



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://waikato.researchgateway.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

PRACTICE MAKES THE DIFFERENCE: THE EFFECT OF
RATE-BUILDING AND RATE-CONTROLLED
PRACTICE ON RETENTION

A thesis
submitted in fulfillment
of the requirements for the degree
of
Master of Social Sciences
at the
University of Waikato
by
SUSAN JENNIFER MCGREGOR

University of Waikato

2006

Abstract

Six home-schooled students and one adult participant each initially practiced to accuracy two decks of five previously unknown multiplication facts. The decks were yoked for practice and reinforcement. Once accurate performance was achieved, overpractice was undertaken using custom computer software that allowed either fast (free-operant) or rate-controlled responding. Rate-building practice, to an established fluency performance standard, was used with one deck while practice with the other deck was rate-controlled. The number of times a fact was practiced was the same for both methods. Response rate and accuracy was assessed after training to accuracy, at the end of overpractice and after 4- and 8-weeks of no practice. The assessment at the end of rate-building confirmed that rate-building resulted in fast and accurate responding. It also confirmed that, for the rate-controlled facts, response rates did not meet the fluency performance standard. However, the 4- and 8-week retention assessments showed no consistent differences in accuracy or response rate between the rate-controlled and rate-built decks. After 8 weeks without practice, performance on the rate-built deck was not significantly different to that prior to rate-building. These results suggest that practice to fluency does not lead to superior retention when compared to the same amount of rate-controlled practice. The results also indicate that when a skill is practiced to fluency, a period without practice leads to deterioration, to pre-rate-building levels, of accuracy and response rate. This study highlights the need for research examining the role of maintenance in the effectiveness of fluency based learning like Precision Teaching.

Acknowledgements

My sincerest thanks to my supervisors, Dr Mary Foster, Dr Cath Sumpter and Dr James McEwan. Thank you for all that led up to this thesis, the many lectures and explanations of the intricacies of behaviour analysis. Thank you for your time, you were always available to help, very prompt in reading drafts, remained patient and maintained a sense of humour throughout.

Thank you to Tony, Luke and Meaghan, my family and best friends. Thank you for always caring, always believing, always being able to make me laugh and helping me to keep it all in perspective. Thank you for the innumerable big and little things you did to help me to produce this.

Thank you to Eric Messick who showed me that behaviour analysis really works. Finally, I would like to thank all those who participated in this study and their families. Thank you for your cooperation, your sense of humour, jokes, piano recitals, delicious baking and perseverance, even when you thought you'd never make it. You were fun to work with and I couldn't have done it without you all.

Contents

	Page
Abstract	ii
Acknowledgements	iii
Contents	iv
List of Tables	v
List of Figures	vi
List of Appendices	ix
Introduction	1
Method	32
Results	47
Discussion	77
References	95
Appendices	
Appendix A	102
Appendix B	103
Appendix C	106
Appendix D	107
Appendix E	108
Appendix F	113
Appendix G	114
Appendix H	115

List of Tables

Table		Page
1	Percent inter-observer agreement for all observed sessions for each participant.	47
2	Results for the comparison of the differences between the number of both correct and error responses made in the 1-min timing at each assessment stage.	69

List of Figures

Figure		Page
1	Outline of the experimental procedure.	37
2	The process by which facts are allocated to each of Deck A and Deck B.	41
3	Number of practices of each fact during accuracy training (Phase 1) and equalisation (Phase 2) for the five Deck A and five Deck B facts learned for each participant.	49
4	Number of Deck A and Deck B facts answered correctly during card presentations at each assessment session for each participant.	51
5	Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P1 and P2 respectively.	53
6	Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P3.	54
7	Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P4.	55
8	Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P5.	56

9	Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P6.	57
10	Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P7.	58
11	Response rate on 2 times table facts during a single 1-min timing after each training component and at retention assessments.	61
12	Response rate on Deck A facts (trained to accuracy plus fluency) during a single 1-min timing after each training component and at retention assessments.	62
13	Response rate on Deck B facts (trained to an accuracy criterion only) during a single 1-min timing after each training component and at retention assessments.	62
14	Number of correct and error responses made for each Deck A and Deck B fact answered during both un-timed and timed 4- and 8-week retention assessments for P1, P2, P3, and P4.	65
15	Number of correct and error responses made for each Deck A and Deck B fact answered during both un-timed and timed 4- and 8-week retention assessments for P5, P6 and P7.	66
16	The mean of all participants' percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages.	

17	The mean of all participants' percent frequencies of response latencies for the deck of 2 times table facts over the four assessment stages.	72
18	The 10 th percentile response latencies for each session taken from all trials in that session, during rate-controlled software training on Deck B (rate-controlled) facts.	75
H1	The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P1.	116
H2	The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P2.	117
H3	The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P3.	118
H4	The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P4.	119
H5	The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P5.	120
H6	The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P6.	121
H7	The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P7.	122

Appendices

Appendix		Page
A	Example of the notice seeking participants posted on the local home-schooling website.	102
B	Information sheet for parents of potential participants.	103
C	Example of a flashcard used in Phase 1 (accuracy training).	106
D	Example of Phase 1 (accuracy training) data collection form.	107
E	Description and specifications of software.	108
F	Example of Phase 2 (software training) data collection form.	113
G	Example of Phase 3 (assessment) data collection form.	114
H	Histograms for each participant, showing the percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages.	115

Success in education is the foundation for many positive outcomes in life.

Unfortunately, for many students, traditional education systems do not produce effective learning outcomes. Precision Teaching (PT) is a behavioural teaching methodology that has shown impressive results, including with students who have been labeled as learning disabled in mainstream education and who have fallen behind two or more grade levels.

The effectiveness of Precision Teaching is demonstrated by the ability of some schools to ensure learning gains consistently, to the point where they offer a money-back guarantee to parents that they will raise students' grade levels by two years each school year (Morningside Academy; Johnson & Layng, 1992). These gains have been shown to maintain even when students return to traditional school settings (Johnson & Layng, 1994, p.177). While the effectiveness of Precision Teaching is clear from its outcomes, there has been insufficient scientific study of the methodology to develop a clear understanding of both why the method is effective and what each of the components of this methodology contribute to its effectiveness.

At the heart of Precision Teaching is the building of response rates on basic skills to fluency (fluency building). It is claimed that learning skills to fluency has positive learning outcomes (Binder, 1996, 2003; Binder, Haughton & Bateman, 2002; Eshleman, 2001; Haughton, 1984; Johnson & Layng, 1992, 1994, 1996; Kubina & Morrison, 2000; Kubina & Starlin, 2003; Lindsley, 1992; White, 1984). This thesis investigates the relationship between response rate, practice and that most fundamental of learning outcomes, retention.

Precision Teaching (PT)

Precision Teaching (PT), pioneered by Ogden Lindsley in the 1970s, is a teaching model and measurement paradigm that incorporates effective practice (Bullara, Kimball & Cooper, 1993). It has been used to produce effective learning outcomes with many teaching strategies (Kubina, Morrison & Lee, 2002; Lindsley, 1992) with both adults and children and with both able and differently abled populations (Beck & Clement, 1991; Binder, 1996; Binder, Haughton & Van Eyk, 1995; Haughton, 1997; Johnson & Layng, 1994; Olander, Collins, McArthur, Watts & McDade, 1986). PT came about through dissatisfaction with traditional accuracy-only measures of learning and a desire to use research methods derived from the experimental analysis of behaviour in the education sector (Lindsley, 1992).

Precision Teaching is not a single teaching program; it is a method of assessing the success of learning precisely and of adapting teaching to the specific needs of individual students. Teachers make decisions to continue or change teaching strategies and to progress learners to a new learning element based on the behaviour of the learner (Beck & Clement, 1991). PT uses response rate as the primary measure of progress towards skill competency. The response rates of component skills are built to fluency before complex composite skills are introduced. Response rate is measured and built during frequent, daily timed trials, typically of 1-min duration. Data are collected and plotted on Standard Celeration Charts (SCC). Celeration charts provide a pictorial representation of learning that allows analysis of the relationship between changes in performance (both response accuracy and rate) and changes in the learning environment (Lindsley, 1992). They also provide a means of identifying effective and ineffective teaching procedures and a protocol for facilitating informed instructional decisions.

Response Rate as a Measure of Behaviour

At the heart of PT is the building of free-operant response rates to fluency and the use of response rate as a measure of behaviour. In basic laboratory studies with animals, Skinner discovered that the rate of free-operant responding (frequency) is a sensitive measure of behaviour (Skinner, 1953, pp. 64-68). Free-operant response rates are achieved when the subject's response rate is not limited in any way and the subject is free to respond at any time. Response rate, or frequency, is defined as the number of responses per unit time (Lindsley, 1992). Frequency has been used widely as a measure of behaviour in laboratory settings (Ferster & Skinner, 1957). However, when basic laboratory findings have been applied to educational settings, behaviour analysts have tended to use accuracy measures expressed as percent correct (Binder, 1996; Lindsley, 1992). Accuracy presents just one dimension of competent performance, telling only whether an individual *can* perform a skill, but not *how well*. The weakness of accuracy-only measures is their inability to distinguish between expert, and accurate but non-expert performance of a skill (Binder, 2003). A complete description of behavior must include a temporal dimension, since behavior occurs in time. For example, consider two children completing addition sums. Both children may complete 'plus one' addition sums with 100% accuracy. However, one child completes sums at a rate of 2 per min and another at a rate of 40 per min. It is clear that one child's performance outweighs the other's.

Although the time dimension is frequently omitted from classroom-based learning, it is commonly included in measures of performance in other contexts. Examples include athletics, music and standardised IQ testing (Chiesa & Robertson, 2000).

Fluency

Building response rates to fluency is a core component of Precision Teaching. In common usage, the term fluency describes behaviour that is accurate, flowing and seemingly effortless (Encarta dictionary, 2005). In behavioural education literature, such as Precision Teaching, fluent behaviour is typically described topographically as “the fluid combination of accuracy plus speed that characterizes competent performance” or “true mastery” (Binder, 1996, p.164). Fluency is defined functionally by Binder (1996) as a ‘metaphor’ describing the observed relationship between response rate and several critical learning outcomes. That is, fluency is defined by its outcomes (Hartnedy, Mozzoni & Fahoum, 2005; Lindsley, 1996). The outcomes most commonly included are retention, endurance and application (Binder, 1996; Binder, 2003; Binder et al., 2002; Eshleman, 2001; Haughton, 1984; Johnson & Layng, 1992, 1994, 1996; Kubina & Morrison, 2000; Kubina & Starlin, 2003; Lindsley, 1992; White, 1984). Some authors suggest that building response rates to fluency also produces the outcomes of stability (stable performance during distraction) and adduction (Johnson & Layng, 1996). Adduction refers to a special case of application where a composite repertoire is established without any direct instruction (Binder, 2004). However, these outcomes are less well accepted and are not the focus of this study.

Retention refers to the accurate performance of a learned behaviour long after training has ended. Endurance refers to the ability to perform a behaviour at a competent rate for an extended period of time without a significant decrease in response rate or increase in error rate. The definition of endurance for a particular skill should be related to real-world requirements (Binder, 1996; Binder et al., 1995; Kubina & Morrison, 2000). For example, we may say that a typist exhibits endurance when they maintain a competent typing rate for most of their working day. Application refers to

the ease with which fluent component behaviours are combined and applied in the performance of a composite skill (Binder, 2004; Johnson & Layng, 1992; Kubina, 2005).

The relationship between the critical learning outcomes and response rate is captured in Haughton's acronym REAPS which stands for: retention, endurance and application performance standards. Performance standards, or fluency aims, specify the particular skill, the modality by which the stimuli are sensed and by which the response is made (learning channel) and the accuracy (quality) and rate (quantity) of the specific response for fluency. Binder et al. (2002) reported "widely accepted" fluency performance standards for twenty-one basic skills (p. 8). The standard given for 'see/say' 'answers to basic multiplication facts' is '70-100/min' with 'near-zero errors'.

A range of correct responses per min is used to denote fluency. This range covers response rates "within which most learners seem to" show the outcomes of retention, endurance and application (Binder, et al., 2002, p. 8). Precision teaching has been determining and refining fluency performance standards since the 1960s. Precision teaching researchers have determined performance standards for some skills. These performance standards are said to be consistently associated with the following critical learning outcomes: retention, endurance, application (Haughton, 1997; Johnson & Layng, 1994, 1996; Kubina et al., 2002; Kubina & Starlin, 2003; Lindsley, 1996; Luyben, Hipworth & Pappas, 2003). However, Kubina (2005) notes that meeting a fluency performance standard is not the *cause* of the critical learning outcomes. Rather, attaining fluency provides an indicator of skilled fluent performance which is characterised by retention, endurance and application.

Fluency researchers do not use consistent definitions of the critical learning outcomes. Definitions of retention are discussed in more detail later. Binder (2004, p. 284) extends the definition of endurance by adding the criterion that behaviour remain stable, even in the face of distractions that “can potentially compete for stimulus control.” Doughty, Chase and O’Shields (2004) suggest that endurance refers to the persistence of high response rates in the face of distraction (e.g., a noisy classroom). However, Binder (2004) cautions that the use of distracters, such as ambient noise, that do not compete for stimulus control do not provide an appropriate assessment of endurance.

Doughty et al. (2004) and Weber and Cowardin (1994) incorrectly define application as being synonymous with generalisation, performance of a previously learned skill in an untrained setting, or with another instructor or with non-training stimuli (Binder, 2005; Kubina, 2005). Johnson and Layng (2002) note that when achieving performance standards does not result in application, then application may be taught by creating new environmental contexts for the skill while maintaining fluent response rates (generalisation). Kubina et al. (2002) note that while some authors suggest that developing fluent performance in children with autism enhances the application of learned behaviours to novel situations (generalisation), this relationship has yet to be proven. Both Binder (2004) and Kubina (2005) caution that using incorrect and inconsistent definitions leads to invalid research and may contribute to a lack of understanding of the results.

Although most researchers suggest that retention, endurance and application appear when fluency performance standards are met, Johnson and Layng (1992; 1994) and Kubina et al. (2002) state that endurance and application do not necessarily occur unaided as a result of achieving fluency aims. For example, learners may need to build

endurance by practicing the skill at the fluency rate for increasingly longer periods of time. Precision Teaching is based on the assumption that complex composite performance is simply the product of a number of simpler component elements. These simpler elements are learned to frequencies specified by performance standards. However, Johnson and Layng (1994) note that application may not occur unaided when these performance standards are met. Games and activities may need to be arranged to promote adduction of these simple elements into more complex composite skills (Johnson & Layng, 1994).

Fluent responding is by definition fast, flowing and continuous. Johnson and Layng (1996) state that reference to free-operant response situations must be included in a definition of fluency. Binder (1996) and Johnson and Layng (1996) state that free-operant and not discrete-trial responding is an essential feature of fluent performance. Johnson and Layng (1996) distinguish between discrete trial and free-operant performance. In discrete trial responding, the presentation of the discriminative stimulus is controlled by the experimenter. Typically, a stimulus is presented, a response is made and sometimes the response is reinforced. The experimenter then presents the next discriminative stimulus after an interval of time. Free-operant performance, on the other hand, requires that subjects' response rates are not limited in any way; the subject is free to respond at any time (Binder, 1996; Johnson & Layng, 1996; Wolery, Bailey & Sugai, 1988). Typically in laboratory settings, the discriminative stimulus remains the same, e.g., a key to press. In applied educational settings, typical of PT, the discriminative stimulus is usually varied. Examples include the changing words in a text to be read or different multiplication facts presented on a computer screen. In real-world settings, fluent performances are typically characterised by fast and accurate responding to constantly changing discriminative stimuli.

In addition to fluency's relationship to retention, endurance and application, fluent behaviours are advantageous for several reasons. Firstly, fluent behaviours can compete more effectively with previously learned behaviours that serve the same function and achieve the same reinforcers (Wolery et al., 1988). For example, both crawling and walking serve the same function. They allow an individual to move from one place to another. When infants first learn to walk they typically revert to crawling when they wish to get somewhere fast. Crawling is a fluent behaviour, but walking is not yet. When children become fluent at walking they will stop crawling. Fluent walking can effectively compete with crawling. Secondly, when individuals perform behaviours fluently they have more opportunities for positive social reinforcement. For example, when a child reads fluently they will be able to read stories to other children, be able to help others and pass on information from text to others. The child is also less likely to experience the negative social consequences of being unable to read. Further, some behaviours must be performed both accurately and fluently to be useful. For example, a child may be taught the actions involved in swimming accurately but these actions must be performed with sufficient frequency for the child to remain afloat and move through the water.

Fluency Training (Rate-building)

Fluency training is a means of building response rates to those of competent performers (Chiesa & Robertson, 2000). In precision teaching, a two stage model of learning a new skill is often proposed: the acquisition stage and the practice stage (Binder, 2003; Miller & Heward, 1992). During the acquisition stage, students learn how to perform a skill accurately with the focus on quality of performance. The ensuing practice stage focuses on the quantitative aspects of performance. During this stage,

students repeatedly practice skills during frequent timed trials while attempting to maintain accuracy. Response rates during timed trials are built to a fluency performance standard. This process is typically called fluency training (Hartnedy et al., 2005).

Johnson and Layng (1994) state that some learners can build fluency while still acquiring skills; others need to complete accuracy training first. Individual performance data indicates to teachers how learners should progress through this sequence. Haughton (1997) suggests that for some skills it is more beneficial to build rate first and improve accuracy later, e.g., handwriting.

Setting Appropriate Fluency Aims (Performance Standards)

Doughty et al. (2004) and Haughton (1997) suggest that fluency building produces *high* rates of responding (emphasis mine). Binder cautions the use of the term “high response rates” and suggests the use of the term “competent” or “normal ranges of response rate” (Binder, 2004, p. 282). In real life situations, behaviour has a natural speed at which it is most functional to an individual (Ivarie, 1986; Kerr, Smyth & McDowell, 2003). It is this functional response rate that should be represented by performance standards. However, Haughton (1997) states that, to achieve fluency outcomes, some practice must occur at frequencies higher than those measured in competent performance.

Much Precision Teaching/fluency research is aimed at establishing specific educational performance standards (Haughton, 1997). While several methods have been proposed, all agree that standards should produce skills that occur at the rate, and for the duration required, in real life situations (i.e., are functional). They should also ensure retention, endurance and application (Binder et al., 1995).

While Binder (1996), Doughty et al. (2004) and Haughton (1997) suggest that aims be empirically based, performance standards are typically based on inductive reasoning (Lindsley, 1992). Celeration charts from large samples of students' performance on a particular skill have been used to set performance standards. Consistencies in the range of performance frequencies whose attainment has been followed by retention, endurance and application can be identified in these charts (Binder et al., 2002; Haughton, 1980; Peladeau, Forget, & Gagne, 2003). It should be noted that this evidence is from uncontrolled situations. Results attained from celeration charts lead to hypotheses that can be validated via empirical studies. These empirical studies typically involve building the response rates of participants to those around the hypothesised performance standard range and then evaluating learning outcomes. For example, Evans, Merger and Evans (1983) and Evans and Evans (1985) investigated the relationship between attaining several different frequency criteria on a component skill (saying short 'a' and consonant sounds) and subsequent changes in performance during acquisition of the next related skill (saying consonant-vowel-consonant, CVC, trigrams with the letter 'a' as the central vowel). After performance frequencies were reached during daily timed trials, participants completed a series of post-test 1-min timings saying CVC trigrams. Gains in response rate from both pre-test to post-test timings and over post-test timings were compared to give an indication of the ease with which the more complex skill was acquired. Their results suggested that, for this task, component skill frequencies of 80-90 responses per minute optimised acquisition.

Researchers should be aware that the performance standards necessary to ensure each of the fluency outcomes (retention, endurance and application) may be different. For example, lower rates may ensure retention of math facts, but higher rates may be necessary to ensure application (Binder, 1996).

Wolery et al. (1988) surveyed published fluency aims. They found substantial variation in fluency aims for single component skills. For example, published aims for see/say multiplication facts up to the 9 times table, range from 40 correct answers with less than 3 errors to 90 correct answers with no errors. This lack of consistency represents a serious limitation in fluency research (Kubina, 2005). Until all researchers use the same fluency aims, it is difficult to compare results. Binder et al. (2002) present a set of fluency aims for a range of basic skills that they say are widely accepted among the precision teaching community. Consistent use of such a set of performance standards would enable researchers to compare results effectively.

The Importance of Practice

There is much literature suggesting that practice is one of the most influential variables in both the acquisition and retention of skills (Binder, 1996; Doughty et al., 2004; Haughton, 1997; Johnson & Layng, 1996). Few teachers would disagree that newly acquired skills should be practiced. Indeed, it is common for teachers to provide practice sheets to be completed at school and as homework. Further, homilies such as ‘practice makes perfect’ have been repeated by teachers and parents for generations. However, while educators agree that practice is important, there is little agreement on what form this practice should take and how long it should continue (Miller & Heward, 1992). Researchers have found that the number of practice trials required to reach 100% accuracy differs among individuals (Haughton, 1997; Ivarie, 1986).

Practice involves repeating a response over time with the aim of improving that response (Kubina, 2005). Binder (1996, p. 179) defines practice as “the repetition of a given response class after it has been accurately established in a repertoire.” Fischer et al.’s study (as cited in Peladeau et al., 2003) indicates that the amount of time spent on

academic tasks is highly correlated with academic achievement. But simply increasing time spent on academic tasks does not necessarily predict academic improvement. Some academic activities contribute more than others to the acquisition of skills and related academic success (Greenwood, Delquadri & Hall, 1984). Greenwood et al. (1984) found that the number of opportunities to respond, e.g., question asked (antecedent) followed by response (behaviour) followed by feedback on response (correct/incorrect), was a better predictor of success than time spent on academic tasks alone. The sequence of question, response and feedback is a type of controlled practice.

Educationalists often describe traditional practice (rote learning) as outmoded and unproductive (Binder, 1996). Rote learning often conjures a picture of long periods of time spent practicing the same thing. Binder et al. (1995) and Binder (1996) suggest that practice for long periods of time may be punishing and is likely to be associated with decreased performance and task avoidance behaviour, particularly when response rate is well below fluency performance standards. Practice within the PT framework typically involves responding as many times as possible for short 1-min (or less) timings.

Educationalists also suggest that rote learning is ineffective in aiding the acquisition of the more complex composite skills needed in our environments today (Binder, 1996). However, fluency research has suggested that this assertion is incorrect and that skills practiced to fluency will in fact increase the ease with which more complex composite skills are acquired (Binder, 1996). Binder suggests that this line of thought may have come about because practice, where performance is measured only with accuracy criteria, receives little reinforcement. The measurement of frequency that is part of fluency training allows improvements in performance during practice to be identified by teachers and thereby provides an opportunity for reinforcement.

Fluency building is one means of providing a large number of practice opportunities. It is commonly understood that fluency is achieved through repeated practice (Binder, 1996; Wolery et al., 1988). As an example of the primary role of practice in PT consider the Morningside Academy. At Morningside, 70% of classroom time is spent fluency building (practicing) and approximately 10% each on charting, testing and acquiring/establishing skills. This is in contrast to a typical school classroom, where 70% of the time is spent establishing skills, 10% testing and only 20% practicing (Johnson & Layng, 1994, p. 191).

The Role of Deliberate Practice

Research into the area of *expert performance* (for example Olympic athletes or world class musicians) has found a relationship between a specific type of practice, termed deliberate practice, and the critical learning outcomes proposed to occur with fluency. Ericsson, Krampe and Tesch-Romer (1993) conducted a retrospective analysis of the progress and practice histories of expert versus competent but non-expert musicians to identify the factors contributing to their performance. They found a number of consistencies. Firstly, pre-existing individual differences play a weak role in the development of expert performance. Secondly, experts perform both more accurately and *faster* than non-experts. Thirdly, repetition per se, does not lead to maximal performance. Finally, and of particular note to this thesis, they found that the level of performance attained was directly correlated to the amount of what they term, extended *deliberate practice*. Deliberate practice is characterised by highly structured practice activities that are explicitly directed towards the goal of improving performance (Ericsson et al., 1993). This deliberate practice requires: a brief period of explicit instruction on a skill where prerequisites have already been met; provision of many

response (practice) opportunities; the receipt of immediate and informative feedback on performance; remediation training for errors; and careful monitoring of performance. Further, they state that instructors need to sequence training tasks from simpler component skills to more complex composite tasks. Performers should move on to more complex tasks only when prerequisites are mastered.

Ericsson and Charness (1994) reviewed studies examining the histories of expert performers in a range of domains and also found that experts' training was characterised by high levels of deliberate practice. In summary, although they are not in the PT research area, Ericsson et al. (1993) and Ericsson and Charness (1994) support the claims of PT. The deliberate practice that they find is necessary to achieve expert performance is consistent with the nature of practice in Precision Teaching. Secondly, they suggest that extensive deliberate practice leads to expert performance which is characterised in part by endurance and application and also retention which may persist even when the skill is not used for several years.

