable to measure reading ages below 6 years 5 months.

To test for reading of words in real texts *(application),* three passages selected from books written for children between the ages of 7 and 9 were presented to each participant following the PT programme. Each passage was printed on A4 paper as they appeared in the books from which they came. The first passage, taken from *I Was A Teenage Goldfish* (Shipton, 1996; reading level 7-8 years), contained 224 words, 60% of which were in the 1-220 list. The second passage, taken from *Legendary Places* (Broker, 2000; reading level 8-8.5 years), contained 225 words, 60% of which were in the 1-220 list. The third passage, taken from *The Magic Boathouse* (Llewellyn, 1994; reading level of 8.5-9 years), contained 222 words, 62% of which were in the 1-220 list.

Procedure

The children attended approximately three 20-minute sessions a week over a 10-week period (excluding school breaks). During these sessions the PT children conducted timed practice sprints aimed at building frequency in reading common words. The TAU children were given the same time, and each supported by a teaching assistant (TA), to work on reading; during these sessions the children would pick a book and read this with the TA who was allocated to them. There was no other instruction given to the TAs working with the TAU children and the session typically comprised the TA listening to the child read, and correcting the child when they read words incorrectly. Beyond this we did not take any further measures on the TAU children apart from the standardised school tests as previously described.

For the PT children, on average, three time probes were taken per session (range 1 to 6). Each child sat at a separate table in the same classroom and worked with either a TA or a researcher. At the beginning of a typical session each child collected their personal folder containing their SAFMEDS cards, word sheets, charts, data sheets, and a digital timer.

During the programme, the child either read words from SAFMEDS or word sheets. These were alternated in the following order: SAFMEDS 1-220; word sheets containing words 1-220; SAFMEDS 221-420; word sheets containing words 221-420. SAFMEDS were split into smaller packs consisting of between 50 to 100 cards.

Phase I: SAFMEDS. Initially, each child was issued with the first pack of SAFMEDS. The AIMS for reading SAFMEDS was ≥70 correct with no more than 2 errors per minute (denoted as 70/2). Before commencing each practice sprint the child was reminded of the frequency aim and encouraged to ‘try to get a personal best’, or a ‘PB’. In this manner the teaching quickly focused the reinforcement contingencies and the child's attention on self-competition. This aspect of the PT approach encourages a child to experience success whatever their performance relative to others in the class.

The child was asked to read each card as quickly as possible, skip responses that were too difficult, and not to give up during a timing period. The child was asked to shuffle the pack and the timer was then set (to either 1 minute or 30 seconds). Usually the timings for each child began as a 1 minute timing and this was only altered to a different time period if he or she failed to progress. The child was then instructed to start reading once the timer had initially sounded and to continue reading until they heard the timer sound to signal the end of the timing period. As is common in PT procedures the child was provided with more opportunities to respond (words to read) than could be completed in the time period; in this way the procedure did not place an artificial ceiling, or Fluency Blocker (Binder, 1996), on the frequency of responses possible within the allocated time. The TA praised the child following all practice sprints.

During the practice sprint the TA or researcher instructed the child to place all words read correctly in a pile on his or her right hand side, and all the words read incorrectly or skipped (the child did not know how to read the word) in a separate pile on his or her left hand side. Within the sessions all incorrect responses were referred to as 'Learning Opportunities' or LOs.

For the error correction procedures, the TA wrote down all the LOs and the child practiced them for approximately a minute following each practice sprint. Error correction involved either decoding the word with the aid of the TA and then practicing this several times out loud, or, for irregular words, spelling and writing the word out on the child's white board and being required to read the word on each occasion.

Phase II: Word sheets. Once the child had reached the frequency AIM for the 220 SAFMEDS, he or she moved on to the word sheets that contained the same 220 words printed in a random order. The child read across the page from left to right as they would in reading a book. Each time the child attended a session he or she was issued with new randomised word sheets. The AIM range for the words sheets was 120-80 correct for three consecutive timings (Kubina, 2002). The TA had an identical sheet and followed the child as they read, marking any LOs with a pencil mark.

Data collection. Following each practice sprint, the child then counted the number of correct and LO responses (involving either SAFMEDS or a word sheet). Under the supervision of the TA, the child then recorded each score pair (i.e., number of corrects and number of LOs) on the data sheet and then charted the best score pair of that session on the corresponding calendar day on the SCC. (Each score pair was calculated as rate per minute before being charted on the SCC; for example, the score was multiplied by 2 if a 30 second time probe was used, so that all scores were displayed as count per minute data).

Learning pictures and decision criteria. The SCC presents a learning picture after several timings. The researcher assessed these daily in order to make decisions about progress. Different learning pictures revealed areas of strength and weakness, and allowed instructors to adapt instructional materials for each child accordingly. Thus, the programme was individualized dependent on the learning picture emerging from the child's daily performances over time (i.e., learning or *celeration)*. If the child was progressing and increasing the rate of reading we continued with the practice. If they were failing to progress we changed some aspect of the teaching, either the number of cards or the timing period was shortened (e.g., to 30 seconds). Children were moved on to the next learning task when they had maintained AIM across three successive attempts.

Measuring fluency. Fluency was measured following the 10-week programme by testing retention (PT children only), endurance, stability, and application; additionally all children completed the GRT II. The retention test occurred on the first available opportunity between 5 and 9 weeks from the last day the child worked on word sheets containing the words 1-220. Endurance, application, and stability, were tested within three school days of the end of the teaching programme.

Retention. Retention was tested using the first 220 words from the Dolch list. Each child underwent two 1-min timings, and, as before, his or her best score was charted.

Endurance. Endurance was calculated as 3 times the duration of the typical time probes (i.e., 3 minutes). The child was presented with every word previously presented during the programme using three randomised word sheets. The child was required to read as many words as possible within 3 minutes without a break.

Stability. During this test the child read as many words as possible from a randomised word sheet containing words 1-220 in a busy classroom with a radio playing in the background for a 1-min timing.

Application. Each of the three passages chosen was suitable for a different reading level: 7-8 years, 8-8.5 years, and 8.5-9 years. The passages contained 224 words, 225 words, and 222 words, respectively. During the test, the child was instructed to read aloud each passage as if he or she were reading a book on their own; however, the child was not instructed to read the passage as fast as they could; the application tests thus varied in timing depending on how long the child took to finish the passage.

Inter observer reliability. We checked the reliability of the data in two ways. Decisions to move a child to a new learning task once they had reached their aim for that task were verified in the presence of a TA and one researcher. All of the tests for learning outcomes were double coded and inter observer agreement was calculated as 97.5% across all the tests.

Results

Individual Data

All five participants' reading frequency of the selected words showed improvement over the 10-week intervention. Figure 2.1 shows a typical example of the daily best score pair (corrects and LOs) for a participants' reading performance during the PT programme and on post-programme fluency measures as plotted on a Daily per minute standard celeration chart. We used the SCC to monitor, or measure, the change in performance in response to the practice sprints.

Fluency Measures

Pre- and post-measures of reading frequency. Figure 2.2 shows the difference between the participants' frequency scores obtained on the first day of the PT programme, their end frequency scores, and the highest score during the programme. For the PT children the end score was their retention score, and for the TAU children it was taken at the end of the teaching programme. It can be seen that each participant's reading frequency increased following the PT programme. With the exception of Participant 5, each participant read over twice as many words per minute during the retention test compared to his or her score obtained on the first day of the PT programme, thus demonstrating significant increases in rate when reading high frequency words, even following a period without direct practice opportunities. As can be seen from the data for the TAU children, engaging in their usual reading support had little effect on their reading rates of the high frequency words. Generally, the PT participants who produced higher rates of correct responses during the programme produced higher rates of correct responses during maintenance tests, except for Participant 4 who scored higher than Participant 3 on the test.

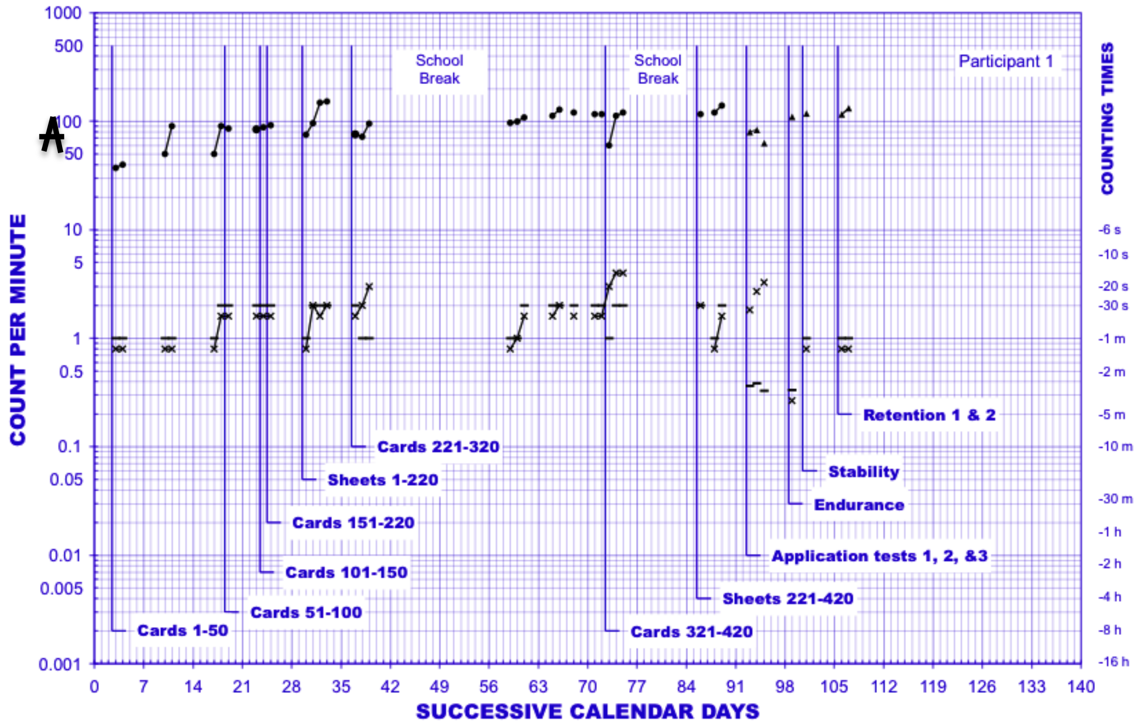


Figure 2.1. A daily per minute standard celeration chart showing the best daily score pairs for Participant 1. The Y-axis indicates frequency (count per minute) on a multiply-divide scale ranging from .001 per minute (1 in 16 hours) to 1000 per minute. The X-axis is a calendar scale indicating day 1 – 140. Correct responses per minute are denoted by dots and incorrect responses by Xs. The black dashed lines (-) represent the counting period (timing floor) and denote the length of the practice sprint: a line on the ‘1’ indicates the counting period is 1 minute; a line on the ‘2’ indicates the counting period was 30 seconds (i.e., 2 x 30 seconds in 1 minute). All data were based on count per minute data. A score below the counting floor represents zero responses.

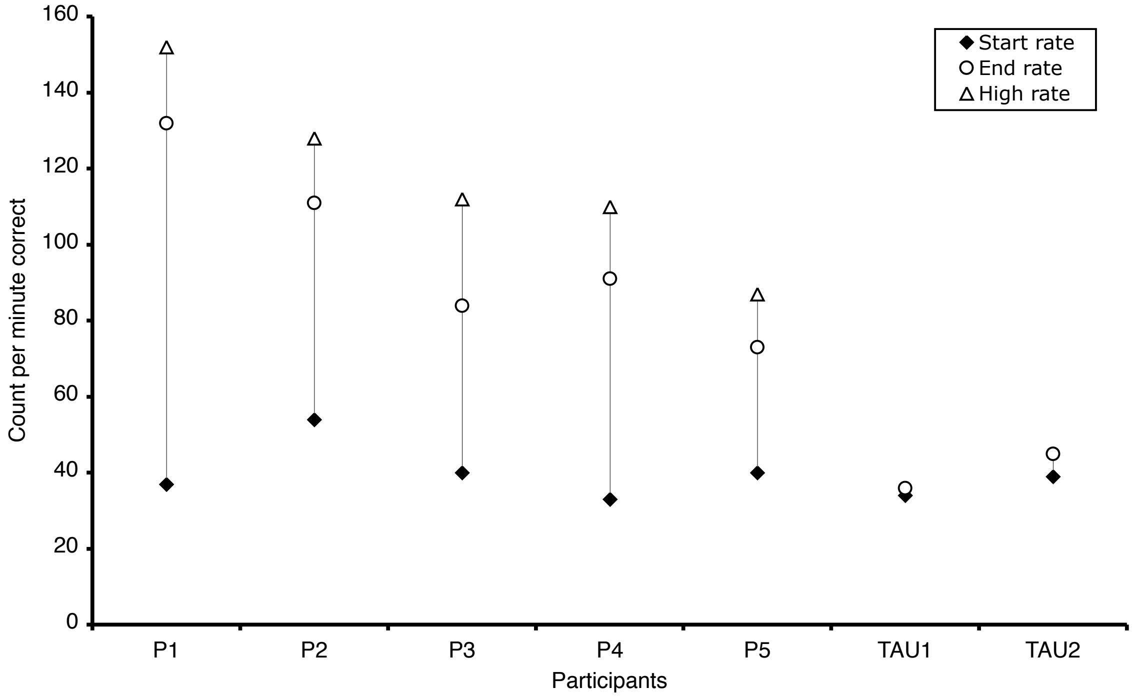


Figure 2.2. For each participant (P = PT; TAU = Treatment as usual) the fluency scores (correct only) for the start rate, the end rate, and for the PT children only, the highest rate achieved during the intervention.

Endurance. Table 2.2 shows the correct and LO frequency obtained during the 3-minute endurance test. With the exception of Participant 4, the children who obtained higher frequencies of word reading during the PT programme also obtained higher frequencies during the endurance test.

The PT children read, on average, 4 times more words correctly per minute than the TAU children during the endurance test. The PT children also read at a faster rate relative to the TAU children: PT children read between 57.3 and 110.7 correct responses per minute (range 53.4), whilst in comparison TAU children read between 17.7 and 28 correct responses per minute (range 10.3). Even though the PT children read faster, their error rates were still lower than the TAU children. The TAU children read at a slow rate relative to the PT children but also made many errors: PT children obtained between 0.2 and 4 errors per minute (range 3.8), whilst in comparison TAU children read between 4.7 and 38.7 errors per minute (range 34.0).

Stability. The children who obtained higher frequencies during the PT programme attained more stable learning on the test. The PT children read, on average, 5 times more words correctly during the stability test than the TAU children. The PT children also read at a faster rate relative to the TAU children. With the exception of Participant 2, the PT children made fewer errors than the TAU children even though they were reading faster: PT children read between 76 and 116 correct responses per minute (range 40) whilst the TAU children read between 2 and 37 correct responses per minute (range 35). PT children obtained between 0 and 7 errors per minute (range 7) and the TAU children read between 5 and 47 errors per minute (range 42).

Table 2.2

Shows for the PT (P) and Treatment as usual (TAU) children, the correct and LO rate scores on both the endurance and stability tests (corrects/ LOs), the GRT II scores reflecting reading ages at pre- and post-programme (years:months); and, for the five PT children only, the number of practice sprints, the total active time (hours:minutes), and the average active time per week (minutes).

PT (P) TAU	Endurance (corr./ LOs)	Stability (corr. / LOs)	Pre-GRT II (Y:M)	Post-GRT II (Y:M)	No. practice sprints	Total time (hr:min)	Av. time / week (min)
P1	110.7/0.2	116.0/0	7:03	6:10	91	2:37	16
P2	96.7/3.3	107.0/7	6:09	7:09	62	1:50	11
P3	70.7/1.3	90.0/1	6:10	6:10	81	2:29	15
P4	92.0/1.0	87.0/4	<6:05*	7:01	43	1:17	8
P5	57.3/4.0	76.0/2	<6:05*	<6:05*	61	1:45	11
TAU1	28.0/4.7	37.0/5	<6:05*	<6:05*	–	–	–
TAU2	17.7/38.7	2.0/47	6:07	<6:05*	–	–	–

*The test is unable to measure reading ages below 6 years 5 months.

GRT II. Pre-programme GRT II scores reflected reading ages ranging across the participants from below 6:05 (years:months) to 7:03 (see Table 2.2). Mean scores for the PT children increased from 6:09 (a precise pre-programme mean score is prevented due to the test's inability to measure a reading age below 6 years 5 months) to 7:00 post-programme. If Participant 1 is excluded from the analysis, the PT children would have increased on average from 6:08 to 7:01. The TAU children's mean reading age actually decreased from 6:07 to 6:06, or below. The PT children's mean reading age increased by at least 3 months, or 5 months if Participant 1's scores are excluded. In contrast, the TAU children's mean reading age decreased by at least 1 month.

Application tests. These texts were previously unseen and contained approximately 60% of the words that had been targeted during the PT programme (Tests 1, 2, and 3 comprised 60%, 60%, and 62% of the words presented in the PT programme, respectively). Figure 2.3 highlights the difference between the PT and TAU children when we isolated the words that had been targeted in the PT programme for frequency building, termed *target words*. Figure 2.3 shows both correct *and* incorrect responses for target words included in all three tests. The PT children read all words previously presented in the PT programme correctly (see Figure 2.3), except for Participant 5 who read one of these words incorrectly (the word *saw*). In contrast, the TAU children read 370 target words incorrectly. TAU Participant 1 read 17%, 9%, and 15% of these words incorrectly on application tests 1, 2, and 3, respectively. TAU Participant 2 read 60%, 74%, and 51% of these words incorrectly on application tests 1, 2, and 3, respectively. Thus the PT children showed they could read these words when they occurred in the context of a novel, age appropriate text.

Figure 2.4 highlights the difference between the PT and TAU children when we isolated the words that had been not been targeted, termed *non-target words*. Application tests 1, 2, and 3 contained 40%, 40%, and 38% of non-target words, respectively. The PT children made more errors (LOs) when reading non-target words than target words. Participant 3 obtained the most number of LOs for non-target words, reading 11%, 13%, and 23% of them incorrectly on application tests 1, 2, and 3, respectively. The TAU children obtained, on average, 15, 10, and 8 times more LOs for *all* words per minute on application tests 1, 2, and 3 respectively than the PT children. The PT children read at a much faster rate than the 2 TAU children, yet their error rate was considerably lower.

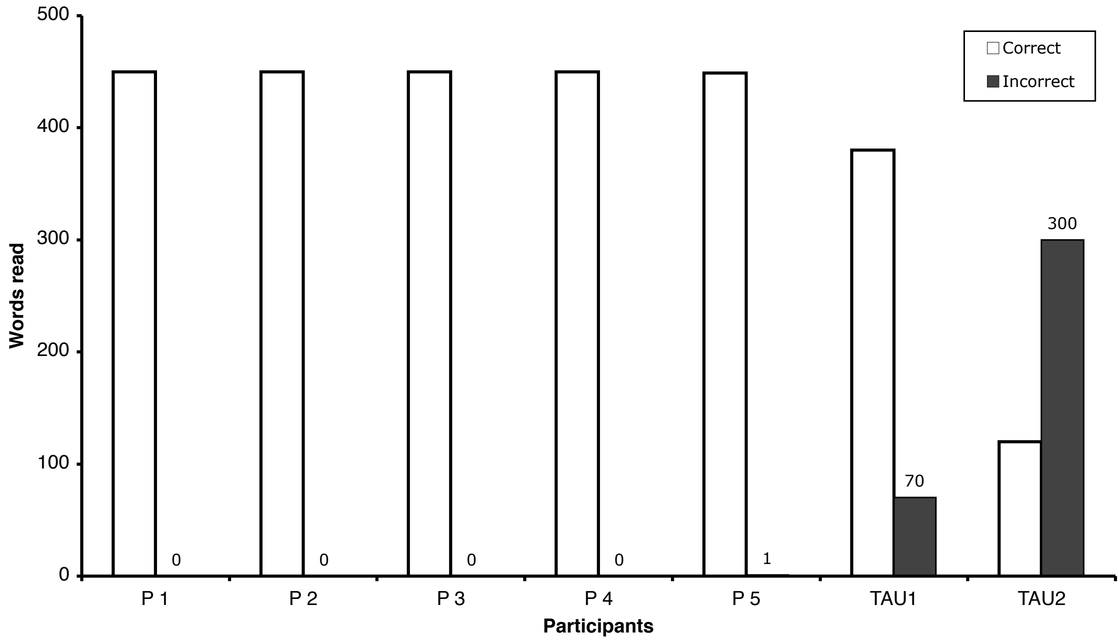


Figure 2.3. Number of target words (i.e., those covered in the PT reading programme) read correctly and incorrectly by each of the seven participants (P = PT; TAU = Treatment as usual) across the three application tests. (NB: Y-axis scale is from 0 -500).

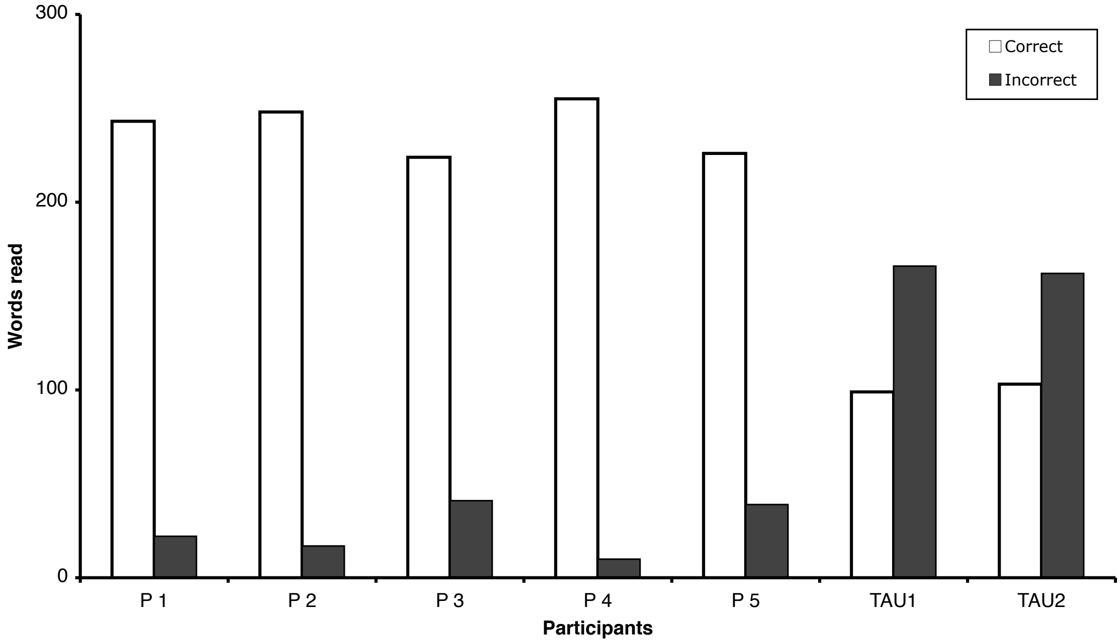


Figure 2.4. Number of non-target words (i.e., those not covered in the PT reading programme), read correctly and incorrectly by each of the seven participants (P = PT; TAU = Treatment as usual) across the three application tests. (NB: Y-axis scale is from 0 -300)

Time engaged in intervention. The amount of time each PT participant spent engaged in the programme was calculated by adding the total duration of time each pupil carried out time probes to the total number of time probes they undertook (assuming for each time probe taken, that a minute was spent practicing LOs; see Table 2.2). The average weekly time spent engaging with the targeted words was 12 minutes (range 8 to 16 minutes) per week for 10 weeks. Participant 1 spent the most time actively engaged with the PT words. This participant also obtained the highest frequencies on the retention, application, endurance, and stability tests, and achieved the highest frequency during the programme.