A major point of difference between Precision Teaching and deliberate practice is that, while deliberate practice aims to produce performance that is fluent (flowing and accurate), it does not include explicit rate-building to a fluency performance standard. With the commonalities between the outcomes of deliberate practice and Precision Teaching, it is reasonable to ask the question whether it is the amount of practice alone that is responsible for the positive results of fluency.

Overlearning and Automaticity

There is further corroborating evidence in other non-behavioural research areas for the effects of practice on the critical learning outcomes proposed to occur with fluency (Dougherty & Johnston, 1996; Ivarie, 1986). Dougherty and Johnston (1996) and Peladeau et al. (2003) suggest that three conceptualisations of performance used in

different research areas: overlearning, fluency and automaticity, are closely related and may describe the same phenomenon.

The term overlearning is used to refer to both a procedure for practice and a 'degree' of learning (Dougherty & Johnston, 1996). As a procedure, overlearning involves practicing a skill a specific number of times beyond the point where 100% accuracy is achieved. As a degree of learning, overlearning refers to the improved retention (assessed after a non-practice period) and generalisation of skills that has been found to occur with extra practice (Driskell, Willis & Cooper, 1992).

Overlearning literature uses accuracy-only measures to assess performance. The limitation of accuracy-only performance measures makes it impossible to measure the effects of consecutive overlearning trials (Dougherty & Johnston, 1996). Instead, researchers have found indicators of improved performance with practice beyond 100% in outcomes such as enhanced retention and generalisation of skills.

Binder (1996) reviewed overlearning research and found that, in some studies where instrumentation was available, response latencies were measured in progressive overlearning trials. Where measured, these latencies were found to consistently decrease with overpractice. Osgood (1946) found that lower latencies predicted superior retention (as cited in Binder, 1996).

Fluency training also involves practicing a skill beyond 100% accuracy. Binder (1996) distinguishes between overlearning and fluency by suggesting that overlearning describes the *process* of extra practice beyond 100% accuracy and fluency is the *measurable goal* of this extended practice. The main difference between fluency training and overpractice is that practice beyond 100% accuracy in fluency training continues until a specific rate and accuracy criterion is reached.

Automaticity is an outcome that may be observed when a skill is practiced beyond 100% correct and is defined as “expert, easeful and efficient” performance of a task without the need for conscious attention to the task (Dougherty & Johnston, 1996, p. 290). For example, driving a car may be seen as automatic when a driver can hold a conversation with a passenger while driving, without any decrement in driving performance. Further, conscious attention is said to have a detrimental effect on task performance (Dougherty & Johnston, 1996).

There are similarities between fluency training and training to automaticity (Dougherty & Johnston, 1996). Automaticity is measured, like fluency, by the rate of correct responses or by response latency. Dougherty and Johnston (1996) reviewed automaticity research and found that automaticity researchers suggest that when behaviour is practiced until automaticity occurs, short and long-term retention, endurance, ‘resistance to distraction’ and application will be improved. Retention, endurance, and application are also claimed as outcomes of fluency.

Binder, Haughton and Bateman (2002) and Dougherty and Johnston (1996) suggest that overlearning, fluency and automaticity are, at the least, very closely related and may refer to the same ‘behavioural phenomenon’. Overlearning, automaticity and fluency all require practice beyond 100% accuracy and all are associated with improved retention. It may be that the only difference between them is the means of measuring training effectiveness (percent correct versus response rate/latency) and differences in outcome may reflect the differences in the amount of practice beyond 100% correct.

All three methods deploy practice to achieve learning outcomes, including retention. Automaticity and fluency training both involve practice until a performance aim is met. However, unlike overlearning and automaticity, fluency training has the additional requirement that practice should involve rate-building in a free-operant

environment. Precision teachers suggest that since overlearning and automaticity research do not measure response rate, they are attributing enhanced outcomes, similar to those claimed by precision teachers, to extra practice beyond accuracy alone. However, when these outcomes are seen, if response rate had been measured at the end of practice sessions, it may well have been fluent.

Retention

The most fundamental of the learning outcomes claimed for fluency is retention. To assess retention an understanding of how it is defined within fluency training literature is needed. In common usage, retention is equated to remembering. That is, producing an equally accurate response after a non-practice period. Retention is an essential dimension of useful learning. Retention is not a *thing* that you either have or don't have; it is a comparison of accuracy and/or response rate over time.

Haughton (1997) explains retention of accuracy. He states that "high degrees of accuracy" are retained over non-practice periods when skills have been practiced to "relatively" low response frequencies (Haughton, 1997, p. 77). Haughton refers to his 1976 study which found that the accuracy of maths facts learned at school remained high over the 2-month summer vacation regardless of the performance frequency attained at the end of the school year. That is, students who had reached higher performance frequencies were not more accurate on retention tests than their classmates who had reached lower frequencies. Performance measured in terms of accuracy alone was indistinguishable. Kruger (1929) studied the relationship between amount of practice and subsequent retention levels and found that the number of practice trials required to reach accuracy could not be used to predict the number of extra practice trials required to attain a given level of retention (cited in Binder, 1996).

Eshleman (2001) and Doughty et al. (2004) define retention in terms of response rate. Retention is the stability of (Eshleman, 2001) or the persistence of (Doughty et al., 2004), the rate of correct and error responses attained when a performance standard is met, over a subsequent non-practice period. Binder (1996, p. 164) defines retention as: “the relation between behavior frequencies at two points in time, between which the individual has no opportunity to emit the behavior.” These definitions of retention are based on rate rather than accuracy-alone, which is consistent with PT philosophy. Intuitively, one would expect the frequency of correct responses to decrease over time, and the frequency of incorrect responses to increase. Retention can be quantified as the degree to which these frequencies change over periods of non-practice (Binder, 1996; Doughty et al., 2004; Eshleman, 2001) or the rate at which these changes occur (Eshleman, 2001). For example, the fluency aim for see/say answers to basic multiplication facts is 70-100 correct responses per minute with less than two errors (Binder et al., 2002). If the rate and accuracy attained after practice was also attained after a 2-week non-practice period, this would be an example of perfect retention over a 2-week time period.

Other authors avoid a precise definition of retention, explaining retention as remembering or not forgetting (Johnson & Layng, 1994, 1996; Kubina & Morrison, 2000). Johnson and Layng (1994) suggest that retention refers to behaviour “not forgotten” or “remembered” after *significant* periods of non-practice. Johnson and Layng (1994, p. 285) suggest a *significant* retention period to be “a month or more.” They also suggest that having students reach the fluency performance standard for a skill *predicts* that, after one month with no practice, students would answer facts *at* the performance standard *without* any errors. There is however, no consistent definition of what constitutes an appropriate retention period. Research into retention after

rate-building has used many different retention periods. These include 10 days (Shirley & Pennypacker, 1994); 3 weeks (Ormrod & Spivey, 1990); 1, 2, and 3 months (Ivarie, 1986); 4 months (Young, West & Crawford, 1985; Bucklin, Dickenson & Brethower, 2000) and 8 months (Olander et al., 1986). Further, while most authors assess retention during timed tests in terms of the rate of correct responses, some do not (Peladeau et al., 2003). None of these authors document the reasons for their chosen retention period. One would assume that if learning is to be useful, the retention period would be one which would allow individuals to meet real-world requirements which are specific to a particular skill.

Maintenance is a concept related to retention. Binder defines maintenance, and distinguishes it from retention: “Maintenance, on the other hand, refers to the relation between a behavior’s frequency at two points in time, between which the individual *has an opportunity* to emit the behavior to produce reinforcement in the natural environment” (Binder, 1996, p. 164 emphasis mine).

Kelly (1996) investigated the relationship in mastery learning between fluency and the maintenance of sight reading words over a 6½ month post-training period. In a series of experiments, young children in a special education school class learned to read groups of words to either an accuracy-only or an accuracy plus fluency criterion. After training, the words were maintained through maintenance probes (typically weekly), and presumably, normal classroom use. This is clearly a maintenance situation as Binder (1996) defines maintenance. In her third experiment, Kelly controlled for the amount of instructional time the 2 participants received on words learned with and without the fluency criterion. Kelly found that sight reading behaviour was maintained more accurately when words were learned to a fluency criterion. Words learned to fluency were maintained at 90-100% accuracy across time while those learned without

fluency ranged from 0-60% accuracy. She suggests that achieving accuracy plus fluency may lead to words being better maintained than achieving accuracy alone.

In Kelly's study, participants' responses to sight words were always preceded by teacher-presented instructions during the learning phases. This is not a free-operant arrangement. Participant response rate was limited by the rate at which the teachers presented instructions. Binder (personal communication, April 30, 2005) notes that fluency is only attained when "frequency is not constrained by anything but the organism's ability to respond." Further, Binder (personal communication, April 30, 2005) notes that "When procedures prevent us from measuring above a certain frequency of response, we really can't use that measure to predict outcomes."

There is some confusion in the research literature between retention and maintenance. Some use maintenance and retention synonymously while others use one term or the other to describe one of the outcomes of fluency (Hartnedy et al., 2005; Luyben et al., 2003; Miller & Heward, 1992; Weber & Cowardin, 1994; Weiss, 2001). Binder's (1996) definition of maintenance clearly differentiates it from retention. Further, emphasising that there is a difference between retention and maintenance, Binder warns that the frequencies required to produce retention and maintenance may not necessarily be the same and require empirical study. Sometimes confusion between retention and maintenance has caused misunderstanding of prior research. For example, Binder misinterprets Kelly's (1996) doctoral dissertation as support for higher response rates producing retention. In referring to Kelly (1996), Binder states that "Kelly (1995) used a within-subjects yoked design to separate the effect of mere repetition from that of achieving more rapid responding, and supported the conclusion that achieving more rapid performance yields greater retention" (Binder, 1996, p. 175). However, while

Kelly studied rate-building, she assessed maintenance, as per Binder's definition, not retention.

Evidence Supporting Retention as an Outcome of fluency

While PT has been shown to be effective at enhancing learning, many have noted the need for more empirical study of the presumed functional relationship between reaching fluency and the learning outcomes of retention, endurance and application (Binder, 1996; Doughty et al., 2004; Peladeau et al., 2003; Shirley & Pennypacker, 1994). Despite the obvious importance of retention, only a small percentage of applied research has empirically investigated the relationship between achieving fluent responding and retention. Of those studies that do exist, few have controlled for practice. The studies that are often cited as providing support for fluency producing enhanced retention fall into five main categories. These studies are described in these five categories next.

Training to different response rates. In some studies, participants were trained to several different response rates and retention was assessed after a non-practice period (Ivarie, 1986; White, 1984). Subjects received more practice on the skills learned to the higher response rate. The response rates attained on retention tests were compared with response rates at the completion of training. Typically, results indicate superior retention for the skill trained to the higher response rate. While these studies provide support for the proposal that amount of practice is positively correlated with enhanced retention, they do not compare the differential effects of two types of practice, overpractice that involves rate-building to fluency and unpaced overpractice. For example, White (1984), in a single case study, found superior retention of response rate when the participant learned a see/say word task to a higher performance standard.

Ivarie (1986), in a group study, investigated whether retention was optimised by learning a skill to one of two response rates. Assessing retention at 1, 2 and 3-months, she found superior performance, in terms of response rate, at retention assessments for those who learned to the higher rate.

While it is clear from these studies that training to higher response rates resulted in improved retention, it is not clear whether this is purely an effect of attaining high response rates. Some other component of training, particularly the extra practice required to achieve the high response rates may have contributed to improved retention.

Training to a fluency performance standard. In some studies, rather than training to various response rates, skills were practiced to specific fluency performance standards before retention was assessed. For example, Bullara et al. (1993), in a single case study, found 'acceptable' retention (a drop in corrects from 74 to 56 per minute while maintaining 100% accuracy) after 3 months when a skill was learned to a fluency performance standard. In Young et al.'s (1985) study, 2 participants learned two tasks to fluency performance standards. A 4-month retention assessment revealed mixed results, showing maintenance of response rate for one skill and a decrease (the paper does not report by how much) in response rate for the other skill. These studies show (weak) support for the claim that attaining a fluency performance standard delivers adequate retention. However, like the previously discussed group of studies, they do not allow clarification as to whether retention is an effect of attaining high response rates or of undertaking extensive overpractice.

Comparison with accuracy training. Another group of studies compares the effect on retention of rate-building training versus training to accuracy-alone. Typically two skills are trained to accuracy. Training of one skill then ends but practice of the second skill continues. In some studies, rate-building ends when the participants reach a specific fluency performance standard. For example, Bucklin et al. (2000) trained 29 college students on two component skills of a stimulus equivalence task. Half of the students ended practice when they reached accuracy. A single 1-min timing revealed that they had not reached fluency. The other students continued practicing until they met a fluency performance standard. Retention was assessed by comparing percent accuracy, not response rate, after a 16-week non-practice period. They found a statistically significant smaller drop in accuracy for the group that learned to accuracy plus fluency. However, they note that, since “fluency trainees practiced the component skills more than accuracy trainees ... the results cannot be attributed solely to the response rate requirement inherent in fluency training” (Bucklin et al., 2000, p. 160).

In similar studies, rate-building occurred but participants were not required to meet a performance standard. Berquam (1981) and Olander et al. (1986) describe studies of this type. Note that they use the term ‘fluency’ loosely to refer to rate-building, but do not use a performance standard. Olander et al. (1986) assessed the retention of skills learned by nursing students using either rate-building or traditional lectures. Retention was assessed after an 8-month interval with no training. They found that those who received rate-building were 1.8 times more accurate and were more fluent than those taught traditionally.

Berquam (1981) studied 34 students in the third grade. Students were divided into two groups and trained on a paired associates task (see trigram/write corresponding 1-digit number). In the first phase of the experiment, all students learned the task for

five days, at which point response rate was measured. This was followed by a 2-week retention interval. They then relearned the task. This was followed by another 2-week retention interval. Berquam found that the higher the response rate attained at the end of training, the better the retention of response rate and accuracy two weeks later.

A second experiment was designed to compare the effects of extra practice and rate-building practice. One group of students received practice specifically designed to increase the rate of correct responses (described as fluency training), while the other group received unpaced practice. The fluency training group performed better on retention tests (as assessed by rate of correct responses per min) than did the group who received unpaced practice. The unpaced practice group was then given rate-building practice. Their performance on retention tests improved. He concluded that fluency training provided an efficient method of improving retention.

Berquam (1981) and Olander et al. (1986) did not employ a fluency performance standard. Binder (1996) notes that fluency is not just going faster per se, but is defined as meeting a specific response rate and accuracy requirement (performance standard) and observing retention, endurance and application. All the studies described in this group did not control for practice. Given that rate-building is a form of practice (Kubina, 2005), participants received more practice on the skill where rate was built. The positive retention effects they note with rate-building may have been produced by the extra practice.

Controlling for practice without fluency performance standards. A number of studies, while controlling for practice, did not train to fluency performance standards. For example, Ormrod and Spivey (1990) taught undergraduate students spelling words to either: (a) accuracy; (b) accuracy with an additional ten ‘unpaced’ (non-rate-building)

practice trials; or (c) accuracy with an additional ten 'paced' (rate-building) practice trials. Retention was assessed by comparing the group's average post-training accuracy to the group's average accuracy after three weeks of no practice. They found a statistically significant difference between the means of the accuracy only and both the groups with extra practice, suggesting that extra practice beyond accuracy enhances retention. However, they found no statistically significant difference between the means of the paced and unpaced extra practice groups, suggesting that rate-building practice was no more effective than unpaced practice in enhancing retention. However, Peladeau et al. (2003) reviewed Ormrod and Spivey's (1990) study and note that there was a difference in favour of the fluent group (0.36 standard deviations) and suggest that failure to find a statistically significant difference may be due to their small sample size. While both of these groups received the same amount of practice, response rate was not measured at the end of training for any group. In addition, the paper does not report that a fluency performance standard was met during paced practice.

Peladeau et al. (2003) investigated the effects of differences in the type and amount of practice on long-term retention of quantitative methods. A class of 108 first-year college students were matched on pre-test abilities and randomly assigned to one of three experimental groups: (a) learning to an 85% correct mastery criterion alone; (b) achieving an 85% correct mastery criterion and then overlearning with a focus on maintaining accuracy for up to 5 weeks; and (c) achieving an 85% correct mastery criterion and then overlearning for 5 weeks with a focus on building response rates. They also had a special comparison group of students who did not practice enough to achieve 85% correct mastery. Each overlearning group received the same amount of practice. Accuracy was assessed through exams at the end of the course and 6 months later for a sub-sample of 83 students. Half of the final exam and retention test

questions were drawn from practice questions and the other half were questions of a similar format with different data values.

In comparing final exam scores and 6-month retention test scores they found that practicing to mastery resulted in significantly higher scores than not achieving mastery. Overlearning, whether via fluency building or just extra practice trials, resulted in higher scores than the mastery-only group. This is contrary to other research that found that the positive effect of overlearning on retention fades after 6 weeks (Driskell et al., 1992; Rohrer, Taylor, Pashler, Wixted, & Cepeda, 2004).

Peladeau et al. (2003) found that rate-building (without a fluency performance standard) did not provide any additional benefit over the same amount of non-rate-building practice in terms of either final exam or 6-month retention test scores. This contradicts the conclusions of the earlier work of Berquam (1981) who found higher performance in the group who received rate-building.

Ormrod and Spivey (1990) and Peladeau et al., (2003) suggest that when practice is controlled for, rate-building does not produce superior retention than unpaced (non-rate-building) overpractice. However, if we consider the studies in the light of Binder's (2004) cautions, we can not draw conclusions relating to retention as an outcome of fluency because they neither specified nor trained the rate-building group to a fluency performance standard. A further difficulty with these studies is that they did not check the response rates of any groups other than the rate-building group at the end of training. Other groups may have reached similar response rates. The need to probe response rates post-training and before retention tests is stressed by Binder (2004).

Controlling for practice with a fluency performance standard. Shirley and Pennypacker (1994) compared retention, after rate-building practice beyond accuracy to a fluency performance standard, with the same amount of rate-building practice that did not meet the fluency performance standard, while controlling for practice. They compared the effectiveness of four training methods, using a within-subjects experimental design, on the retention of spelling words: (a) typical classroom training; (b) daily training for a period of one week; (c) training to an accuracy criterion; and (d) training to both an accuracy and fluency criterion.

In Phase 3 of their experiment, participants practiced two yoked lists of ten spelling words daily until the participant reached 100% accuracy on one list. In Phase 4, participants practiced two yoked lists of 10 spelling words daily until, during a daily timed performance trial, both lists were completed with 100% accuracy and one list was completed at a fluent rate. The fluency performance standard was set as the rate at which each participant wrote the letters of their own name during a 1-min timing. In both phases, participants ceased practice on both groups of spelling words when these criteria were met, received two new word lists and repeated this procedure. Hence, each list of words was practiced the same number of times as its yoked partner.

Retention accuracy data were collected during two follow-up sessions ten days after the performance criterion was reached. Shirley and Pennypacker (1994) found that more frequent practice sessions (daily versus tri-weekly) resulted in greater learning evidenced by better retention. This finding provides support for the role of practice in both the acquisition and retention of spelling words. For one participant, the words learned to the more stringent performance criterion (daily practice versus less frequent practice; 100% accuracy versus not reaching accuracy; and fluency plus accuracy versus accuracy alone) in each phase were retained better (in terms of accuracy). In Phase 3,

the other participant showed better retention on the group of words not learned to accuracy than on the group learned to accuracy. In Phase 4 he showed better retention on the words learned to fluency than on those that did not reach fluency. Shirley and Pennypacker (1994, p. 86) note that, while data from Phase 3 (the comparison of training to fluency and accuracy with training to accuracy alone) suggests that practicing to a fluency performance standard produces better retention, “this conclusion is not supported by comparing phases three and four.” In summary, Shirley and Pennypacker’s (1994) study provides inconclusive support for the claim that meeting a fluency plus accuracy criterion produces superior retention than meeting an accuracy criterion alone if the same amount of practice occurs.

There are several reasons why this study may not have found compelling evidence for the efficacy of rate-building to a fluency performance standard compared to the same amount of rate-building that did not meet a fluency performance standard. The authors suggest that requiring participants to both meet a higher fluency criterion and maintain the fluency rate over three consecutive sessions before ending training may have produced more consistent and larger effects. A second reason for the lack of difference between rate-building practice with and without a fluency performance standard may be that Shirley and Pennypacker (1994) did not limit the response rate of the yoked word list that did not reach fluency. As a consequence, there may have been little difference between the rates at which the two spelling lists were answered. From their paper, one assumes that it is a race between which of the two yoked lists in a pair the participant will reach fluency on first. Since response rate was not limited on either list within a yoked pair, one would assume that when participants reached fluency on one list they were close to fluency on the other list. The final reason they may not have

found a difference is that fluency training may not produce significantly superior retention than the same amount of non-rate-building practice.

Purpose of this Study

Given the lack of controlled research and inconclusive results it is pre-emptive to suggest that retention, one of the proposed outcomes of rate-building procedures is the result of rate-building alone (Bucklin et al., 2000; Doughty et al., 2004). Attaining rate criteria typically involves practice beyond 100% correct. The effectiveness of rate-building procedures, as opposed to training to accuracy criteria alone, may be due to the greater number of exposures to practice trials and/or reinforcement rate associated with rate-building training rather than achieving the fluency performance standard *per se*.

Doughty et al. (2004) suggest the need for research involving “isolation of the effects of high rates of responding from increased practice (i.e., number of exposures) or a denser schedule of reinforcement” (p. 20). Understanding the relationships between practice, high-rate responding and learning outcomes is foundational to an understanding of the effectiveness of Precision Teaching. If equal learning outcomes are achieved from both rate-building to a fluency performance standard and an equal amount of non-rate-building practice, the need for attaining fluency is called into question. Rate-building techniques may just provide a more efficient method of achieving this practice, providing the same gains in less time, thus freeing time for other learning (Binder, 2004; Peladeau et al., 2003). However, there may be other costs associated with high response rate training (e.g., student fatigue) which can be avoided or reduced if increased practice is as effective as fluency training.

Given these findings, the aim of the current study was to determine whether rate-building to a fluency performance standard produced better retention than the same

number of practices without rate-building. In this study, practice was defined as *seeing* a multiplication question, *saying* an answer and *seeing* the answer, regardless of whether the participants answer was correct or incorrect. Using a within-subjects experimental design, participants learned two sets of previously unknown multiplication facts to either an accuracy plus fluency performance standard or an accuracy criterion alone. Sets were yoked so that both sets of facts were practiced an equal number of times. Retention was assessed at both four and eight weeks after training and practice ended. The experiment ran in four main phases. In the first phase, students learned previously unknown multiplication facts to accuracy. The second phase equalised the number of practices of each fact. In the third phase, the ten original facts were divided into two decks of facts. One deck was then used in free-operant rate-building training to an established fluency performance standard, while the other was used in rate-controlled training. During rate-controlled training, response rate was kept well below the fluency performance standard. In the final phase, retention and response rate was assessed at zero, four and eight weeks post training for both decks of multiplication facts.

Methodological Considerations

The practice of rate-building does not conform to a standard protocol. For example, rate-building sessions may occur daily or less frequently, with one or many timings per session. Trials could occur with or without feedback or reinforcement (Kubina, 2005). A rate-building protocol that incorporates daily sessions with multiple trials, feedback, and reinforcement was chosen for this study because it is typical of approaches used within the Precision Teaching model.

There are a number of methodological considerations. Firstly, there is variability in how long and how much practice is required to reach fluency. To take account of

individual differences between participants a within-subjects experimental design was used. Secondly, to take account of possible prior learning, participants were pre-tested on possible facts to be learned and only facts they were unable to answer were included. Thirdly, to control for the possibility that some facts will be learned to accuracy faster than others, participants continued to practice all facts until 100% accuracy was achieved. Fourth, to control for practice, facts were divided into two groups, one to be overpracticed to a fluency criterion and the other to be overpracticed without rate-building but keeping the amount of practice the same for both decks. Fifth, reinforcement amount was kept constant for each group of facts. Finally, freely emitted response rates for both decks were measured at several points in the experiment. Measurement at the end of training ensured that fluency had not developed incidentally on the facts that were practiced without rate-building.

Method

Participants and Setting

Ethical approval for this research was obtained from the Department of Psychology Research and Ethics Committee, at the University of Waikato. Participants were recruited by placing a notice on a home-schooling web site accessed by those in the local area. This notice briefly outlined the nature of the research and gave the researcher's contact details (the notice is included in Appendix A). When the researcher received expressions of interest, a fuller description of the experiment and what was involved in participation was sent to the inquirer (an example is presented in Appendix B). Those interested then contacted the researcher who met with them to discuss the experimental goals and describe sessions. Informed consent was obtained from all participants prior to their inclusion in this research. Those over 16 years gave written consent. In the case of child participants, written consent was obtained from a parent or caregiver of each child. Subsequent to parental consent, the researcher described to the child what was involved in the study and children gave verbal consent. All participants were given the right to withdraw from the experiment at any stage. Each participant was provided with a summary of the experimental findings at the conclusion of the experiment.

Five children, one adolescent and one adult with no diagnosed learning disabilities participated in this study. There were 4 girls aged six (P2), eight (P5), eleven (P3) and seventeen (P6) years; 2 boys aged eight (P1) and eleven (P4) years all of whom were home-schooled; and 1 adult aged 20 (P7) years. All training and experimental sessions occurred daily (Monday through Friday) in the participants' homes. Sessions lasted approximately 40 min for those aged less than 12 years and 15 min for the others.

Longer sessions were scheduled for younger participants to allow for a 25-min game time with the experimenter at the conclusion of the experimental component.

Task Selected

The task selected for study was learning multiplication facts. This task was chosen as recall of basic multiplication facts (one to 10 times tables) is an aim of the New Zealand mathematics curriculum number section entitled 'exploring computation and estimation' (Ministry of Education, 2005).

Multiplication facts from the three, four, six, seven, eight and nine times tables that resulted in 2-digit answers were selected as the training task. Facts were omitted if they involved the operands one, two, five, ten and eleven as these were deemed easier. One fact from each pair of facts involving the same operands was also omitted. For example, the fact 3×4 was included but 4×3 was omitted. There are 20 facts that meet these inclusion criteria.

Facts that the participants answered correctly during a pre-test were also omitted. If this resulted in less than 10 facts for a particular participant, facts from the 12, 13 and 14 times tables resulting in 2-digit answers were included. This provided an additional 13 facts to choose from and allowed 10 unknown facts to be chosen for all participants.

Pre-test of Pre-requisite Skill

During the course of this study, participants learned to answer multiplication facts accurately and fluently. If participants could not read one and two-digit numbers quickly and accurately (a pre-requisite skill) it would limit their ability to meet the rate criterion set for the experimental task. For this reason, each participant was assessed as to the rate at which they accurately read 1- and 2-digit numbers (see number/say

number) prior to their inclusion in the study. The experimenter presented participants with flashcards of all the one and two digit numbers that were used in the experiment. Participants were asked to read each number out loud as the card was presented to them. If participants failed to read all numbers correctly or their response latencies were greater than 1 s they would not have been included in the study. However, all potential participants met these criteria.

Apparatus

The first learning phase, learning to answer facts to an accuracy criterion, involved the use of a set of ten 7 x 12-cm laminated flashcards for each participant. Each flashcard had a multiplication question printed on one side and the answer printed on the reverse side (an example is presented in Appendix C). A data collection form was used to record each individual fact presented and participants' responses (correct or incorrect) during accuracy training (an example is presented in Appendix D).

The second learning phase involved learning to answer half the facts to an accuracy plus fluency criterion. Rate-building software that was designed specifically for the purposes of this study was used. Appendix E provides a complete description of this software. The software allows the experimenter to create virtual decks of cards containing multiplication questions and answers. These multiplication facts were repeatedly presented by the computer, in a random order, to the participants during practice trials. This ensured that participants encountered each fact in no predictable sequence and with an equal probability. The computer also tracked the performance of individuals, recording the number of cards presented, the number of problems answered, the length of a trial and response latency.

Computer-based presentation was chosen because it presents facts and displays answers in a standardised fashion. All computer-based sessions were conducted on a Compaq Armada 1750 laptop computer running Windows XP. A manual data collection form was used to record the number of correct and incorrect responses for each presentation of Deck A or Deck B facts (an example is presented in Appendix F).

A Sony digital video camera was used to record two sessions from each of experimental Phases 1 and 3 and each retention assessment for each participant. These videos were later viewed by a trained observer and used to conduct an inter-observer reliability check.

Several reinforcers were used through the course of the study. For child participants these included: praise, small sweets, and tokens (valued at 25 cents each) that could be accumulated and exchanged for toy items of various prices, the minimum of these being 8 tokens. Children also spent a minimum of 25 mins each session playing a game of the child's choice with the experimenter. The adult participant received larger sweets and praise but no further reinforcement. Since the adult participant specifically indicated a wish to participate, the learning of the times tables themselves was considered an appropriate incentive.

Procedure

A within-subjects experimental design was used. The experiment ran in four main phases. In the first phase, accuracy training, students were taught previously unknown multiplication facts to accuracy. The second phase, equalisation of practice, equalised the number of practices of each fact. In the third phase, rate-building and rate-controlled training, the ten original facts were divided into two decks of cards. One deck was then used in free-operant rate-building training to a fluency criterion, while the other was

used in rate-controlled training. In the final phase, assessment, retention and response rate was assessed at zero, four and eight weeks post training. The experimental method is described below and is summarised in Figure 1.

Phase 1: Accuracy training. Accuracy training was conducted over multiple sessions and continued until an accuracy criterion (described below) was reached. During the first session, the participant was presented with multiplication facts from a subset of the facts between 3×4 and 9×9 , in a random order. The facts were presented on 7×12 -cm flashcards. The participant was asked to either give an answer to the problem or say “I don’t know.” The first ten unknown facts compiled for each participant were used for that participant for the rest of the experiment. If less than ten facts were unknown, participants were presented with a second set of flashcards, also in a random order, with a subset of facts from the 12, 13 and 14 times tables.

In subsequent sessions, participants learned to give accurate answers to these ten multiplication facts. During the first of these sessions, the experimenter provided a demonstration of the procedure using simple 2 times table multiplication facts that were not included in the participant’s ten facts. Next, the participant was presented with two of the flash cards selected from the ten to be learned. They were given two minutes to look at the facts and the answers to those facts. The experimenter then laid the cards out in a row, answer down. The participant was asked to give an answer to the first fact and then turned over the card to check their answer. If their answer was correct, they moved onto the next card in the row and repeated this process for the second card. If the answer given was incorrect, an error correction procedure was used. This involved looking at the correct answer, turning the card over again so that the problem was shown and

Pre-test		Test participants on pre-requisite skill: reading of 1- and 2-digit numbers.
Phase 1: Accuracy training		Compile set of 10 unknown multiplication facts by testing participants.
		Train to an accuracy criterion on all 10 facts using flashcards. Continue until accuracy criterion of 3 correct consecutive trials for 3 consecutive sessions was reached.
Phase 2: Equalisation		Provide extra practice to equalise the number of practices of each fact.
Phase 3: Rate-building and rate-controlled training		Allocate facts to either Deck A or Deck B such that overpractice differences between Decks A and B are minimised.
		Post-accuracy training assessment of answer rate for three 1-min timings on: a) 2 times table facts b) Deck A facts c) Deck B facts
		Use software to train Deck A facts to a fluency criterion and to present Deck B facts the same number of times but at a controlled rate. Training continues until the participant achieves the fluency criterion of three consecutive sessions with rates of 70-100 correct facts per minute and no more than two errors in the rate-building component.
Phase 4: Assessment	Post-software training assessment	Post-software training assessment includes four components: a) test on all ten flashcards b) 1-min timing on 2 times table facts c) 1-min timing on Deck A facts d) 1-min timing on Deck B facts
		4-week period of no training or practice
	4-week retention assessment	Retention assessment includes four components: a) test on all ten flashcards b) 1-min timing on 2 times table facts c) 1-min timing on Deck A facts d) 1-min timing on Deck B facts
		4-week period of no training or practice
8-week retention assessment	Retention assessment includes four components: a) test on all ten flashcards b) 1-min timing on 2 times table facts c) 1-min timing on Deck A facts d) 1-min timing on Deck B facts	

Figure 1. Outline of the experimental procedure.

beginning the procedure again (with the first card). If an incorrect answer was given to either card, the participants always started again from the first card. Once the participant correctly answered both cards, this completed a trial and they chose a sweet. The cards were shuffled and laid out again for a new trial. Participants continued until they completed three correct trials in a row or until the first sweet was obtained after the session time had reached 15 min, whichever came first. For the adult participant, the session then ended. For child participants, the experimenter played a game of the child's choice with the child. The length of this play was the time remaining in the 40-min session. If the participant failed to follow the procedure, the experimenter provided verbal prompts. Participants were requested not to practice the facts learned between sessions.

At the start of every session the experimenter asked each participant whether they had practiced since the last session. If they reported that they had not, all participants were praised and child participants were given one token (worth 25 cents). Tokens (also given for other responses later in the experiment) accumulated over multiple sessions and could be exchanged for toy items at the end of any session. In each session, the procedure of the previous session was repeated, with the same cards, unless the participant had reached the accuracy criterion in the previous session (three consecutive correct trials). In this case two new cards were introduced. When introducing new cards, the same procedure was used as with the first two cards, except that once the participant completed two correct trials in a row these cards were added to the previously learned cards. After cards were added, new trials were formed from the combined deck of cards. For example, once four cards had been learned, all four cards were laid out in a row and a trial consisted of answering all four cards correctly, moving from left to right. Once more than five cards were in the deck, only five cards, drawn at random, were laid out

in a row in a single trial. The remaining cards were always used in the next trial. If there were less than ten cards in the total deck, a random selection of cards from the previous trial were also included to make up the total of five cards in a trial. For example, if there were eight cards in the deck (numbered 1 to 8 for the sake of this example) and Cards 1, 3, 5, 6 and 8 were drawn for the first trial, the second trial would include Cards 2, 4 and 7 plus two cards, drawn at random from 1, 3, 5, 6 and 8. Each session continued until the participant completed three consecutive correct trials in a row or until the first sweet was obtained after the session time had reached 15 min, whichever came first. Training continued in this way until all ten cards were learned and the participant reached the terminal accuracy training criterion of three consecutive errorless trials on each of three consecutive days.

Steps were taken to limit rate-building during this phase. Requiring participants to always turn cards over once an answer was given and then turn cards again limited response rate. Participants were given verbal instruction to 'go slowly' when answering cards if they moved through the cards at a high rate. The experimenter shuffled the cards and laid them out, limiting the rate at which new trials were presented.

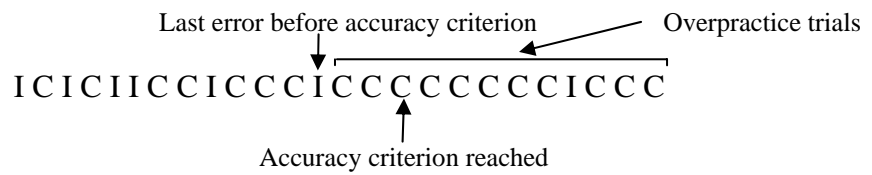
For each participant, a session early and late in this phase was videotaped. Videotapes were viewed by another observer who was trained by the experimenter to record each practice of a card and whether responses were correct or incorrect on data sheets identical to those used by the experimenter. Percentage agreement was calculated by dividing number of agreements by number of agreements plus disagreements and multiplying by 100.

Phase 2: Equalisation. Once the terminal accuracy criterion was reached, further sessions were used to equalise the number of practices of each fact. That is, to provide enough extra practice of each fact so that all facts within a participant's set of 10 facts were practiced the same number of times. As with the later sessions in the accuracy training phase, cards were presented in five card trials. Participants were asked to answer each fact and then to check their answer. Even when facts have been learned to accuracy, occasional incorrect answers do occur. In this phase of the experiment, return to the first card in a trial did not occur when an incorrect answer was given; the next card was always attempted. Equalisation sessions had between 10 and 15 trials. Participants chose a sweet at the end of each trial, even if an error was made. Each session continued until the first sweet was obtained after the session time had reached 15 min. The Equalisation Phase ended when all of a participant's 10 facts had been practiced an equal number of times.

During Phases 1 and 2, a record was kept of the facts presented for each trial and whether they were answered correctly. During the Equalisation Phase the same steps were taken to limit rate-building as were used to limit rate-building in Phase 1.

Phase 3: Rate-building and rate-controlled training. Following Phase 2, each participant's ten cards were divided into two decks of five cards each. Facts were allocated between the two decks so that the number of overpractice trials was approximately the same for each deck. Overpractice has been defined as practice beyond the point where facts were answered accurately (Dougherty & Johnston, 1996). In this study, the precise definition of overpractice used is: the number of times a fact was practiced after the last error before the accuracy criterion was reached. For

example, in the following sequence of answers (corrects (C) and incorrects (I)) to a fact, there are 12 overpractice trials.



Facts were allocated between the decks by first ordering them based on the number of overpractice trials. The facts were then allocated alternately to each deck and the difference between the total number of overpractice trials for each deck was then calculated. The difference was then minimised by swapping cards at the same rank between decks. This process is as shown in the example in Figure 2.

After the facts were allocated to either Deck A or Deck B, participants were trained in a single session to use the software developed for this experiment. All training sessions involved a deck of five simple 2 times table facts. Initially, the experimenter

Fact	OP	Fact	OP	Deck A		Deck B		Deck A		Deck B	
7 x 8	80	7 x 8	80	Fact	OP	Fact	OP	Fact	OP	Fact	OP
4 x 9	80	4 x 9	80	7 x 8	80	4 x 9	80	7 x 8	80	4 x 9	80
6 x 7	77	6 x 7	77	6 x 7	77	8 x 9	76	6 x 7	77	8 x 9	76
8 x 9	76	8 x 9	76	6 x 6	72	7 x 7	56	7 x 7	56	6 x 6	72
7 x 7	56	6 x 6	72	6 x 8	46	8 x 8	37	6 x 8	46	8 x 8	37
6 x 6	72	7 x 7	56	9 x 9	27	4 x 7	22	9 x 9	27	4 x 7	22
8 x 8	37	6 x 8	46	Total	302	Total	271	Total	286	Total	287
6 x 8	46	8 x 8	37								
9 x 9	27	9 x 9	27								
4 x 7	22	4 x 7	22								
(a)	(b)	(c)	(d)								
Initial No. of overpractice (OP) trials for each fact	Facts ordered by No. of overpractice (OP) trials	Initial allocation between decks	Pair at rank 3 swapped to minimise the difference in the total number of overpractice (OP) trials for each deck								

Figure 2. The process by which facts are allocated to each of Deck A and Deck B.

explained and demonstrated use of the software. The software required users to click the left mouse button as they said the answer to a multiplication fact presented on the computer screen in a large green highlighted box. When the participant demonstrated an understanding of what was required they practiced both timed 1-min and rate-controlled computer trials using the same 2 times table facts.

The software was configured so that, during rate-controlled (Deck B) trials, a minimum 3-s delay was imposed between presentations of facts. As a consequence, participants could not build response rates beyond 20 responses per min. During rate-building (Deck A) trials, new cards were presented immediately after an answer was given to the previous card without limiting the response rate. Apart from the inclusion or omission of a pause, all card exposures were presented in an equivalent manner by the software.

The 2 times table trials were repeated until participants used the software with ease. A baseline measurement of answer rate was then taken during three single 1-min timings, for each of the following three decks: the deck of five facts from the 2 times table, Deck A and Deck B. These computer trials provided a baseline for accuracy and response-rate data and a check that fluency had not developed during accuracy training. The order of presentation of Deck A and Deck B facts was alternated between participants to control for possible order effects. For all subsequent assessments, the order of presentation of Deck A and Deck B facts used on this pre-test was repeated. Prior to each 1-min trial, each participant was instructed to “Try to answer as many facts correctly as fast as you can.” No prompts or instructions were given during trials. At the end of each 1-min trial, participants were told the number of correct and incorrect responses they made. In all computer trials, participants were given the opportunity to

self-correct. If they did so before the answer appeared on the computer screen this was recorded as a correct response. The software recorded the facts answered during a trial and the number of incorrect responses was counted by the experimenter.

Subsequent sessions using the software built response rate on Deck A facts to the established fluency performance standard of 70-100 correct responses per minute with no more than two errors (Binder et al., 2002). These sessions also provided an equal number of rate-controlled practices on Deck B facts. During each experimental session, participants were presented with between one and three pairs of trials. Each pair of trials consisted of a Deck A (rate-building) trial followed by a Deck B (rate-controlled) trial. The number of facts presented during each Deck A (rate-building) trial determined the number of facts to be presented during the subsequent Deck B (rate-controlled) trial. In this way the number of Deck A and B facts presented in each session was yoked. Up to three pairs of trials were conducted in a single session. When a participant exceeded their previous best rate in a rate-building trial, the session ended with the associated rate-controlled trial. When a participant did not beat their previous best rate they completed another pair of trials, to a maximum of three pairs. When the session ended, the child participants played a game of their choice with the experimenter for the remainder of the 40-min session. Sessions continued in this manner until the participant reached the fluency criterion for Deck A facts.

During all software sessions similar conditions were maintained to those in the demonstration session. That is, participants were given no prompts or instructions during trials. Prior to rate-building trials participants were given an instruction like: "Answer as many facts correctly as fast as you can." Prior to rate-controlled trials participants were given an instruction like: "Try to answer as many facts correctly as

you can.” Praise and informational feedback (rate and number of correct and incorrect responses) was given at the completion of each trial.

If a participant’s response rate did not increase over three consecutive sessions, short timings (sprints) of 10, 20 or 30 s were employed. Participants initially attempted to increase their response rate to 12 correct responses per 10-s trial. On achieving this, the trial length was increased to 20 s with a goal of 24 correct responses, and then to 30 s with a goal of 36 correct responses. Probe 1-min timings were conducted when response rates met each goal to determine if the response rate achieved in a short timing could be maintained for a 1-min period.

The software was affected by occasional “key bounce” events. Key bounce events cause the software to register two key presses, close together, when only one was made by the participant. When this occurred in a rate-building trial, the experimenter noted the extra response and removed it from the data at the end of the session. When this occurred during a rate-controlled trial, there was no effect on the number of facts answered, or the rate at which they were answered because of the rate-limiting that was in effect. That is, the extra key event would be treated as a rapid response to the new fact and would trigger the software to delay the presentation of the following fact. During this delay, the participant could answer normally. Any additional mouse key presses were ignored by the software during this period. Key bounce events do cause the software to record an invalid low latency response.

Child participants earned tokens in both rate-building and rate-controlled trials. During rate-building trials, all participants were encouraged for their effort if they did not beat their previous best rate. If they exceeded their previous best rate participants were congratulated and children received a token. During rate-controlled trials,

participants were encouraged for their effort after each trial. A token was also awarded for completing all rate-controlled trials in a session without errors.

Participants could receive an unequal number of tokens for each deck of facts. When this occurred over several sessions, equalisation of the number of tokens received for each deck of facts was achieved by the experimenter awarding 'bonus' tokens for some aspect of the participant's performance. For example, if, over two consecutive sessions, a participant received one token for beating their previous best score (for Deck A trials) and none for Deck B trials, the experimenter would provide a bonus token following the final Deck B trial. The participants were not told that any rule was affecting delivery of bonus tokens.

For each participant, a session early and late in this phase was videotaped. Videotapes were viewed by another observer and percentage agreement was calculated in the same manner as described in Phase 1.

Phase 4: Assessment. In the session after the fluency criterion was met, a post-software training assessment was conducted. This assessment consisted of presentation of each of the ten flashcards, one at a time, in a random order. Participants were asked to say the answer to each problem as it was presented by the experimenter. Flashcard presentation was followed by three 1-min timings using the software without rate limiting: one for the 2 times table facts, one for Deck A facts and one for Deck B facts. All software trials included the answers, as was the case for all software trials to this point. These trials provided an opportunity to measure response rates on all three decks of facts and to ensure that fluency had not developed on Deck B facts which had been presented in the rate-controlled format. During assessment, the 2 times table facts were always presented first. For 4 of the participants, Deck A facts were presented next

followed by Deck B facts. For the remaining 3 participants Deck B facts were presented before Deck A facts. During this assessment, participants received no feedback from the experimenter or reinforcement for accuracy or rates achieved. However, they did receive reinforcement, in the form of praise and a token, for completing the 10-card presentation and for each of the three 1-min timings. A data collection form was used to record the order of presentation of each of the three decks of facts and the number of correct and incorrect responses made during each 1-min timing (an example is presented in Appendix G).

Two retention assessments were conducted at four and eight weeks after the post-software training assessment. Participants and their parents were reminded that participants should not practice the multiplication facts over this 8-week period. A modified version of the rate-building software that did not display answers to facts was used for these retention assessments. In all other respects, these retention assessments were the same as the post-software training assessment. No other testing was conducted between the post-software training and the 8-week retention assessment so as to avoid possible maintenance effects introduced by testing.

The 4- and 8-week retention assessment sessions for each participant were videotaped. Videotapes were viewed by another observer and percentage agreement was calculated in the same manner as described in Phases 1 and 3.

Results

Inter-observer Reliability

Video recordings were used to conduct inter-observer reliability checks. These were performed on the recordings of two sessions from each instructional component (Phase 1, accuracy training; and Phase 3 rate-building and rate-controlled training) and each retention assessment for all participants. A trained observer viewed the videos and recorded correct and incorrect participant responses to presented multiplication facts in the same way as the experimenter. Inter-observer agreement was calculated for each observed session by dividing the total number of agreements for responses made to presented facts by the number of agreements plus disagreements and multiplying by 100% (Cooper, Heron & Heward, 1987).

Table 1 presents inter-observer agreement for each retention check, the mean for the dually observed sessions for each instructional component and the mean for all observed sessions for each participant. Reliability for all observed sessions was well above the 80% agreement criterion. Percent agreements for observed sessions ranged from 96.4% to 100%. The mean inter-observer agreement over all observations ranged from 98.7% (P2 and P6) to 99.8% (P3).

Table 1. *Percent inter-observer agreement for all observed sessions for each participant.*

Participant	Inter-observer Agreement (%)				Mean Over All Observations
	Phase 1 Accuracy Training (mean)	Phase 3 Software Training (mean)	Phase 4		
			4-week Retention Assessment	8-week Retention Assessment	
P1	100	99.6	99	100	99.6
P2	96.5	99.4	100	99	98.7
P3	100	100	99	100	99.8
P4	97.2	100	100	100	99.3
P5	100	98.5	100	100	99.6
P6	96.4	99.2	100	99	98.7
P7	98.8	100	99.3	100	99.5

Number of Practices of Each Fact During Phases 1, 2 and 3

Figure 3 shows stacked bar graphs plotting the number of practices of each fact during the three learning phases. The stack of bars show, for each participant: the number of practices until the last error was made during accuracy training (Phase 1, shown in black); the number of overpractices during accuracy training and equalisation (Phases 1 and 2, shown in grey); and the number of overpractices during software training (Phase 3, shown in white). Practices of Deck A facts, those learned to an accuracy plus fluency criterion, are plotted at the left of each graph and practices of Deck B facts, those learned to an accuracy criterion alone, are plotted at the right.

The number of times a fact was practiced before the last error was made during accuracy training (Phase 1) varied within and between participants (black bars), from no practices (e.g., P1 8 x 12 fact) to ninety practices (P2 6 x 7 fact). The number of overpractices required to meet the accuracy criterion (three consecutive correct trials across three consecutive sessions) and complete equalisation for each fact varied between participants, from 58 for P6 to 115 for P2. At the end of equalisation, each participant had practiced each of their 10 facts an equal number of times. This is shown by a dotted line on each graph.

All participants continued to overpractice their 10 facts during software training. Practice for both decks of facts ended when the fluency criterion was met for Deck A. The number of overpractices of each fact required to meet the fluency criterion (three consecutive sessions with rates of 70-100 correct responses per minute and no more than two errors) varied between participants, from 244 for P1 to 731 for P3. This is shown on the graphs as the range of heights of the white bars. Fluency trials were time limited so, in any individual trial, not all facts were practiced the same number of times.