Discussion of Chapter 2

The aim of this study was to evaluate the effectiveness of a brief intervention using practice sprints and PT procedures to increase the reading frequency of common words. All seven children in the study read significantly below their average peers and were at serious risk, not only of experiencing continuing problems in reading but also facing future academic failure. Both the PT and the TAU children worked on reading skills for the same amount of time during the study. The difference between the two was the type of reading activities engaged in during the sessions: the PT children concentrated on frequency-building exercises with common words and self-paced progression (based on decisions made by the researchers analysing each child's progress using the SCC), while the TAU children received standard one-to-one reading support from a TA.

For all seven children we took measures both pre- and post-intervention on frequency of reading the Dolch list words, standardised measures using the GRT II, and measures of endurance, and stability. Additionally, for the PT children only, we measured the active engagement time, and retention of their reading frequency approximately five weeks after they had initially reached frequency aims on the Dolch list words.

All of the PT children increased their reading word frequency on the target words significantly, thus demonstrating that brief practice sprints offer an effective, efficient, and easily administered procedure for helping children with significant reading problems. The increase in frequency also resulted in concomitant improvements on outcome measures of fluency. Indeed, from the measures we took the intervention seemed to have positive effects on all aspects of reading performance. The data from our TAU children, who received their standard form of reading support seemed to show little effect on their reading frequency; this of course may have been expected for two reasons: first, because this is the form of reading experience that they had been exposed to prior to the study and it had obviously not been successful for them; and second, the standard teaching approach, as with many schools, did not focus on short practice sprints that concentrate on building rate of reading high frequency words.

The application tests consisted of prose taken from real books and were designed to test real life reading skills. When we analysed the difference between the PT and TAU children on their success on reading both the target words and non-target words we found that the PT children read almost every target word correctly in real prose; on average the TAU children read only 69% of these words correctly. For the non-target words, this difference in error rate was also evident; the PT children made significantly fewer errors on non-target words than did the TAU children, who were clearly struggling to read material graded for children from 7 – 9 years old.

When we compared how the children did across the two groups of words we found that the PT children made many more errors on the non-target words than they did on the target words. We could speculate why this may be the case. One possibility is that the non-target words were more difficult to read than the target words and thus we would expect this disparity, especially given the problems these children were having with reading. It was certainly the case that the non-target words were relatively less frequent in reading material. However, the data from the TAU children would suggest this is not a clear cut issue because they had made a similar percentage of errors on both target and non-target words. It seems likely then that the reason the PT children could read the target words is simply because they had learned and practiced them.

A second possibility is that children who are able to read the most frequent words fluently would find reading in general easier than those who struggled even to read common words that appear regularly in prose (e.g., where, because, while). It may have been that the PT children had started to engage more with the reading process in general. Indeed, anecdotal written reports from the TAs who supported these children had listed that Participant 2 and Participant 4 had noticeably improved their reading skills in other lessons, and Participant 4 had volunteered for the first time to read aloud in her regular classroom. The benefits of such a programme to children struggling to acquire reading skills are evident in the tables and figures presented in this paper. However, they do not show the PT participants' changing relationship to their reading as the programme progressed. Before the programme's commencement, it was evident from both witnessing the children reading and from their reading scores on standardised tests that their reading was extremely poor, and that engaging with reading materials was difficult and frustrating for these children; such conditions are likely to establish escape and / or avoidance of reading as a powerful reinforcer.

When we consider how this intervention may have had its effect, there are five issues. First, the procedure helps the child focus on small manageable units of information to be learned. In this way it may help to make the learning situation less aversive for children who have a long history of failure with reading and more generally within the school setting. Second, because frequency-building gives the opportunity for massed practice it gives children with learning issues time to practice–time that the normal classroom pace would not typically allow. These procedures thus encourage high active student responses (Barbetta, Heron, & Heward, 1993). Third, the PT procedures encourage self- rather than peer-competition; for children who are struggling to learn it can be a liberating experience not to be compared, either explicitly or implicitly, with other often more successful peers. Fourth, the PT procedures adapt to an individual child's learning and so ensure that the child is making progress and not stagnating without improvement—this is motivating for both the children and their teacher. And finally, the teaching procedures themselves are different from those the children had previously experienced; they require high levels of activity, children could follow their daily progress on their charts, and they were motivated to beat their own previous personal best scores. They also enjoyed using the SAFMEDS cards as an alternative learning channel for practicing.

The GRT II results concluded that the PT children's mean reading age had increased by 3 months (or 5 months if we exclude Participant 1 from the analysis). In contrast, the TAU children's mean reading age decreased by at least 1 month. Participant 1 was the only child from the PT children whose score decreased on the GRT II standardised reading test. This was an unexpected result because he had obtained higher reading rates than any other child in the intervention. Anecdotally, the teachers had noticed a recent fall in the child's commitment to school and the verbal reports of the tester confirmed that he had not engaged with the test. Whatever the reason for his decrease in scores on the standardised reading tests, this decrease was at odds with both expectations and his demonstrable increase as measured on the other outcome tests.

Further benefits of the PT programme include the minimal time and materials required for its implementation. Raybould and Solity (1988) noted that PT procedures may turn out to be costly both of time and resources. However, the present study suggests that a brief intervention using PT techniques can be very effective in terms of enabling learning but relatively inexpensive in terms of the resources required in both time and materials. The children spent, on average, only 12 minutes per week engaged with the targeted words. The only resources required were straightforward to prepare. These comprised data sheets, word sheets, and SAFMEDS. Master copies were filed, and could be recopied for later use with other children. Other studies support these findings, reporting that frequency-building procedures increase teachers' instructional time, not reduce it (Beck & Clement, 1991; Kubina & Morrison, 2000). In addition, frequency training has been reported to ensure retention, endurance, application of skills and knowledge even in the absence of instruction (Binder, 1996). These outcomes are all likely to reduce teachers' time requirements: relearning, refreshing and reteaching would be necessary if such skills were not retained.

One of the limitations of this study was that we only tested for application of these skills post-intervention; it could be argued that there may have been significant differences between the PT and TAU children prior to the intervention. We decided not to test for application prior to the study because these children were finding reading normal texts of this level extremely challenging and aversive at that point—our primary goal was to first create the space where they could begin to succeed with reading.

Although we talk here of the differences between the PT children and the TAU children there are a number of possible confounds in the present study that would require further research to clarify. We have mentioned that the TAU children received the standard school reading support for the same *time* as the other children were conducting frequency-building practice. From this data we can say that the TAU children did not improve their reading as measured on the tests included here and with the reading instruction they received (i.e., 20 minutes a few times per week for 10 weeks). However we do not have data on the type of learning experience they received during this time. For example, we do not have information on the response rate during these sessions, the rate of reinforcement, or the type of instruction. It could be the case that the children in the PT intervention received higher rates of reinforcement and conversely the TAU children may have been exposed to aversive conditions associated with a high effort task. In fact this is highly probable. It is thus difficult for us to directly compare the PT children to the TAU children systematically because we did not control or document the experience of the children in the treatment as usual condition. This should be a consideration of future research.

We also need to make clear that this is not a comparison study that sought to compare treatment under one condition with treatment under another, typical of a between group comparison design. This was clearly a single subject design study that used multiple individuals and examined the individual children across the 10-week study. In this case the individual children acted as their own control and the pre- and post-scores across the various measures should be taken on an individual basis. And lastly, because the numbers in the present study were small we cannot say with confidence that the TAU children were similar to the PT children on variables important in reading. What we can comment on is the difference of each child on the pre- and post-tests and how this difference correlates to the particular experience they received.

The data we present here suggest that those children who achieved the highest frequencies of word reading during the intervention were also the children who tended to score highest on our outcome tests of RESA (or fluency). We could conclude that this is evidence that higher frequencies lead to better learning outcomes; however, in the present study we did not control for the number of practice opportunities that each child had. Therefore it is difficult to tease apart whether the important factor in better learning outcomes is rate *per se,* or more opportunities to practice a skill.

Doughty, Chase, and O'Shields (2004) recently initiated a new dialogue about the relationship between rate building and the outcomes of fluent performance (RESA). They argue there is insufficient empirical evidence to support claims made by precision teachers because previous research has not adequately controlled for either amount of practice or rate of reinforcement. From the ongoing debate a number of recommendations have emerged for the design of future research (Binder, 2004; Chase, Doughty, & O'Shields, 2005; Kubina, 2005b). Further research is required to compare learning outcomes following frequency-building procedures with pace-restricted or discrete trial procedures that have the same number of opportunities to respond. However, such research would have to test whether the pace-restricted learning condition had enabled learners to become fluent as an outcome of the learning procedure prior to testing for RESA (Binder, 2004; Johnson & Street, 2004). At present the limited research that has compared the two learning approaches has not controlled for this confound (Bucklin, Dickinson, & Brethower, 2000; Kim, Carr, & Templeton, 2001). Additionally, further research should also examine latency between the presentation of a stimulus and the response from the learner to ascertain whether pace-restricted learning procedures result in a significant reduction in response latency, and thus the potential to respond at a high rate under situations where the discriminative stimulus is not restricted by the instructional procedures; that is, under free operant conditions where there is freedom to present stimuli and freedom to respond from the learner (Barrett, 2003).

The current study involved five children in the intervention who were between the ages of 11 and 12. There is a need to replicate this form of short intervention because it represents an efficient and procedurally straightforward way to help children build frequency on reading words. The procedure should be extended to different age groups and may be a timely intervention for those children who are beginning to show evidence of reading problems earlier in primary education.

The words selected for this research were the 1000 most common in reading literature, and make-up over 80% of all reading literature (Maloney et al., 2001). During the brief intervention the children covered between 220- 420 of these words. If the same level of teaching had continued, we could speculate that these children could have achieved good fluency levels with all 1000 targeted words in a period of between six to eleven months (P1 in less than six months; P2, P3, and P4 in less than eight months; and P5 in less than eleven months). We could further speculate that if they had received the same level of frequency-building procedures for approximately 20 minutes per day each school day, as opposed to a few times per week, we may have expected them to achieve these levels in a period of between four to six months of teaching. We could speculate that a child who is fluent with at least these words, and who had a speaking and listening vocabulary (i.e., understood these words in normal conversation), could successfully negotiate many of the reading tasks required both in and outside of the school environment. This of course will require further research to corroborate.

The significance of the improvement made in the time we had to work with these children should not be underestimated in the context of the goals of this research and the children we sought to help. The children selected for the intervention were experiencing or at serious risk of academic failure and all the concomitant risk factors correlated with this outcome. They were also at a crucial age were illiteracy was an extremely serious situation that effectively prevented them from engaging with the rest of the academic curricula.

This 10-week PT programme was designed to increase the reading skills of 5 pupils with specific reading difficulties. Despite the very brief teaching intervention, each PT child increased reading frequency on targeted words significantly and demonstrated they could read these words when tested under normal reading conditions. This study not only demonstrates the effectiveness of PT and frequency-building procedures, but also that simple learning procedures with a long history of research evidence can enable students, who have previously failed to learn the most crucial academic skill in the standard educational settings, to learn and succeed.

Chapter 3: What's the probability of that? Using SAFMEDS to increase undergraduate success with statistical concepts.[2]

---

[2] This chapter has been published as: Beverley, M., Hughes, J. C., & Hastings, R. P. (2009). What's the probability of that? Using SAFMEDS to increase undergraduate success with statistical concepts. *European Journal of Behavior Analysis, 10*(2), 235-247.

Abstract

Psychology undergraduates are required to study statistics as one of the main components of their degree, and for many this is the most challenging aspect of the curriculum. Undergraduates taking an introductory statistics module participated in this study. Based on the results from their first weekly open-book test, students scoring at the 50th percentile and below were identified and invited to sit a pre-test of questions selected to cover all the statistical content for the semester. Using the results from the pre-test we randomly allocated 55 participants to either a precision teaching SAFMEDS intervention (PT) (24 students) or a "treatment as usual" (TAU) group (31 students). We made SAFMEDS cards with key statistical concepts taken from the study guide of the main statistics textbook and students were instructed how to use the cards and how to chart their learning performance. Additionally participants were invited to attend bi-weekly meetings with student proctors to review individual progress and make instructional decisions. Results indicated that those students in the PT intervention maintained a higher score than the TAU for all of the weekly tests through the semester, and achieved a statistically significant gain in post-test performance almost equivalent to an entire degree class score when compared with the TAU group, $F(1, 53) = 5.23$, $p = .026$, $d = 0.62$.

The ability to interpret mathematical and statistical information at a level that is functional in today's information-laden society is, for many individuals, an essential skill. Figures and statistics are routinely presented via newspapers, television, and the internet, and often in a way intended to influence how the information is received (Tufte, 1983). Nevertheless, the accurate interpretation of this information is vital in making informed choices and decisions, especially within the workplace (Huff, 1993). Despite the importance of such skills, studies have shown that the general level of numeracy in the UK is poor. For instance, the results from a survey carried out in England found that approximately 15 million adults had only entry level numeracy skills (Grinyer, 2005), and results from a survey conducted for the Welsh Assembly Government (2005) illustrated that 53% of adults did not have level one numeracy skills. A concerning fact when you consider that in the UK students take General Certificates of Secondary Education (GCSE) examinations when they are 16 years old and level one numeracy equates to the lowest achievable grade mark without failing (i.e., approximately below the GCSE's G grade standard).

There is some evidence that the level of numeracy skills in new psychology undergraduate students is deteriorating. A recent study comparing cohorts who entered psychology departments a decade apart highlighted declining standards in mathematical abilities (Mulhern & Wylie, 2004), and striking deficiencies in mathematical reasoning have also been identified when comparing across a sample of UK educational institutions (Gnaldi, 2006; Mulhern & Wylie, 2005). Undergraduate students taking science degrees face an even greater requirement to become familiar with data interpretation and to become competent with statistics, so that they are able to use systematic, objective methods for summarising and interpreting quantitative information (Everson, Zieffler, & Garfield, 2008). Students must be able to understand statistics so that they can conduct their own research effectively and interpret and evaluate the research of others. Therefore, students need to become fluent in the vocabulary of basic statistical terms and concepts (Lalonde & Gardner, 1993); something that many students find difficult. The complexities involved in learning about statistics have been likened to those associated with learning a second language (Lalonde & Gardner, 1993; Lazar, 1990).

When learning to play a musical instrument, when training to become a professional athlete, or when learning a second language, repeated practice is the accepted route to an accomplished performance (Binder, 1996; Kubina & Morrison, 2000). Furthermore, a learner may reach a level of mastery that distinguishes their performance from that of an average learner: a fluent performance can be characterised as effortless, almost automatic, accurate responding (Binder, 1996). This type of performance has also been referred to as *automaticity* (Bloom, 1986).

Fluency in basic skills, or *elements*, may be the basis of what we refer to as 'mastery' levels of performance and may be the prerequisite for learning higher-order more complex skills, or *compounds* (Kubina & Morrison, 2000). There are recognised sequences in which to learn each skill element. For example, each step of a sequence in conducting mathematical operations should be mastered before proceeding onto the next (Stein, Silbert, & Carnine, 1997). Problems may occur when a student lacks such mastery in the necessary skill elements that are required for fluent performance on the next level of that skill (Binder, 1996; Kubina & Morrison, 2000). In these circumstances, a learner may be thought not to 'understand' the concept being presented. However, this can better be conceptualised as the learner being *dysfluent* at skill elements of the compound task (Binder, 1993, 2003; Binder et al., 2002; Hughes et al., 2007; McDowell & Keenan, 2001b).

A deficit in skill elements prevents learners from further progression through a curriculum, a term that Binder (1996) referred to as *cumulative dysfluency*. To ensure that a skill is mastered before moving on, specific performance targets or *fluency aims* must be set that allow the learner to demonstrate mastery on the existing skill element prior to moving on to the next. Detailed fluency aim ranges of appropriate rates of accurate responding have been researched and documented for many academic and motor skills (Kubina, 2002). For example, to read effectively a reader should be able to read aloud between 200-250 words of prose per minute and failure to be able to read at this rate may mean that the very process of reading is difficult and aversive to a learner (Hughes et al., 2007).

When the specific aims of fluent performance are achieved, there are benefits that result from this level of performance. The outcomes of fluent performance have been captured in the acronym RESA: Retention, Endurance, Stability, and Application (Fabrizio, 2004; Fabrizio & Moors, 2003; Johnson & Layng, 1992). Retention is the ability to perform the skill fluently after long periods without practice, endurance is the ability to perform the skill for longer periods of time, stability is the ability to perform the skill in the presence of other distractions, and application is the ability to combine previously learnt element sub-skills to perform novel, more complex, compound skills (Kubina & Morrison, 2000).

Statistics teachers face many problems in teaching such a difficult subject area. According to Conners, McCown, and Roskos-Ewoldsen (1998) there are four main problems: variability in the performance of individuals, student motivation, anxiety about statistics, and making learning last. Applied behaviour analytic approaches to teaching have a long history of success in overcoming these and other obstacles to learning (Gardner et al., 1994; Heward, 2005; Heward et al., 2005; Johnson & Layng, 1994; Layng, Twyman, & Stikeleather, 2003; Moran & Malott, 2004). As Frederick and Hummel (2004) comment, effective educational practices—those that produce good educational outcomes—adhere to certain principles, all of which are relevant to these four challenges:

> Effective instruction begins with clearly stated behavioral objectives; provides accurate, competent models; provides many opportunities for active responding;

> delivers immediate feedback about the accuracy of responses; allows self pacing; teaches to mastery; reinforces accurate responding; and frequently and directly measures responding that is explicitly tied to behavioral objectives, using the outcomes of those measurements to make instructional decisions. (p. 11)

Precision Teaching (PT) has proved to be an effective intervention that can monitor the effectiveness of any teaching method and also evaluate the outcomes of learning—RESA. PT has over four decades of evidence supporting its use (Beck & Clement, 1991); and in terms of meeting the eight principles of effective instruction (Fredrick & Hummel, 2004), PT and frequency-building procedures do extremely well (Beck & Clement, 1991; Binder, 1988; Binder et al., 2002; Chiesa & Robertson, 2000; Hughes et al., 2007; Kubina & Morrison, 2000; McDade, 2005). PT procedures ensure that all skills have clearly stated behavioural objectives, provide many opportunities for responding, often give immediate feedback about the accuracy of each response, and allow learners to work at their own pace until they meet predefined levels of performance—fluency aims. Learners who follow PT and fluency-based procedures often use short, timed practice sprints to collect data on their performance. These timed practice sessions are usually carried out daily. They record the number of correct and incorrect responses per minute (Learning Opportunities; LO) and then plot these data on a *Standard Celeration Chart* (SCC). The SCC and its use have been described in detail in a number of previous publications (e.g., Calkin, 2005; Kubina & Yurich, in press; West et al., 1990; White, 1986; White & Neely, 2004) but essentially its most obvious use allows both the learner and teacher to easily observe whether effective learning is taking place. After only a few days of plotting data, a SCC produces a learning picture that enables data-driven instructional decisions to be made in order to improve an individual's learning (Claypool-Frey, 2009a; Lindsley, 1995a; McGreevy, 1983).

SAFMEDS (*S*ay *A*ll *F*ast *M*inute *E*very *D*ay *S*huffled) are a practice and assessment procedure developed to help students learn and build fluency on key facts (Graf & Lindsley, 2002). SAFMEDS are typically used to help students become fluent in definitions and basic concepts, thus making complex learning and later synthesis of concepts more likely. Further details of SAFMEDS usage have been described in earlier publications (Claypool-Frey, 2009b; Graf & Lindsley, 2002; Vieitez, 2003).

In the current study we explored the effects of a brief intervention using SAFMEDS cards within a PT and frequency-building framework designed to build high frequency performance in key statistical definitions and concepts in a group of first year psychology undergraduates. We examined whether student performance on basic statistics facts (elements) would influence their performance on the post-test (compound).

Methods

Participants

Initially we sought informed consent from an entire first year statistics class containing approximately 340 students. From those who gave consent, we identified those who scored at the 50^{th} percentile and below as determined by their first weekly class test scores, and invited them to sit a pre-test that consisted of questions selected to cover all the statistical content for the semester. We randomly allocated students to either the Precision Teaching (PT) or Treatment as Usual (TAU) group using a stratified random sampling procedure (Gravetter & Forzano, 2009): to ensure that the groups were balanced we allocated students to blocks of equally performing students and randomly assigned from each of those blocks to the two groups (each block contained an even number from which to randomly assign). Participants' ages ranged from 18-30. Initially the participants (n = 66) were split into two equal size groups, however by the end of the study attrition rates had reduced the number of participants (n = 55), to 24 in the PT group and 31 in the TAU group. Participants earned course credits for taking part in the study.

Stimuli & Materials

We constructed a pre-test that comprised 50 randomly selected questions from the core text's bank of test questions (Gravetter & Wallnau, 2007). The tests consisted of multiple-choice questions (MCQ) from all the content textbook chapters to be covered that semester. We prepared two packs of SAFMEDS cards (each containing 80 cards) for each participant in the PT group. The packs covered course content for the entire semester (Pack 1: weeks 3-5 of the course & Pack 2: weeks 6-11) and were based on materials taken from the core text's study guide (Gravetter & Wallnau, 2004). We also gave participants data collection sheets with a 10-week 3-cycle Standard Celeration Chart (SCC) for charting their learning, and digital timers to use during practice sprints and timings.

Design

For the purpose of group data analysis, we used a mixed design with one between-group factor (Group: PT & TAU) and one repeated measures factor (Test: pre- & post-test) to allow the comparison of group performance between the pre- and post-tests.

As is typical when using behavioural research methodologies, we examined each participant's ongoing graphed data to make intervention decisions on an individual basis.

Procedure

All students who had been invited to participate in the study took the initial 50-item pre-test and were given 45 minutes to complete. Following the initial pre-test, students who were allocated to the TAU group continued with their standard lectures and their normal preparation for their weekly open-book test. Students who were allocated to the PT group attended lectures as normal but were also instructed in the use of SAFMEDS and how to chart learning performance. Students from both the PT and the TAU group had the opportunity to attend a two-hour lecture and a one-hour revision session each week. However, approximately 30 minutes of each 2-hour lecture were used to administer the weekly class tests from week 3 onwards. Students in both groups received the same tests under the same conditions within the class setting. The same lecturer presented all lectures and revision sessions on the course.

Precision teaching procedure. We held an initial one-hour session for the entire PT group during which we tutored them in PT methods, the rationale for our approach, and provided them with examples of the research evidence that supports the use of PT methods. During the session we emphasised the importance of data collection, using the SCC to monitor learning progress, and how to effectively use SAFMEDS cards (see below). The PT group was split into subgroups each with their own proctor—an undergraduate student helping with the research. Participants were instructed to conduct 1-minute timings three times per day, collect data for each timing, and plot their best score for each day onto their SCC. This would take them approximately six minutes per day. We also asked them to meet twice weekly in their subgroup, so that their proctor could verify their SAFMEDS and charting procedure and help them decide on strategies that may aid their learning. Whilst they were involved in the PT intervention they continued to take their open-book weekly tests.

At the end of the semester, participants from both groups completed a post-test; this post-test was the same test they had previously taken as a pre-test.

SAFMEDS procedures. We constructed our packs so that there was a statistical definition on the front of each card with a key term of that definition missing. On the reverse of the card, the missing term was printed (see Figure 3.1 for example of card content). This then required the learner to *see* the definition on the front of the card, and then to *say aloud* the missing term that was printed on the reverse of the card. They then would turn the card over for immediate feedback and place the cards onto the table in one of two piles—on the right pile if the answer was correct and on the left if it was not. Following each 1-minute timing, participants would count up the number of correct and incorrect responses (learning opportunities) and record these scores on their data sheet. Their best scores for each day would then be plotted on the SCC. Although most of the SAFMEDS had terms and definitions, some had a statistical symbol on the front with the name or the symbol's interpretation on the back. None of the SAFMEDS cards was based on questions from the pre- and post-test, only key terms and definitions.