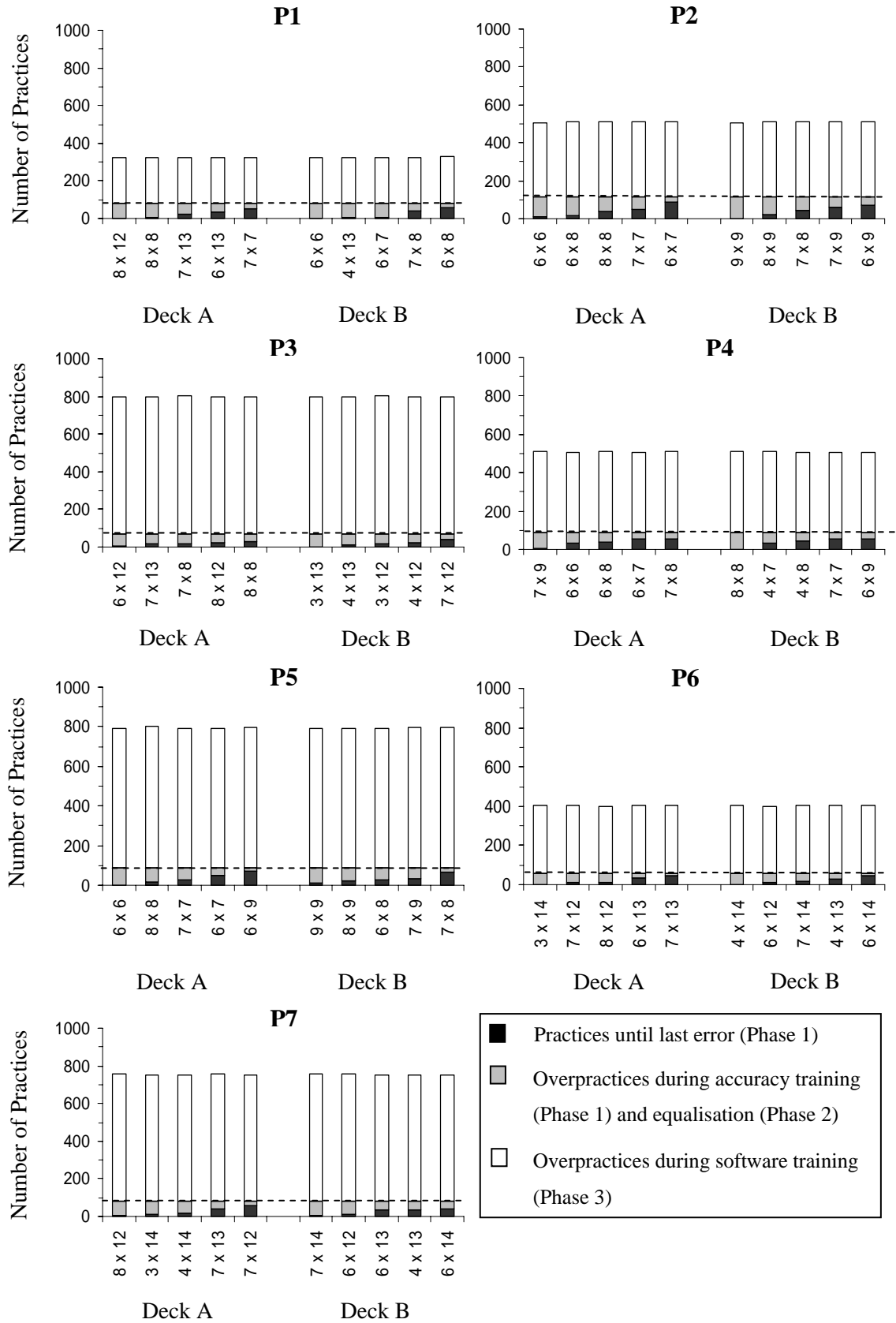


Figure 3. Number of practices of each fact during accuracy training (Phase 1) and equalisation (Phase 2) for the five Deck A and five Deck B facts learned for each participant.

Within a single participant's data, the number of overpractices of each fact during software training (Phase 3) varied over a small range (729-734 for P3 to 704-714 for P5). However, for each participant, the total number of overpractices for each deck (Deck A and Deck B) was equal.

Accuracy During Un-timed Assessments

Figure 4 shows histograms plotting the number of facts from each deck that were answered correctly during un-timed card presentations at each assessment stage. That is, pre-test (before any intervention), post-accuracy training, immediately post-software training and at four and eight weeks after software training, for each participant. The light grey bars represent results for Deck A facts, those learned to an accuracy plus fluency criterion. The dark grey bars represent results for Deck B facts, those learned to an accuracy criterion alone.

For all participants, accuracy on both Deck A and Deck B facts at the pre-test was 0%. Accuracy on both Deck A and Deck B facts at the post-accuracy training and the post-software training assessments was always maximal (100%). At the 4- and 8-week retention assessments, all participants' accuracy was at least four out of five facts correct for each deck. However, participants' accuracy varied within this range. During the 4-week retention assessments, accuracy on both Deck A (accuracy plus fluency) and Deck B facts (accuracy alone) remained maximal (five out of five facts correct) for P2, P4, P6 and P7. At the 8-week retention assessments, accuracy of both Deck A and Deck B facts remained maximal for P4 and P5. However, for P2, while the accuracy of Deck A facts remained maximal, the accuracy of Deck B facts decreased to four out of five correct. For P7, the reverse was true, accuracy of Deck B facts remained maximal and accuracy of Deck A facts decreased to four out of five correct.

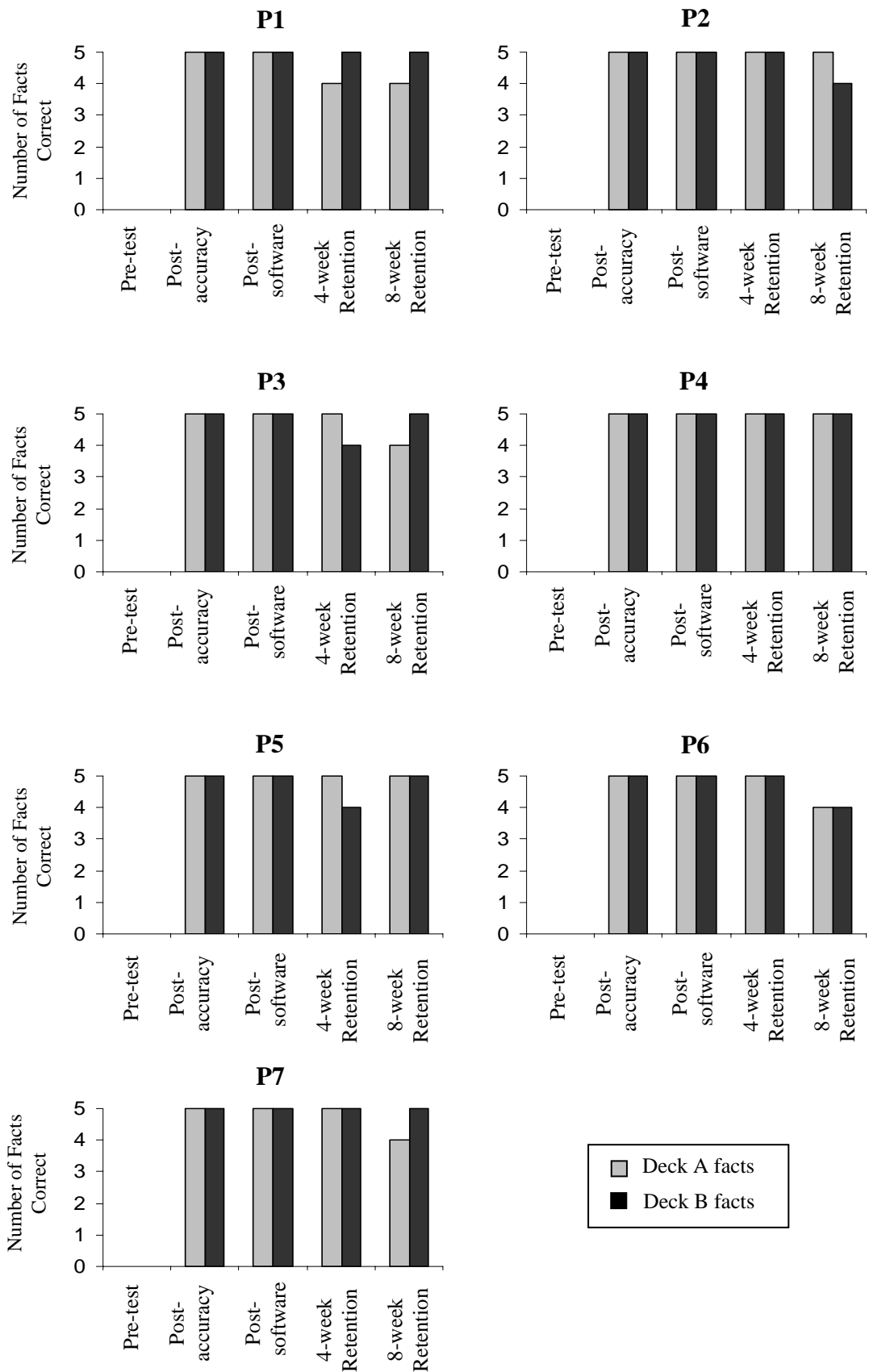


Figure 4. Number of Deck A and Deck B facts answered correctly during card presentations at each assessment session for each participant.

All other participants' accuracy decreased on either Deck A or Deck B facts at the 4-week retention assessment. For P3 and P5, accuracy on Deck A facts remained maximal, while accuracy on Deck B facts decreased to four out of five correct. At the 8-week retention assessment, P3 showed a reversal in accuracy for the two decks of facts. For P5, accuracy on Deck B facts increased and was maximal for both decks of facts. For the remaining participant (P1), at the 4-week retention assessment, accuracy of Deck A facts (accuracy plus fluency) decreased to four out of five correct, while accuracy on Deck B facts remained maximal. This accuracy was maintained at the 8-week retention assessment.

Accuracy and Response Rate During Timed 1-min Assessments

Figures 5 through to 10 show celeration charts for each participant. The celeration charts show the rate of both correct and error responses for 1-min timings over the sessions where response rate was assessed during software training and assessment (Phases 3 and 4). Software training (Phase 3) included a single post-accuracy training assessment for each deck of facts and for a deck of 2 times table facts. The software training phase included multiple sessions where rate was built on Deck A facts. Assessment (Phase 4) included a single post-software training assessment, a single 4-week retention assessment and a single 8-week retention assessment for each deck of facts including the 2 times table.

The number of correct responses to the 2 times table, Deck A and Deck B facts are denoted by the blue diamonds, green squares and red triangles respectively. The number of errors for each deck of facts is labeled with an 'x' of the corresponding colour.

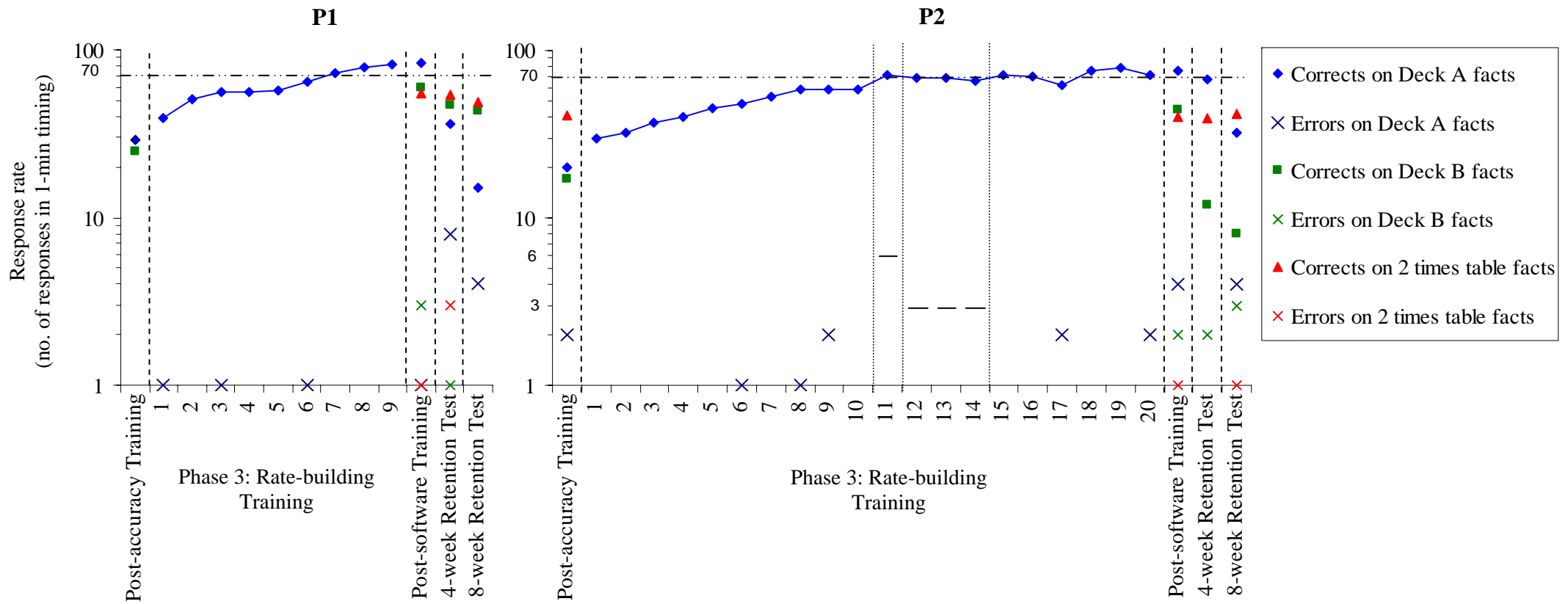


Figure 5. Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed During Phases 3 and 4 for P1 and P2 respectively.

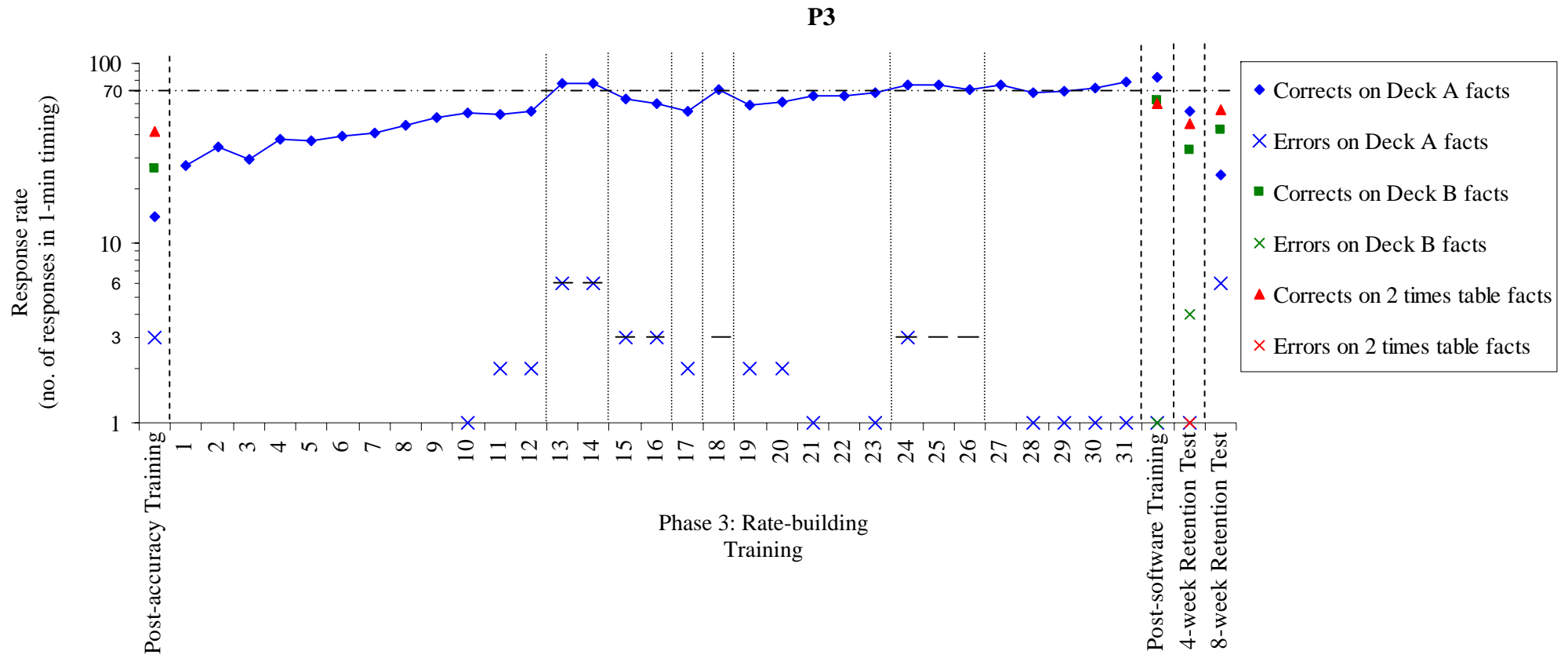


Figure 6. Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P3.

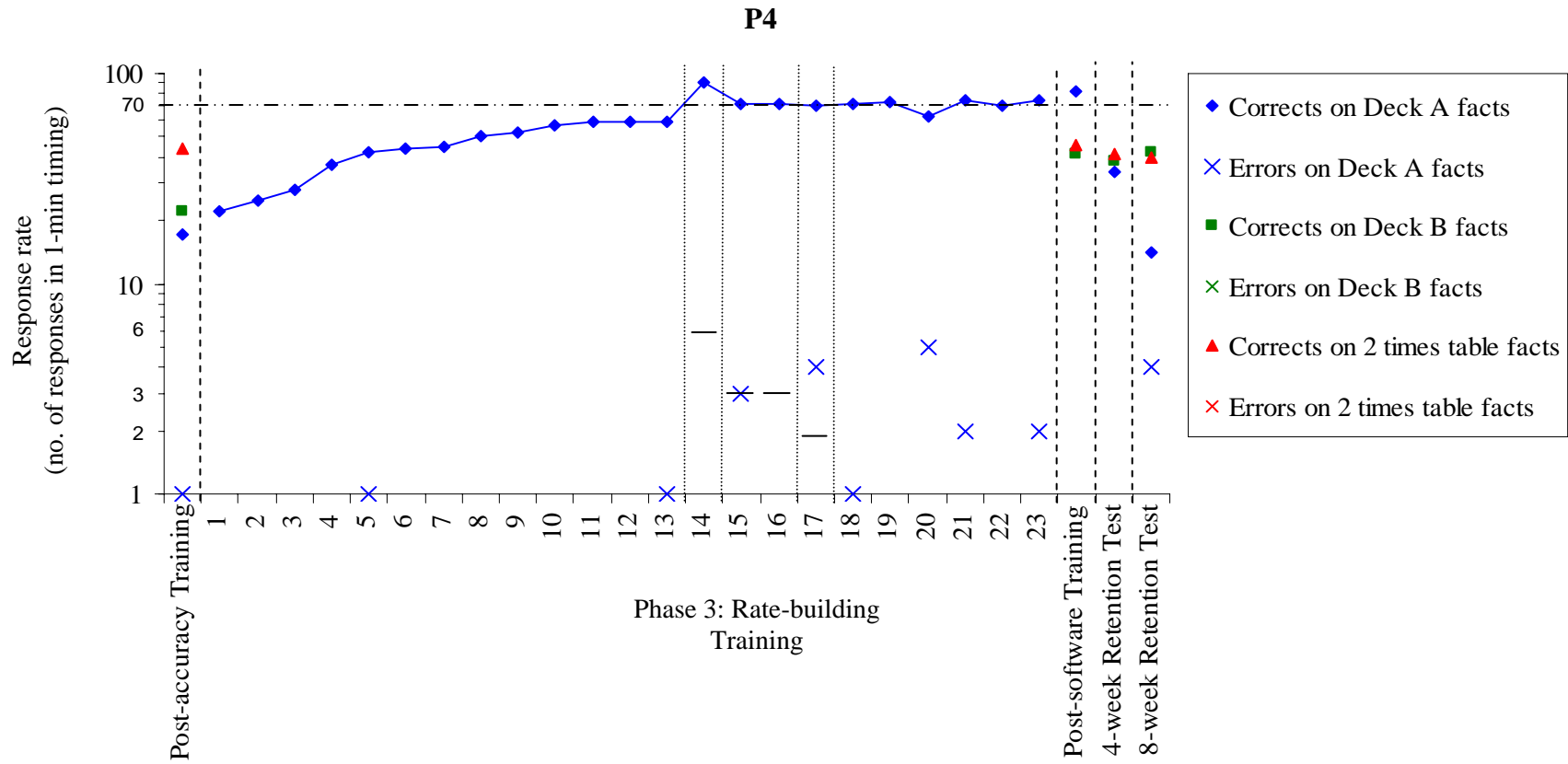


Figure 7. Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P4.

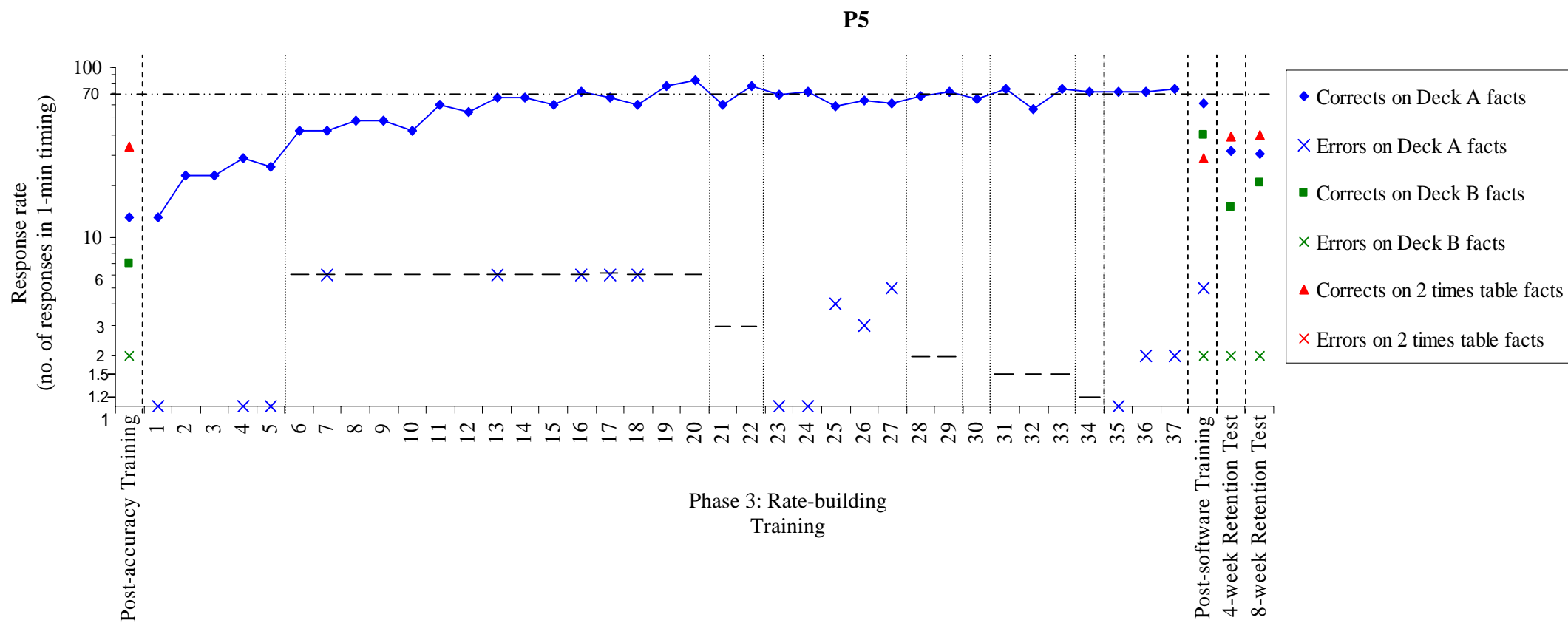


Figure 8. Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P5.

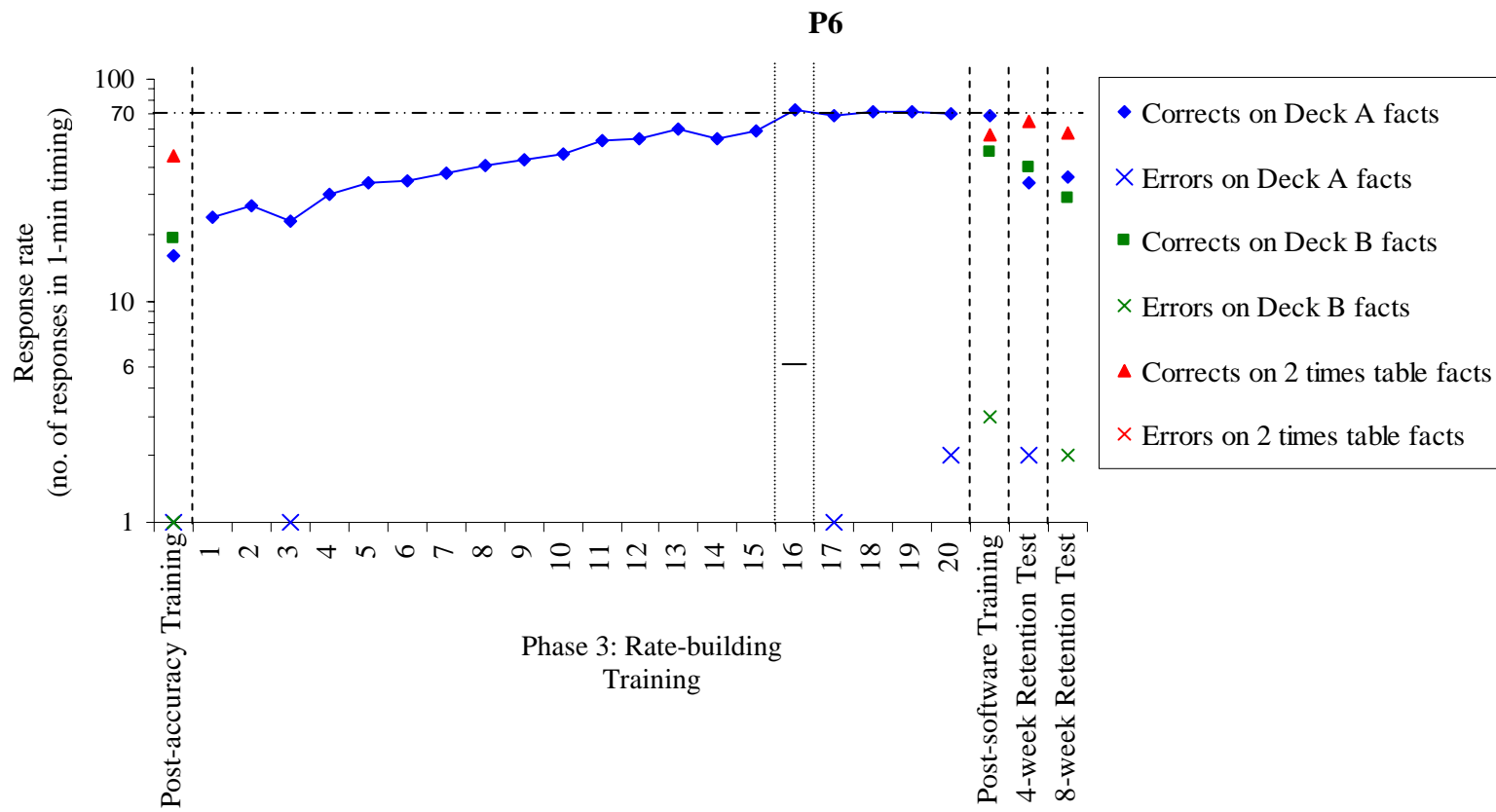


Figure 9. Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P6.

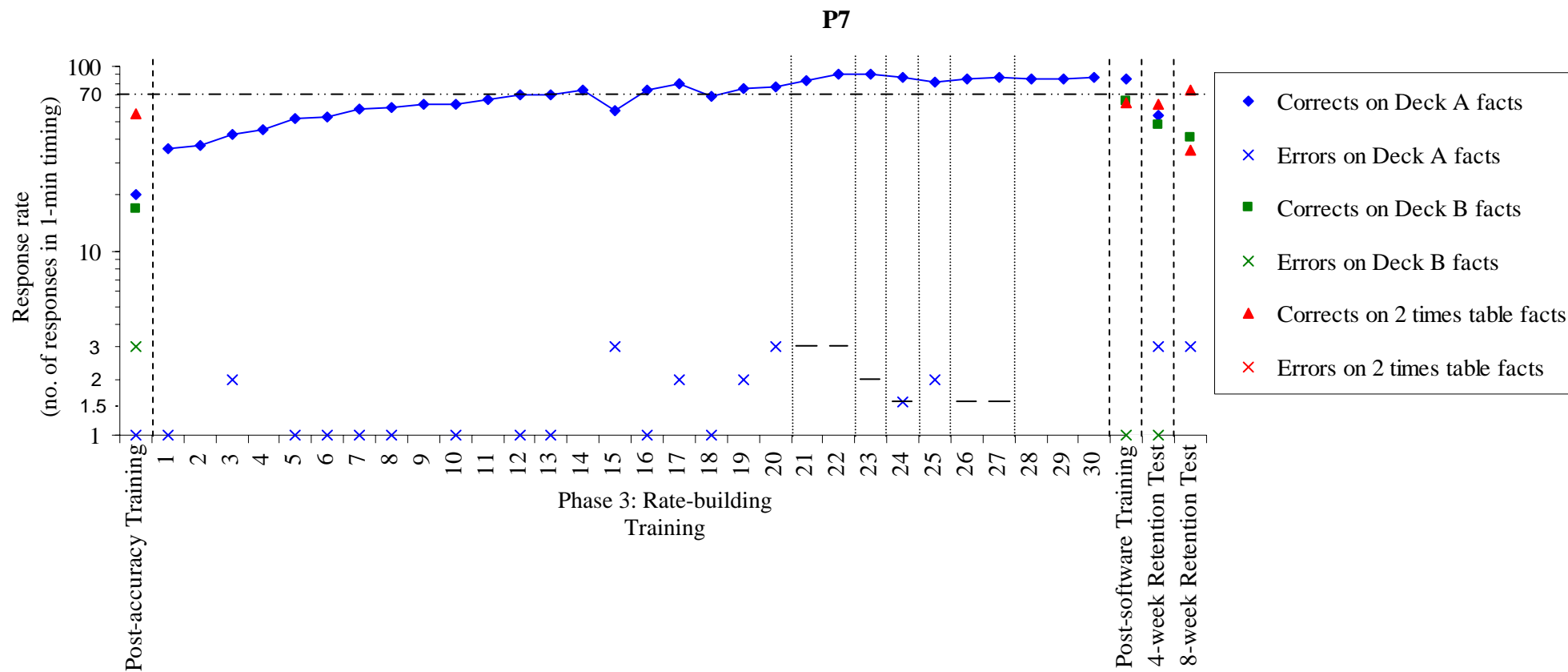


Figure 10. Rate of both correct and error responses, number of facts answered during a 1-min timing, over the sessions where response rate was assessed during Phases 3 and 4 for P7.

On each chart, the dotted horizontal line placed at the 70 responses per minute value denotes the minimum fluency aim for all participants on Deck A facts. Dotted vertical lines during rate-building training on Deck A facts represent changes in timing floors. Timing floors indicate the time period during which facts were presented and responses could be made and were counted. When timing floors other than 1-min were used, response rates were normalised to rate per minute for presentation. Dashed vertical lines (outside rate-building training) indicate a change to another assessment, e.g., the change from post-software training assessment to the 4-week retention assessment. Short horizontal lines during rate-building training with Deck A facts represent 10-, 20-, 30- and 40-s timing floors applied for that session. By convention, these lines are plotted at the y-value which corresponds to the fraction of 60 that the timing floor represents. For example, a 20-s timing floor is one third of the 60-s goal and is therefore plotted at $y = 3$.

For all participants, the rate of correct responses on Deck A facts (denoted by blue diamonds) increased over a number of rate-building sessions until each participant met the fluency criterion. The number of sessions required to meet this fluency criterion varied among participants from 9 for P1 to 37 for P5. All but one participant (P1) required the introduction of sprints, denoted by changes in timing floor, to achieve the fluency aim. For five out of seven participants, response rate on Deck A facts remained at or above 70 correct responses per min at the post-software training assessment. For the other 2 participants, P5 and P6, response rate fell below the fluency criterion at the post-software training assessment session. For all participants, response rate to Deck A facts decreased over the retention assessment period.

The maximum number of error responses per minute to Deck A facts during rate-building varied within and across participants in no consistent pattern. The error

rate ranged from 1 error per minute for P1 to 6 errors per minute for P3 and P5. Note that the process of normalising short trials (e.g., 10-s timings) to the 1-min standard results in a multiplier effect for error rate.

The rate of correct responses to Deck B facts (denoted by green squares and trained to an accuracy criterion alone) also increased during software training for all participants. However, the participants' response rates did not increase to the same levels as those attained on Deck A facts (those trained to accuracy plus fluency). Over the retention assessment period, response rates to Deck B facts tended to decrease. However, for P3 and P5, response rates decreased at the 4-week retention assessment and increased again at the 8-week assessment. This increase was 24% of the decrease at Week 4 for P5 and 34% for P3. P4 showed only small changes in response rates to Deck B facts over the assessment period. Over this period, the rate of error responses for Deck B facts remained low for all participants (0 to 3 errors per min).

For all participants, the rate of both correct and error responses to the 2 times table facts (denoted by red triangles) did not change consistently over the assessment period. Error responses remained low, at either no errors or 1 error per min. The increase in response rate to Deck A and Deck B facts did not generalise to the 2 times table facts.

For ease of comparison, the response rate data for the four assessment sessions (post-accuracy training, post-software training and 4- and 8-week retention assessments) are re-plotted in Figure 11 for the 2 times table facts, Figure 12 for Deck A facts and Figure 13 for Deck B facts. These figures show changes in the rate of correct and error responses during 1-min timings at each of the four assessment sessions. Solid lines show the rate of correct responses at each assessment session. The rate of error responses are denoted by dotted lines.

Figure 11 shows the changes in response rate for the deck of 2 times table facts for each participant. There is no systematic trend in the rate of either correct or error responses per min across assessments within any participant's data set. Participants with fast initial correct response rates tended to maintain these fast response rates and those with initial slower response rates tended to maintain slower response rates across assessments. The rate of error responses remained low (0 to 3) for all participants across the assessments.

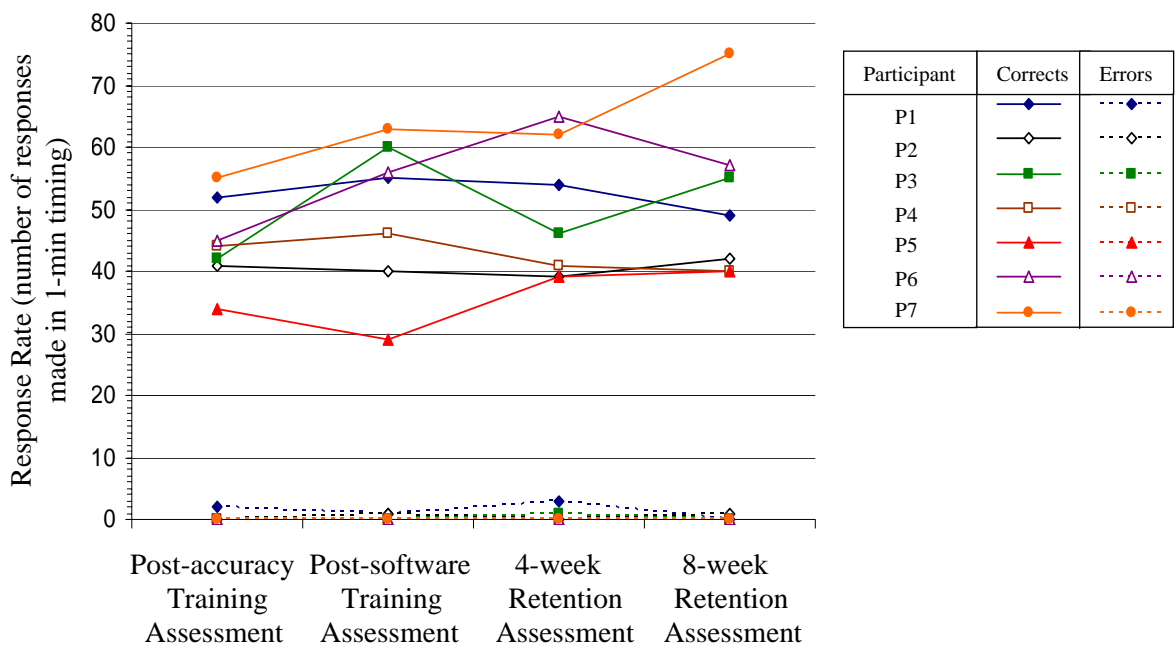


Figure 11. Response rate on 2 times table facts during a single 1-min timing after each training component and at retention assessments.

Figures 12 and 13 show changes in each participant's response rates for the Deck A facts (accuracy plus fluency) and Deck B facts (accuracy criterion alone), respectively. These figures show bitonic response rate functions where response rates for each participant on each deck of facts starts slow, at a similar level for each participant, and increases following the software training experimental phase.

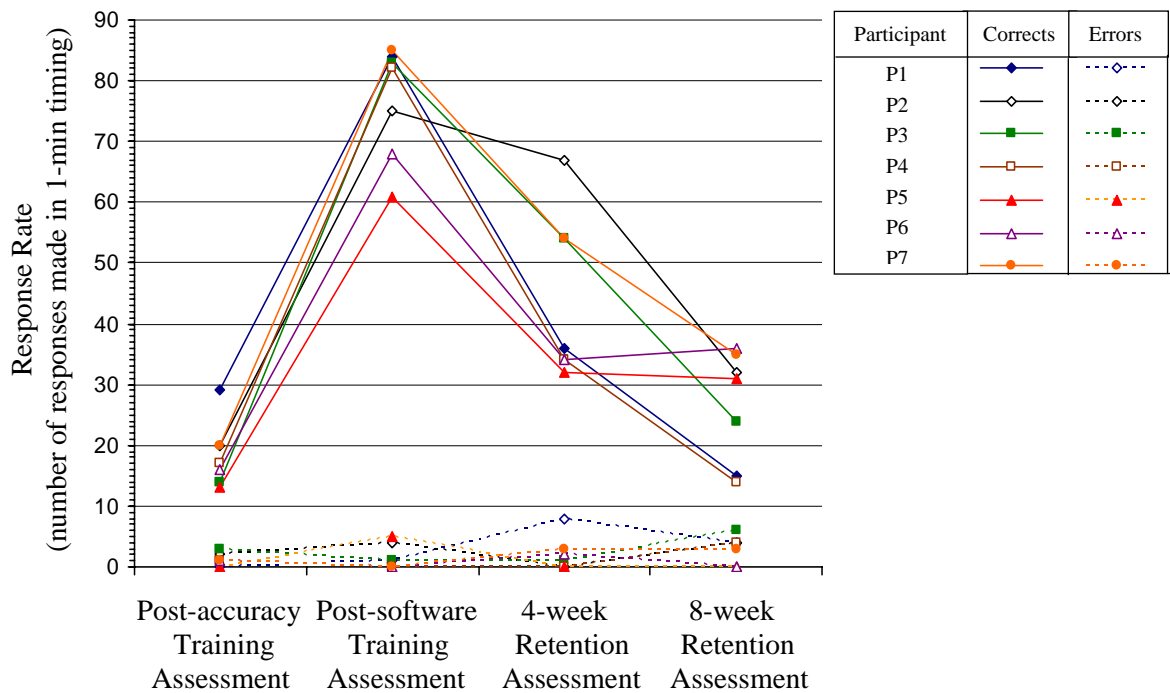


Figure 12. Response rate on Deck A facts (trained to accuracy plus fluency) during a single 1-min timing after each training component and at retention assessments.

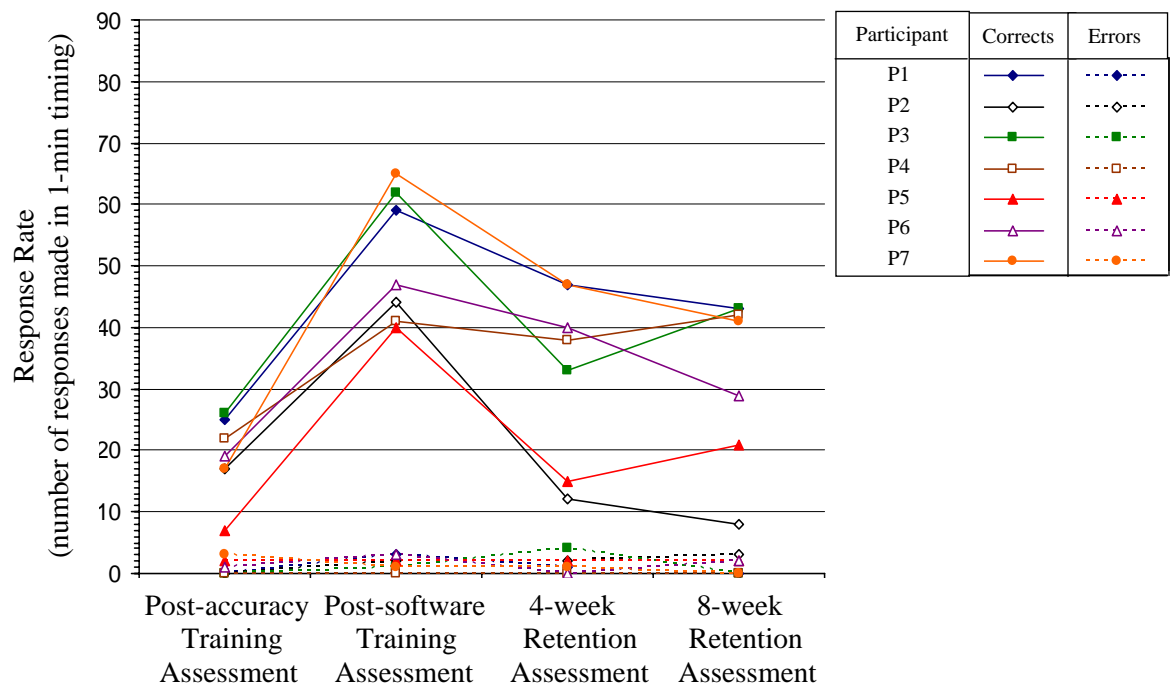


Figure 13. Response rate on Deck B facts (trained to an accuracy criterion only) during a single 1-min timing after each training component and at retention assessments.

At the post-accuracy training assessment, the individual response rates ranged from 13 (P5) to 29 (P1) correct responses per min for Deck A and from 7 (P5) to 26 (P3) responses per min for Deck B. Some participants responded faster to Deck A facts (P1, P2, P5 and P7) and some to Deck B facts (P3, P4 and P6). The rates of correct responses attained at the post-software training assessment were, for each participant, faster for Deck A facts (ranging from 61 to 85 correct responses per min) than for Deck B facts (ranging from 40 to 65 correct responses per min). Increases in response rates to Deck A facts, from the post-accuracy training assessment to the post-software training assessment, ranged from 290% (P1) to 592% (P3). For Deck B facts, these increases ranged from 186% (P4) to 571% (P5).

The participants' response rates to both Deck A and B facts decreased over the 4-week retention period for all participants. However, there is a trend for the response rates to decrease more steeply for Deck A facts (accuracy plus fluency). For all but one participant (the exception being P6), the response rates to Deck A facts continued to decrease over the 4 to 8-week retention period. P6 showed a small increase in rate between the 4-week and 8-week retention assessments. The rates of correct responses to Deck B facts (accuracy-only) decreased across each retention assessment for P1, P2, P6 and P7. The rate of correct responses decreased for the remaining participants, P3, P4, and P5 at the 4-week retention assessment but their response rates increased at the 8-week retention assessment. Notwithstanding this increase, the final rate for each of these participants was slower than the rate measured at the end of software training.

Errors During Timed and Un-timed Retention Assessments

Figures 5 to 13 show responses to a deck of facts as a whole. To enable comparison of correct and error responses to individual facts in each deck during the two retention assessments, the participants' responses to each fact are plotted in Figure 14 for P1, P2, P3 and P4 and Figure 15 for P5, P6 and P7. These figures show two stacked bar graphs for each participant, one for the 4-week timed retention assessment and below this, one for the 8-week timed retention assessment. The light grey bars represent correct responses and the dark grey bars represent error responses. At the top of each graph, a tick above a fact represents a correct response made during the single un-timed presentation of the fact at the retention assessment. A cross represents an error response.

The number of facts answered during the 1-min timed assessments was, for some participants, greater for Deck A and, for others, it was greater for Deck B, with no clear difference between the decks. There is no relationship between which deck a fact came from and whether there were errors in the un-timed or timed assessment trials. Consider, for example, errors during un-timed assessments as indicated by the ticks and crosses. P1, at both retention assessments, P3 at the 4-week retention assessment and P7 at the 8-week retention assessment, made more errors on Deck A facts than Deck B facts. Conversely, P5 at the 4-week and P2 and P3 at the 8-week retention assessment made more errors on Deck B facts. At all other assessment periods, the participants made the same number of errors on both decks. Specifically, P6 made one error on each deck at the 8-week retention assessment and all other participants made no errors on either deck.

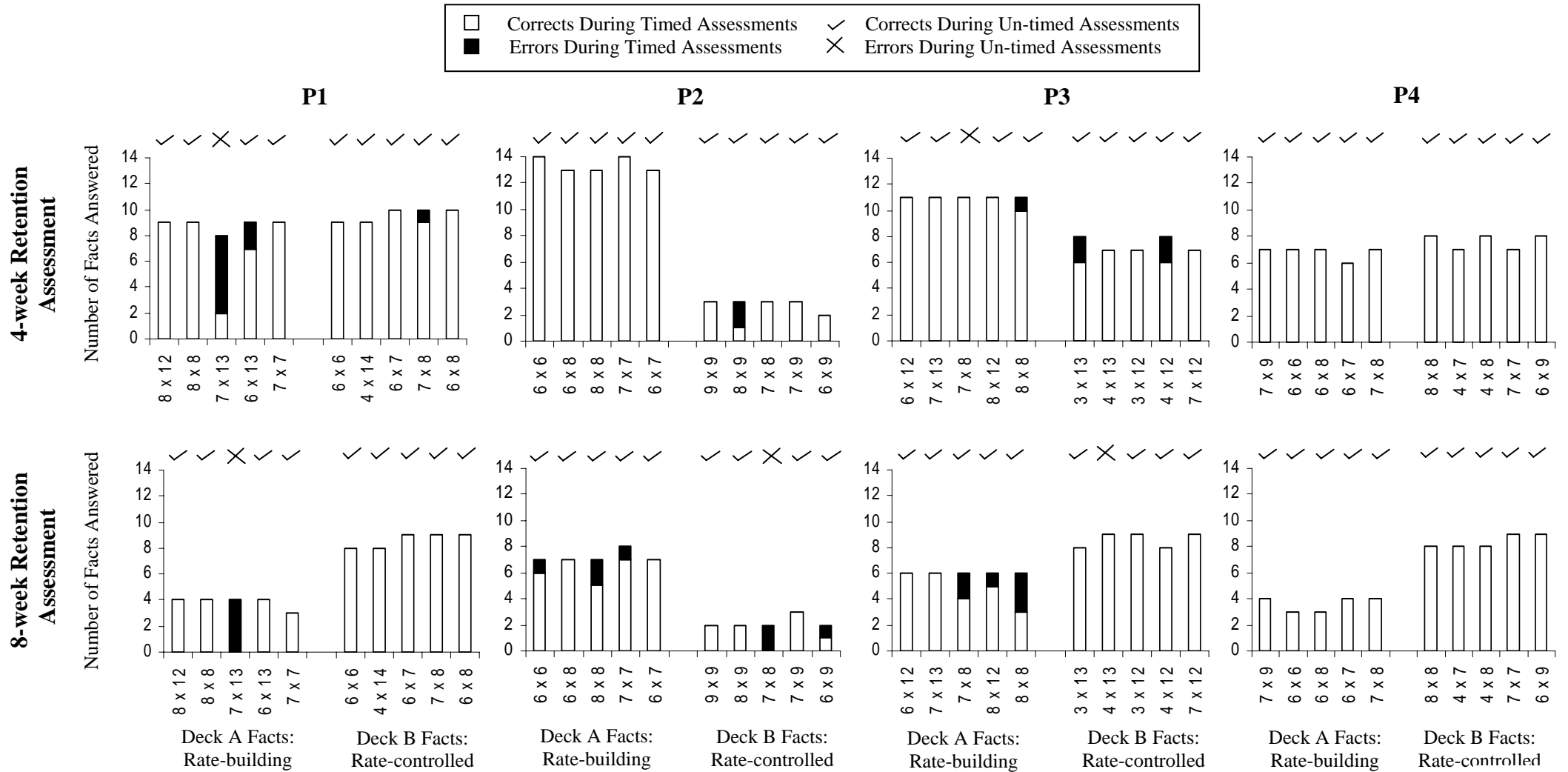


Figure 14. Number of correct and error responses made for each Deck A and Deck B fact answered during both un-timed and timed 4- and 8-week retention assessments for P1, P2, P3, and P4.