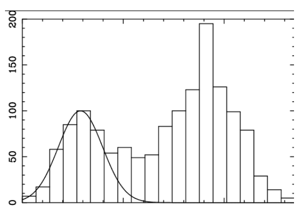
Front of Card	Back of Card
standard deviation is the _____ _____ of the variance	square root
 _____ distribution	bimodal

Figure 3.1. Example of the front and back of two of the SAFMEDS cards from Pack 1

Initially we issued each participant with the first pack of SAFMEDS. This pack covered terms and definitions from the chapters that they had covered in weeks 3 and 4, and that they would be tested on in the following week's test (week 5). After the test in week 5, we gave them the second pack designed to cover content from chapters for the rest of the course. From this point they would conduct timings for both of the packs working towards the pre-determined fluency aim range, and continue either until the end of the study or when they reached aim with either or both packs.

If participants reached an aim that was considered fluent with a pack, they continued working with the other pack whilst reviewing their mastered pack at weekly intervals to ensure retention of content. The fluency aim for the SAFMEDS cards was 60-40 correct with no more than 2 errors per minute (Kubina, 2002).

Interventions. Examples of possible interventions were: changes in timing period or number of timings conducted, slice-backs in number of cards worked with, and introduction of error correction procedures. For the error correction procedure, for example, we instructed the students that before conducting their daily timings they should first conduct an untimed pack review. During this process they would review each card in the pack, as they would do with a normal timing, placing corrects in one pile and learning opportunities in another. Then they would take the pile of learning opportunities and review these again following the same procedure. They would continue to repeat this error correction procedure until they had *said out loud* one correct response for every card in their pack. They would then conduct their daily timings as usual and plot their best score of the day.

Results

An ANCOVA (controlling for any differences between the groups at pre-test) was conducted on the data to compare the mean percentage correct scores at post-test. This showed a significant difference in favour of the PT group with a medium effect size, $F(1, 53) = 5.23$, $p = .026$, $d = 0.62$ (see Figure 3.2).

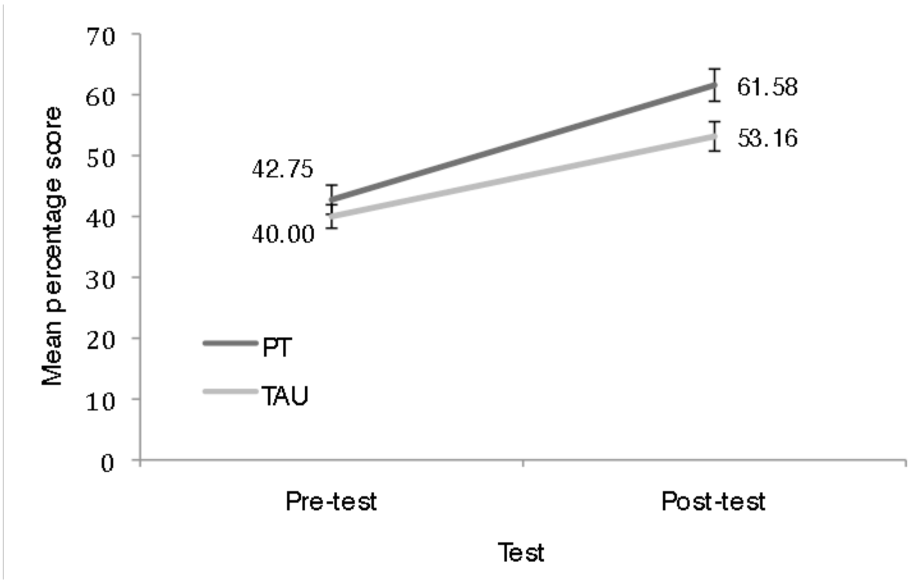


Figure 3.2. Pre- and post-test mean percentage scores for the PT and TAU groups. Error bars represent ± 1 standard error of the mean. Covariates appearing in the model (pre-test) were evaluated as 41.20.

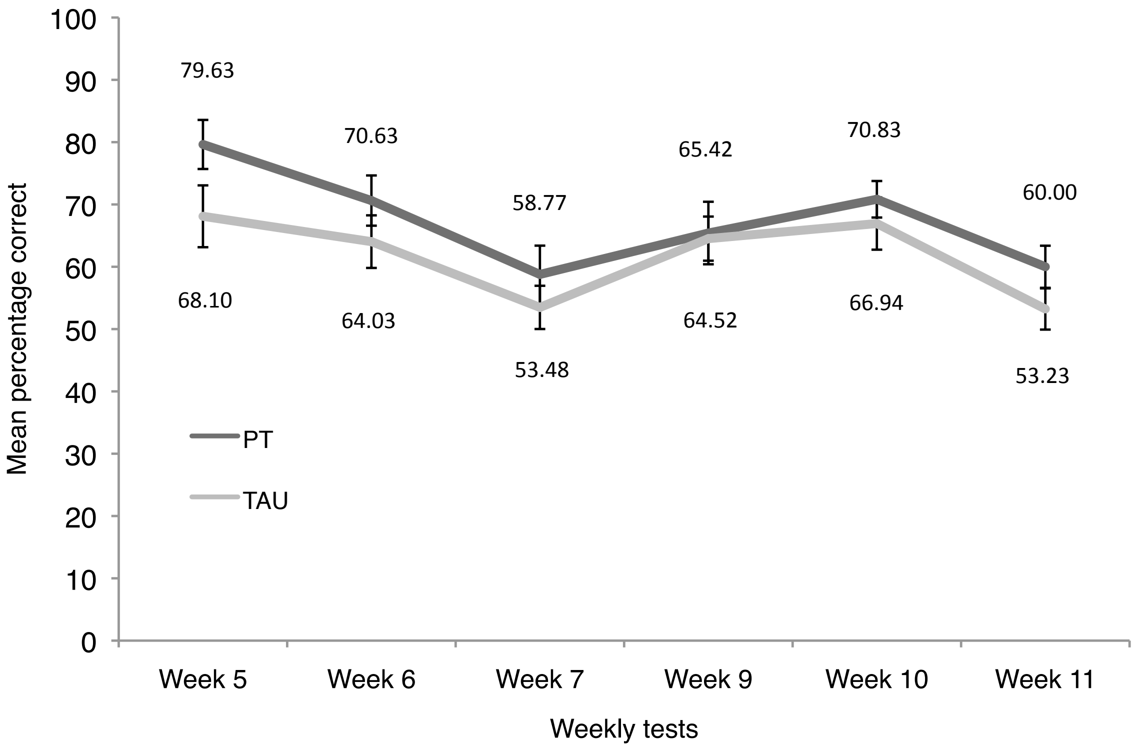


Figure 3.3. Mean percentage correct scores for both the PT and TAU groups for each weekly test. Error bars represent ± 1 standard error of the mean.

Figure 3.3 shows the performance of the PT and TAU groups on each of the weekly open-book tests taken in weeks 5, 6, 7, 9, 10, and 11. The PT group consistently had higher mean percentage scores for all of the weekly tests. Figure 3 clearly shows that the PT group's mean score was over 10% higher than that of the TAU group for the first test after beginning the SAFMEDS (week 5). This was after only one week of working with the first pack of cards.

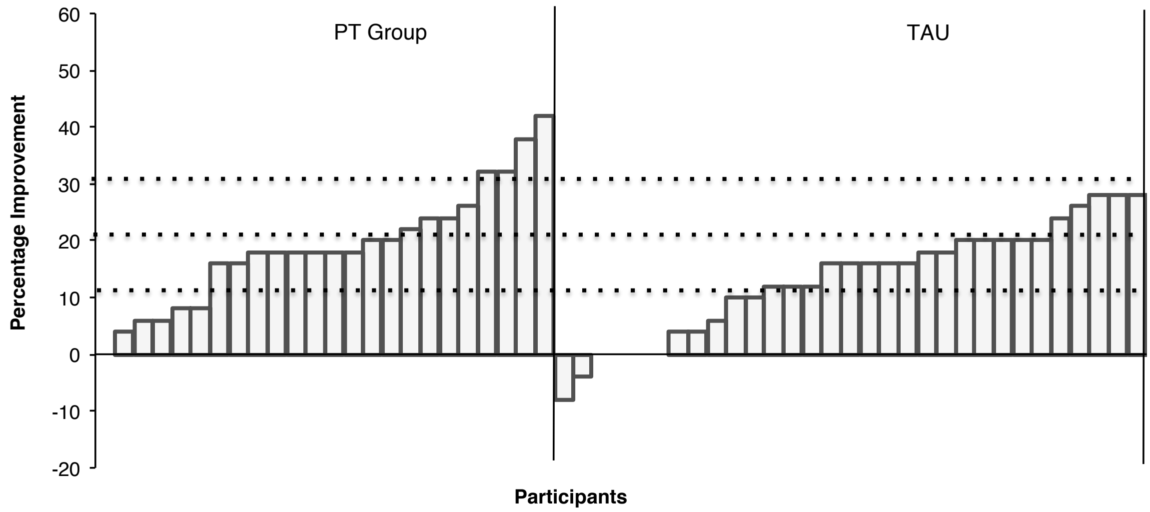


Figure 3.4. Displays the improvement gain from pre- to post-test for each participant from both the PT and the TAU groups. The black dotted lines indicate the position of the pooled standard deviation at pre-test for one, two, and three standard deviations

Figure 3.4 shows the percentage improvement from pre- to post-test scores for the PT and the TAU groups at a more individual level. It can be seen that almost all participants in the PT group improved on their score from pre- to post-test with no students showing a deterioration in performance (improvement range = 0 to 42); whilst in the TAU group six participants made either no gain or showed a deterioration in performance between pre- and post-test (improvement range = -8 to 28). Eight participants (33.33%) from the PT group increased their score by more than 2 standard deviations, whereas only 5 (16%) of the TAU group did so. Furthermore, from these 8 PT participants 4 increased their scores at or above 3 standard deviations.

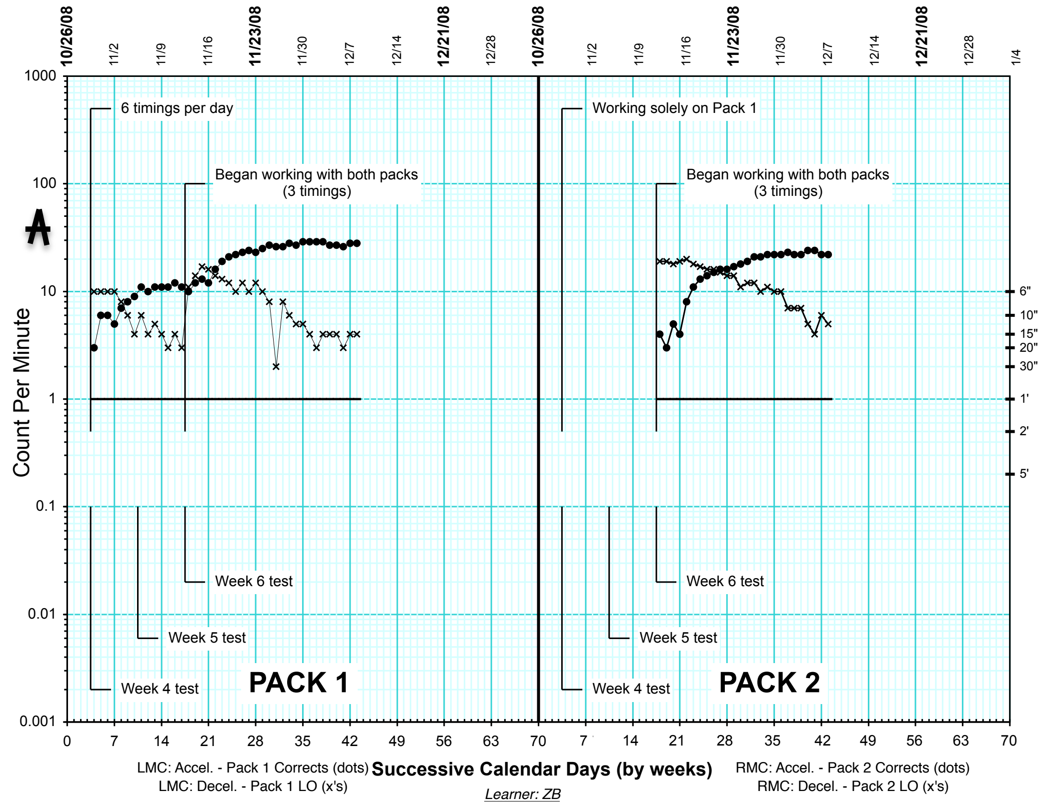


Figure 3.5. A daily per minute split standard celeration chart (Born, Harder, & White, 2003) showing the best daily score pairs for one individual from the PT group for Pack 1 (Left chart) and Pack 2 (Right chart). The Y-axis indicates frequency (count per minute) on a multiply-divide scale from .001 per minute (1 in 16 hours) to 1000 per minute. The X-axis is a calendar scale indicating up to 10 weeks of possible data for SAFMEDS Pack 1 (Left half of chart) and SAFMEDS Pack 2 (Right half of chart). The start date is the first Sunday before the study began (26-10-08) for both halves of the chart and each half of the chart can accommodate 70 days of data. The black dashed lines (-) represent the counting period (timing floor) and denote the length of the practice sprint: a line on the ‘1’ indicates the counting period is 1 minute. A score below the counting floor indicates zero responses.

Figure 3.5 illustrates the use and interpretation of a SCC at the individual level for one participant from the PT group. This learner began conducting six timings a day working with Pack 1, and recording their best daily score, after they had already taken the week four statistics test. As can be seen from Figure 3.5 there are no gaps in the data with the participant even conducting timing sprints over the weekend.

During the period of working solely with Pack 1 the charted data displays a classic *crossover jaws* learning picture, at first errors (LO) are high and corrects are low, however after four days of practice timings, the errors begin to decrease as the corrects increase.

When the learner begins to work with both packs there is an apparent increase in LO for Pack 1 as they jump up above corrects again. However the LO soon begin to decrease again over each subsequent week. The first timing for Pack 1 was 3 correct and 10 LO (13 cards attempted in one minute) and for the first 2 weeks the learner averaged approximately 15 cards worked through in a 1-minute timing.

For the following two weeks the learner was working with both packs (conducting 3 timings with each pack). The first timing with Pack 2 was 4 correct and 19 LO (23 cards attempted in one minute), and the average cards worked through in a 1-minute timing were approximately 31 and 28 for packs one and two respectively.

Discussion of Chapter 3

The aim of this study was to evaluate the effectiveness of a brief intervention using SAFMEDS cards within a framework of PT, and frequency-building procedures designed to build mastery performance in key statistical definitions and concepts in a group of first year psychology undergraduates. The students selected had scored at the 50th percentile and below at their initial weekly statistics test, so were those students who were performing below the average.

There was no statistically significant difference when comparing mean test scores between the two groups at pre-test. However, at post-test (after controlling for pre-test score differences) the mean score was significantly higher for the PT group in comparison to the TAU. As expected both groups improved during the semester, however, the PT group improved by almost one entire grade level above the performance of the TAU group (i.e., from a C- to a B-); the categorical marking system is a letter grade system ranging from F (Fail) to A (Distinction), with each of these having three sub-categories (e.g., B-, B, B+).

When comparing group performance on the weekly tests (weeks 5, 6, 7, 9, 10, & 11) the trend was consistently higher for the PT group (see Figure 3). The difference between the two groups was more noticeable for the first test that participants took after the initial week of practicing with the first pack of SAFMEDS (week 5). The PT group outperformed the TAU group by an average of over 10% for that week, which provides additional support for the PT group having become more versed in the basic definitions and terminologies. Recall that the test in week five was still concerned with the early chapters of the core text and success in this test relied heavily on students having mastered the earlier taught content.

Although we did not collect data to validate the class tests that were used as pre- and post-test measures, they are clearly statistic questions written by respected authors whose book is now in the seventh edition and used as a required text in many psychology departments. Furthermore the instructor test bank items were based on the curriculum, all students initially performed poorly at baseline, and the same test was used for both pre- and post-test measures.

When comparing pre- to post-test improvement at an individual level it was found that twice as many from the PT group increased their score by over two standard deviations in comparison to the TAU group, and half of these achieved or bettered an increase of three standard deviations (see Figure 4). We would regard this type of improvement as statistically and educationally significant.

Considering the example SCC for one individual presented in the results, it can be seen that throughout this brief intervention the participant was consistent in conducting regular timings and recording their best daily score for each pack. Their corrects continued to increase and their LO decreased, although they did not quite reach the set fluency aims. At the point that the second pack was introduced the LO seem to take a *jump up* (rising above the number of corrects) for Pack 1; although at this same point in time for Pack 2 there is a classic picture of beginning learning in that LO are high and corrects are low. This then progresses into a *crossover jaws* learning picture as the other pack did.

The explanation for the jump up for Pack 1 would be that although the LO did increase temporarily, this was essentially because the learner had become more practiced and comfortable with the procedure and increased the number of cards they attempted during sessions. This is something that would not show up so clearly if using the traditional percent correct measure of academic success. The learning picture for Pack 1 for the first two weeks is a typical picture of a learner who is concentrating on achieving correct answers, irrespective of how long it takes them to think of the response. During this period the learner's average response rate is less than 15 responses per minute (1 every four seconds). Following this initial two-week period the average response rate was approximately 30 cards per minute (1 response every 2 seconds) for both packs.

Some issues of the research methodologies merit discussion. In the present design, the daily use of SAFMEDS could not be monitored with complete accuracy; we had to rely on self-report of the PT participants at their bi-weekly meetings; although we did have participants conduct timings at each meeting (in the presence of their student proctor) to ensure their recorded data matched their performance data. Neither were data collected to monitor attendance at weekly lectures or the time spent by each participant in study and preparation for the weekly tests.

In addition, because the first test in week 3 was used to identify participants for possible inclusion in the study, students did not receive their first pack of cards until after the test in week 4; therefore they only had one week in which to work with the first pack until they sat the next test in week 5, following which they received their second pack of SAFMEDS cards. It would have been preferred to pre-test the entire year and then invite participants to take part in the study, so that they could aim for a fluent performance in the first pack before sitting their first test in week 3. However, this would demand considerably more resources than were at our disposal.

Whilst all possible control checks were followed to ensure consistency in the definitions and concepts in the SAFMEDS packs, there were some perhaps unavoidable differences between some of the cards. As we included statistical symbols for some cards and more complex definitions and concepts for others it is likely that different cards could be responded to at different latencies; this could have had an adverse effects on an individual's ability to perform equally well on all cards in a pack. Nevertheless in considering the SCC for this individual there is very little variation in their performance from day to day.

Further research would benefit from some changes to the methodology to overcome some of the problems we encountered during this study. The students in the PT group had incentives to study that were not available to the TAU group. Future research should ensure that a comparison group also receive instruction on using key-term flashcards but no instruction on how to time or chart their performance. Additionally, the results found could be more robustly attributed to the effects of the SAFMEDS intervention if this comparison group were to not only revise using standard methods, but also meet up with a student proctor bi-weekly (as the PT group did in this intervention).

The advantages in conducting frequency-based interventions via the internet are many. We could have SAFMEDS presented via an internet-based system, that would allow easy tracking of effective engagement of the participants, tracking both the number of timings conducted and also each of the timings scores, as well as automatically charting and displaying their progress. The programme could be designed to be even more adaptive to each individual learner, dropping cards temporarily from their pack when the participant had mastered them (i.e., consistently shows correct responding and low response latencies for that card) and readmitting them to the pack at set periods to ensure that retention is maintained. This would allow the more difficult cards for each learner to be presented more often. Finally, a web-based intervention would allow the automatic measurement of not just participant's rate of responding but also the measurement of response latency and the outcomes of fluency defined by RESA (i.e., Retention, Endurance, Stability, & Application) (Fabrizio, 2004; Fabrizio & Moors, 2003; Johnson & Layng, 1992). Of course by definition someone who is fluent would demonstrate low response latencies; it would be interesting to look at the relationship between response latencies and these outcomes of fluent responding (i.e., RESA) more systematically.

If we were to also add contingencies so that students would gain some credit towards the module assessments for participating, this could increase their motivation for active engagement with the procedure. Future research could also benefit from recording both groups attendance at the weekly statistic lectures, interviewing participants about their revision strategies, and measuring statistics anxiety to discover whether the PT and frequency-building procedures may reduce anxiety levels (Conners et al., 1998).

This study set out to improve undergraduate statistic performance using a combination of PT and frequency-building procedures. These procedures included daily timed practice sprints using SAFMEDS cards, charting participant's best daily score on SCC, and using these plotted data to make instructional decisions to help participants to progress towards their goal at an optimal rate. The study provides some additional support for the merits of such interventions to enhance learning that are effective, can be easily implemented, and are of relatively low cost. Such interventions have been successfully used previously (Beck & Clement, 1991; Chiesa & Robertson, 2000; Hughes et al., 2007) and are suitable to be used with any age group and with almost any curriculum. Although in this study we targeted psychology undergraduates undertaking a statistics module, if more widely adopted, these techniques could help to diminish the skill deficits that a great number of both children and adults currently experience.

Chapter 4: Improving essential numeracy skills in primary school children using a brief fluency-building intervention: A randomised control trial[3]

[3] This chapter is based on a paper that has been submitted for publication: Beverley, M., Hughes, J. C., & Hastings, R. P. (2012). Improving essential numeracy skills in primary school children using a brief fluency-building intervention: A randomised control trial.

Abstract

Numeracy skills are essential if children are to succeed academically. We investigated the efficacy of short fluency-building exercises to target a number of essential numeracy skills (writing digits and simple addition problems). Children from two mainstream primary schools (aged 9-10 were allocated randomly to either a fluency-building intervention group (FBI; n = 19) or a control group (C; n = 10). The intervention group engaged in daily fluency-building practice that consisted of writing and reciting essential numeracy facts. The control group children received their standard numeracy instruction. Following the brief, 5-week intervention, the children in the FBI group performed significantly better ($p < .05$) on three of the four numeracy elements with large effect sizes ($d = 1.25\text{-}1.67$). Results are discussed in the context of incorporating simple essential skills fluency-building exercises into mainstream classes, and how the implementation need not be costly in terms of staff time or resources.

Deficits in essential skills have been demonstrated to have serious implications for an individual's future success. For example, Parsons (2002) conducted a UK longitudinal survey the results of which indicated that poor literacy or numeracy skills directly increased the likelihood of future criminal offending. Additionally, a recent UK government report (Department for Education, 2010) highlights that the general standard of basic skills has declined significantly over the past few years when compared to the performance of children in other countries (Bradshaw, Ager, Burge, & Wheater, 2010). Across Europe, deficits in literacy and numeracy skills have been identified, and typical recommendations focus on more use of learning and instructional practices that can increase levels of achievement whilst also reducing individual variation in performance (Organisation for Economic Co-operation and Development, OECD, 2010).

Slavin and Lake (2008) conducted a meta-analysis of 87 studies (36 of which used random assignment to intervention groups) to ascertain which type of teaching approach is most beneficial in helping children learn maths skills. From their analysis, Slavin and Lake generated three categories of maths approaches: curriculum changes, supplementing curricular with computer assisted instruction (CAI), and changes in classroom practices (both in terms of instruction and classroom management). Slavin and Lake concluded that the evidence strongly supports changes in classroom practices as most likely to be effective compared with changes in either technology or curricula alone. The evidence review suggested that instructional practices and learning strategies that can be easily implemented in mainstream educational settings may be particularly effective.

There are a number of principles of effective educational practice that increase the likelihood of successful learning in the classroom, such as ensuring competent models, high active responding from the learner, immediate feedback, mastery learning, and motivational environments (Binder, 1996, 2003; Ericsson, Krampe, & Tesch-Römer, 1993; Fredrick &

Hummel, 2004; U. S. Department of Education, 2009). Additionally a valid system of monitoring whether individual children are achieving planned gains in learning objectives has proven to be an essential measure in order to measure progress (Fuchs, 2004; Lindsley, 1995a).

With regard to the second point above, most schools in the UK are required to administer standardized tests at set times throughout a child's education. These assessments are typically used to compare their performance with that of their peers and thus monitor their progress. However, these tests are too infrequent to guide meaningful change in the delivery of the curriculum (Fuchs, 2004; Fuchs & Fuchs, 1993; Johnson, 2008; Johnson & Layng, 1994). To successfully track individual progress, measurements of learning must be taken more regularly, and educators need be trained to be able to use this information to rapidly react and adapt to each child's individual needs (Fuchs & Fuchs, 1993; Lindsley, 2010; National Centre for Response to Intervention, 2010; Reynolds & Shaywitz, 2009; West et al., 1990; White, 2000)

There are a number of different approaches that measure learning more frequently to monitor individual academic progress for all learners (Bramlett, Cates, Savina, & Lauinger, 2010; Moors, 2010; Reynolds & Shaywitz, 2009; Vaughn et al., 2008; White, 2002). Response to Intervention (RtI) is implemented using a three-level system that consists of primary (Tier 1), secondary (Tier 2), and tertiary (Tier 3) interventions (National Centre for Response to Intervention, 2010). A similar graduated response approach has also been introduced in the UK to provide more intensive literacy and numeracy interventions for children according to their need (Department for Education and Skills, 2002; DfEE, 1999; Dowker, 2009). RtI has a growing evidence base (Moors, 2010; Reynolds & Shaywitz, 2009; White, 2002) and is similar in some key ways to other approaches to measuring learning that have been used successfully within academic settings and have a long and growing evidence base, such as Curriculum-Based Measurement (CBM) and Precision Teaching (PT).

Within a CBM approach (Deno, Fuchs, Marston, & Shin, 2001; Fuchs, 2004) measures of academic improvement are taken at least monthly, and proponents of PT typically use daily measurements as indices of improvement (Alper & White, 1971; Binder, 1990; Johnson, 2008; White, 2000). All three, (RtI, CBM, & PT) approaches, irrespective of the frequency at which they are administered, can act as efficient “academic thermometers” (Shinn & Barmonto, 1997, p. 1) to monitor each child’s progress within a specific curriculum domain (e.g., maths). Although there is little research directly comparing the relative efficacy of these approaches (Binder, 1990; Fuchs, 2004), the more regular measurements provided by the PT approach and the recording of these measurements on a Standard Celeration Chart (SCC) has been shown to provide motivational feedback to learners, especially when they are taught to record and chart their own learning (Bower, 1985; Lindsley, 1995a).

PT is a measuring system that helps teachers ensure that *every child in a class* maintains rapid and successful learning. This approach has had considerable success across a number of educational settings and academic areas, with children in mainstream schools (Beck & Clement, 1991; Chiesa & Robertson, 2000; Hughes et al., 2007; Miller & Calkin, 1997), undergraduate students (Beverley et al., 2009), children with autism, (Kerr, Smyth, & McDowell, 2003; Kubina et al., 2002; Kubina & Wolf, 2005; Zambolin et al., 2004), and other special educational needs (Solis, Derby, & McLaughlin, 2003). Combined with regular teaching, PT can represent a powerful accelerated learning approach, when effectively amalgamated with other evidenced-based methods of instruction (such as Direct Instruction) to provide highly effective learning environments (Binder, 1990; Binder & Watkins, 1990; Kubina et al., 2009; Morelle, Morrell, & Kubina, 1995; Sante, McLaughlin, & Weber, 2001). One key aspect of the PT approach is to encourage students to build fluent performances in essential skills. Previous studies have successfully used these approaches to improve individual maths performance, however, most of these were restricted to small *n* designs (Hayden & McLaughlin, 2004; Raggio & Bitgood, 1982; Sweeney et al., 2001). Two previous studies did apply these PT strategies and tactics in a more traditional group study design with successful outcomes (Chiesa & Robertson, 2000; Fitzgerald & Garcia, 2006).

Chiesa and Robertson (2000) investigated the effects of a brief fluency-building with elements of long-division over a 12-week period. The intervention group only practiced elements of long division rather than fully completing long division problems. In contrast, all of the children in the control group spent time practicing long division problems. Nevertheless it was children in the intervention group who outperformed all but one of their peers on tests of long division.

In the present study, we have explored the effect of a brief intervention targeting fluency-building of essential numeracy skills in the intervention group, measuring each child's daily improvement as would be typically done within a PT framework. We examined whether student performance on basic numeracy facts (components) would influence their performance on the post-test (composite) in comparison with a control group, which received the standard mathematics instruction. Furthermore, we tested for effects beyond the areas practiced to provide support for the emergence of novel composite skills, which are reliant on the fluent performance of prerequisite component tasks.

Method

Participants and Setting

Mainstream schools in the local area were approached to see if they wished to take part in research on essential numeracy skills. From the schools approached, two schools agreed to take part. The study took place in these two mainstream "primary" schools in North Wales, UK. Twenty-nine children (aged 9-10) who had been identified by their class teachers to be struggling with maths were eligible to take part in the study. Nineteen of these children were allocated randomly (using weighted randomized selection based on a 2:1 model) to the fluency-building intervention group (FBI; Females = 12, Males = 7), and 10 were allocated to the control group (C; Female = 6, Males = 4). The criteria used by teachers for children's inclusion in the study was the children's performance on the usual standardised annual achievement tests administered within the school, as well as individual classroom performance data. Because all children selected were struggling in numeracy skills, we randomized a larger proportion of children to the intervention group. All of the children were typically developing, and none had any identified special educational needs.

Apparatus and Materials

Standard worksheets were used for both pre- and post-tests (see Appendix B). Each worksheet contained more maths problems than any child could complete in the timing period to ensure that no artificial ceiling was placed on their performance. Each worksheet covered one of four levels of numeracy problems (see Table 4.1). The same types of worksheets were used to generate random practice sheets for use during the individual intervention periods. However, only the first two skill levels were practised during the intervention (see Table 4.1; the intervention group did not practise subtraction at any time during the study). In addition, *flashcards* were produced for children to conduct timed practice sprints during the intervention period (see *Procedure*). Digital timers were used to time practice sprints and data collection sheets were used to record daily data. Recorded data consisted of the date, duration of session, number of correct and errors per minute timing, personal best scores, and total timings conducted during the session. The children charted these data so that both the research team and each child could readily view their daily performance on each practice slice (Writing numbers and Single digit addition), allowing progress to be effectively monitored. Children were trained by the research team to record and plot their best scores of the day for each maths skill.

Table 4.1

Details and description of each skill and its associated AIM for the treatment group (PT) during the brief intervention.

Pre- and post-test essential skills tested	Description	AIM per minute	Examples of skill order
1] Writing numbers	Digits 0-9 written in a continuous stream.	120-100 digits written correctly with ≤ 2 errors	Writing single digits 0 – 9. Writing double digits 10-19, etc.
2] Single digit addition with answers of ≤ 10.	Computation of single plus single digit addition problems	80-60 correct answers with ≤2 errors	Adding digits with answers of < 10.
3] Single digit addition problems with answers of ≥ 10 and ≤ 19.	*Computation of single plus single (or double) digit addition problems	80-60 correct answers with ≤2 errors	Adding digits with answers of > 10 but ≤ 19, etc.
4] Single digit subtraction problems with answers between 0-6	*		

Note. An * in the description column denotes skills that were never practised during the intervention.

Design

The study used a mixed 2 X 2 design with one between group variable (Group: Fluency-building intervention vs. Control) and one repeated measures variable (Time of test: pre-test vs. post-test).

Procedure

Daily sessions took place in one of the school classrooms, supervised by the researcher team. Ethical procedures were followed throughout the course of the study. Ethical approval was obtained from both Bangor University's Ethics Board and each School's Board of Governors, and informed consent was obtained from all children's parents or guardians.

Pre-testing and pre-training. All children underwent both pre- and post-testing on the four selected essential skills. The FBI group received the intervention in place of their standard maths lesson, whilst all the control children continued with their standard maths lesson in the school. Initial pre-test performance was such that all FBI children began with digit writing and the simplest addition problems, which would provide confirmatory evidence that the children were performing below average and were likely to experience problems with more complex mathematical operations.

Before beginning the intervention, all FBI children were coached to enable them to not only record their data for each timing within the session, but also to graph their own progress (Bower, 1985; Maloney, 1993). We did this by teaching all the FBI children to chart. Furthermore, the researchers verified all charts to ensure the accurate recording and plotting of data. Teaching children to graph allowed them to observe their own learning improve as shown in the emergent learning pictures on the SCC (Claypool-Frey, 2009a; Lindsley, 1995a).

PT Intervention. At the beginning of the session children would collect their work materials and proceed to their desks to work in pairs on their practice sprints. In a typical daily session, children would begin with a tool-skill warm-up task, writing digits 0-9. This prepared them for their daily timings and also enabled them to practice and increase their fluency at writing digits—an essential pre-requisite skill for any written maths computation. They would then proceed to conduct three timed sprints from the maths worksheets, following which they would conduct the same number of timed sprints on their flashcards. Each child was allowed the flexibility to alternate between these two practice materials, as long as they completed their daily timings for each level of the maths problem they were working on. To prevent serial learning effects, the order of presentation for all practice sheets was randomized and the flashcards were shuffled before each sprint.

Practice sheets had maths problems for the level of difficulty at which the children were working and had sufficient space for children to write their answers. Answer keys were provided so that children could check their answers and record their data after each timed sprint. Initially, each timed sprint was typically 60s in duration but if the emerging learning picture on the SCC indicated, this timing could be shortened to 30s or 20s as an intervention for learning (Calkin, 2005; Johnston, 1970; Lindsley, 1995a; Pennypacker et al., 2003; White & Neely, 2004). The decision to shorten timings would be made if the child's emerging *learning picture* on the SCC indicated this would be a useful intervention (e.g., a child may significantly slow down towards the end of the 60s, which would indicate they had a problem with endurance and suggest a shorter time period may work better for that child).

Flashcards had the maths problems (again for the level of difficulty at which they were working) written on one side of the card and the answer written on the reverse. The cards were consistent in their presentation, with the front side always displaying the numbers to add (e.g., 4 + 5 = __, or 5 + 4 = __), and the reverse side showing the answer to the math addition problem. Children worked in pairs with one child conducting the practice sprint and the other checking their answers. The child would shuffle the pack, then read the maths problem (silently), say the answer (aloud) and check their answer by turning the card to receive either corrective or positive feedback. Their partner counted and verified the number of *correct responses* and *learning opportunities* (LO) per timing, logged this information on the data sheet, and then charted the best score of the day on the learning chart. Sessions took place at the same time each day.

Each time a child beat their personal best score they received a sticker, which was placed on the next space on their reward chart. At pre-determined points along this chart they could select an appropriate academic reward, such as a pencil, eraser, or notebook. There were 56 points on the reward chart with gifts awarded at spaces 4, 13, 25, 39, and 56.

The FBI children received the intervention over a consecutive five-week period with each session lasting approximately 20 min. At the end of the intervention phase, the pre-test was re-administered as a post-test to all children.

Data Analysis Approach

In addition to carrying out ANCOVA group-based statistical analyses of the results with associated effect sizes, we explored two indices that enable consideration of outcomes at an individual level. To evaluate change, we adopted an index that has been successfully used in the medical literature and more recently has been used in special education. The Reliable Change Index (RCI) originated from psychotherapy outcome research (Jacobson & Truax, 1991), and recently it has been used to evaluate the clinical significance of outcomes in autism educational interventions (Eldevik et al., 2010; Remington et al., 2007). The RCI identifies by how much an individual score needs to change to be significant at .05 level and takes into account typical variation in the scores in the population and stability of the measure over time. Improvement of performance at or above this index can be regarded as educationally significant gains.

To ensure that this test was conservative the RCI was calculated using the means and standard deviations of the entire groups' scores at pre-test to calculate the SE. Multiplying the SE by 1.96 provides a measure of magnitude of change required to be reliable at the $p < .05$ level (Evans, Margison, & Barkham, 1998; Jacobson & Truax, 1991; Zahra & Hedge, 2010). For calculation purposes, stability of the test scores were estimated by calculating the correlation for these scores in the control group between pre- and post-tests, as this would provide a more accurate measure of stability as the control group had not received the intervention.

Using individual level outcomes such as those generated via the RCI, an effect size more commonly used in medical interventions (Number Needed to Treat; NNT) can be used to communicate success rates of interventions in a clear manner. The NNT represents the number of children that would require to take part (be treated) in an intervention to have one more success or one less failure than would have been the outcome if all children had not received the intervention or received an alternative intervention (Kraemer et al., 2003).

Results

Mean scores on each numeracy skill pre and post-test for the intervention and control groups are shown in Table 4.2. ANCOVAs comparing the groups' post-test scores and controlling for pre-test performance were conducted. The results indicate that for the first two skills that were practised by the FBI group (Writing digits and Addition <10) there was a significant difference at post-test compared to the control group. For the third skill (Addition ≥10 ≤19; which was not practised by either group) there was no significant difference at post-test. However, for the fourth skill (Subtraction <10), which again had not been practised by either group, the FBI group showed a significantly better performance in comparison to the control group at post-test. All three of the skills that showed a significant difference between the FBI and control groups had associated large effect sizes.

Table 4.2

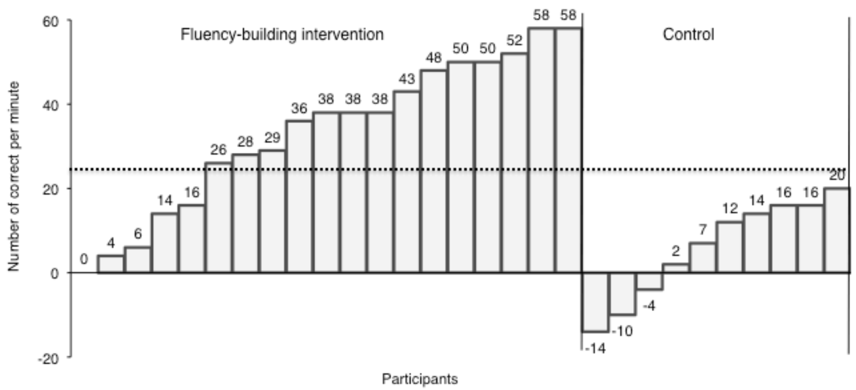
Mean numeracy skill scores at pre and post-test for intervention and control groups and results of ANCOVA analysis. Effect size d was calculated comparing the change scores of the two groups for each skill using the pooled pre-test standard deviation.

	Fluency-building Intervention (FBI)		Control group (C)		ANCOVA		
	Pre-test	Post-test	Pre-test	Post-test			
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*F (1, 26)*	*p*	*d*
Writing digits	72.37 (20.83)	105.63 (17.81)	72.30 (7.78)	78.20 (16.42)	22.87	< .001	1.67
Addition < 10	26.37 (9.19)	49.05 (9.12)	29.80 (9.25)	34.50 (8.19)	19.38	< .001	1.60
Addition >10 < 20	16.68 (7.78)	23.16 (9.28)	19.80 (9.54)	26.60 (8.04)	0.20	.662	-0.05
Subtraction <10	19.47 (7.25)	35.89 (12.95)	24.50 (9.14)	29.10 (12.49)	9.80	.004	1.25

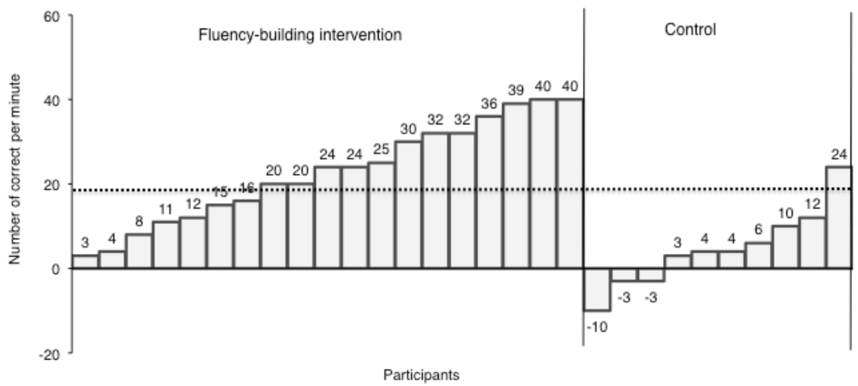
As the Pre-test for Writing digits had large variation for the FBI group, we ran the analysis again after removal of outliers (*n*=5)—outliers were defined as any score above the highest score or below the lowest score for the Control group at pre-test. No effect on the pattern of results was found – the group effect was still statistically significant.

Figure 4.1 highlights the results from the RCI analysis on the individual children's data. Every child in the intervention group either maintained or showed improvement at post-test for writing digits and addition <10. Additionally, for subtraction (not practised) all but one child in the FBI group improved at post-test and 12 children exceeded the criteria to achieve reliable change. Figure 4.1 also shows the percentages of children who achieved reliable change from the intervention and control groups for each of the maths skills from the beginning of the study until post-test. It can be seen that for the two maths skills that were practised by the FBI children the range of children who achieved reliable change criteria for the significant skills was between 63% to 74%, whereas for the control children it was between 0% to 20%.

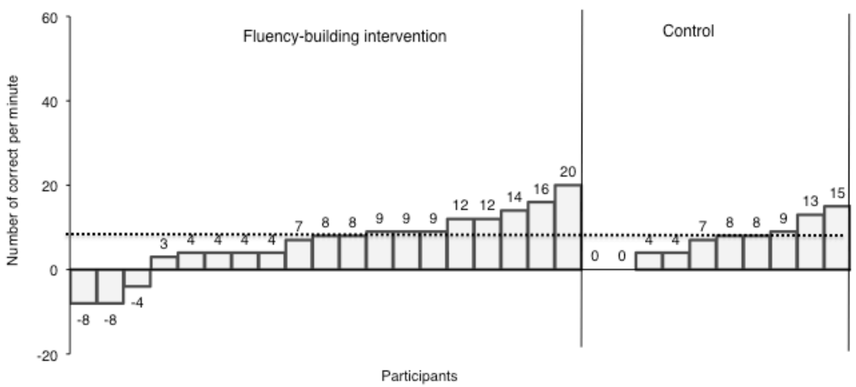
The NNT for the different skills were: (1) Writing digits, NNT = 1.36, 95% CI [1.1 — 1.9], (2) Addition <10, NNT = 1.88, 95% CI [1.2 — 4.1], (3) Addition >10<20, NNT = 8.26, 95% CI [-4.8 — 15.2, and (4) Subtraction, NNT = 2.31, 95% CI [1.3 — 9.8]. This shows a positive outcome for the NNT (i.e., because the numbers are low) for all but addition >10<20. These low NNT numbers represent the number of children that would require to take part (be treated/educated) in an intervention to have one more success than would have been the outcome if all children had not received the intervention (Kraemer et al., 2003). In simple terms, for approximately every two children receiving the intervention one child achieved individual change at a significant level than would have been expected if the children had been in the control group.



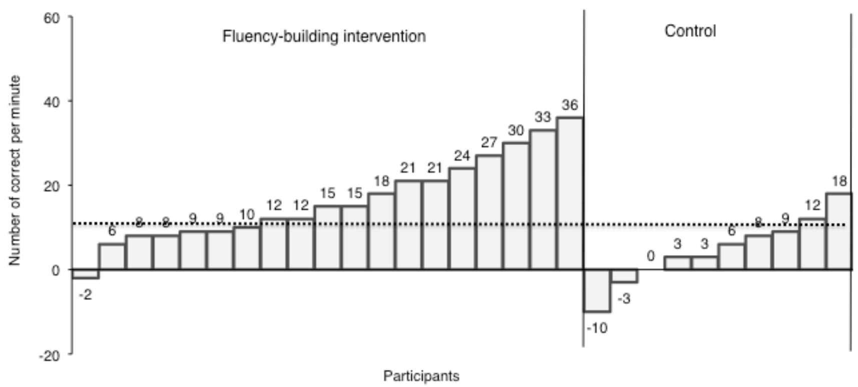
4.1 (a) Writing digits. RCI achieved by 14; FBI = 14 (74%), C = 0 (0%)



4.1 (b) Addition ≤ 10. RCI achieved by 13; FBI = 12 (63%), C = 1 (10%).



4.1 (c) Addition > 10 ≤ 20. RCI achieved by 11; FBI = 8 (42%), C = 3 (30%)



4.1 (d) Subtraction ≤ 10. RCI achieved by 14; FBI = 12 (63%), C = 2 (20%)

Figure 4.1. Charts comparing the individual performance of participants from both the FBI and control groups on the four different maths skills.

Scores above zero are gains in performance from pre- to post-test, scores at or below zero are maintenance or losses in performance respectively.

Scores above the dotted line indicate that the improvement is reliable at the individual level as measured by the *Reliable Change Index* (RCI).

Discussion of Chapter 4

The aims of this study were: (a) to evaluate the effectiveness of a brief intervention using fluency-building practice sprints to build mastery of essential maths facts with children whose performance on numeracy problems was falling behind their peers, and (b) to ascertain whether this mastery would impact their ability to solve more complex untrained numeracy problems that relied on the acquisition of fluent performance repertoires in prerequisite skills.

The results from this brief intervention provide evidence of positive results in relation to both of these aims. The intervention group performed significantly better at post-test on basic numeracy tasks that they had practised (writing digits and single-digit addition). In contrast, no difference was found between the intervention and control group's performances at post-test on one of the maths task that had not been practised (double-digit addition). More importantly, although neither of the groups had practised basic subtraction problems the intervention group still performed significantly better at post-test than the control group. These findings concur with previous research (e.g., Chiesa & Robertson, 2000; Kubina & Yurich, 2012; McDowell & Keenan, 2001a) that showed how brief fluency-building practice could impact on target composite skill performance, even when the target skill was never practiced and only components of the target skill were practiced.

We can postulate why some transfer to non-practised skills might occur. In mathematical operations, there is an inverse relationship between addition and subtraction (Stein, Silbert, & Carnine, 2005). Therefore, it could be that increasing children's performance to mastery on basic addition facts also increased their performance on related subtraction problems. These concepts are also discussed within the cognitive literature, in that an increase in procedural fluency may have facilitated improved access to the addition-subtraction inverse concept (Briars & Siegler, 1984; Cowan et al., 2011; Gelman & Meck, 1983; Greeno et al., 1984; Rittle-Johnson & Siegler, 1998).

Additionally, if we consider the results at a more individual level, the data show that the majority of children in the intervention group met or exceeded the criteria required to achieve an educationally significant change (as assessed using the RCI). These findings, along with the low NNT score, would support the overall effectiveness of this brief fluency-building intervention. Furthermore, the findings are also consistent with the effectiveness of implementing more frequent measurements of each individual's learning to ensure adequate progress as would be used in interventions such as RtI, CBM, or PT (Slavin & Lake, 2008).

This brief intervention took place over just five weeks. Because of this relatively short duration, the children only received practice opportunities on lower level essential maths skills. It would be interesting for researchers in future to undertake a longer period of work on fluency-building including a broader scope of essential skills to ascertain how much maths skills could be further improved. We used an "education as usual" control condition. However, we did not control for the personalised attention that children in the FBI group received. Future controlled trials should incorporate an attention control condition. In addition, it would be desirable to use additional standardized maths tests that would give more reliable results from which to interpret the effects of the intervention.

Future research should also be conducted to more systematically investigate the effects that building component tasks to an appropriate fluency aim has on composite (untrained) task performance (as demonstrated in the present study when the children performed significantly better on the untrained subtraction task). There is a growing body of research indicating that working on component skills fluency has direct benefits for more complex tasks made up from those components (Johnson & Street, 2004; Kubina, 2002, 2005a; Kubina & Morrison, 2000; McDowell & Keenan, 2001a). Conceptually, this is an interesting area for future research and suggests that variation in performance across individuals may be better accounted for by what Binder refers to as *cummulative dysfluency*

(Binder, 1993, 1996). From a Precision Teaching perspective, it would be fruitless to attempt to teach at the composite skills level if the perquisite component skills were not first mastered. Intuitively, we may desire to aim our teaching directly at the skill with which a child is struggling. However, component skill fluency would suggest a possible alternative explanation for some aspects of academic failure—that the component skills necessary to do the more complex task are weak. If this is the case, intervention should be aimed at the component skill level, rather than at the more complex skill with which the child is actually exhibiting problems.