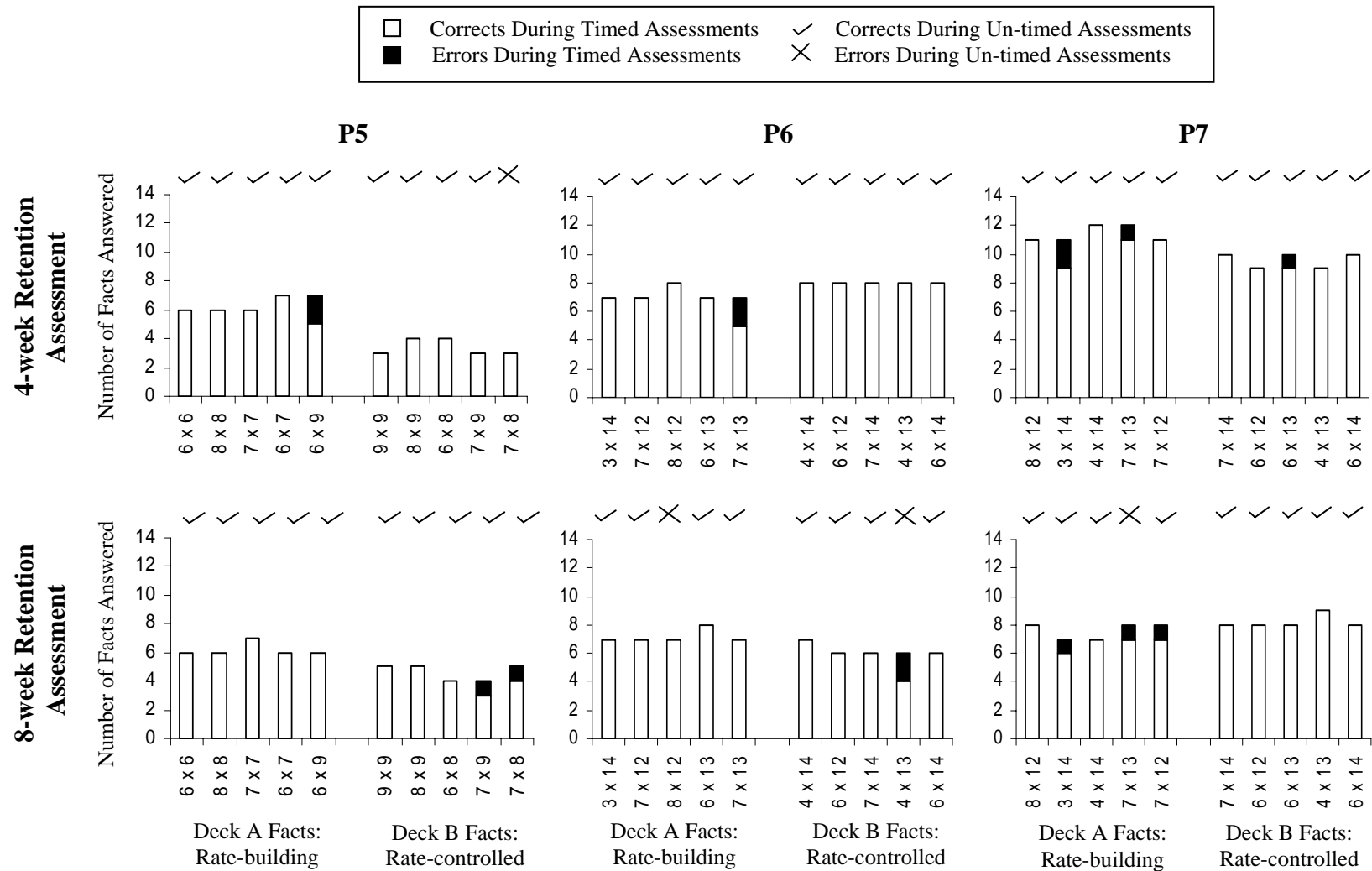


Figure 15. Number of correct and error responses made for each Deck A and Deck B fact answered during both un-timed and timed 4- and 8-week retention assessments for P5, P6 and P7.

During the timed 4-week retention assessment, shown as bars in Figures 14 and 15, P1, P5, P6 and P7 made more errors on Deck A facts and P2 and P3 made more errors on Deck B facts. At the 8-week retention assessment P1, P2, P3 and P7 made more errors on Deck A facts and P5 and P6 made more errors on Deck B facts. P4 made no errors on either deck of facts at either the 4-week or 8-week retention assessments.

In Figures 14 and 15, for each fact, the combined height of the respective grey and black bars indicates the number of times each fact was answered in total (either correctly or incorrectly) in a 1-min timing. During the 1-min timed 4-week retention assessments, P2, P3, P5 and P7 answered more Deck A facts (rate-building) than Deck B facts (rate-controlled). That is, they exhibited a faster response rate to Deck A facts. The difference between the number of Deck A and Deck B facts answered ranged from 19% to 380% (19% for P7, 49% for P3, 88% for P5 and 380% for P2). P1, P4 and P6 answered more Deck B facts than Deck A facts, exhibiting a faster response rate to Deck B facts. The difference between the number of Deck A and B facts answered was in the range of 11% to 12%. Note that these differences are smaller than the differences for participants who answered more Deck A facts than Deck B facts. The results for the 8-week timed assessments also showed no consistent difference between decks. P2, P5 and P6 answered more Deck A facts than Deck B facts. The differences between the number of Deck A and B facts answered were larger for P2 (230%) and smaller for P5 (35%) and P6 (16%). P1, P3, P4 and P7 answered more Deck B facts. The differences between the numbers of facts answered in each deck were larger for P1 (126%) and P4 (130%) and smaller for P3 (43%) and P7 (8%).

Results of the 1-min free-operant timing at each of the assessment stages after post-accuracy training were compared to the results for the same deck of facts at the post-accuracy training assessment, using a dependent t test. The effect sizes were calculated using Cohen's measure of effect size, d (Aron & Aron, 2003, p. 257). Both the number of correct responses and the number of error responses were compared. Similar comparisons were made between Deck A and Deck B at each assessment (post-accuracy, post-software, 4- and 8-week retention assessments). All dependent t tests were two-tailed. There were 6 degrees of freedom ($n = 7$) and the significance level chosen was $p < .05$. The dependent t test and effect size results are shown in Table 2. Results that were significant at the $p < .05$ level are marked with an asterix.

For the 2 times table facts, there was no significant difference between either correct or error response rates measured at the post-accuracy training assessment and those measured at any of the subsequent assessment stages (post-software training and the 4- and 8-week retention assessments). Deck A facts were answered correctly significantly faster post-software training and at the 4-week retention assessment than at the post-accuracy training assessment. Over these same assessments, there was no significant difference in the rate at which error responses were made. For Deck A facts, there was also no significant difference between the rate of correct or error responses at the post-accuracy training and the 8-week assessment.

Deck B facts were answered correctly significantly faster at post-software training, at the 4-week retention assessment and at the 8-week retention assessment than at the post-accuracy training assessment. The effect sizes for all the t tests that were significant would be classified as large by Cohen (Aron & Aron, 2003, p. 257). Comparisons of response rates for Deck A facts and Deck B facts for the post-software training assessment found that correct responses for Deck A were significantly faster

than for Deck B, with a large effect size. For all other assessments, the rate of both correct and error responses on Deck A and Deck B were not significantly different.

Table 2. *Results for the comparison of the differences between the number of both correct and error responses made in the 1-min timing at each assessment stage.*

Assessment Comparisons		Calculated t Value (<i>t</i>)		Effect Size (<i>d</i>)	
		Correct	Error	Correct	Error
2 Times Table Facts	Post-accuracy to Post-software	1.750	0.616	0.66	0.25
	Post-accuracy to 4-week Retention	1.630	1.549	0.62	0.59
	Post-accuracy to 8-week Retention	1.919	-0.162	0.73	-0.06
Deck A Facts	Post-accuracy to Post-software	19.640 *	0.305	7.42	0.12
	Post-accuracy to 4-week Retention	4.754 *	0.651	1.80	0.25
	Post-accuracy to 8-week Retention	1.773	2.772 *	0.67	1.05
Deck B Facts	Post-accuracy to Post-software	9.441 *	0.650	3.57	0.27
	Post-accuracy to 4-week Retention	3.203 *	0.760	1.21	0.29
	Post-accuracy to 8-week Retention	3.281 *	0.213	1.24	0.08
Deck A to Deck B	Deck A post-accuracy to Deck B post-accuracy	-0.236	0.400	-0.09	0.15
	Deck A post-software to Deck B post-software	8.766 *	-0.179	3.31	-0.07
	Deck A 4-week to Deck B 4-week Retention	1.317	0.437	0.50	0.17
	Deck A 8-week to Deck B 8-week Retention	-0.746	1.702	-0.28	0.64

* $p < .05$

Changes in the Frequency of Response Latencies Over Timed Assessments

The response latency for each fact presented during 1-min timings was recorded by the software. Changes in the distribution of response latencies to all facts, whether the facts were answered correctly or incorrectly, were compared by plotting the percent frequency of each response latency for individual participants as histograms. Percent frequency was calculated by allocating the response latency for each fact answered in a 1-min timing to a response latency range bin. For example, response latencies falling between 0 and 500 ms fell in the '500' bin and response latencies above 500 and up to 1000 ms fell in the '1000' bin. The number of latencies falling in each bin range was totaled. These totals were expressed as the percentage of the total facts answered in the 1-min timing. Separate graphs are presented for each deck of facts (Decks A and B and the deck of 2 times table facts) at each assessment stage (post-accuracy training; post-software training; and at the 4-week and 8-week retention assessments).

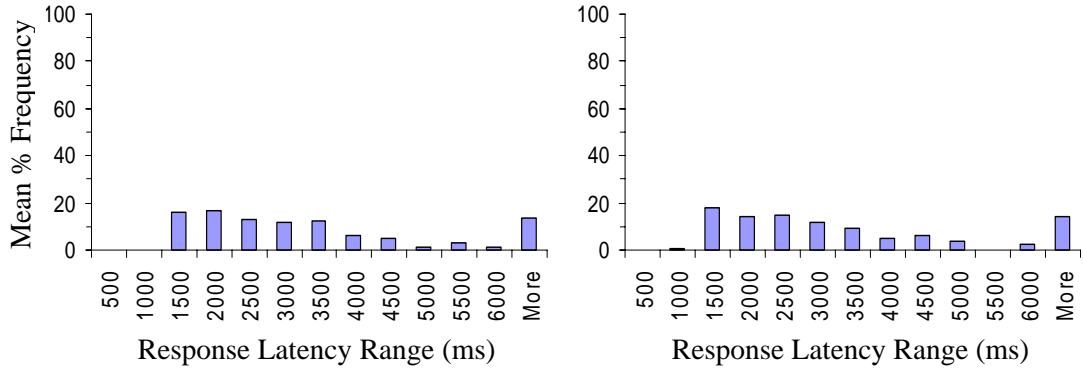
For the sake of presenting a concise summary, the mean of the participants' percent frequencies of response latencies are presented in Figures 16 and 17. These means were calculated by averaging all participants' percent frequencies for each response latency range. The mean provides a useful summary of the main features of the individual data. For detailed analysis, readers are referred to the graphs for individual participants in Appendix H.

Response latencies (ms) are presented as ranges on the x-axis. For example, response latencies falling between 0 and 500 ms are plotted in the '500' range and response latencies above 500 and up to 1000 are plotted in the '1000' range. The mean percent frequency over all participants is plotted on the y-axis.

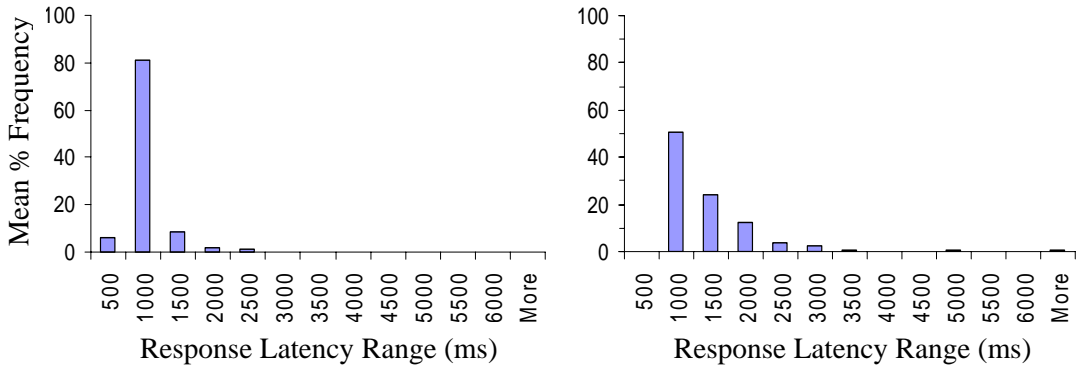
Deck A Facts

Deck B Facts

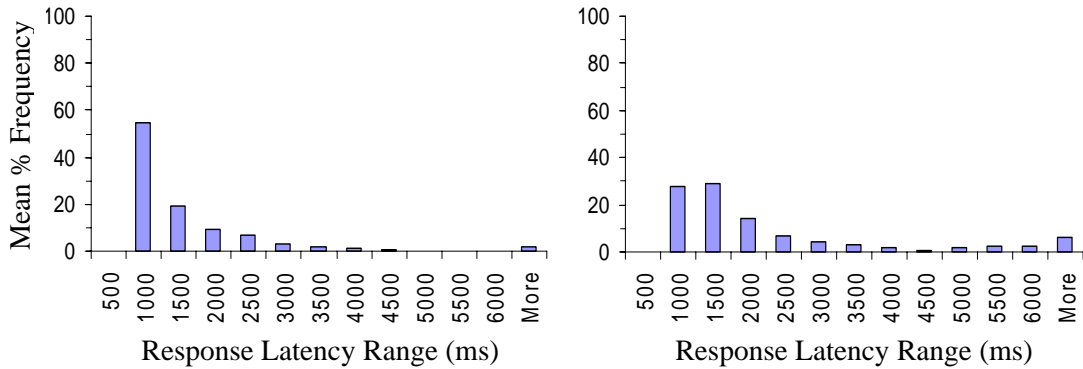
Post-accuracy Training Assessment



Post-software Training Assessment



4-Week Retention Assessment



8-Week Retention Assessment

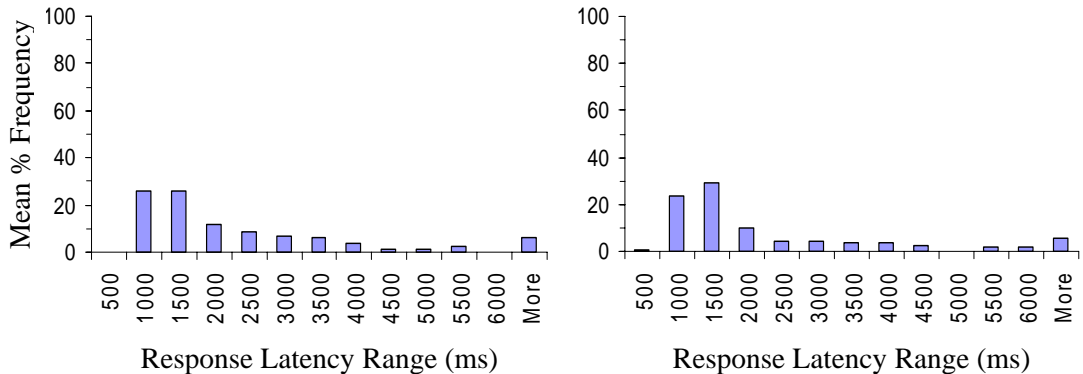


Figure 16. The mean of all participants' percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages.

2 Times Table Facts

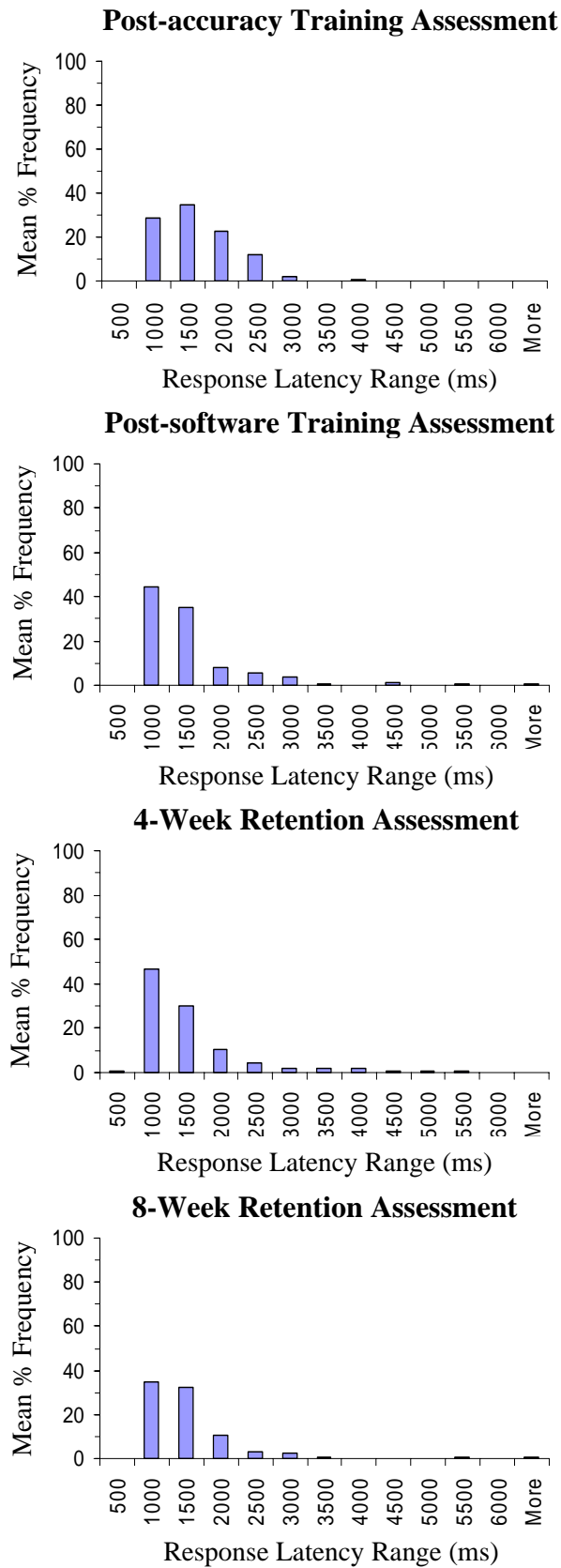


Figure 17. The mean of all participants' percent frequencies of response latencies for the deck of 2 times table facts over the four assessment stages.

The top pair of graphs in Figure 16 show mean percent frequencies of response latencies at the end of accuracy training for both Deck A and Deck B facts. For example, the mean of all participants' percent frequencies in the 1500-ms bin is 16% for Deck A facts and 17% for Deck B facts. No latencies fell at or below 1000 ms for Deck A facts or at or below 500 ms for Deck B facts. Less than 1% of the latencies for Deck B facts fell in the 500- to 1000-ms range. For both Deck A and Deck B facts, the mean percent frequency of response latencies is spread widely from the 1000-ms bin to the more than 6000-ms bin, with higher mean percent frequencies at lower response latencies within this range.

After using the software to build response rates on Deck A facts to fluency, response latencies to both Deck A and Deck B facts shortened and were distributed over a narrower range. This is shown in the second pair of graphs in Figure 16. This change in distribution is less marked for Deck B facts which, during software training, were practiced an equivalent number of times without rate-building. The highest percentage of frequencies falls in the 1000-ms bin for both Deck A and Deck B facts. For both Deck A and Deck B facts the mean percent frequency shows an approximate exponential decay over increasing response latencies. That is, most responses (including both correct and error responses) were made within 1000 ms, fewer took over 1000 ms but less than 1500 ms and successively fewer still as the response latency interval increases. There is a trend for the mean percent frequencies to be spread over a wider range for Deck B facts than for Deck A facts.

After four weeks of no practice on either deck of facts the spread of the mean percent frequency of response latencies increased for both decks of facts. Response latencies for Deck B facts were spread over a wider range than those measured for Deck A facts. This is shown in the third pair of graphs in Figure 16. After a further four weeks

of no practice, the spread of the mean percent frequencies of the response latencies had further increased for both decks of facts and was similar to the wide spread seen at the end of accuracy training. This is shown in the bottom pair of graphs in Figure 16.

Figure 17 shows the mean of the participants' percent frequencies of response latencies for the deck of 2 times table facts. At each assessment, a similar pattern of response latencies is apparent. That is, most response latencies (including both correct and error responses) are short (< 1500 ms) with the mean percent frequency showing approximate exponential decay as response latencies get longer.

10th Percentile Response Latencies During Rate-controlled Trials

Figure 18 shows 10th percentile latencies for each participant. Consecutive sessions are plotted on the x-axes. The y-axes show the 10th percentile point of the response latencies (including both correct and error responses) for that session. Minimum response latencies give an indication of the participants' ability to respond. The 10th percentile is indicative of the participants' minimum achievable response latency, despite the presence of some outliers in the data. These outliers result from occasional "key bounce" events with the mouse key. These key bounce events caused some inaccurate, short latencies to be recorded by the software. These were few in number and did not affect the 10th percentile.

Participants P1, P3, P5, P6 and P7 show an overall downward trend in 10th percentile response latency. These participants' results include some periods of three or fewer sessions where latencies increased before resuming the downward trend. For these participants, the 10th percentile of response latencies at the start of software training range from 981 ms (P1) to 2103 ms (P5) and decrease to between 461 ms (P7) and 951 ms (P3).

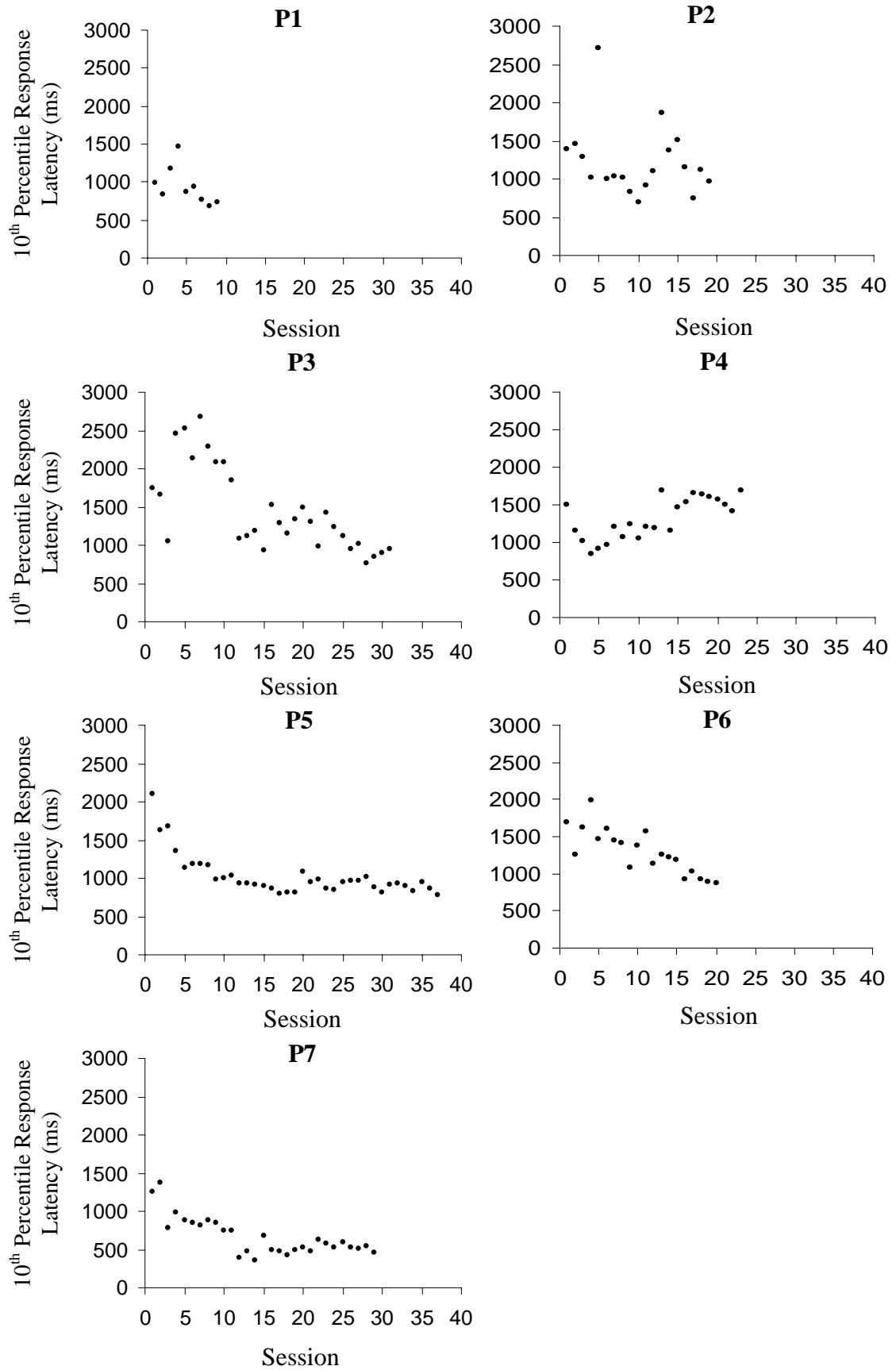


Figure 18. The 10th percentile response latencies for each session taken from all trials in that session, during rate-controlled software training on Deck B facts.

The remaining participants (P2 and P4) also show a downward trend for most consecutive sessions with periods of downward trend separated by periods of increasing response latencies for three or fewer sessions. However, for these two participants, the sum of the increases in 10th percentile response latencies results in a final 10th percentile response latency that is larger than the other participants'. The initial 10th percentile response latencies were 1382 ms (P2) and 1503 ms (P4). These latencies had decreased to 961 ms for P2 and increased to 1682 ms for P4, at the end of software training.

Participants' Practice Preferences

At the end of the final retention assessment, each participant was asked which practice method they "liked doing best": fast trials (timed rate-building trials) or slow trials (rate-controlled trials). P5 and P6 reported a preference for rate-controlled trials. All other participants reported a preference for rate-building trials. Each participant was then asked to give reasons for their preference. P5 reported that "I always get really worried when I'm going to do a fast one" and P6 that "I find it really stressful going fast." The other participants, who reported a preference for rate-building trials, gave answers that included the following: "It gets it done quicker", "I like going fast", "I like trying to beat my score", and "The slow ones take ages." However, it should be noted that the following participant behaviours were observed by the experimenter prior to timed trials. Participants typically wriggled around in their chair, prepared to start a trial and then stopped and re-prepared, and made statements like: "I feel scared/nervous", "I'm worried I don't think I can beat it/go faster", "I don't want to do this" or "I have sweaty palms."

Discussion

The primary purpose of this study was to determine whether rate-building to an established fluency performance standard provided better 4- and 8-week retention than the same amount of practice conducted in a rate-controlled manner. The study extended previous research by controlling for amount of practice when comparing non-rate-building practice to training to fluency. The widely accepted fluency performance standard of 70-100 correct responses per 1-min trial (Binder et al., 2002) was used. The total number of reinforcers attained over all sessions for each deck of facts was also equalised. Each participant improved the rate at which they stated correct answers to all the multiplication facts that they learned, regardless of the method of practice. Fluency training led to significantly faster rates of correct responding, immediately after software training, than the same amount of rate-controlled practice. Participants completed an average of 54% more correct responses per min to Deck A (rate-built) facts than to Deck B (rate-controlled) facts.

Many authors, including Berquam, (1981), Binder (1996), Johnson and Layng (1996) and Olander et al. (1986), suggest that fluency training leads to faster response rates or better accuracy on retention tests than training to accuracy alone. This suggestion, however, was based on studies that did not control for practice. This study controlled for practice and does not support this suggestion. Across the 4- and 8-week retention periods, rate-building to a fluency performance standard did not lead to consistently faster correct response rates or lower error rates during timed 1-min retention assessments. At retention assessments, some participants performed more accurately on the facts rate-built to a fluency performance standard and some performed more accurately on the facts practiced the same

number of times in a rate-controlled manner. There was no clear bias for either rate-building (Deck A) or rate-controlled (Deck B) practice.

The results reported here are consistent with the single earlier study that controlled for practice and trained to a fluency performance standard (Shirley & Pennypacker, 1994). In Shirley and Pennypacker's study, participants practiced two lists of 10 spelling words. Lists were yoked so that each list was practiced the same number of times. Practice on both lists continued until both lists were completed with 100% accuracy and one list was completed at a fluent rate. Shirley and Pennypacker did not limit response rate for the list of words not learned to the fluency performance standard. Nor did they report the response rate for this list of words when the other list reached fluency and practice ended. As a consequence, their results, unlike the present results, could have been explained by a lack of difference in the final response rate of the two skills compared. Unlike Shirley and Pennypacker (1994), the current study included both rate-building and rate-controlled training while controlling for practice. This was achieved using a computer. This study also included measurement of response rate on both the rate-controlled and rate-built facts at the end of fluency training to ensure that the rate-controlled facts had not reached the fluency performance standard. How practice was controlled in this study and the reasons for using a computer to build response rates are discussed next.

In the current study, practice occurred within experimental sessions only. This provided a means of controlling the amount of practice and of ensuring the consistent quality of that practice. Child participants in this study were educated at home and did not attend a regular school. Parents agreed not to include multiplication in school work for the duration of the study, including the follow-up period. They were asked to remind their child not to practice multiplication outside of experimental sessions. The purpose of the study

and the importance of not practicing between sessions were explained to all participants. Participants were given no record of facts learned and they agreed not to practice or rehearse facts or answers between sessions. Child participants were also told that if they thought of the facts or their answers, they should distract themselves by ‘thinking of,’ or doing, something else, for example singing a song or reciting a rhyme. In addition, children received a token at the start of each session if they reported that they had not practiced, including rehearsing facts and/or answers, since the last session. While tokens were provided based on reports rather than direct observation, parents confirmed that their child had not practiced between sessions. For example, all parents reported that their children were quick to identify and state that they “Couldn’t look at that page since it had multiplication sums on it” if any multiplication inadvertently appeared in school work.

Throughout training, every practice of each fact was recorded. There were two types of practice: practice before a fact was consistently answered accurately and practice after the last error was made (overpractice). Overpractice is practice beyond 100% accuracy. During accuracy training, participants received a greater number of practice exposures to facts presented early in training than to those introduced in later sessions. The focus of this study involved controlling for practice, so this difference needed to be equalised. The equalisation phase provided extra practice so that, by the end of this phase, each participant had practiced each of their 10 facts the same number of times.

While each fact had been practiced the same number of times by the end of equalisation, some facts had received more overpractice. To control for overpractice between the two decks of facts (Deck A and Deck B), facts were allocated between decks so as to minimise the difference between the total number of overpractice trials in each deck. With all participants, there were small differences between the decks (ranging from

3 to 12) in the total number of overpractice trials. During subsequent overpractice, yoking of the decks ensured that each participant received the same total number of practices on Deck A facts as they received on Deck B facts. However, because the software used in this study randomised the order of the facts within each presentation of the five facts, small differences in the number of practices of any individual fact within a deck can and did occur. This is the case when the number of facts answered within a computer trial was not a multiple of five. For example, if seven facts were answered, the first five facts presented may have been 6×8 , 8×9 , 7×7 , 4×6 and 9×9 (in that order). The facts would then have been shuffled and the next two facts answered may have been 8×9 and 9×9 . The facts 8×9 and 9×9 were each practiced twice in this trial and the remaining facts were practiced only once. For each participant, accumulating these differences over all rate-building trials resulted in small differences in the number of practice trials of each fact. These differences never amounted to more than 0.3% of total practices (Figure 3).

Computer-based presentation of facts during Phase 3 (rate-building and rate-controlled training) was chosen because it provided standardised presentation of facts and answers, randomisation of sequences of multiplication facts and accurate timing of sessions. The computer also provided an interface that both encouraged the development of fluent performance and provided a means of effectively limiting response rate. The software also allowed for the automatic and accurate collection of data such as response latencies.

Both Binder (1996) and Wolery, Bailey and Sugai (1988) suggest that computers may be used as fluency-building tools. However, they caution that software must not limit response rates. Most fluency research employing computers has used a see-fact/type-answer learning channel (Luyben et al., 2003; McDade, 1998; Peladeau et al., 2003). When

using this learning channel, typing-rate imposes a ceiling on response rate (Binder, 1996). For example, Luyben et al. (2003) studied the effects of computer-aided instruction on academic performance using the ThinkFast™ fluency-building software. They found that the maximum possible response rates participants achieved were well below published fluency performance standards because of the constraints imposed by typing speed. ThinkFast™ also allows users to choose to say, rather than type, the answer and click a key as the answer is said. However, even in this mode, ThinkFast™ limits response rates by inserting delays between card exposures and also between when a response is made and when the correct answer is displayed. New cards are not presented until the correct answer has been displayed. This is not a free-operant situation. Binder (personal communication, April 20, 2005) warns that “If our procedures prevent us from measuring above a certain frequency of response, we really can't use that measure to predict outcomes.”

The software developed for this study was designed to support free-operant responding. When running on the computer used for this study, the software can present facts at rates up to 1200 facts per min. As a consequence, the computer does not prevent rate-building to the established fluency performance standard (70-100 correct responses per min). The software uses a see-fact/say-answer learning channel. Using this learning channel did, however, require the experimenter to detect and count error responses. This introduced the possibility of human error in data collection. However, the interobserver reliability checks indicated that data recording was at least 98% accurate.

The use of computer presentation proved to be practical, precise and effective. Even though there was a wide range of ages, the participants found the software easy to use and they required very little training to be able to use it.

For all participants, rate-building to a fluency criterion took more practice trials than training to accuracy. The bounds are shown by Participants P1 and P3. P1 required 2.7 times, and P3 11 times, more practice to reach the fluency criterion than it took them to learn all facts to accuracy and to reach equalisation. This suggests that meeting a fluency performance standard requires much practice beyond accuracy. However, it should be noted that, for all participants, this was their first exposure to rate-building. It may be that it takes more practice to reach fluency on the first skill that is rate-built, but that 'learning to go fast' is a component skill. Johnson (personal communication, November 22, 2005) of Morningside Academy is of this view; he states that “kids learn to learn how to accelerate.” If learning to respond quickly is a component skill, learning to respond fluently to one set of facts may improve a common component and thereby influence the rate at which participants respond to other sets of facts.

In this study, the inclusion of the 2 times table facts provided a control for the effects of improvements in any common component skills within the experimental method that might exist. The deck of 2 times table facts was the same for all participants. Each of the five facts included in this set was answered correctly by all participants at the start of the experiment. The participants did not practice these facts, either during experimental sessions or outside of sessions with the exception of the inherent practice associated with the assessments. The response rates for the 2 times table facts were slower than the rate achieved at the end of software training for Deck A (rate-built) facts. This indicates that response rate for the 2 times table facts was not limited by a ceiling and that this rate could have improved with practice. There was no systematic trend in the rate of either correct or error responses per min across assessments for the 2 times table facts for any participant.

This suggests that learning to ‘go fast’ to the facts learned to fluency (Deck A facts) did not impact the response rate to other maths facts (2 times table and Deck B facts).

This study found no consistent relationship between the number of practice trials required to reach accuracy and the number of overpractice trials required to reach the fluency performance standard. For example, P3 and P6 required a similar number of practice trials (67 and 58 respectively) to reach accuracy. However, P3 required approximately twice as many subsequent overpractice trials as P6 to reach the fluency performance standard for Deck A facts.

Informally, we think of retention as remembering a response after a period of non-practice. Other researchers have studied the effect of different response rates on retention of accuracy. Some research indicates that attaining faster end-of-training response rates is not correlated (Haughton, 1976), or is not consistently correlated (Shirley & Pennypacker, 1994), with more accurate performance at retention assessments. The contrary findings of Bucklin et al. (2000), Olander et al. (1986) and Ormrod and Spivey (1990), who found superior accuracy on retention tests for skills learned to higher response rates, could be explained by their lack of control for practice. Participants in their studies received more practice on the skills learned to higher response rates. Peladeau et al. (2003) address this confound in their study. They compared the retention of skills practiced to accuracy-only, with the retention of skills overpracticed the same number of times in both rate-building and non-rate-building formats. They found that overpractice improved retention compared with practice that ended when accuracy was attained, but the same amount of overpractice in different formats produced indistinguishable retention.

The current study extends the work reported in Peladeau et al. (2003) by training to an established fluency performance standard. The un-timed retention tests using flashcards

provided an opportunity to assess whether the answers to the facts learned were accurately remembered after a period of non-practice. Training to a fluency performance standard did not produce superior performance when performance was assessed by accuracy-only (percent correct) measures. The retention data shows that neither deck produced fewer errors across all participants at any of the retention assessments. This result suggests that, in terms of accuracy alone, there is no clear benefit to building response rates to a fluency performance standard when compared with the same amount of rate-controlled practice, if no maintenance occurs.

One possible confound in this study is that the change in presentation mode during practice and assessments may have altered the stimulus control over correct responses. While practice during the first two experimental phases involved presentation of facts on flashcards, in the ensuing computer-based practice (Phase 3), the stimuli presented for the same facts looked different. That is, facts were presented on a screen rather than a flashcard and appeared in a green highlighted box. At assessments accuracy was assessed using flashcards. The accuracy of the response to a fact during un-timed flashcard assessments was not necessarily the same as when the fact was presented again in the same session during subsequent computer-based timed retention assessments. For example, at the 4-week retention assessment, the fact 6 x 8 may have been answered incorrectly in the un-timed test but then answered correctly on all presentations in the timed assessment. The reverse also occurred, where a fact was answered correctly in the un-timed assessment but incorrectly for one or more presentations in the timed assessment. Some researchers believe that training to a fluency performance standard enhances generalisation (Bonser, 2002; Young, West, Howard, & Whitney, 1986). The inconsistencies in accuracy between the different

modes of presentation may indicate that learning was not generalised across different stimulus presentation modes.

Comparing the response rate for facts from each of the three decks of facts (2 times table facts, Deck A facts and Deck B facts) at each assessment period (post-accuracy training, post-software training and the 4- and 8-week retention assessments) reveals some consistencies. At the end of accuracy training, and prior to any rate-building practice, the participants' rate of correct responses on both Deck A and Deck B facts were similar. At the post-software training assessment, participants responded faster on Deck A facts than on Deck B facts. This is expected given that practice during software training had been specifically directed at increasing response rates to Deck A facts. Notably, response rates to Deck B facts also increased by between 186% and 571%. For P3 and P7, the response rate to Deck B facts at the end of software training was 62 and 65 correct responses with 1 and 2 errors per min respectively, close to the fluency performance standard of 70-100 correct responses per minute with no more than two errors. The control provided by the 2 times table facts (as previously discussed) makes it clear that this increase in response rate to Deck B facts is not due to an improved general ability to answer facts quickly. This suggests that response rates increase with overpractice which is consistent with Binder (1996). The response rates of some participants were close to the fluency performance standard on the rate-controlled deck. This may raise the concern that the results for these participants, for the rate-controlled deck, would be similar to their results for the rate-built deck. This is unlikely to be the case for several reasons. First, their rate for Deck A facts (83 and 85 correct responses per min for P3 and P7 respectively) well exceeded the minimum rate required to meet the fluency performance standard. That is, there was a clear difference in their performance on the two decks of facts. More importantly, fluency

researchers suggest that reaching a fluency performance standard is a binary effect; they suggest that being close to, but under, the range expressed in the standard will not optimise the beneficial outcomes of fluency (Binder, 1996). In spite of their high rate of correct responses on Deck B, the retention results of these two participants (P3 and P7) are consistent with those of the other participants, as discussed below.

At the 4-week retention assessment, 3 out of 7 participants answered Deck A facts correctly faster than they answered Deck B facts correctly. The remaining 4 participants answered Deck B facts faster. Where Deck B facts were answered faster than Deck A facts, the differences between the rates at which the two decks of facts were answered was smaller (in the range of 11-12%) than for the participants who answered Deck A facts faster (in the range of 19-380%). At the 8-week retention assessment, 3 out of 7 participants answered Deck A facts correctly faster than Deck B facts and 4 answered Deck B facts correctly faster. The response rates measured for both Deck A and Deck B facts at the 8-week retention assessment do not differ significantly. These results raise doubt about the need for practice to be at high rate. They also suggest that reaching a fluency performance standard does not produce better retention. This is in contrast with previous research (Bucklin et al., 2000; Bullara et al., 1993; Young et al., 1985). The most probable explanation for the difference between this study and the aforementioned ones is the effective removal of practice between training and retention assessments. This was possible because the participants were home-schooled. Most other studies took place in a regular school setting or are based on data collected in a school setting. In such cases, there is normally ongoing learning and application of the component skills which provides maintenance of the skills learned.

The only assessment stage at which there was a significant difference between the rates of correct responses between Deck A and Deck B was immediately post-software training. There were no significant differences between the rates of error responses on the decks at any assessment stage. Response rates, for both decks, fell over the retention assessment phase. After 4 weeks of no practice, the rates of correct responses for both decks were significantly faster than at the post-accuracy training assessment but there was no significant difference in the error rate. However, after 8 weeks of no practice, the rates of correct responses on Deck A facts were not significantly faster than the rates achieved at the post-accuracy training assessment. The rate of error responses was significantly faster. The rates of correct responses on Deck B facts at the 8-week assessment were significantly faster than the rate achieved at the post-accuracy training assessment, with a large effect size. There was no significant difference in the rate of error responses.

The results for the rates of both correct and error responses suggest that rate-building to a fluency performance standard does not enhance retention of rate or accuracy, compared with the same number of rate-controlled practices. In contrast, rate-controlled learning delivered a faster rate of correct responses and a slower rate of error responses after 8 weeks of no practice when compared with post-accuracy training. An explanation for better retention after 8 weeks for rate-controlled learning may be that the extra time taken for each presentation of each fact enhances learning. Alternatively, there may be a negative effect associated with rate-building practice.

Reviewing the retention results of the two participants with high response rates on the rate-controlled deck at the end of software training, indicates that their results are consistent with those of the other participants. Like the other participants, these participants showed no clear difference in terms of which deck was answered faster or with higher accuracy.

In the un-timed accuracy test, P3 made fewer errors on Deck A at the 4-week retention assessment and on Deck B at the 8-week retention assessment. P7 only had one error which was made on Deck A at the 8-week retention assessment. At the timed retention assessments, both participants were faster on Deck A facts at the 4-week retention assessment and Deck B facts at the 8-week retention assessment. P3 made fewer errors on Deck A facts at the 4-week retention assessment and fewer on Deck B at the 8-week retention assessment. P7 made fewer errors on Deck B facts at both the 4- and 8-week retention assessments. The difference in performance between the two decks of facts for P3 and P7 was compared to the difference for the other participants by considering their rank order. P3 had the 4th smallest difference at the 4-week retention assessment. At the 8-week assessment, P3 ranked 5th. P7 ranked 3rd at the 4-week retention assessment and 1st at the 8-week retention assessment. Like the results for the other participants, P3 and P7 showed no systematic difference in accuracy or response rate across the retention assessments.

Bucklin et al. (2000) suggest the use of relative-loss measures, rather than comparisons of absolute response rates, to assess changes in response rate and accuracy over periods of non-practice. They suggest this comparison because relative-loss measures make the differences in post-training response rates clear. For 6 participants, the relative response rates decreased more for the rate-built facts than for the rate-controlled facts between the post-software training assessment and the 4-week retention assessment and between the 4-week retention assessment and the 8-week retention assessment. This was true for the first four-week retention period for the remaining participant. Thus, when response rates were built to a fluency performance standard, there was a greater decrease in relative response rate over both retention periods than when rate-controlled practice was

used. This result should be understood in the context of the faster response rate achieved for rate-built facts at the post-software training assessment than for rate-controlled facts.

If education is to be efficient in producing skills, teachers need to be able to measure changes in performance with practice beyond accuracy. Measurement allows educators to identify effective teaching strategies and to maximize students' progress by determining when to end practice. Kruger (1929) found that the number of practice trials is not a useful measure for these ends. This study has found that, practice to a fluency performance standard is not a good predictor of retention after 8-weeks of no practice. Further, whether retention is assessed by accuracy or response rate, either as an absolute rate or as relative loss, retention is not enhanced by rate-building to a fluency performance standard over the same amount of rate-controlled practice.

Binder (1996) reviewed the use of time-based measures of human learning outside the precision teaching and operant conditioning research areas. He found that, where instrumentation was available, overlearning researchers have used latency measures to assess the progressive effects of practice beyond 100% accuracy. These researchers found that latencies become progressively shorter with practice (Binder, 1996). The results of this study provide support for this finding. While the focus of this study was the effect of practice and response rate on retention, not the effect of response latency on retention, response latencies were collected.

For free-operant responding no new insights are gained from response latency because response latency and response rate measure the same variable. With rate-built facts presented in a free-operant manner, response latency is the reciprocal of response rate. All timed assessments in this study (post-accuracy training, post-software training and 4- and

8-week retention assessments) allowed free-operant responding. Therefore, the discussion to this point with respect to response rates also applies to response latencies.

When computer-based rate-controlled practice is used, rate is not a variable. However, in this study, participants could respond to a fact at any time they wished after it was presented. The software measured and recorded this response latency. Participants then waited for the next fact to be presented. Figure 16 shows the mean of all participants' percent frequencies of response latencies for Deck A and Deck B facts over the four assessment stages. The overall trend for most participants was for response latency to shorten with successive overlearning trials, whether trials were presented in a rate-building or rate-controlled format. For both decks of facts, the progression towards shorter response latencies is visible in the first two rows of histograms in Figure 18, which summarise the results for the post-accuracy training assessment and the post-software training assessment. The figure shows a reduced range and increased frequency of shorter response latencies. This occurred in rate-controlled trials even though accuracy was reinforced rather than reductions in response latency. For both rate-built and rate-controlled facts, response latency then lengthened over the 4- and 8-week retention periods.

Many responses in the rate-controlled trials were not at, or near, the minimum response latency. In this study, consequences may have interfered with a natural tendency for response latencies to decrease with practice. The consequences in effect during rate-controlled trials reinforced accurate, rather than short latency, responding.

Anecdotally, most participants frequently responded slowly on rate-controlled trials, often stating an answer and then restating the answer before pressing the mouse key. The single adult participant, P7, did not receive tokens for accuracy. This participant typically responded quickly to all facts, both rate-built and rate-controlled, and reported that he

wanted to finish all trials as quickly as possible. For this participant, reinforcement may have been completing the task.

The 10th percentile response latency (Figure 18) provides an indication of the participant's ability to respond quickly. Using the 10th percentile, rather than the minimum, removes the effect of some outliers in the data caused by key-bounce events. The overall trend for most participants was for 10th percentile response latencies to shorten as practice increased. However, for most participants there were periods where 10th percentile response latencies became longer for a time, before resuming a downward trend (Figure 18). For one participant (P4), the increase was sufficiently large to result in a final 10th percentile response latency that was longer than the latency attained at the beginning of software training. Anecdotally, P4 appeared consistently cautious when giving answers to rate-controlled facts. P4 stated that he "didn't want to make a mistake and not get a token." This possibility is also supported by P4 making only 5 errors in over 400 practices; fewer errors than any other participant.

Response latency provides a metric allowing changes in performance beyond 100% accuracy to be measured. There may be a relationship between response latency and retention. Future research comparing retention of skills learned to different ranges of response latencies may elucidate this relationship and determine 'performance standards' for latencies. This research, unlike the present study, should apply consequences appropriate to comparing the effects of response latencies.

Like the present study, Shirley and Pennypacker (1994) were also unable to conclude that rate-building provides superior retention in terms of rate and accuracy compared with the same amount of non-rate-building practice. There are, however, several differences between the current study and that of Shirley and Pennypacker (1994). As discussed earlier,

those authors did not limit the response rate of the list of words not learned to the fluency performance standard. They also did not report the response rate for this list of words when the other list reached fluency and practice ended. Shirley and Pennypacker (1994) note that requiring the 2 participants in their study to meet the fluency performance standard on only a single occasion may have limited retention. These issues were addressed in the current study which involved more participants ($n = 7$). Response rate was restricted for facts not learned to a fluency performance standard and correct response rates were measured at the end of training. This ensured that fluency was not reached on the rate-controlled facts. Further, the fluency performance standard included a requirement that participants meet the standard for three consecutive sessions before practice ended. Even with these more stringent requirements, this study did not find better retention for facts learned to a fluency performance standard.

As a consequence of completing this study, some areas of further research have been identified. Although the findings of this study appear clear, as with all studies involving a small number of participants, repeating the study with more participants would provide further external validity for the findings. This would be particularly valuable given that some of the findings of this study are in conflict with the widely held beliefs of fluency proponents.

A notable outcome of this study is the degree to which performance (as measured by response rate) decreased over a period of eight weeks without practice. This outcome occurred despite each fact being overpracticed between 244 and 731 times. The particular group of participants meant that the likelihood of practice between sessions was very low. The home-schooling parents and their children were in a position to be able to commit to not include any multiplication in the participants' schooling, or any other activities, during

the period of the study. It may be that, if appropriate maintenance of skills occurred, an enduring difference in response rate between a skill trained to a fluency performance standard and a skill trained with the same amount of rate-controlled practice would have been observed. Investigating this question would be a useful future study.

Even when facts are learned to accuracy, occasional incorrect answers still occur. For example, five out of seven participants, having met the accuracy criterion of three consecutive errorless trials on each of three consecutive days, made between one and three errors during equalisation practice. All participants made errors during subsequent practice using the software. One might assume that stimulus control strengthens with progressive practice. However, comparisons of error rates at the end of accuracy training with those obtained at the end of software training, show error rates to both Deck A and Deck B facts increased for approximately half the participants as response rate increased (error rates to Deck A facts increased for 3 participants and error rates to Deck B facts increased for 4 participants). Fluency performance standards include an accuracy component. In this study, the widely accepted practice of accepting no more than two errors per 1-min timing was used. However, there is currently no clear research justifying the degree of accuracy required as part of the fluency performance standard.

The relationship, if any, between response latencies and learning outcomes such as retention, endurance and application, is currently unclear as noted in the discussion of response latencies in this thesis. Research investigating this relationship could reveal a metric that allows the effect of overpractice on learning outcomes to be quantified.

It is possible to draw some conclusions from this study regarding the role of practice in skill retention. First, extra practice beyond accuracy, whether rate-building or rate-controlled, results in faster response rates and hence shorter response latencies.

Second, neither the results of Shirley and Pennypacker (1994), nor those of this study, provide support for the notion that high rate training, including rate-building to a fluency performance standard, produces better retention than the same amount of rate-controlled practice. This is the case for retention whether it is assessed by accuracy or response rate, either as an absolute rate or as relative loss.

The enhanced retention claimed as an outcome of meeting fluency performance standards, and reported in studies that do not control for practice, may not be due to rate-building to fluency performance standards per se. Instead, it may be the result of the large amounts of extra practice necessary to build response rates to fluency.

Binder (2004, p. 282) suggests that, when considering learning outcomes, instead of comparing the effects of rate-building to fluency with rate-controlled practice that provides the same number of opportunities to respond, “[The better question] is to ask if rate of freely emitted responding better predicts these learning outcomes.” The evidence of this study is that the answer to this question, for retention, is also “no.”