Although this was a brief intervention, part of its strength is that it is economical in its implementation as well as being educationally beneficial. It would be relatively easy to train teachers or classroom assistants in these methods and have them incorporate such procedures into their typical classroom activities without a large cost both in terms of teacher time and other resources (Chiesa & Robertson, 2000; Roberts & Hampton, 2008). This simple methodology has the potential to impact maths performance significantly in any educational setting. However, it is important to note that research has shown that it is not sufficient to simply train teachers in these methods, they must also receive further support and coaching to create sustainable change within organisations (Georgiades & Phillimore, 1975; Johnson & Street, 2004; Roberts & Hampton, 2008).

Our data add to the research that supports the notion that measuring learning in a classroom setting need not be cumbersome and time consuming, especially if children are directly involved in tracking and taking some of the responsibility for their own learning. This can be achieved simply by teaching children how to time, record, and chart their own learning from the outset (Lindsley, 1995a). Other researchers have commented that this aspect of the PT approach also seems to enhance children's confidence and 'ownership' of their learning (Bower, 1985; Lindsley, 1995a; Maloney, 1993). Anecdotally, we found a similar effect.

Chapter 5: Using flashcards to assist language learners to acquire second language vocabulary[4]

[4] This chapter is based on a paper that has been submitted for publication: Beverley, M., Hughes, J. C., & Hastings, R. P. (2012). Using flashcards to assist 2nd language learners to acquire second language vocabulary.

Abstract

Flashcards, in conjunction with graphing learner's progress, were used as an intervention to teach Welsh vocabulary to second language learners in four year-7 classes (11-12-year-olds) in a Welsh secondary school. Another class was randomly allocated to be a Waiting List Control (WLC). From the curriculum, 200 words were selected by the teachers for the children to study during the term. All classes were given a pre-test to establish the number of words that each child already knew at baseline. Children in the four intervention classes were each given packs of Welsh-English flashcards and required to pair up with another child. Both children were instructed to conduct a practice run followed by a 1-minute timing at the start of each 50-minute lesson (3 lessons per week). They then recorded their correct and incorrect responses on data sheets and graphs. The procedure took less than 5 minutes, and the teacher then continued with the scheduled lesson. The WLC class received their normal Welsh classes. The intervention spanned four weeks of the term and was followed with a post-test of the target words for all classes. During the post-test, the intervention children wrote significantly more correct Welsh vocabulary words than the control children, $p < .001$, $ES = 1.54$. The study demonstrates that flashcards are an efficient, inexpensive, and effective method to enhance vocabulary learning with brief exposure within the context of standard lessons.

For any language, vocabulary is a key component when learning to speak, read, and write. Although vocabulary acquisition forms only one element of a comprehensive language learning programme, it nevertheless plays a central role in language acquisition and is of particular importance to beginner language learners (Coady & Huckin, 1997). There has been an increased interest in vocabulary learning since the 1970s (Carter, 1987) and current research supports the position that second-language vocabulary can be learned through four main methods: Direct teaching—teacher explanation or peer teaching; Direct learning—using dictionary or word cards; Incidental learning—guessing from context; or Planned encounters—graded reading or vocabulary exercises (Nation, 2001).

Although learning a second-language through incidental reading has been said to be one of the primary means of the continued development of a learner's vocabulary, according to Huckin and Coady (1999), any learner must already have a sight recognition vocabulary of approximately 3,000 words for this strategy to be effective. Other research also supports the effectiveness of deliberate learning of vocabulary, particularly at the beginning stages of learning (Elgort, 2011; Kang & Golden, 1994), and that learners need to take responsibility for their own learning for it to be successful (Nation, 2008).

A typical passage of text contains a number of different types of vocabulary that fall into four distinct categories (i.e., high-frequency, low-frequency, academic, and technical words; (Nation, 2001). Therefore, for learners of second language vocabulary, it may be crucial to include direct teaching of these high frequency words early in the learning journey so they can begin to negotiate reading and listening to the new language with comprehension. These strategies may also be as equally applied to the general learning of an individual's first-language as to any second-language learning (August, Carlo, Dressler, & Snow, 2005; Mueller Gathercole, Mon Thomas, & Hughes, 2008; NICHD, 2000).

Various approaches have been used to directly teach high-frequency vocabulary including using mobile phone applications (Lu, 2008); language laboratories and other audio-visual media (Vanderplank, 2009); and keyword, mnemonic techniques, and imagery (Beaton, Gruneberg, & Ellis, 1995). Supporting evidence for these techniques varies and suggests that keyword vocabulary learners do not retain much of what they had learnt over time (Wang & Thomas, 1995). For learners to retain second-language vocabulary over extended periods without on-going practice (maintenance), it has been suggested that there must also be a *fluency development* aspect to any instruction, whereby students are enabled to become more proficient with words they already know (Nation, 2008). This requires that educators design the curriculum to allow the learner to repeatedly cover the most useful language items, and practice these items to a fluent level of performance (Binder, 1996; Binder et al., 2002; Binder et al., 1995; Nation, 2008).

The focus for this study was on the learning of language in a mainstream school in the context of second language learning. Precision Teaching (PT) and fluency based instruction tactics and strategies have previously been used successfully to teach academic tasks to many individuals (Binder et al., 2002; Bloom, 1986; Hughes et al., 2007). Although PT methods have often been used with small numbers of participants we wanted to evaluate the use of these procedures within whole classes in the context of standard lessons (Beverley et al., 2009; Roberts & Norwich, 2010).

It has been shown that too often any teaching methods that are introduced in a school setting, are not usually continued once the researchers or other implementing bodies leave the organisation (Georgiades & Phillimore, 1975; Roberts & Hampton, 2008), and especially if the organisation receive no further training or coaching in those methods (Johnson & Street, 2004).

The aims of the study from the school's perspective was to intervene to buck the current trend with existing practices; teachers had noted that children from these classes tended to do poorly on learning new Welsh vocabulary words, even though this was the main intention of the class. Our remit from the school was that throughout the entire study we would try to enable the teachers to be in control and ultimately responsible for the introduction and running of the intervention within each classroom. This, in turn, would allow teachers to learn about these methods through active engagement whilst they applied the intervention in their normal classroom settings. Therefore, whilst the research team designed the intervention and evaluation methodology, the teachers ran the intervention at the beginning of each class, throughout the entire study. It was designed to be a brief intervention that would allow children many opportunities for active responding (Binder, 1996; Fredrick & Hummel, 2004; Heward, 1994; Johnson & Layng, 1996), but would not take up much of the class time, particularly because the teachers were under pressure to ensure that the curriculum was covered in its entirety during the school term. Therefore, the intervention we implemented was a simple, direct way of teaching. It was a short, sharp, focused, fluency-based method, which was not resource intensive but was easy to teach to the children and easy for teachers to implement. We wished to measure whether children who received this brief intervention would be able to correctly write more Welsh vocabulary words (from a randomly sorted list of their English equivalents) at post-test than the children who did not receive the intervention.

Method

Participants and Setting

The study took place with five classes of school children within one mainstream secondary school in North Wales. There were 95 children in the sample, with 50 males and 45 females. All children were between 11-12 years old. One class from the five selected was randomly allocated to be the Waiting List Control group (WLC; *n* = 16), the other four classes (*n* = 79) were allocated to receive the intervention: Intervention Groups 1 (I1; *n* = 25), 2 (I2; *n* = 16), 3 (I3; *n* = 18), and 4 (I4; *n* = 20).

Materials

The pre- and post-test Welsh vocabulary items were taken directly from the school curriculum for that semester. Two hundred Welsh vocabulary words were used that were derived from the following categories: School Uniform (34 words); Colours (20 words); Descriptors (54 words); School Mealtimes (64 words); and General (28 words).

These 200 words were further divided into three separate subtests that would be administered to children over three consecutive days (and would match the vocabulary content for the three flashcard packs that would be used during the intervention). Subtest one (*n* = 67, 34 School Uniform, 7 Colours, 26 Descriptors); subtest 2 (*n* = 67, 32 School Dinners, 7 Colours, 14 Descriptors, 14 General); subtest 3 (*n* = 66, 32 School Dinners, 6 Colours, 14 Descriptors, 14 General). Digital timers were provided so that children could accurately time their sessions at the beginning of every class. The children graphed their data on Standard Celeration Charts (SCC). These charts are semi-logarithmic; they have a calendar x-axis and a logarithmic y-axis. This allows daily progress to be charted and proportional changes in performance to be graphically displayed (see for example Calkin, 2005).

Design

Because the children had already been pre-assigned to school classes, we used a quasi-experimental design. The design therefore had two factors: one between-group factor (Group: WLC vs. I1, I2, I3, and I4) and one within-group factor (Test: pre- and post-test).

Procedure

Pre- and post-testing. Pre-testing for all groups took place over three days during the first week of class, using one of the three subtests for each day. Children were instructed to write the Welsh word in the space next to the English word, not to worry if they did not know a word, but simply to move on to the next word. They were further instructed to bring their completed sheets to the front of the class as soon as they had finished, so that no children would be able to rehearse their responses prior to the start of the study. The post-test used the same 200 words but was administered to children as one complete test; all the words were randomly sorted. Children followed the same procedure for post-tests as detailed above for the pre-tests.

Intervention. The week following pre-testing (week 2), each child in each of the intervention groups was given a pack of flashcards at the start of each lesson, beginning with Pack 1. They also received data sheets and graphs to record and chart their daily data. The fluency aim for each pack was a frequency of ≥ 50 correct responses (hits) per minute with ≤ 2 errors (misses). If a child met fluency aims for three consecutive days, they would progress onto the next pack of flashcards.

All children worked in pairs. At the beginning of each class the children would collect their work folder that contained all of their individual materials, and get their cards ready with the English writing facing toward them to prepare to carry out a warm-up practice session. Before they began, the teacher would remind them not to worry about knowing the words during the sessions, because this was not a test.

The first child of the pair would shuffle the pack to ensure that cards were in a random order and to work through the entire pack of flashcards as quickly as possible. They were instructed to *see* the English word on the front of the card (read silently), and then to *say aloud* (speak) the Welsh equivalent of that English word. After they had said the Welsh word for the English word on the card they would flip the card over to check if they had said the correct Welsh word. If they were correct, they would move on to the next card. If they were not correct, they would say the Welsh word aloud. This meant that through participation in the practice session they would have said the Welsh word correctly once for every card in the pack.

The first child of the pair would then be ready to carry out a one-minute timing. They would set their timer and get ready to begin their timing by shuffling the cards once again. As with the warm-up procedure, they would read the English word from the front of the card silently and speak the Welsh word aloud. They would then flip the card over for immediate feedback, either positive or corrective. Cards to which they had responded correctly (hits) were placed in a pile to their left hand side, whereas cards to which they had responded incorrectly (misses) were placed in a pile to their right hand side. Following this, the children counted both the number of hits and misses and recorded these on their Daily Scores Tables, and recorded their data on the SCC. A dot (•) was plotted for hits and an 'x' for misses. Once this was done, children would go through their misses again following the same procedure. This was repeated until they had made each card a hit (i.e., said one correct response to every card). At the end of this procedure, the second child from the pair would repeat the whole procedure—warm up and one-minute timing. The intervention was carried out daily, three-times per week, and lasted for four weeks overall (weeks 2 to 5 of the study).

Results

ANCOVA Analysis

A one-way ANCOVA was used to compare the post-test scores for the combined scores of the Intervention Groups (n = 79) with the Waiting List Control group (n = 16). The analysis revealed a significant effect for group, $F(1, 92) = 46.25$, $p < .001$, with the Intervention groups writing significantly more words correctly than the Waiting List Control group. Table 5.1 illustrates the associated means and standard deviations for the two groups. Using the Cochrane Method, an effect size was calculated using the *change scores* (post-test – pre-test) and SDs of the change scores for the Intervention and Waiting List Control groups (Higgins & Green, 2011, editors). The resulting effect size was large, *ES* = 1.88, 95% CI 1.28-2.48]

Table 5.1

Means and Standard Deviations for the combined Intervention Groups compared to the Control Group. Also illustrated in the table is the mean gain from pre- to post-test in the number of Welsh words written correctly.

	Pre-test	Post-test	Gain
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Waiting List Control	35.13 (20.12)	34.81 (22.58)	-0.32 (10.88)
Intervention	37.03 (24.24)	74.57 (32.98)	37.53 (21.50)

Each intervention group's data were also analysed using a *paired sample t*-test to compare differences in performance between pre- and post-test for the number of words correct per minute. It was found that all intervention groups pre- and post-test mean scores were statistically significant at least at a .001 significance level. Intervention Group 1 gained 36.96, $t(24) = 10.42$; Intervention Group 2 gained 37.47, $t(15) = 7.52$; Intervention Group 3 gained 20.89, $t(17) = 4.71$; Intervention Group 4 gained 52.87, $t(19) = 12.51$. These results clearly indicate that it was not any single intervention group that was driving the overall effect. Positive change was found in each of the four intervention classes.

Analysis of Individual Change

To enable the measurement of individual change improvement scores were calculated for each child; the pre-test scores for each child were subtracted from their post-test scores to provide us with a clear measure that illustrates whether children had improved, maintained, or deteriorated in performance from pre- to post-test.

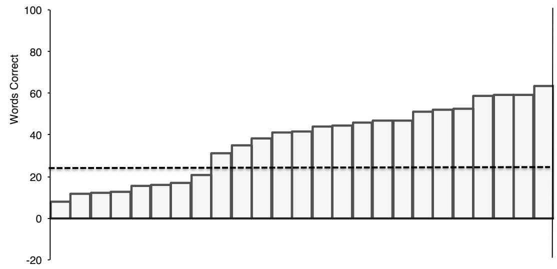
Using these improvement scores we then calculated how much improvement was needed to meet the criteria for reliable change. The RCI is calculated using the means and standard deviations of the groups at pre- and post-test to calculate the SE of the difference. Multiplying this difference by 1.96 provides a measure of magnitude of change required to be reliable at the $p < .05$ chance level (Evans et al., 1998; Jacobson & Truax, 1991; Zahra & Hedge, 2010). Thus a positive increase in change above this RCI criteria indicates an educationally significant increase in performance, whilst no change or deterioration of score indicates maintenance or decrease in performance respectfully.

We could further calculate the number of children from the Intervention Groups and Waiting List Control group for whom the intervention had been successful and represent this with the Numbers Needed to Treat (NNT; Altman, 1998; Barrowman, 2002; Bender, 2001; Lesaffre & Pledger, 1999; Pinson & Gray, 2003). The NNT provides us with a measure of the effectiveness of the intervention in terms of the number of the children required to receive the intervention in order to have one additional child achieve a positive outcome that would not have occurred if the child had received the comparison treatment. Therefore low NNT are preferred to high NNT.

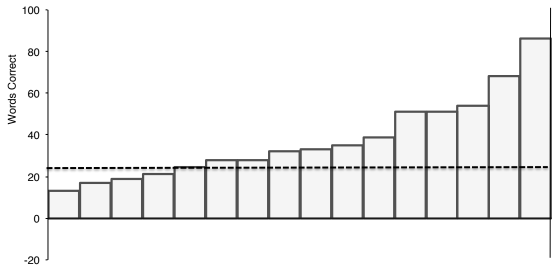
Figure 5.1e highlights the results from the RCI analysis on the WLC group's data. It clearly shows that no child in this group met the criteria to achieve reliable change (23.07 words correct or greater) when comparing their performance. Whilst nine children did improve, the other seven showed a decrease in performance. The poorest performing child achieved a deterioration of 24 words less than at pre-test.

Figure 5.1 (a-d) shows the performance of all children (n = 79) from the four intervention groups. It shows that 56 children (70.89%) in total achieved the RCI criteria, with only two of the remaining children showing a decrease in performance.

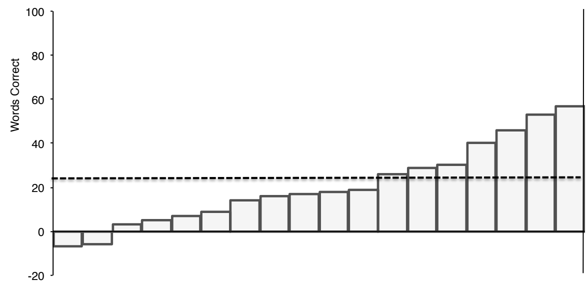
When considering the overall outcomes for all the children, the NNT is 2 (95% CI [1.2, 1.6]). This means that for every two children who received the intervention, one additional child responded positively to the treatment than would have if they had received teaching as usual (i.e., what was provided for the WLC group).



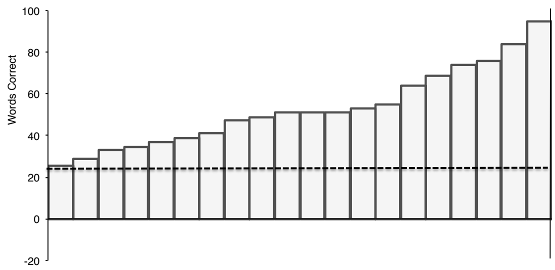
5.1a] Intervention Group 1 (n = 25); Reliable change achieved by 17 children



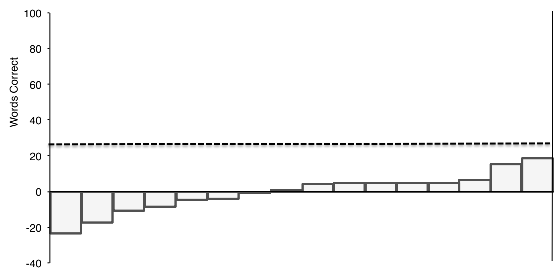
5.1b] Intervention Group 2 (*n* = 16); RCI achieved by 12 children



5.1c] Intervention Group 3 (*n* = 18); RCI achieved by 7 children



5.1d] Intervention Group 4 (*n* = 20); RCI achieved by 20 children



5.1e] Waiting List Control Group (*n* = 16); RCI achieved by 0 children

Figure 5.1. Shows charts for each of the intervention groups and the control group respectively. Scores above zero are gains in performance from pre- to post-test in the number of Welsh words written correctly, scores at or below zero are maintenance or losses in performance respectively. Scores above the dotted line indicate that the improvement is reliable at the individual level as measured by the *Reliable Change Index* (RCI).

Discussion of Chapter 5

The impetus and main aim for introducing this study into the school setting was that teachers had informed the researchers that children from these classes tended to do poorly on learning new Welsh vocabulary words, even though one of the main aims of the class was to achieve this very outcome.

The findings from this study illustrate support for this brief intervention that employed a direct method of teaching second-language vocabulary that was easy to implement, did not require specialised equipment, and was not costly in terms of time or other resources. The children in each of the intervention classes improved, showing that it was not just one of the intervention classes that was driving the effect; rather that the intervention's effect was distributed through all the intervention classes. This is again highlighted in the large number of individual children who benefitted reliably from receiving the intervention.

A large group effect had been found when comparing the combined intervention groups with the control illustrating the effectiveness across intervention groups. Additionally, the low NNT from this study's results indicate the success of the intervention (for those who had received it). Although there is no existing research providing comparison NNT figures for second-language vocabulary interventions, these low numbers do compare favourably to NNT rates for treatments for major depression (NNT between 3-5) and bulimia nervosa (NNT = 9) (Pinson & Gray, 2003).

A secondary aim of the study had been to give responsibility to the teachers to implement and manage the study from the outset. This was designed to increase the likelihood that the methods could continue to be used in future classes without relying on support from the research team, as previous research suggests that enabling staff will lead to increased chance of the successful methods being used in future (Georgiades & Phillimore, 1975; Roberts & Norwich, 2010).

Despite these impressive results, there are a number of limitations of the present research. The initial pre-tests were administered as three separate subtests over a consecutive three-day period (which tallied with the three flashcard pack contents). However, the final post-test was given as one single test in just one sitting. This limitation in experimental control was due to the balance we were striving to achieve between running a research study in a real world setting whilst still enabling the teachers to be responsible for the implementation and management of the intervention within each classroom.

This present study failed to carry out follow up testing of children at a period after post-tests; otherwise we could have explored whether the gains for the intervention children continued to maintain over time in comparison to the Waiting List Control children.

From the results, we are not able to conclude whether the obtained effect then acts as a pivotal language skill for better language learning, future research should attempt to better control in both consistency and standardisation of the measures used, and ensure that the outcome measures include tests of application of vocabulary in real world settings, such as in writing stories or conversational skills.

This current study did not try to separate the effects of the SAFMEDS intervention from the process of charting and monitoring individual progress. Previous research has shown that children charting their own learning can have motivational effects (Bower, 1985; Lindsley, 1995a; Maloney, 1993), and that “Going fast and beating one’s previous score has gamelike qualities and motivates children to do better.” (Kubina et al., 2002, p. 243). Future research might attempt to pick apart these effects, perhaps by having three separate conditions: SAFMEDS, SAFMEDS plus charting, and a Waiting List Control.

The teaching methodology employed for this study was inexpensive to implement, both in terms of time and resources. Initially, whilst children were learning the procedure the time taken was longer, but once children were conversant with the procedure it took approximately 5 minutes of lesson time at the start of each 50 minute class. The children were responsible for carrying out the entire procedure in their pairs, requiring only teacher supervision, so it was not demanding on teacher time.

Anecdotally, children appeared to enjoy this method of learning, as they would frequently already have begun working through their timings in pairs even before the teacher had called the class to order to begin the classroom activities. This was also emphasised as certain children were requesting that they be allowed to take their cards home to practice (something that our initial research design would not allow). Again, anecdotally the teachers reported an improvement in the punctuality of the children—if children were not there at the beginning of the class they would miss both their timing and the opportunity to beat their previous personal best score.

As this research took place in a real world setting, and empowered the teachers to be the main instigators of the entire intervention, it lacked the control that could be achieved with a more rigorous experimental design; but nevertheless, we were able to measure the effects of the intervention at the individual level by using Reliable Change Index (RCI) methodology. There is an opportunity in the future for this type of research to be conducted as a basic experiment and look specifically at the length of time it takes to become fluent on each set of flashcards; how effective this fluent learning is in aiding recall after periods of time without practice, and what are the precise range of rates of responding that result in the proposed outcomes of fluency: Retention, Endurance, Stability, and Application (RESA; Binder, 1996). It is probable that only when these outcomes of fluency are apparent in a learner that the learnt vocabulary will be generalised into other second-language behaviours:

for example, using appropriate learned vocabulary in conversation, reading, and writing. Additionally, future research could explore the effectiveness of these techniques when applied to first-language vocabulary teaching, as well as being applied more broadly in second-language settings.

It would be perhaps beneficial for this research to be replicated in a real world setting, but try to gain an improvement in experimental control within that setting. For instance, if we had greater researcher contact in the initial stages of the study, we could then fade out this contact over time with the eventual aim of providing monthly coaching support. We would postulate that this increase in the level training and on-going support would enable teachers to better, and more confidently, incorporate these strategies and tactics into their normal everyday teaching activities. Nevertheless this procedure has wide applicability and would appear to tick all the boxes; it is cheap, brief and extremely effective—in fact effective for nearly every child in this study.

Chapter 6: Discussion of Thesis

Broad Overview

The need for effective research-proven methods of teaching to ensure that all children learn efficiently is vitally important to every child's future and indeed to society as a whole (Barrett et al., 1991). Papers 1, 3, and 4 contained within this thesis have sought to conduct research into evidence-based methods of education, which have a long track history of success, investigate them empirically, and then disseminate these findings to a wider audience of educators within both mainstream and special needs educational settings. Within this process the aim has been to equip teachers to effectively develop their research skills, so that they can take these methods proven through *evidence-based practice* and use these effectively to collect *practice-based evidence* within their own classroom environment. Consequently, whilst engaging in this process they can use the measures gathered to ensure that their learners are progressing; and where they are not progressing at an adequate pace, intervene in a timely manner.

Chapter Analysis

Paper 1 was a single case design study. Five children received the intervention targeted at improving their accurate reading fluency. The study used two instructional stimuli: SAFMEDS and randomised word sheets. The results demonstrated the effectiveness of the intervention, which provided small-targeted units of practice to develop fluency in reading selected vocabulary. Although the intervention group only spent a short period of time each day independently practicing their reading, they nevertheless made significant gains in reading ability—both in terms of the practiced word lists and unpractised application test. An example of application of skills resulting in spontaneous learning (Binder, 1996; Johnson & Layng, 1994)

Paper 2 was a pre-post test group design, but endeavoured to incorporate important elements of single case experimental designs within the study. All undergraduates taking part in the intervention charted their daily progress on SCC and used these data to ascertain individual learning growth and to intervene with strategies to increase their learning when the data indicated they should. Results showed that the intervention group outperformed the TAU group on two measures—performance on a randomly selected MCQ at post-test, and on their weekly tests that were part of their module assessment. The intervention group's performance was one grade level higher in comparison to the TAU group at post-test, and on the weekly tests they outperformed the TAU group by an average of 10%. The results provide support for the concept that fluency with basic statistic terminology (component skill) can improve the performance of higher order (composite) skills and, as we demonstrated in Paper 1, that the intervention need not be costly in terms of teaching time or resources.

Paper 3 was conducted in a mainstream primary school setting, and investigated whether a brief fluency-building intervention could significantly improve essential maths skills within this population, and additionally whether such regular practice in simple addition skills (component) could also have impact on higher level maths skills (subtraction) that are the reciprocal skill (composite skill). In this research we again sought to extend the analysis in terms of incorporating principles of analysis from both group and single case designs, and moreover to develop the way that the results could be communicated with a broader audience in a meaningful way. To this end the RCI and NNT measures were introduced to the analysis to allow comparison not only in terms of group performance, but also in terms of the effects on individual learners, and the *educational significance* of those effects (see Paper 3—Data Analysis Approach).

Paper 4 continued the theme from paper 3 in the use of both group and single case design and in the use of the RCI and NNT; but also endeavoured to give more autonomy to the teachers in the schools. We sought to equip them with the training and support needed to allow them to conduct the study on their own, with minimum intervention from the research team. The purpose of this study (as with much real life or applied research in schools) was to address a problem that the teachers were experiencing at the time. They had noted that pupils in general were not learning new vocabulary words essential for them to learn the curriculum. We therefore implemented a brief fluency-building intervention designed to aid them with second language vocabulary learning. Results were encouraging in demonstrating the significant impact that this short focussed practice could enhance children's development of second-language vocabulary. Although the children only engaged in this practice for 15 minutes per week (over four weeks), again we showed that this (resource light, but effective) approach could demonstrate educationally significant change. Clearly, if brief amounts of fluency-building can have such a beneficial effect, then it would be prudent to incorporate this as part of a school's everyday classroom practice. Later in the discussion I explore the difficulties of bringing about organisational change that is necessary to embed such methods within the normal teaching practices.

An important aspect of all the papers in this thesis is that the research was conducted across different types of educational setting and with different populations. Papers 1, 3, & 4 were conducted in mainstream schools (primary and secondary), and paper 2 was conducted within a university setting with undergraduates as the targeted group. Combining these four studies the theme of this thesis highlights the effectiveness of the strategies and tactics of PT and fluency-based interventions across settings and populations. Furthermore, they emphasise the efficacy of such interventions, especially in light of their limited demands in terms of cost of implementation, time taken from the available teaching time, and perhaps even more importantly, the ease with which teaching staff can be trained to use these approaches.

The emphasis, especially in the later studies in this thesis, has been focussed on enabling the teachers to receive training that allows them to further develop the skills for themselves, in order that they are autonomous in their ability to deliver their teaching and monitor its effectiveness. It is arguably only through this approach that these methods can be seamlessly incorporated into daily teaching practices, and increase the likelihood that these methods will be used in future classes after the study has ended. Moreover, in order to implement such methods in a classroom setting, they must not be costly both in terms of implementation and also the amount of time they take during the normal school day.

At the time I started this research we were not aware of any schools locally that were using or had received training on PT or FBI methods. During the period of conducting this research, the number of schools involved has expanded and at the point of writing this most of the schools in the area have received training from us in the use of PT and FBI methods, and eleven settings are actively using these methods across a wide variety of curriculum areas to help children learn better. Of these eleven, six are special educational needs (SEN) schools and five are mainstream schools situated in all three of the counties surrounding Bangor University: Gwynedd, Conwy, and Ynys Môn (see Appendix C).

Additionally all the research conducted for this thesis has sought to add further supporting evidence to the current database of the effectiveness of PT and fluency-building procedures, and in the case of Paper 2, to use these methods to benefit the students I teach as part of my university research methods and statistics teaching. The use of SAFMEDS has transferred across many events that are now run within the school: 6th form taster days; undergraduate statistics classes, and a master programme in Applied Behaviour Analysis. Their use also forms a major part of the assessment in my Year 3 education class—Evidence-based Behavioural Methods in Education. In this module students are asked to coach another student with their learning project. Data from the learner is plotted on SCC and used to make decisions about interventions to improve the efficiency of their learning. Charts are shared in class as mini chart shares that local teachers, behaviour analysts, and professionals also attend (see Appendix I), and at the end of the module all students share their learning project in a three-minute presentation (see Appendices J & K).

Methodological Limitations of the Research

As all of the research was conducted within real-life settings, inevitably we were not able to control and manipulate variables as might have been possible within a laboratory. Nevertheless, whilst this may be considered a methodological weakness, it is also a particular strength of the studies, as in every setting we were seeking to have a positive impact on learners and allow them to gain significant improvement in their fluency of essential skills. A central theme of this thesis was to transfer knowledge and procedures into applied settings. As well as positively impacting the learners, we also sought to empower the teachers by providing them with effective, simply implemented, and cost efficient methods that could be introduced into their classrooms with minimal effort and yet have maximum impact. Through the implementation of the research studies I strove to enable existing classroom settings to incorporate elements of PT procedures with the aim that teachers might choose to continue to use the procedures once they had been demonstrated to be effective and efficacious. In this context the papers in the current thesis represent examples of applied research; a key aspect of all four studies was the implementation of effective practice into real-world settings.

Paper 1 had the limitation that as a single case design, we are thus limited in how much we are able to generalise our findings, but nevertheless for the children in the intervention group the fluency-building procedures significantly enhanced their reading ability of targeted and non-targeted words.

In paper 2 we were not able to monitor the daily use of SAFMEDS by the intervention group, although we did reliability checks by having the participants conduct timings at their bi-weekly meetings. Additionally we were not able to monitor study time or motivation of either the intervention or TAU group (education as usual condition); and we perhaps could have also had TAU students meet with a student proctor every two weeks, as the intervention group did. In general when we have a TAU group perhaps we should try in future to incorporate an *attention control condition* that receives the same social contingencies as the intervention groups.

Paper 3 had a limitation in that we had teachers select the children to be included in the study, defined by their perceived needs. Although we then randomly assigned children to either the TAU or intervention group, it would have been preferable to screen the entire class and allocate based on performance on a pre-test. However this again is a consequence of balancing between laboratory or real-life settings.

Paper 4 had a limitation in the difference of how the pre- and post-test were administered. At pre-test the students received three separate mini tests, at post-test they received the entire test at one session. Therefore the analysis may have been somewhat compromised in that we could not match the amount of SAFMEDS packs each child had worked through with the final results. Closer control of this in future research could improve this aspect ; however it also needs to be noted that this probably resulted in the analysis being conservative in estimates of improvement. The number of words that had been covered through SAFMEDS varied between children in the intervention group, so the post-test included many words that they had not actually studied in this way; it is highly likely that had we been able to isolate the analysis to reflect the pre-post differences with only the cards that were studied our findings would have been even more impressive.

Future Directions

Future research could enhance the methodology used in paper 2, perhaps by having both groups of students have access to the same SAFMEDS (flashcards) but only have the intervention group receive training in the correct methods of SAFMEDS and charting procedures, or perhaps even have three groups: flashcards no instruction, SAFMEDS, and SAFMEDS plus SCC charting to attempt to discern further whether the fluency-building procedures benefit further from the visual feedback from the charting of learning. Future research might also benefit from a web-based intervention that would allow the more refined measurements of response latency and outcomes of fluency: RESA. In collaboration with a colleague from Akershus University, Oslo, Norway, we are currently piloting basic research in this area. We intend to use a stimulus equivalence paradigm to attempt to discriminate differences in learning outcomes between groups who learn to accuracy or fluency.

In paper 3 we added evidence to the relationship between composite and component parts of essential skills; in some ways similar to the findings of previous research (Chiesa & Robertson, 2000). Future research should continue to research the effects of building component skills to fluency and investigate the effects this can have on composite (untrained) skills. It would be interesting for instance to compare the effectiveness of traditional maths teaching (addition and subtraction) with the more economical teaching of fact families (where addition and subtraction are part of one family of facts).

As I have progressed through the process of conducting research within school settings as part of this thesis, and introduced (through training) teachers to the systems and strategies of fluency-building and PT, I have added further support of the effectiveness of these methods and the impact that they could have in education settings for both mainstream and SEN pupils, and adults of any age range (see Appendices C-I). Outside the remit of my PhD, I have recently carried out a pilot project using PT systems and strategies with adult learners learning Welsh, and have also been using these methods to increase my fluency in speaking Welsh. It has become a focus of both the research and myself as a personal goal to continue to disseminate these methods to any educational setting that could benefit from instruction.

Reflecting on these studies that rely on teachers to deliver interventions, it is clear that more research needs to be carried out to investigate why there is still resistance to teachers implementing these practices into their regular teaching. Research supports the effectiveness of many methods of Applied Behaviour Analysis when used within educational settings (Heward, 2005; Heward et al., 2005; Moran & Malott, 2004; Vargas, 2009), and both Skinner (1984) and Lindsley (1992b) have lamented why there seems to be a resistance to adopt these available, effective, and often resource light practices. Additionally we are aware of the 'knowledge-to-practice gap' (Barbash, 2011; Lamb, 2001) that may prevent things being readily accepted into usage; yet, also we know that if certain conditions are in place, their combination can result in rapid change. The question this raises is how can we bring about this combination of conditions that can result in, what might be regarded as a *Tipping Point* (Gladwell, 2000)?

Roberts and Hampton (2008) propose that the literature supports four broad aspects of *change practice* that can help the adoption of effective practices within systems: (a) building successes, (b) developing supportive groups, (c) involving management, and (d) professional development and support. Reflecting on the process of the research carried out within the local schools that we have formed partnerships with; the following components have perhaps worked together to slowly bring about such change.

Local teachers from schools we have partnerships with currently present at both an annual ABA conference within the university, and through chart shares in my year three undergraduate module (see Appendix I). Here we celebrate the effective use of PT within their schools and hear about the ways in which these methods have helped their clients and also improved their methods of practice (points a & b). We have provided, and continue to provide, professional training and support to teachers. Whilst receiving further training there are always ample opportunities to discuss their own successes with teachers from other schools (points b & d). However, none of this could have happened without support from the school heads; this is a crucial part of any attempt to integrate effective practices within schools (point c).

Perhaps the key element is to have at least one trained person who works within many schools within the area, and who is willing to continue to champion their use within schools. Undoubtedly, they also require the support from the head teacher, and the staff members must by their own volition adopt these methods for use in classrooms. However, if all these aspects of change are in place, then slowly these changes can impact on the environment, and there is an increased opportunity for other teachers to observe first hand the benefits of such practices and adopt them in the future. Rather than taking a common approach when training others in a specific approach, termed *the myth of the hero innovator* (Georgiades & Phillimore, 1975), it is more about gradual systemic change within the establishment that can bring about changes in effective practices. As has been shown before, traditional forms of training cannot bring about systematic change within an organisation, a combination of factors must be in place for this to effectively happen (Georgiades & Phillimore, 1975; Roberts & Hampton, 2008). We can provide teachers with the tools that can facilitate change but it is down to a combination of the factors listed above that come together to support the gradual implementation of effective practice and systemic change.

Current and Imminent Initiatives

We are continuing to be able to influence the adoption of both Headsprout Early Reading Programme and Headsprout Reading Comprehension within local schools. Although this was not directly part of my thesis, project students of mine have been carrying out research in this area for a number of years, and encouragingly the schools are now valuing the benefits of these programmes to such a degree that they are now buying their own licences, and incorporating Headsprout into their everyday curriculum (for example, nine schools around Bangor are now conducting research and evaluation research with us using Headsprout programmes and all are purchasing their own licences to work with their children).

Additionally we are in the process of carrying out a survey to attempt to discover more about possible causes of resistance to implementation within SEN settings. We aim to investigate the barriers that teachers feel they cannot overcome when teaching literacy skills to children with learning difficulties.

Clearly there also needs to be a drive to further disseminate these practices to the wider public and to communicate their benefits in as direct a way as possible. Unfortunately we need to remember that there exists a *knowledge to practice gap* that can delay the implementation of effective practice, even when existing evidence supports such practices.

Nevertheless, it is wonderful to view the extent to which these methods have been adopted into various settings across North Wales, and how they continue to be used effectively for the teaching of many skills. These skills are not only academic. SAFMEDS have been used as an aid to encourage more social interactions. A simple procedure resulted in children within a SEN setting to learn the names of other students and staff in their school setting, which in turn allowed them to increase the number of occasions that they greeted and talked to each other throughout the school day, and thus vastly increase the number of self initiated interactions with others. This is merely an example of their use, which is apparent when you see evidence of the widespread use of SAFMEDS within an educational setting (see Figure 6.1). This also provides an example of how schools must adopt a very organised implementation of PT methods, as materials need to be readily available for use and easily stored away for future sessions.

More recently I have been involved in further discussions with one of the local schools and, we are planning to launch a *Learning Centre* that will serve three crucial educationally beneficial purposes within one local school: (a) provide interventions in essential skills for targeted pupils in the school (i.e., children currently struggling and at risk of academic failure); (b) train school staff such that they can take these procedures into practice (i.e., skilling up current educators); and (c) provide practical training and work experience for students who are studying on my year three module Evidence-based Behavioural Methods in Education (i.e., the future teachers and educators). I developed the module at the outset of this thesis and to date have had approximately 350-400 students who have benefitted in receiving training in PT and other evidence-based approaches. Many of these students have gone on to become teachers, and I have heard from them how they have incorporated these methods in their teaching.

Figure 6.1. SAFMEDS storage station in one of the North Wales SEN settings.

The Learning Centre will be free of charge to struggling learners and in the future we aim to expand the number of children we can intervene with and the number of students who will gain practical experience within the centre. We aim to do this by restructuring the current year three module that runs across one semester to instead run over two semesters and incorporate a practical element of teaching children who need to develop their fluency in essential skills. This will be the first module in my psychology department to focus on developing practical skills in undergraduates within special educational settings. Additionally, having been given access to *Aimchart* by Professor Chuck Merbitz, we are planning to begin to roll this out throughout the schools as a means to encourage the sharing of data across schools; not only in terms of viewing data from many children simultaneously, but also in terms of sharing insights into effective interventions, so that teachers can begin to collaborate on making effective decisions. To clarify, *Aimchart* is a web-based data collection application, which can display learner's data on a SCC. It allows many people to view many learner's data and so can enhance the decision making progress across locations, as long as teachers have Internet access. This has the possibility of increasing the available supervision for teachers who may be new to PT, and be able to look and make decisions on data even when not in the same locality.

The progress we have made in North Wales continues to encourage my focus, and I hope that it will continue to do so for many more years to come and influence the educational practice within as many schools as possible. The practice of charting individual learner's progress and using these measures to guide their practice and the instruction that these learners receive is a vital addition to any curriculum. The more that educators can be influenced to use *evidence-based practice* and to monitor the learning of the pupils they teach—their *practice-based evidence*—the more that individual learners will be allowed to benefit from the methods that have been demonstrated successful. It is an on-going aim to provide teachers with the tools that have been developed from behaviour analysis to monitor their classroom practices and to allow them to take objective measurements to guide their practice and thus introduce the *Fourth R* into their classrooms—RESEARCH (Haring, Lovitt, Eaton, & Hansen, 1978).

This is a crucial ongoing aim, as only when measurements are used to guide instruction can we adopt a truly scientific approach to teaching that can aid all learners. As Carroll quite correctly stated “The concept of ‘underachievement’ does not automatically imply the possibility of remediation any more than the concept of illness does. Some patients never get well, and some underachievers remain underachievers.” (1963, p. 731). This was written 50 years ago.

To this aim I will continue to give my committed support. By carrying out research for my thesis, I have had opportunities to effect some change into the practice within local schools. Whilst this is relatively small success thus far, it is nevertheless in real terms socially significant success for the children whose education it has influenced. I particularly keep in mind what Fred Keller reported Burrhus Fredrick Skinner to have said when asked what can be done to promote better education.

> "Well, I guess we just keep nibbling." I take that to mean to keep on working in a small way, keep on promoting good things. When you see something good taking place, reinforce it if you can. When you see something going in the right direction, praise it. Anytime you see a model school that looks as if it's applying good behavioral principles, give it your support. Every time you hear of somebody doing something good, drop him or her a line and say, "Thank you very much for what you're doing." I believe the process is something like shaping. Don't expect many big changes to take place. There's not going to be any revolution. But maybe, if we all

keep on nibbling, we can change education. I don't know of any other way. (Heward & Dunne, 1993, p. 340)

Therefore I will continue to do as advised and *keep on nibbling*.

References

ALBSU. (1993). The cost to industry. Basic skills and the UK workforce. Retrieved June 12, 2003, from http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED356397

Alper, T., & White, O. R. (1971). Precision teaching: A tool for the school psychologist and teacher. *Journal of School Psychology, 9*(4), 445-454.

Altman, D. G. (1998). Confidence intervals for the number needed to treat. *BMJ, 317*(7168), 1309-1312.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English Language Learners. *Learning Disabilities Research & Practice, 20*(1), 50-57. doi: 10.1111/j.1540-5826.2005.00120.x

Baguley, T. (2012). Can we be confident with our statistics? *The Psychologist, 25*(2), 128-129. doi: 10.3758/s13428-011-0123-7

Barbash, S. (2011). Clear Teaching: with Direct Instruction, Siegfried Engelman discovered a better way of teaching. Retrieved from http://education-consumers.org/ClearTeaching.htm

Barber, M. (1997). *The learning game: Arguments for an educational revolution*. London: Indigo.

Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition, maintenance, and generalization of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*(1), 26.

Barrett, B. H. (1979). Communitization and the measured message of normal behavior. In R. York & E. Edgar (Eds.), *Teaching the severely handicapped* (Vol. 4, pp. 300-318). Columbus, Ohio: Special Press.

Barrett, B. H., Beck, R., Binder, C., Cook, D. A., Engelmann, S., Greer, R. D., . . . Watkins, C. L. (1991). The right to effective education. *The Behavior Analyst, 14*(1), 79-82.

Barrowman, N. J. (2002). Missing the point (estimate)? Confidence intervals for the number needed to treat. *Journal of the Association of Computing Machinery, 166*(13), 1676-1677.

Baxter, J., & Frederickson, N. (2005). Every Child Matters: Can educational psychology contribute to radical reform? *Educational Psychology in Practice, 21*(2), 87-102.

Beaton, A., Gruneberg, M., & Ellis, N. (1995). Retention of foreign vocabulary learned using the keyword method: a ten-year follow-up. *Second Language Research, 11*(2), 112-120. doi: 10.1177/026765839501100203

Beck, R., & Clement, R. (1991). The Great Falls precision teaching project: An historical examination. *Journal of Precision Teaching, 8*(2), 8-12.

Bender, R. (2001). Calculating confidence intervals for the number needed to treat. *Controlled Clinical Trials, 22*(2), 102-110.

Beverley, M., Hughes, J. C., & Hastings, R. P. (2009). What's the probability of that? Using SAFMEDS to increase undergraduate success with statistical concepts. *European Journal of Behavior Analysis, 10*(2), 235-247.

Bijou, S. W. (1970). What psychology has to offer education—now. *Journal of Applied Behavior Analysis, 3*(1), 3-65. doi: 10.1901/jaba.1970.3-65

Binder, C. (1988). Precision Teaching: Measuring and attaining exemplary academic achievement [Electronic Version]. *Youth Policy Journal, 10*(7), 12-15. Retrieved from http://www.fluency.org

Binder, C. (1990). Precision teaching and curriculum based measurement. *Journal of Precision Teaching, 7*(2), 33-35. Retrieved from http://www.fluency.org/

Binder, C. (1991). Marketing measurably effective instructional methods. *Journal of Behavioral Education, 1*(3), 317-328. Retrieved from http://www.fluency.org/

Binder, C. (1993). Behavioral fluency: A new paradigm. Retrieved March 21, 2006, from http://curry.edschool.virginiad.edu/go/specialed/papers/Binder-et-al_Fluency.pdf

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*(2), 163-197.

Binder, C. (2001). Measurement: A few important ideas [Electronic version]. *Performance Improvement Quarterly, 40*(3), 20-28. Retrieved from http://www.fluency.org

Binder, C. (2003). Doesn't everybody need fluency? *Performance Improvement Quarterly, 42*(3), 14-20. doi: 10.1002/pfi.4930420304

Binder, C. (2004). A refocus on response-rate measurement: Comment on Doughty, Chase, and O'Shields. *The Behavior Analyst, 27*(2), 281-286.

Binder, C., Haughton, E., & Bateman, B. (2002). Fluency: Achieving true mastery in the learning process. Retrieved March 21, 2006, from http://www.fluency.org/Binder_Haughton_Bateman.pdf

Binder, C., Haughton, E. C., & Van Eyk, D. (1995). Increasing endurance by building fluency: Precision teaching attention span. *Journal of Precision Teaching, 12*(2), 29-34.

Binder, C., & Watkins, C. L. (1990). Precision Teaching and Direct Instruction: Measurably superior instructional technology in schools [Electronic version]. *Performance Improvement Quarterly, 3*(4), 74-76.

Bloom, B. S. (1986). The hands and the feet of genius: Automaticity. *Educational Leadership, 43*(5), 70-77.

Born, S., Harder, S. R., & White, O. R. (2003). Standard Celeration Chart 6x6 v1-2: A Microsoft Excel Chart Template. [Computer software] Cambridge, MN: Non commercial software. Retrieved May 2, 2008, from http://harderchartingtemplates.pbworks.com/

Bower, R. (1985). Self-charting: Giving kids a chance. *Journal of Precision Teaching, 6*(1), 14-20.

Boyce, T. E. (2003). Moving from precision teaching to precision measurement: We need to just DO IT. *European Journal of Behavior Analysis, 4*, 59-63.

Bradshaw, J., Ager, R., Burge, B., & Wheater, R. (2010). *PISA 2009: Achievement of 15-year-olds in England*. Slough: NFER.

Bramlett, R., Cates, G. L., Savina, E., & Lauinger, B. (2010). Assessing effectiveness and efficiency of academic interventions in school psychology journals: 1995–2005. *Psychology in the Schools, 47*(2), 114-125. doi: 10.1002/pits.20457

Briars, D., & Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology, 20*(4), 607-618. doi: 10.1037/0012-1649.20.4.607

Broker, S. (2000). *Legendary places*. Berkshire: Lands End Publishing.

Brusling, C. (2005). *Evidence-based practice in teaching and teacher education: What is it? What is the rationale? What is the criticism? Where to go now?* Paper presented at the Professional Development of Teachers in a Lifelong Perspective: Teacher Education, Knowledge Production and Institutional Reform, November, 17-18. 2005., Centre for Higher Education Greater Copenhagen in collaboration with OECD, Copenhagen, November, 17-18. 2005.

Bucklin, B. R., Dickinson, A. M., & Brethower, D. M. (2000). A comparison of the effects of fluency training and accuracy training on application and retention. *Performance Improvement Quarterly, 13*(3), 140-163. doi: 10.1111/j.1937-8327.2000.tb00180.x

Calkin, A. B. (2003). Some comments on precision teaching. *European Journal of Behavior Analysis, 4*(1 & 2), 1-4.

Calkin, A. B. (2005). Precision teaching: The standard celeration chart. *The Behavior Analyst Today, 6*(4), 207-213.

Carnine, D. (2000). *Why education experts resist effective practices (And what it would take to make education more like medicine)*. Washington, DC: Thomas. B. Fordham Foundation.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 723-733.

Carter, R. (1987). Vocabulary and second/foreign language teaching. *Language Teaching, 20*, 3-16. doi: 10.1017/S0261444800004225

Cawley, J. F., Miller, J. H., & Carr, S. C. (1990). An examination of the reading performance of students with mild educational handicaps. *Journal of Learning Disabilities, 23*, 284-290.

CEM. (2007). Middle Years Information System (MidYIS) Retrieved from http://www.cem.canterbury.ac.nz/midyis/

Chase, P. N., Doughty, S. S., & O'Shields, E. M. (2005). Focus on response rate is important but not sufficient: A reply. *The Behavior Analyst, 28*(2), 163-168.

Chiesa, M., & Robertson, A. (2000). Precision teaching and fluency training: Making maths easier for pupils and teachers. *Educational Psychology in Practice, 16*(3), 297-310.

Claypool-Frey, R. (2009a). Learning pictures and change effects. Retrieved 15 July, 2009, from http://precisionteaching.pbworks.com/Learning-Pictures-and-Change-Effects

Claypool-Frey, R. (2009b). SAFMEDS. Retrieved June 9, 2009, from http://precisionteaching.pbworks.com/SAFMEDS

Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge, UK: Cambridge University Press.

Conners, F. A., McCown, S. M., & Roskos-Ewoldsen, B. (1998). Unique challenges in teaching undergraduates statistics. *Teaching of Psychology, 25*(1), 40-42.

Cowan, R., Donlan, C., Shepherd, D.-L., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology, 103*(4), 786-803. doi: 10.1037/a0024556

Cumming, G. (2012). *The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. Hove, East Sussex: Taylor and Francis Group.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*(4), 507-524.

Department for Education. (2010). *The importance of teaching - The Schools White Paper 2010*. (CM 7980). The Stationary Office Limited Retrieved from http://publications.education.gov.uk/eOrderingDownload/CM-7980.pdf.

Department for Education and Skills. (2002). *Special Educational Needs: Code of Practice*. (DfES/581/2001). Retrieved from https://http://www.education.gov.uk/publications/eOrderingDownload/DfES 0581 200MIG2228.pdf.

DfEE. (1999). The national numeracy strategy: Framework for teaching mathematics, reception to year 6. London: Department for Education and Employment.

DfES. (2003). Targetting support: Choosing and implementing interventions for children with significant literacy difficulties Vol. 2007. Retrieved from http://www.standards.dfes.gov.uk/literacy

Dolch, E. W. (1948). *Problems in reading*. Champaign IL: Garrard Press.

Doughty, S. S., Chase, P. N., & O'Shields, E. M. (2004). Effects of rate building on fluent performance: A review and commentary. *The Behavior Analyst, 27*(1), 7-23.

Dowker, A. (2009). *What works for children with mathematical difficulties?* Nottingham: DfES Publications.

Doyle, W. (1983). Academic work. *Review of Educational Research, 53*, 159-200.

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2010). Using participant data to extend the evidence base for Intensive Behavioral Intervention for children with autism. *American Journal on Intellectual and Developmental Disabilities, 115*(5), 381-405. doi: 10.1352/1944-7558-115.5.381.

Eldevik, S., Hastings, R. P., Jahr, E., & Hughes, J. C. (2012). Outcomes of behavioral intervention for children with autism in mainstream pre-school settings. *Journal of Autism and Developmental Disorders, 42*(2), 210-220. doi: 10.1007/s10803-011-1234-9

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367-413. doi: 10.1111/j.1467-9922.2010.00613.x

Ericsson, K. A., Krampe, R., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406. doi: 10.1037/0033-295x.100.3.363

Eshleman, J. W. (1990). The history and future of precision teaching. *Journal of Precision Teaching and Celeration, 7*(2), 18-27.

Eshleman, J. W. (1992). A behavioral measurement parable. *Journal of Precision Teaching and Celeration, 9*(1), 6-19.

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health, 1*(3), 70-72.

Everson, M., Zieffler, A., & Garfield, J. (2008). Implementing new reform guidelines in teaching introductory college statistics courses. *Teaching Statistics, 30*(3), 66-70.

Fabrizio, M. (2004). *The application of precision teaching with children with Autism: Perceived problems with possible solutions [DVD]*, 30th Annual ABA Convention, Boston, Masachusetts.

Fabrizio, M., & Moors, A. L. (2003). Evaluating mastery: Measuring instructional outcomes for children with autism. *European Journal of Behavior Analysis, 4*(1 & 2), 23-36.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications Limited.

Fitzgerald, D. L., & Garcia, H. I. (2006). Precision Teaching in developmental mathematics: Accelerating basic skills. *Journal of Precision Teaching and Celeration, 22*, 11-28.

Fox, L., Carta, J., Strain, P., Dunlap, G., & Hemmeter, M. L. (2009). Response to Intervention and the pyramid model. Retrieved from http://www.challengingbehavior.org/do/resources/documents/rti_pyramid_web.pdf

Fredrick, L. D., Deltz, S. L., Bryceland, J. A., & Hummel, J. H. (2002). *Behavior analysis, education, and effective schooling*. Reno, Nevada: Context Press.

Fredrick, L. D., & Hummel, J. H. (2004). Reviewing the outcomes and principles of effective instruction. In D. J. Moran & R. W. Malott (Eds.), *Evidence-based educational methods* (pp. 9-22). San Diego, California: Elsevier Academic Press.

Fuchs, L. S. (2004). The past, present and future of curriculum-based measurement research. *School Psychology Review, 33*(2), 188-192.

Fuchs, L. S., & Fuchs, D. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*(1), 1-30.

Gardner, R., Sainato, D. M., Cooper, J. O., Heron, T. E., Heward, W. L., Eshleman, J. W., & Grossi, T. A. (Eds.). (1994). *Behavior analysis in education: Focus on measurably superior instruction*. Pacific Grove, California: Brooks/Cole Publishing Company.

Geary, D. C., & Brown, S. C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology, 27*(3), 398-406. doi: 10.1037/0012-1649.27.3.398

Gelman, R., & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition, 13*(3), 343-359. doi: 10.1016/0010-0277(83)90014-8

Georgiades, N. J., & Phillimore, L. (1975). The myth of the hero-innovator and alternative strategies for organisational change. In C. C. Kiernan & F. P. Woodford (Eds.), *Behaviour modification with the severley retarded* (pp. 313-319). North Holland: Elsevier Academic Press.

Gladwell, M. (2000). *The tipping point*. London: Abacus.

Gnaldi, M. (2006). The relationship between poor numerical abilities and subsequent difficulty in accumulating statistical knowledge. *Teaching Statistics, 28*(2), 49-53. doi: 10.1111/j.1467-9639.2006.00240.x

Graf, S., & Lindsley, O. R. (2002). *Standard Celeration Charting*. Poland, Ohio: Graf Implements.

Gravetter, F. J., & Forzano, L. B. (2009). *Research methods for the behavioral sciences*. Belmont, CA: Wadsworth Cengage Learning.

Gravetter, F. J., & Wallnau, L. B. (2004). *Study guide for Gravetter and Wallnau's statistics for the behavioral sciences* (6th ed.). Belmont, CA: Wadsworth.

Gravetter, F. J., & Wallnau, L. B. (2007). *Statistics for the behavioral sciences* (7th ed.). Belmont, CA: Wadsworth.

Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology, 16*(1), 94-143. doi: 10.1016/0010-0285(84)90005-7

Grinyer, J. (2005). *Literacy, numeracy and the labour market: Further analysis of the skills for life survey*. Nottingham: DfES Publications.

Gryiec, M., Grandy, S., & McLaughlin, T. F. (2004). The effects of the copy, cover, and compare procedure in spelling with an elementary student with fetal alcohol syndrome. *Journal of Precision Teaching and Celeration, 20*(1), 2-8.

Hargreaves, D. H. (1997). In defence of research for evidence−based teaching: A rejoinder to Martyn Hammersley. *British Educational Research Journal, 23*(4), 405-419. doi: 10.1080/0141192970230402

Haring, N. G., & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton & C. L. Hansen (Eds.), *The fourth R: Research in the classroom* (pp. 23-40). Columbus, Ohio: Charles E. Merrill.

Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hansen, C. L. (1978). *The fourth R: Research in the classroom*. Columbus, Ohio: Charles E. Merrill.

Haughton, E. C. (1972). Aims: Growing and sharing. In J. B. Jordan & L. S. Robbins (Eds.), *Let's try doing something else kind of thing*. Arlington, VA: The Council for Exceptional Children.

Haughton, E. C. (1980). Practicing practices: Learning by activity. *Journal of Precision Teaching, 1*(3), 3-20.

Hayden, J., & McLaughlin, T. F. (2004). The effects of cover, copy, and compare and flash card drill on correct rate of math facts for a middle school student with learning disabilities. *Journal of Precision Teaching and Celeration, 20*(1), 17-21.

Heward, W. L. (1994). Three low-tech strategies for increasing the frequency of active student response during group instruction. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 283-320). Pacific Grove, California: Brooks/Cole Publishing Company.

Heward, W. L. (2005). Reasons applied behaviour analysis is good for education and why those reasons have been insufficient. In W. L. Heward, T. E. Heron, N. A. Neef, S. M. Peterson, D. M. Sainato, G. Cartledge, R. Gardner III, L. D. Peterson, S. B. Hershe & J. C. Dardig (Eds.), *Focus on behavior analysis in education: achievements, challenges, and opportunities* (pp. 316-348). Upper Saddle River, New Jersey: Pearson Education.

Heward, W. L., & Dunne, D. J. (1993). For students of Behavior Analysis: A teleconference with Professor Fred S. Keller. *The Behavior Analyst 16*(2), 341-345.

Heward, W. L., Heron, T. E., Neef, N. A., Peterson, S. M., Sainato, D. M., Cartledge, G., . . . Dardig, J. C. (Eds.). (2005). *Focus on behavior analysis in education: Achievements, challenges, and opportunities*. Upper Saddle River, New Jersey: Pearson Education.

Higgins, J. P. T., & Green, S. (2011, editors). Cochrane handbook for systematic reviews of interventions, from http://www.cochrane-handbook.org.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S. L., & Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education *Exceptional Children, 71*(2), 137-148.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 21*(02), 181-193.

Huff, D. (1993). *How to lie with statistics*. London: Norton & Company.

Hughes, J. C., Beverley, M., & Whitehead, J. (2007). Using precision teaching to increase the fluency of word reading with problem readers. *European Journal of Behavior Analysis, 8*(2), 221-238.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19.

Johnson, K. R. (2008). Precision Teaching. In N. J. Salkind (Ed.), *Encyclopedia of Educational Psychology* (Vol. 2, pp. 809-813). Thousand Oaks: Sage Publications.

Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist, 47*(11), 1475-1490. doi: 10.1037/0003-066X.47.11.1475

Johnson, K. R., & Layng, T. V. J. (1994). The Morningside model of generative instruction. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman & T. A. Grossi (Eds.), *Behavior Analysis in education: Focus on measurably superior instruction* (pp. 173-197). Pacific Grove, California: Brooks/Cole Publishing Company.

Johnson, K. R., & Layng, T. V. J. (1996). On terms and procedures: Fluency. *The Behavior Analyst, 19*(2), 281-288.

Johnson, K. R., & Street, E. M. (2004). *The Morningside model of generative instruction: What it means to leave no child behind*. Concord, MA: Cambridge Centre for Behavioral Studies.

Johnston, J. M. (1970). A universal behavior graph paper. *Journal of Applied Behavior Analysis, 3*(4), 271-272.

Johnston, J. M., & Pennypacker, H. S. (1971). A behavioral approach to college teaching. *American Psychologist, 26*, 219-244. doi: 10.1037/h0031241

Johnston, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research* (3rd ed.). New York: Routledge.

Jordan, N. C., & Montani, T. O. (1997). Cognitive arithmetic and problem solving: A comparison of children with specific and general mathematics difficulties. *Journal of Learning Disabilities, 30*, 624-634. doi: 10.1177/002221949703000606

Kang, H.-W., & Golden, A. (1994). Vocabulary learning and instruction in a second or foreign language. *International Journal of Applied Linguistics, 4*(1), 57-77. doi: 10.1111/j.1473-4192.1994.tb00055.x

Kazdin, A. E. (2010). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). USA: Oxford University Press.

Kennedy, C. (2005). *Single-case designs for educational research*. New York: Pearson Education.

Kerr, K., Smyth, P., & McDowell, C. (2003). Precision Teaching with children with autism: Helping design effective programmes. *Early Child Development and Care, 173*(4), 399-410.

Kim, C., Carr, J. E., & Templeton, A. (2001). Effects of fluency building on performance over "long" durations and in the presence of a social stimulus. *Journal of Precision Teaching and Celeration, 17*(2), 7-26.

King-Sears, M. E. (1994). *Curriculum-based assessment in special education*. San Diego, CA: Singular Publishing Group.

Kozloff, M. A. (2005). Fads in general education: Fad, fraud, and folly. In J. W. Jacobson, R. M. Foxx & J. A. Mulick (Eds.), *Controversial Therapies for Developmental Disabilities* (pp. 159-173). Mahwah, NJ: Lawrence Erlbaum Associates.

Kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J., & Harmon, R. J. (2003). Measures of clinical significance. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 1524-1529.

Kubina, R. M. (2002). Fluency aims for academic behavior: Performance standards or fluency aims. Retrieved August 2, 2006, from http://www.precisionteachingresource.net/psfluencyaims.pdf

Kubina, R. M. (2005a). In response: The relations among fluency, rate building, and practice: A response to Doughty, Chase, and O'Shields (2004). *The Behavior Analyst, 28*(1), 73-76.

Kubina, R. M. (2005b). The relations among fluency, rate building, and practice. *The Behavior Analyst, 28*(1), 73-76.

Kubina, R. M., Commons, M., & Heckard, B. (2009). Using precision teaching with direct instruction in a summer school program. *Journal of Direct Instruction, 9*(1), 1-12.

Kubina, R. M., & Morrison, R. S. (2000). Fluency in education. *Behavior and Social Issues, 10*, 83-99.

Kubina, R. M., Morrison, R. S., & Lee, D. L. (2002). Benefits of adding precision teaching to behavioral interventions for students with Autism. *Behavioral Interventions, 17*, 233-246.

Kubina, R. M., & Starlin, C. M. (2003). Reading with precision. *European Journal of Behavior Analysis, 4*, 13-22.

Kubina, R. M., & Wolf, P. (2005). Potential applications of behavioral fluency for students with Autism. *Exceptionality, 13*(1), 35-44.

Kubina, R. M., & Yurich, K. K. L. (2012). *The Precision Teaching Book*. Lemont, PA: Greatness Achieved.

Kubina, R. M., & Yurich, K. K. L. (in press). Using the Standard Celeration Chart for learners with autism and other developmental disabilities. In M. Cohens (Ed.), *Bridging the gap for learners with autism spectrum disorders and learning disabilities: Precision Teaching and fluency based instruction*. New York: Springer.

Lalonde, R. N., & Gardner, R. C. (1993). Statistics as a second language? A model for predicting performance in psychology students. *Canadian Journal of Behavioural Science, 25*(1), 108-125.

Lamb, J. (2001). *Preserving the self in the south seas, 1680-1840*: University of Chicago Press.

Layng, T. V. J., Twyman, J. S., & Stikeleather, G. (2003). Headsprout Early Reading: Reliably teaching children to read. *Behavioral Technology Today, 3*, 7-20.

Lazar, A. (1990). Statistics courses in social psychology. *Journal of Teaching in Social Work, 4*(1), 17-30.

Lesaffre, E., & Pledger, G. (1999). A note on the number needed to treat. *Controlled Clinical Trials, 20*(5), 439-447.

Lindsley, O. R. (1964). Direct measurement and prosthesis of retarded behaviour. *Journal of Education, 147*, 62-81.

Lindsley, O. R. (1972). From Skinner to precision teaching: The child knows best. In J. B. Jordan & L. S. Robbins (Eds.), *Let's try doing something else kind of thing* (pp. 1-11). Arlington, VA: The Council for Exceptional Children.

Lindsley, O. R. (1991). From technical jargon to plain English for application. *Journal of Applied Behavior Analysis, 24*, 449-458.

Lindsley, O. R. (1992a). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis, 25*(1), 51-57.

Lindsley, O. R. (1992b). Why aren't effective teaching tools widely adopted? *Journal of Applied Behavior Analysis, 25*, 21-26.

Lindsley, O. R. (1995a). Precision teaching: By teachers for children. *Journal of Precision Teaching, 12*(2), 9-17.

Lindsley, O. R. (1995b). Ten products of fluency. *Journal of Precision Teaching and Celeration, 13*(1), 2-11.

Lindsley, O. R. (1997a). Performance is easy to monitor and hard to measure. In R. Kaufman, S. Thiagarajan & P. MacGillis (Eds.), *The guidebook for performance improvement: Working with individuals and organizations* (pp. 519-559). San Francisco, CA: Jossey Bass.

Lindsley, O. R. (1997b). Precise instructional design: Guidelines from Precision Teaching. In C. R. Dills & A. J. Romiszowski (Eds.), *Instructional development paradigms* (pp. 537-554). NJ: Educational Technology Publications.

Lindsley, O. R. (2002). Our Harvard pigeon, rat, dog, and human lab. *Journal of the Experimental Analysis of Behavior, 77*(3), 385-387.

Lindsley, O. R. (2010). *Skinner on measurement*. Kansas City: KA: Behavior Research Company.

Llewellyn, S. (1994). *The magic boathouse*. London: Walker Books Ltd.

Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of Computer Assisted Learning, 24*(6), 515-525. doi: 10.1111/j.1365-2729.2008.00289.x

Maloney, M. (1993). Teaching children to chart. *Journal of Precision Teaching, 10*(2), 18-19.

Maloney, M. (1998). *Teach your children well: A solution to some of North America's educational problems*. Cambridge, MA: Cambridge Center for Behavioral Studies.

Maloney, M., Brearly, L., & Preece, J. (2001). *Teach your children to read well: Level 1A, Grades K2 instructor's manual*. Belleville, Ontario: Teach Your Children Well Press.

McDade, C. (2005, September 6). *Celerating academic skills: Basic or advanced [Workshop Handouts].* Paper presented at the European Association for Behavior Analysis, Gdansk, Poland.

McDade, C., Brown, J. M., & Goggins, L. A. (1983). It's about time: The relationship between accuracy and fluency. *Journal of Precision Teaching and Celeration, 10*(2), 40-42.

McDowell, C., & Keenan, M. (2001a). Cumulative Dysfluency: Still evident in our classrooms, despite what we know. *Journal of Precision Teaching and Celeration, 17*(2), 1-6.

McDowell, C., & Keenan, M. (2001b). Developing fluency and endurance in a child diagnosed with attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis, 34*(3), 345-348.

McGreevy, P. (1983). *Teaching and learning in plain English: An introduction to Precision Teaching and Precision Tutoring*. Kansas City, MO: Plain English Publications (University of Missouri).

Mercer, C. D., Mercer, A. R., & Evans, S. (2000). The use of frequency in establishing instructional aims. *Journal of Precision Teaching and Celeration, 17*(1), 1-8.

Miller, J. E., & Calkin, A. B. (1997). Using precision teaching in a secondary science class. *Journal of Precision Teaching and Celeration, 15*(1), 44-51.

Moors, A. L. (2010). Integrating frequency-based mathematics instruction with a Multi-Level Assessment System to enhance response to Intervention frameworks. *The Behavior Analyst Today, 11*(4), 226-244.

Moran, D. J., & Malott, R. W. (Eds.). (2004). *Evidence-based educational methods*. San Diego, California: Elsevier Academic Press.

Morelle, M., Morrell, R., & Kubina, R. M. (1995). Using Precision Teaching to enhance direct instruction sight-word reading. *Journal of Precision Teaching and Celeration, 13*, 47-55.

Mosteller, F. (1981). Innovation and evaluation. *Science, 211*, 881-886.

Mueller Gathercole, V. C., Mon Thomas, E., & Hughes, E. (2008). Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism, 11*(6), 678-720. doi: 10.1080/13670050802149283

Mulhern, G., & Wylie, J. (2004). Changing levels of numeracy and other core mathematical skills among psychology undergraduates between 1992 and 2002. *British Journal of Psychology, 95*, 355-370.

Mulhern, G., & Wylie, J. (2005). Assessing numeracy and other mathematical skills in psychology students as a basis for learning statistics. Retrieved September 10, 2007, from http://www.psychology.heacademy.ac.uk

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge Univerisity Press.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. New York: Heinle-Cengage Learning.

National Centre for Response to Intervention. (2010). Essential Components of RTI - A closer look at response to intervention. Retrieved from http://www.rti4success.org/images/stories/pdfs/rtiessentialcomponents_051310.pdf

National Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (Report of the National Reading Panel)*. Bethesda, MA: Author.

Neal, D. (1981). The data−based instructional procedures of Precision Teaching. *Education Psychology, 1*(4), 289-304. doi: 10.1080/0144341810010402

NFER-Nelson. (2000). *Group Reading Test II (6-14)*. Chiswick: Author.

Organisation for Economic Co-operation and Development. (OECD, 2010). Mathematics teaching and learning strategies in PISA. Retrieved from http://www.pisa.oecd.org

Parsons, S. (2002). *Basic skills and crime: Findings from a study of adults born in 1958 and 1970.* London: Centre for Longitudinal Studies, Institution of Education: The Basic Skills Agency.

Pennypacker, H. S., Gutierrez Jr., A., & Lindsley, O. R. (2003). *Handbook of the Standard Celeration Chart: Deluxe edition*. Concord, MA: Cambridge Center for the Behavioral Sciences.

Pinson, L., & Gray, G. E. (2003). Number Needed to Treat: Underused measure of treatment effect. *Psychiatric Services, 54*(2), 145-146.

Potts, L., Eshleman, J. W., & Cooper, J. O. (1993). Ogden R. Lindsley and the historical development of PrecisionTeaching. *The Behavior Analyst, 16*, 177-189.

Raggio, S., & Bitgood, S. C. (1982). The effect of number of math drills per day on math performance. *Journal of Precision Teaching and Celeration, 16*(2), 58-60.

Raybould, E. C. (1984). Precision teaching and children with learning difficulties: Perspectives, principles and practice. In D. Fontana (Ed.), *Behaviourism and learning theory in Education* (pp. 43-74): Scottish Academic Press.

Raybould, E. C., & Solity, J. E. (1988). More questions on precision teaching. *British Journal of Special Education, 15*(2), 59-61.

Remington, B., Hastings, R. P., Kovshoff, H., degli Espinosa, F., Jahr, E., Brown, T., . . . Ward, N. (2007). Early intensive behavioral intervention: Outcomes for children with autism and their parents after two years. *American Journal on Mental Retardation, 112*(6), 418-438.

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to Intervention: Ready or not? Or, from wait-to-fail to watch-them-fail. *School Psychology Quarterly, 24*(2), 130-145. doi: 10.1037/a0016158

Rittle-Johnson, B., & Siegler, R. S. (1998). The relationship between conceptual and procedural knowledge in learning mathematics: A review. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 75-110). Hove, UK: Psychology Press.

Roberts, W., & Hampton, E. (2008). Evaluating and sustaining Precision Teaching interventions in schools. *Debate*(127), 23-32.

Roberts, W., & Norwich, B. (2010). Using Precision Teaching to enhance the word reading skills and self-concept of secondary students: a role for professional educational psychologists. *Education Psychology in Practice, 26*(3), 279-298.

Salvia, J., & Hughes, C. (1990). *Curriculum-based assessment: Testing what is taught*. New York: Macmillan.

Sante, A. D., McLaughlin, T. F., & Weber, K. P. (2001). The use and evaluation of a direct instruction flash card strategy on multiplication math facts mastery with two students with developmental difficulties and attention hyperactivity disorder. *Journal of Precision Teaching and Celeration, 17*(2), 68-75.

Shinn, M. R., & Barmonto, S. (1997). Advanced applications of curriculum-based measurement: "Big Ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced Applications of Curriculum-Based Measurement* (2nd ed., pp. 1-31). London: The Guilford Press.

Shipton, P. (1996). *I was a teenage goldfish*. Buckinghamshire: Ginn & Company.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York: Basic Books.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1984). The shame of American education. *American Psychologist, 39*(9), 947-954.

Slavin, R. E. (1989). PET and the pendulum: Faddism in education and how to stop it. *Phi Delta Capan, 70*, 752-758.

Slavin, R. E. (2008). Evidence-based reform in education: what will it take? *European Educational Research Journal, 7*(1), 124-128.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*(3), 427-515. doi: 10.3102/0034654308317473

Snider, V. E. (2006). *Myths and misconceptions about teaching: What really happens in the classroom*. Lanham, Maryland: Rowman & Littlefield Education.

Snider, V. E., & Roehl, R. (2007). Teachers' beliefs about pedagogy and related issues. *Psychology in the Schools, 44*(8), 873-886. Retrieved from http://0-www3.interscience.wiley.com.unicat.bangor.ac.uk/journal/116325647/issue doi:10.1002/pits.20272

Solis, T., Derby, K. M., & McLaughlin, T., F. (2003). The effects of precision teaching techniques and functional communication training on problem behavior for a 12-yesr old male with autism. *International Journal of Special Education, 18*(1), 49-54.

Stein, M., Silbert, J., & Carnine, D. (1997). *Designing effective mathematics instruction: A direct instruction approach* (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Stein, M., Silbert, J., & Carnine, D. (2005). *Designing effective mathematics instruction: A direct instruction approach* (4th ed.). Upper Saddle River, New Jersey: Prentice Hall.

Stuart, M. (2006). Open dialogue: Author's response to peer commentary. *The Psychology of Education Review, 30*(2), 44-50.

Sweeney, S. R., Sweeney, W. J., & Malanga, P. (2001). The effects of one-minute warm up procedures on addition one-minute fluency timings. *Journal of Precision Teaching and Celeration, 17*(2), 76-88.

The Basic Skills Agency. (2001). *The extent of the problem: Basic skills in Wales*. London: Author.

The Basic Skills Agency & The National Assembly for Wales. (2001). The national basic skills strategy for Wales Basic Skills Agency (Ed.) Retrieved from http://www.basic-skills.co.uk

Tufte, E. R. (1983). *The visual display of quantative information*. Cheshire: Graphics Press.

U. S. Department of Education. (2002). *No child left behind: A desktop reference*. U. S. Department of Education, Office of Elementary and Secondary Education. Washington, DC: Author.

U. S. Department of Education. (2004). *No child left behind: A teacher's tool kit*. Washington, DC: Author.

U. S. Department of Education. (2009). *Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools*. (NCEE 2009-4060). National Centre for Education Evaluation and Regional Assistance Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practiceguides/rti_math_pg_042109.pdf.

Vanderplank, R. (2009). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching, 43*, 1-37. doi: 10.1017/s0261444809990267

Vargas, J. S. (2009). *Behavior analysis for effective teaching*. New York: Routledge.

Vaughn, S., Fletcher, J. M., Francis, D. J., Denton, C. A., Wanzek, J., Wexler, J., . . . Romain, M., A. (2008). Response to Intervention with older students with reading difficulties. *Learning and Individual Difference, 18*(3), 338-345. doi: 10.1016/j.lindif.2008.05.001

Vieitez, D. E. (2003, May). *Precision teaching and SAFMEDS in a college course.* Paper presented at the Association for Behavior Analysis 29th Annual Convention San Francisco.

Vonnegut Jr., K. (1975). *The Sirens of Titan*. London: Coronet Books, Hodder and Stoughton.

Wang, A. Y., & Thomas, M. H. (1995). Effects of keyword on long-term retention: Help or hindrance? *Journal of Educational Psychology, 87*(3), 468-475. doi: 10.1037/0022-0663.87.3.468

Welsh Assembly Government. (2005). Words talk - numbers count: The Welsh Assembly Government's strategy to improve basic literacy and numeracy in Wales. Retrieved September 29, 2005, from http://www.learning.wales.gov.uk

West, R. P., Young, K. R., & Spooner, F. (1990). Precision teaching: An introduction. *Teaching Exceptional Children, 22*(3), 4-9.

White, O. R. (1986). Precision teaching-precision learning. *Exceptional Children, Special Issue, 52*(6), 522-534. Retrieved from http://courses.washington.edu/edspe510/510_Readings.htm

White, O. R. (2000). Lindsley and Precision Teaching Retrieved 9 January, 2006, from Athabascau University Psychology server) http://courses.washington.edu/edspe510/510_Readings.htm

White, O. R. (2002). A focus on the individual: Single-subject evaluations of response to intervention. In S. Rosenfield & V. Berninger (Eds.), *Implementing Evidence-Based Interventions in School Settings* (pp. 531-558). New York: Oxford University Press.

White, O. R., & Neeley, M. (2003). An overview of Standard Celeration Chart conventions and practices [Electronic version] Retrieved August 22, 2007, from http://courses.washington.edu/edspe510/510_Readings.htm

White, O. R., & Neely, M. (2004). The Chart Book: An overview of Standard Celeration Chart conventions and practices. Retrieved from http://courses.washington.edu/edspe510/510_Readings.htm

Willingham, D. (2009). Is it true that some people just can't do math? *American Educator, Winter 2009-2010*, 14-19.

Zahra, D., & Hedge, C. (2010). The Reliable Change Index: Why isn't it more popular in psychology? *PsyPAG Quarterly*(76), 14-20.

Zambolin, K., Fabrizio, M., & Isley, S. (2004). Teaching a child with autism to answer informal questions using precision teaching. *Journal of Precision Teaching and Celeration, 20*(1), 22-25.

TABLE OF APPENDICES

Appendix B—Example Maths Work Sheet

Add-sing-sing<10

Ysgol / School

This sheet has addition problems. Try to answer the problems as quickly as you can. Use any method to work out the problem. There is space provided for you to write.

1 + 4 = ____	1 + 5 = ____
4 + 1 = ____	2 + 4 = ____
2 + 4 = ____	3 + 4 = ____
2 + 4 = ____	5 + 4 = ____
1 + 4 = ____	5 + 2 = ____
1 + 4 = ____	5 + 5 = ____
2 + 4 = ____	5 + 5 = ____
2 + 2 = ____	3 + 5 = ____
2 + 5 = ____	3 + 5 = ____
3 + 2 = ____	5 + 5 = ____
5 + 3 = ____	3 + 4 = ____
5 + 5 = ____	5 + 5 = ____
1 + 3 = ____	3 + 3 = ____

REEPS Fluency Sheets

Appendix A—Example Word Sheet

PS-1-100: SHEET 1

Name: Skill:
Date: Aims:
TA/Teacher:
Time

Score	Corect	LO's	Notes
Try 1 Try 2 Try 3 Try 4			

am	trim	fast	thing	tin	rod	odd	7
at	cast	he	ring	hand	hot	cot	14
mat	hat	mist	ran	did	shot	get	21
me	ham	mast	she	dad	from	got	28
sat	hit	mitt	shift	cash	this	drag	35
rat	hid	it	its	is	miss	green	42
ram	the	this	seem	as	send	grin	49
mad	three	acts	fish	are	off	grim	56
sit	maths	stiff	and	gas	in	mend	63
see	fat	need	stand	tag	add	fend	70
sad	fad	than	thin	gash	often		76
rid	feed	man	can	set	song		82
cat	fist	sheet	cats	end	free		88
dim	if	shaft	him	met	rim		94
tam	that	sing	had	rest	on		100
15	30	45	60	75	90	100	

Appendix C—Examples from Local Schools Actively Using PT & FBI: Two Testimonials

Testimonial 1

The professional support and guidance I have received from Mike Beverley over the past 6 years has been invaluable. I first met Mike when he was one of my lecturers on the MSc in ABA course at Bangor University. Since graduating in 2010 I have maintained a close professional relationship with him and other colleagues at Bangor University.

I am a senior specialist speech and language therapist responsible for 134 statemented children with communication and learning difficulties in 20 different classrooms in 6 different schools. I am in close regular contact with both the children, the 80 plus members of educational staff who support them as well as their parents and other allied professionals. This means that I have a great deal of influence over curriculum development and delivery and staff training development in these schools. I carry out on average two formal training sessions a month in different schools.

Mike has enabled me and my colleagues to achieve many things, the following being the most notable:

1.) Design interventions effectively, by operationalising the steps involved using behavioural descriptions and indicators
2.) Dramatically increasing the rate at which pupils acquire and master new skills using Precision Teaching techniques.
3.) Measure pupils' progress on a day-to-day basis using the Standard Celeration chart. This allows staff to respond quickly and appropriately to a pupil's learning profile within days and sometimes hours.
4.) Utilise evidence based curricula and interventions in the class, such as *Headsprout* reading products.
5.) Bring theory and practice closer together. There are now numerous research projects taking place in the schools in which I work. I now regularly participate in training and lecturing and presenting at the university, as do several of my school-based colleagues who have come to know Mike through me. Mike is also a valuable and influential member of several educational Professional Learning Communities

The above are but a brief mention of the myriad ways in which Mike has contributed both to improving my own service delivery and that of my colleagues in the field of education. We have documented elsewhere the remarkable pupil outcomes that have come about as a result of all that he has guided us with.

The close multi-disciplinary working and relationship with the university is unique in many ways. Gwynedd LEA and the Betsi Cadwaladr University Health Board have embraced behavioural interventions in a way that no other Health Board or LEA has to date. In addition most of the work is carried out through the medium of Welsh. It is an additional sign of Mike's boundless energy and commitment to the field of evidence based educational practice that he has learnt Welsh himself and is able to converse competently and hold his own in Welsh medium formal meetings.

His influence has transformed the lives of so many pupils, many of whom he has never even met and he is held in enormous regard by all of his colleagues, particularly those who work in schools and who owe him so much.

He more than lives up to the following saying, taken from The Mabinogion.

"A fo ben, bid bont"

He who would be a leader, must be a bridge

Mike has led all of us to a place where expectations for the children we support have been raised higher than we would have believed possible. Moreover, they have been fulfilled.

Bethan Mair Williams

30/11/12

Testimonial 2

Ysgol Glancegin
Maesgeirchen
Bangor
Gwynedd
LL57 1ST
01248 353097
pennaeth@glancegin.gwynedd.sch.uk

Pennaeth: Miss B. Morris-Jones
Dirprwy Bennaeth: Mr G.W. Jones

To: Mike Beverley

Ysgol Glancegin have benefitted greatly from the PLC set up between the school and the School of Psychology at Bangor University over the past eighteen months.

Under the guidance of Mike Beverley and Dr. Carl Hughes a number of intervention projects have been running that have impacted on both pupils and staff. These have included Headsprout Early Reading, a 'Three Good Things / Happiness' intervention and most significantly, the introduction of the school to Precision Teaching techniques.

Using SAFMED's was new to staff and teachers at the school and training was arranged. The intervention was mainly going to be the responsibility of the teaching assistants and it was decided not only to include staff working to support ALN pupils but also in the Foundation Phase.

At KS2 where children with additional learning needs were targeted, SAFMED's were integrated in to the school's SEN intervention scheme with high frequency words used according to the individual's development. The staff was impressed with how quickly pupil's fluency improved and the impact that their own successes had on their confidence and self esteem.

In the Foundation Phase the effect was slightly different. Staff there concentrated on single letters with pupils and the teachers realized that the department had been teaching letter sounds in a very ineffective way and this resulted in a change throughout the department.

SAFMED's continue to be used effectively throughout the school and it is planned to use them with numeracy.

B. Morris-Jones

Appendix D—Email for Advice on SCC

Bethan Williams

Subject: Could we have some help with SCCs please?

Date: 27 June 2012 16:38:34 GMT+01:00

To: Michael Beverley

Cc:

Hi Mike

As I think you may know I am co-supervising Dawn Owen's MSc research project.

Dawn is copied into this email.

The research project is based at Ysgol and we are looking at the effect graphing PECS usage in 3 classrooms has on frequency of PECS use. It is going very well indeed and should yield extremely useful data both for researchers and educators.

The LEA are very interested in this as they invest a huge amount of money in PECS training in Gwynedd and Mon. Also, the thesis will be written in Welsh and Dawn will be giving presentations on her work to LEA staff.

Until now, PECS data has been collected on a daily SCC (ie one with no timings)

Naturally we want to transfer to SCCs where we can show the 6 hour floor (ie school day) for which we need to dust off the dreaded rate finder.

I'll be helping Dawn but we hoped you might be able to cast an eye over our graphs to ensure that we are on the right track.

We will therefore be meeting in Wheldon at 9:00 am on July 9th.

Will you be there and if so, can we drop in on you?

We'll probably be working on this until lunchtime.

Bethan

Appendix E

Email for aid with Morningside Materials 1—November, 2011

Martin Boagey

Subject: Re: Math facts

Date: 17 November 2011 23:10:08 GMT

To: Michael Beverley

Hello Michael,
Thank you for getting in touch. I didn't like to email as I am sure that you are very busy.
I am the deputy head of School, Hartlepool. I am in the process of writing my Masters dissertation and a big part of that is Precision Teaching. Obviously, when reading the literature, Morningside crops up lots of times. As an introduction, I am buying the Maths materials from Morningside Press to see if we can incorporate this approach into our practise. Are the materials pretty straightforward? The 'teachers manual' is out of print for the moment so I am hoping that they are.
It would be very helpful to get in touch with you when the books arrive if that is ok? That will almost certainly take a lot of time having experienced post from USA.

Martin

On Thu, Nov 17, 2011 at 9:57 AM, Michael Beverley wrote:
Dear Martin,
Just to confirm that if you want to contact me with any queries, I am more than happy for you to do so.
Mike
From: Kris Melroe <kris@morningsideacademy.org
Date: 2 November 2011 16:34:13 GMT
To: Martin Boagey
Cc: Michael Beverley <m.beverley@bangor.ac.uk>
Subject: Math facts

Martin,

It was nice to speak with you this AM. Please look on-line to order material. You can pay for your order with a credit card. Depending on the cost we can add extra charges for mailing at a later point. We will see what it costs. Do you want us to send this fast or regular mail?

I am enclosing my powerpoint that is used to train staff on math facts. This should help some.... it is a powerpoint and doesn't have everything you need to know on the slides.

I am also CC this to a brilliant man in Wales who has done a lot of work with PT and math facts. Michael might be able to answer your questions or and offer some support. Perhaps you could visit his site. He has done part of his doctoral work in this area. He has also attended our Summer School
Institute.

PS
Michael.... I hope it is OK with you to give your email to Martin
Kris Melroe
Work: 206-709-9500
Fax: 206-709-4611
Cell: 206-941-2111

Morningside Teachers' Academy
201 Westlake Ave. North
Seattle, WA 98109

Email for aid with Morningside Materials 2—March, 2012

Martin Boagey
Subject: Re: Training Day on Thursday 5th April
Date: 30 March 2012 20:40:06 GMT+01:00
To: Michael Beverley m.beverley@bangor.ac.uk

Hello Mike,

Thanks for keeping me in mind. Thursday is our last day at school and my boss has planned a series of meetings with parents, governors and senior leaders to redesign our evaluation procedures. I'm meant to be in all of them - a bad day to say the least! I will certainly run it past him on Monday - if he's in a good mood I might get away with it. Is it in Newcastle?

Hope that you are doing well. My research is going well; I'm just pondering what could be done with a bit of a stall in one of the children's learning. These charts certainly make you ponder.

All the best,

Martin

On Friday, March 30, 2012, Michael Beverley wrote:
Hi Martin,

This news may be a bit late, but you know the PT training day we are giving; well, I have checked and Steve Noone is happy for you to attend if you would like. It may not be particularly in your direct area. If you did want to come then I can find the best contact details for you.

I think the day is to run from about 9:30 to 3:30, but I am not completely sure.

Anyway if there is a chance you could make it then reply to this email, otherwise there is always the chance you could come to some PT training in Wales in the future.

Mike

Appendix F—Trip to view School's SCC—November, 2012

From: "Bethan Williams"
Date: 5 November 2012 18:55:26 GMT
To: "Michael Beverley"
Cc:
Subject: Your proposed trip with me to Hafod Lon to view SCCs

Hi Mike
We had a chat in school today following your very kind offer to come and see the wonderful work in Dosbarth at Ysgol and offer us valuable advice on the SCCs.

Both Donna and the classroom staff would appreciate this very much and if convenient the suggestion was an informal get together over a panad on either a Tuesday or a Thursday morning. We could do this in the class while the children are present as they are responsible enough to entertain themselves and would also love a visitor to admire their hard work.

Inevitably, Christmas duties call from now on and there are also supply teachers in this class (both of whom have done a brilliant job) until the end of term so we thought that Spring term would be great.

Could you email us all back with some dates on Tuesday and Thursday mornings, which would suit you? I will come with you and can drive you down if you like.
Catrin could you relay these to and get them to confirm what would be good for them as soon as you here from mike and copy us all in too?

Diolch
Edrych ymlaen yn fawr iawn Mike!

Bethan

Appendix G—PT Future Training Collaboration—November, 2012

Bethan Williams (BCUHB - Speech Therapy)

Subject: Training on Precision Teaching and using the Standard Celeration Chart

Date: 15 November 2012 15:57:44 GMT

To: Carl Hughes Michael Beverley

Cc:

Hello Carl and Mike

Further to my earlier email conversations with you both, I am taking this opportunity to put you in contact with two colleagues I met at the Clinical Effectiveness Symposium, namely Dr who is a consultant in Mental Health and Dr who is a clinical psychologist.

Are there opportunities for them to join the post grad. students at Bangor for this training?
They both know that I train school staff to use the SCC but obviously they would benefit from the more formal training offered within the context of the module.

I mentioned to that there are a great many students looking for research projects as he was interested in the potential for the SCC to monitor and track the enormous information and admin overload he has to bear.

I also told a little about Abigail Calkins work with PT and 'inners'.

I hope that we will be able to help each other and that this may lead to some fruitful collaborations.

Regards

Bethan

Appendix H—Invitation to Teach Precision Teaching in Glyndŵr University—November, 2012

From: Michael Beverley
Sent: 12 November 2012 19:23
To: Sally Baker
Subject: Re: 6th Form Day
Hi Sally,
No that would be fine to pass on my details. *Psychology in Action* is a good working title. Is it a first year module, and do you have any further details of what the topics covered are?
Best and thanks again.
Mike

On 12 Nov 2012, at 15:22, Sally Baker wrote:

Dear All
I am glad it was a successful day, as ever thank you for your hospitality. I have checked my diary and I am available on the 11th. Michael would you mind if I passed your details to my colleague Phill, he was really interested in your *Precision Teaching* for his Psychology in Action module? (it is a semester 2 module)
Best regards
Sally

From: Michael Beverley
Sent: 12 November 2012 15:03
To:
Subject: 6th Form Day

Dear all

Thanks so much for presenting today and making it such a success. Feedback has been very positive.

I hope that if we have similar events in the future, you might consider contributing again.

Appendix I—Email re Recent Chartshare—December, 2012

Bethan Williams

Subject: Re: Chartshare

Date: 6 December 2012 21:17:53 GMT

To: Michael Beverley

Cc: Emily Jehanne Tyler

Hi Mike

Great pleasure as always. Really pleased that Sian could come. We would have had one of the staff from but in the PT class is ill and therefore couldn't spare the other assistant,

They would be happy to present next time around though, as are we all!

Bethan

From: Michael Beverley

Sent: Thursday, December 06, 2012 8:11 PM

To: Bethan Mair Williams

Cc: Emily Jehanne Tyler

Subject: Chartshare

Thanks so much to you, Ceri & Sian,

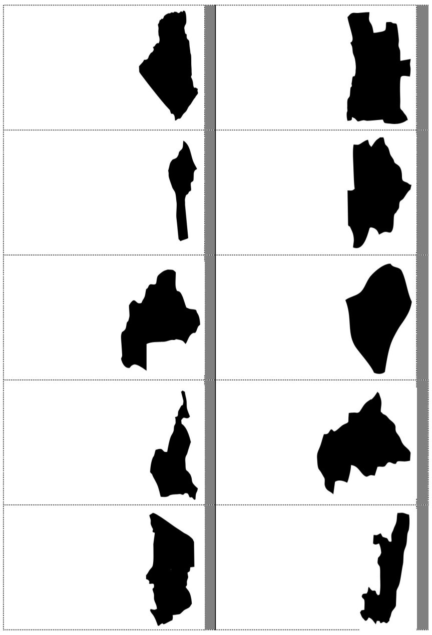
If you could pass our thanks on to them, much appreciated. Already had feedback from students as to how much they enjoyed the session.

Emily & I had a great time too too. We should do more next year.

Mike

Appendix J—SAFMEDS from the PLP3003 Module Learning Project

Carl Hughes, Mike Beverley, 2006-2012 Countrys.xlsx



Carl Hughes, Mike Beverley, 2006-2012 Countrys.xlsx

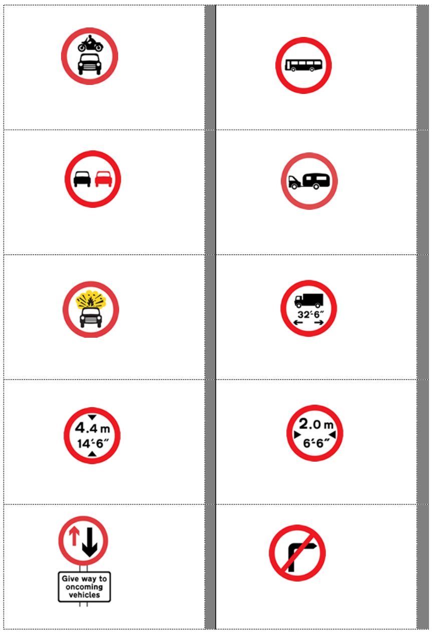
Angola	Algeria
Botswana	Benin
Burundi	Burkina
Central African Republic (CAR)	Cameroon
Congo	Chad

Carl Hughes, Mike Beverley, 2006-2012 Sign language.xls

Light	Like
Lock	Lose
Lots	Lovely
man	me
marvellous	milk

Carl Hughes, Mike Beverley, 2006-2012 Sign language.xls

Carl Hughes, Mike Beverley, 2006-2012 highway code cards1-1.xlsx



Carl Hughes, Mike Beverley, 2006-2012 highway code cards1-1.xlsx

No Buses (over 8 Passangers seats)	No Motor Vehicles
No Towed Caravans	No Overtaking
No vehicle or combination of vehicles over lenght shown	No Vehicles carrying explosives
No vehicles over width shown	No vehilces over height shown
Give priority to vehicles from opposite direction	No right turn

Carl Hughes, Mike Beverley, 2006-2012 Chemical Symbols3.xltx

H	He
Li	Be
B	C
N	O
F	Ne

Carl Hughes, Mike Beverley, 2006-2012 Chemical Symbols3.xltx

Helium	Hydrogen
Beryllium	Lithium
Carbon	Boron
Oxygen	Nitrogen
Neon	Fluorine

Appendix K—SCC from the PLP3003 Module Learning Project