References

- Aron, A., & Aron, E. N. (2003). *Statistics for Psychology* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Ashbaugh, R., & McLaughlin, T. F. (1997). Precisely teaching street names and locations. *Journal of Precision Teaching and Celeration*, *14*, 19-23.
- Beck, R., & Clement, R. (1991). The Great Falls Precision Teaching Project: An historical examination. *Journal of Precision Teaching*, *8*, 8-12.
- Berquam, E. M. (1981). *The relationship between frequency of response and retention, on a paired-associate task*. Unpublished doctoral dissertation, University of Florida, Florida.
- Binder, B. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, *19*, 163-197.
- Binder, C. (2003). Doesn't everybody need fluency? *Performance Improvement*, *42*(3), 14-20.
- Binder, C. (2004). A refocus on response-rate measurement: Comment on Doughty, Chase, and O'Shields. *The Behavior Analyst*, *27*, 281-286.
- Binder, C., Houghton, E., & Bateman, B. (2002). *Fluency: Achieving true mastery in the learning process*. University of Virginia Curry School of Education Website. Retrieved October, 12, 2005, from http://curry.edschool.virginia.edu/sped/projects/ose/papers/Binder-et-al_Fluency.pdf
- Binder, C., Houghton, E., & Van Eyk, D. (1995). Increasing endurance by building fluency: Precision teaching attention span. *Journal of Precision Teaching*, *12*, 29-34.

- Bonser, D. J. (2002). *Behavioural fluency for young children with autism*. Unpublished doctoral dissertation. Murdoch University, Perth, Australia.
- Bucklin, B. R., Dickenson, 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.
- Bullara, D. T., Kimball, J. W., & Cooper, J. O. (1993). An assessment of beginning addition skills following three months without instruction or practice. *Journal of Precision Teaching*, *11*(1), 11-16.
- 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.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied Behavior Analysis*. Columbus: Merrill Pub. Co.
- Dougherty, K. M., & Johnston, J. M. (1996). Overlearning, fluency, and automaticity. *The Behavior Analyst*, *19*, 289-292.
- 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*, 7-23.
- Driskell, J. E., Willis, R., & Cooper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, *77*, 615-622.
- Encarta Dictionary. Retrieved February 12, 2005, from <http://encarta.msn.com/encnet/features/dictionary/dictionaryhome.aspx>

- Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8), 725-747.
- Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.
- Eshleman, J. W. (2001). "Retention" data checks and studies. Standard Chart Commentary Website. Retrieved May, 3, 2005, from <http://members.aol.com/standardcharter/retention.html>
- Evans, S. S., & Evans, W. H. (1985). Frequencies that ensure skill competency. *Journal of Precision Teaching, 4*(2), 25-30
- Evans, S. S., Merger, C. D., & Evans, W. H. (1983). The relationship of frequency to subsequent skill acquisition. *Journal of Precision Teaching, 4* (2), 28-34.
- Ferster, C., & Skinner, B. F. (1957). *Schedules of Reinforcement*. New York: Appleton-Century-Crofts.
- Fischer, C. W., Berliner, D. C., Filby, N. N., Marlianve, R., Cahen, L. S., & Dishaw, M. M. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), *Time to Learn* (pp. 7-32).
- Greenwood, C. R., Delquadri, J., & Hall, R. V. (1984). Opportunities to respond and student academic performance. In W. L. Heward, T. E. Heron, J. Trap-Porter, and D.S. Hill (Eds.), *Focus on Behavior Analysis in Education* (pp. 58-88). Columbus, OH: Merrill.
- Hartnedy, S. L., Mozzoni, M. P., & Fahoum, Y. (2005). The effect of fluency training on math and reading skills in neuropsychiatric diagnosis children: A multiple baseline design. *Behavioral Interventions, 20*, 27-36.

- Haughton, E. C. (1984). Standards: Refining measurement. *Journal of Precision Teaching*, 4, 96-99.
- Haughton, E. C. (1997). Practicing practices: Learning by activity. *Journal of Precision Teaching and Celeration*, 15, 75-91.
- Ivarie, J. J. (1986). Effects of proficiency rates on later performance of a recall and writing behavior. *Remedial and Special Education: RASE*, 7(5), 25-30.
- Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist*, 47(11), 1475-1490.
- 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.
- Johnson, K. R., & Layng, T. V. J. (1996). On terms and procedures: Fluency. *The Behavior Analyst*, 19, 281-288.
- Kelly, R. L. (1996). *A functional analysis of the effects of mastery and fluency on maintenance*. Unpublished doctoral dissertation, Columbia University, New York.
- Kerr, K. P., Smyth, P., & McDowell, C. (2003). Precision teaching children with autism: helping design effective programmes. *Early Child Development and Care*, 173(4), 399-410.
- Koorland, M. A., Keel, M. C., & Ueberhorst, P. (1990). Setting aims for precision learning. *Teaching Exceptional Children*, 22(3), 64-66.

- Kruger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology*, 12, 71-78.
- Kubina, M. J. (2005). The relations among fluency, rate building, and practice: A response to Doughty, Chase, and O'Shields (2004). *The Behavior Analyst*, 28, 73-76.
- Kubina, R. M., & Morrison, R. S. (2000). Fluency in education. *Behavior and Social Issues*, 10, 83-99.
- Kubina, R. M., Morrison, R., & Lee, D. L. (2002). Benefits of adding precision teaching to behavioural interventions for students with autism. *Behavioral Interventions*, 17, 233-246.
- Kubina Jr., R. M., & Starlin, C. M. (2003). Reading with Precision. *European Journal of Behavior Analysis*, 4, 13-21.
- Lindsley, O. R. (1992). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis*, 25(1), 51-57.
- Lindsley, O. R. (1996). Is fluency free-operant response–response chaining? *The Behavior Analyst*, 19(2), 211-224.
- Luyben, P. D., Hipworth, K., & Pappas, T. (2003). Effects of CAI on the academic performance and attitudes of college students. *Teaching of Psychology*, 30(2), 154-158.
- McDade, C. (1998). Retention and application of computerized fluency building. *Journal of Precision Teaching and Celeration*, 15, 74-80.

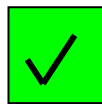
- McDowell, C., & Keenan, M. (2001). Developing fluency and endurance in a child diagnosed with attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis, 3*, 345-348.
- Miller, A. D., & Heward, W. L. (1992). Do your students really know their math facts? Using daily time trials to build fluency. *Intervention in School and Clinic, 28*(2), 98-104.
- Ministry of Education. (2005). *Number: Level 3 achievement objectives*. Te Kete Ipurangi: The Online Learning Centre Website: Retrieved March, 18, 2005 from http://www.tki.org.nz/r/math/curriculum/statement/p40_43_e.php
- Olander, C. P., Collins, D. L., McArthur, B. L., Watts, R. O., & McDade, C. E. (1986). Retention among college students: A comparison of traditional versus precision teaching. *Journal of Precision Teaching, 6*(4), 80-82.
- Ormrod, J. E., & Spivey, N. R. (1990). Overlearning and speeded practice in spelling instruction. *Psychological Reports, 67*, 365-366.
- Osgood, C. E. (1946). Meaningful similarity and interference in learning. *Journal of Experimental Psychology, 36*(4), 277-301.
- Peladeau, N., Forget, J., & Gagne, F. (2003). Effect of paced and unpaced practice on skill application and retention: How much is enough? *American Educational Research Journal, 40*(3), 769-788.
- Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2004). The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19*(3), 361 – 374.
- Shirley, M. J., & Pennypacker, H. S. (1994). The effects of performance criteria on learning and retention of spelling words. *Journal of Precision Teaching, 12*, 73-86.

- Skinner, B. F. (1953). *Science and Human Behavior*. New York: Macmillan.
- Weber, K. P., & Cowardin, J. (1994). Computerized learning of WordPerfect using precision teaching. *Journal of Precision Teaching, 11*, 19-27.
- Weiss, M. J. (2001). Expanding ABA intervention in intensive programs for children with autism: The inclusion of natural environment training and fluency based instruction. *The Behavior Analyst Today, 2*(3), 182-187.
- White, O. R. (1984). Aim*star wars [Setting aims that compete]. *Journal of Precision Teaching, 5*(3), 55-63.
- Wolery, M., Bailey, Jr., D. B., & Sugai, G. M. (1988). *Effective Teaching Principles and Procedures of Applied Behavior Analysis with Exceptional Students*. Needham, Massachusetts: Allyn and Bacon, Inc.
- Young, K. R., West, R. P., & Crawford, A. (1985). The acquisition and maintenance of reading skills by intellectually handicapped deaf students. *Journal of Precision Teaching, 5*(4), 73-86.
- Young, K. R., West, R. P., Howard, V. F., & Whitney, R. (1986). Acquisition, fluency training, generalization, and maintenance of dressing skills of two developmentally disabled children. *Education and Treatment of Children, 9*, 16-29.

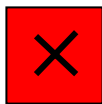
Appendix A

Example of the notice seeking participants posted on the local home-schooling website.

Learning opportunity for kids



Is your child aged between 5 and 12 years?



Do they know the 3, 4, 6, 7, 8, and 9 times table facts?

I am researching different ways that children acquire and retain skills, such as the times tables, as part of my Masters Thesis research at the University of Waikato and am looking for young participants.

During weekday sessions in your home I will help your child learn to give accurate and reliable answers to basic multiplication facts (one of the NZ school syllabus aims). Your child will also gain hands-on experience using a laptop and will receive the added benefit of daily positive one-on-one adult attention at no cost to you.

The sessions should provide a fun way to learn and children will earn both small food and toy items (approved by you) as they learn facts.

Interested in further details?

Phone me on (07) 825 XXXX or email me at
fluency@mcgregor.org.nz
Susan McGregor

Appendix B

Information sheet for parents of potential participants. This sheet describes the experiment and what was involved in participation.



Information sheet for parents and caregivers

What is the study about?

I am undertaking a study of fluency training as part of my Masters Thesis research at the University of Waikato. Fluency training is training beyond the point where facts are answered correctly, to the point where answers are both accurate and fast. As we all know, many of the things we learn are forgotten as time goes by. Research suggests that facts learned to fluency will be retained for longer periods where they are not practiced.

As part of this study I will be teaching children aged 5 to 12 years times table facts. The New Zealand mathematics curriculum number section entitled 'exploring computation and estimation' specifies recall of basic multiplication facts as an aim to be met by the end of level 2 (year 4) when children are typically aged eight to nine years.

What will participation in this study involve?

Being part of this study would involve your child participating in sessions each weekday, Monday to Friday, in your home. Sessions would continue for approximately eight weeks. These sessions should be fun for the children involved! Children will spend a maximum of 15-mins each session learning multiplication facts previously unknown to them. For the remainder of the session, your child will be given the opportunity to engage in play with the researcher at an activity of their choice (e.g., sports or board games). Sessions will last for a maximum of 40-mins. Initially, children will learn to answer multiplication facts accurately. They will practice facts presented on cards and receive a small sweet, when they complete set tasks.

Once they are consistently able to give accurate answers, children will continue to learn to answer some of the facts fluently (fast and accurately) during timed trials. The other facts will be practiced an equal number of times, but not to a fast rate. Facts will be presented on a laptop and children will be given an opportunity to earn tokens each session that may be exchanged for a small food or toy item. Children will be tested on these facts 4 and 8 weeks later to assess which training method resulted in better retention. You will be welcome to be present during sessions with your child and will be asked to approve both food and toy items before the sessions begin.

What prior skills must my child have to participate?

Your child must be able to read two digit numbers fluently.

How will my child benefit from participating in this study?

Your child will be given a head start on meeting one of the New Zealand school syllabus aims. They will learn to give accurate and reliable answers to a set of basic multiplication facts. Your child will also gain experience using a laptop, following instructions and will receive the added benefit of daily positive one-on-one adult attention. Children will be treated with respect and receive appropriate incentives to aid in their learning.

Will it cost me anything?

There is no cost, beyond time, for your child and yourself during the sessions. No practice is required between sessions for either you or your child.

Can we withdraw from the study?

Yes, you can withdraw at any time for any reason.

What will happen to the information collected?

You will be given a copy of the results and will have the opportunity to discuss them with me at the conclusion of the research. Please feel free to discuss any aspect of the study, including your child's progress at any time. The results will be anonymised and combined with those for others participating in the study and presented as part of my thesis, and possibly further research.

Where can I get further information?

If you'd like to be part of the study, or want to know more call me (Susan McGregor) on (07)825 XXXX or email me at fluency@mcgregor.org.nz. Or alternatively you may contact my supervisors.

Thank you

Susan McGregor

Appendix C

Example of a flashcard used in Phase 1 (accuracy training). The multiplication fact is presented on one side of the card and the answer is printed on the reverse side. These images are full size.

$$6 \times 9 =$$

54

Appendix D

Example of Phase 1 (accuracy training) data collection form. This form was used to record whether the participant correctly or incorrectly answered each fact presented.

Participant's name:

Date:	Wed 25/5	Thurs 26/5	Fri 27/5
Session number:	1	2	3
Criterion met? (0)	✓	✓	✓
Reinforcer obtained (0)	✓	✓	✓
6x6	✓	✓	✓
6x7	✓	✓	✓
6x8	✓	✓	✓
6x9	✓	✓	✓
7x7	✓	✓	✓
7x8	✓	✓	✓
7x9	✓	✓	✓
8x8	✓	✓	✓
8x9	✓	✓	✓
9x9	✓	✓	✓
Cards	✓	✓	✓
x = incorrect answer given			
/ = correct answer given			

Date:	Mon 30/5	Tues 31/5	Wed 1/6
Session number:	4	5	6
Criterion met? (0)	✓	✓	✓
Reinforcer obtained (0)	✓	✓	✓
6x6	✓	✓	✓
6x7	✓	✓	✓
6x8	✓	✓	✓
6x9	✓	✓	✓
7x7	✓	✓	✓
7x8	✓	✓	✓
7x9	✓	✓	✓
8x8	✓	✓	✓
8x9	✓	✓	✓
9x9	✓	✓	✓
Cards	✓	✓	✓
x = incorrect answer given			
/ = correct answer given			

Appendix E

Description and specifications of software.

Program Specification

The program presents multiplication facts to the participant then waits for a single left mouse key press to indicate that a verbal answer has been given. The screen contains up to three facts. These are the current fact and up to two previous facts, with the oldest fact at the top, the most recently answered fact next and the current fact at the bottom. 100pt font should be used with a different colour background for the current fact (examples are shown below).

When the mouse key is pressed, to indicate that a fact has been answered, three or four things happen at the same time. First, the facts move up one (the oldest fact being removed and the previous current fact becoming the most recently answered fact). Next a new fact is introduced as the current fact. The next two steps involve a timer, used to control when an answer is displayed adjacent to a fact that has already been answered by the student. The duration of this timer is a configuration parameter but it is expected to be in the order of half a second. The timer is set in step four, but tested in step three. Step three: if, when the now current fact is answered (indicated by a key press), the timer is active, it is cancelled and the answer to the fact that has just become the oldest answered fact (at the top of the screen) is presented. The timer is then (re-)started. If this timer expires before the answer to the now current fact is given (indicated by a key press), the answer to the most recently answered fact (middle fact on the screen) is presented on the screen (alongside the fact). The timer is then (re-)started. Optionally, minimum fact presentation time may be included. If this is the case and the student

presses the key before this time the first and last two events described above happen immediately, but the new fact is not presented until the end of the minimum fact presentation time.

Facts are drawn at random from an input file repeatedly until the time for the trial is complete (this is one run). Facts from the input file should not be reused until all facts have been presented. For example, if there are three facts in the input file, a presentation might be facts 3,1, and 2, but not 1,3,1 (fact 1 is repeated). Participant responses are recorded in an output file.

When the time for the run is complete (this is a configuration parameter) and the current fact has been answered, no new fact is introduced. The normal scrolling and answer presentation occurs. This is presented for a duration (described below) and then a summary screen is displayed. The duration for presentation of the final answer is based on the mean answer rate of the participant for this run. If the total run time divided by the number of facts presented (MEAN_ETIME) is less than the configured delay to present the answer (ANSWER_DELAY below) then the final answer presentation time is MEAN_ETIME. Otherwise, the final answer is presented for $2 * \text{MEAN_ETIME} - \text{ANSWER_DELAY}$. For example, if ANSWER_DELAY is 500 (ms) and the participant answered 130 facts in 60 seconds the final delay will be 461 ms. Alternatively, if the student had only answered 60 facts (in the same 60 seconds) the final delay would have been 1500 ms.

At the end of the run, a summary screen is displayed that contains: the participant's name, the number of facts completed, the total time of the run (which will normally be a little longer than the configured TRIAL_DURATION described below because of the last fact) and the answer rate (total time divided by number of facts answered). There

should be two input options at this time, one to exit and the other to rerun the program from the start with a new sequence of facts.

Configuration file

This file describes the general setup of this run of the program.

Description	Field Name	Example values
Participant's Name	NAME	Susan McGregor
Trial duration (s)	TRIAL_DURATION	60 or 1000
Maximum number of cards in trial	MAX_CARDS	1000 or 15
Input filename	INPUT_FILE	C:\program files\ntf\bob.txt
Output Filename	OUTPUT_FILE	C:\program files\ntf\bob-out.txt
Delay to answer presentation (ms)	ANSWER_DELAY	500 or 0
Minimum fact presentation time (ms)	MIN_FACT_DURATION	0 or 1000

Input file

This file describes the multiplication facts that are presented to the participant.

Element 1 of fact	Element 2 of fact
e.g. 3	6

Which will look like this to participant: 3 x 6 =

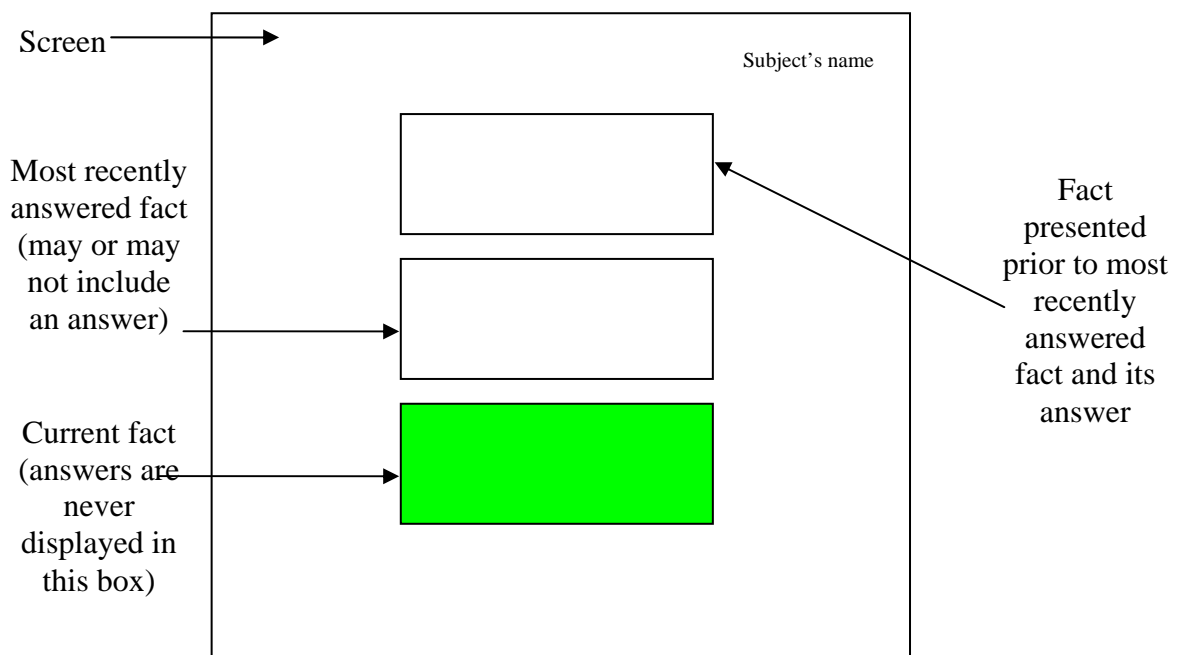
Output file

This file is produced by the program to record the responses made by the participant. The output file begins with a copy of the current configuration file (but in CSV format, not space separated), followed by "RUN_TIME," and the time of day at the start of the run in the format: dd-mmm-yyyy hh:mm:ss, followed by a blank line. This is followed by one line for each fact presented in the CSV format below. This repeats for every card.

Card number	First multiplication operand	Second multiplication operand	Time since start of run that fact is presented (ms)	Time of left mouse key press since fact was presented (ms)	Time since the fact was first presented to when the answer is presented (ms)	Time since the fact is first presented until that the fact is scrolled off the screen (ms)
e.g. 1	3	5	3024	521	1021	1765

At the end of a run a line is added with the following fields: number of cards presented, number of facts answered, length of run (ms), answer rate. If the file already exists before the program is run, the output is appended to the end of the existing file (separated by a blank line).

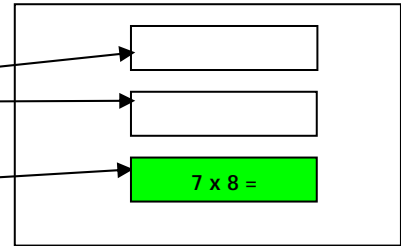
Screen format



User interface

Other boxes start out empty

First problem is displayed in bottom fact box
(this box is always highlighted)

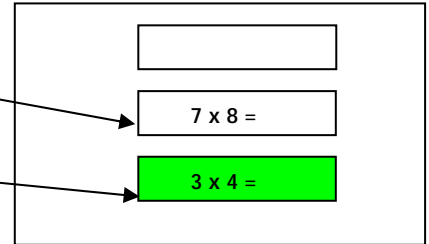


The learner will say an answer out loud e.g., “56” and press the left mouse key to indicate they have provided an answer and three things will occur:

Simultaneously

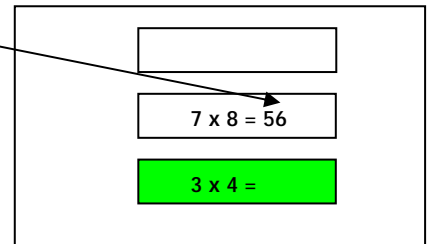
The most recently answered fact will move up to middle box

The 2nd fact (now current fact) will appear in bottom highlighted box



Then:

If the answer to the current fact has not been given by the time the ANSWER_DELAY period is completed the correct answer will be displayed in the middle box next to most recently answered fact



The learner will say an answer to the current fact out loud e.g., “12” and will press the left mouse key to indicate they have provided an answer and 4 things will happen:

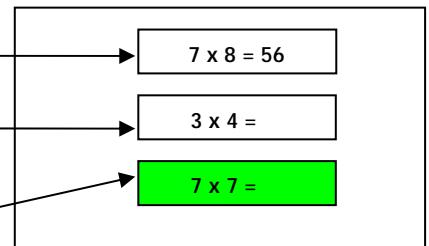
(If this answer is given before the ANSWER_DELAY period is completed, the answer to the fact in the middle box will appear as the fact moves to the top box)

Simultaneously

The middle fact will move up to top box

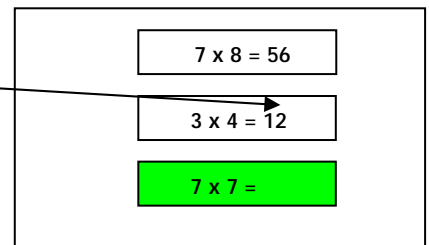
The most recently answered fact will move up to middle box

The 3rd fact (now current fact) will appear in bottom highlighted box



Then:

After the ANSWER_DELAY period is completed the correct answer will be displayed in the box next to most recently answered fact (in the middle box)



The learner will say an answer to the current fact out loud e.g., “49”

The fact in the top box will now disappear and the process above will continue until the time interval specified for the trial ends.

Appendix G

Example of Phase 3 (assessment) data collection form. This form was used to record the number of correct and incorrect answers during each 1-min timing for each deck of facts (2 times tables, Deck A, rate-built facts, and Deck B, rate-controlled facts) at each assessment stage.

Name:
Phase 3: Assessment

Post-accuracy Training Assessment (1-min timings)

Date: Fri 8/7

	2 Times Table	Deck A (rate-building)	Deck B (rate-controlled)
Order of Trial Presentation	1	2	3
Number of facts correct	44	17	22
Number of facts incorrect	0	1	0
Total number of facts completed	44	18	22

Post-software Training Assessment (1-min timings)

Date: Mon 5/9

	2 Times Table	Deck A (rate-building)	Deck B (rate-controlled)
Order of Trial Presentation	1	2	3
Number of facts correct	46	82	41
Number of facts incorrect	0	0	0
Total number of facts completed	46	82	41

4-week Retention Assessment (1-min timings)

Date: Wed 5/10

	2 Times Table	Deck A (rate-building)	Deck B (rate-controlled)
Order of Trial Presentation	1	2	3
Number of facts correct	41	34	38
Number of facts incorrect	0	0	0
Total number of facts completed	41	34	38

8-week Retention Assessment (1-min timings)

Date: Mon 31/10

	2 Times Table	Deck A (rate-building)	Deck B (rate-controlled)
Order of Trial Presentation	1	2	3
Number of facts correct	40	14	42
Number of facts incorrect	0	4	0
Total number of facts completed	40	18	42

Appendix H

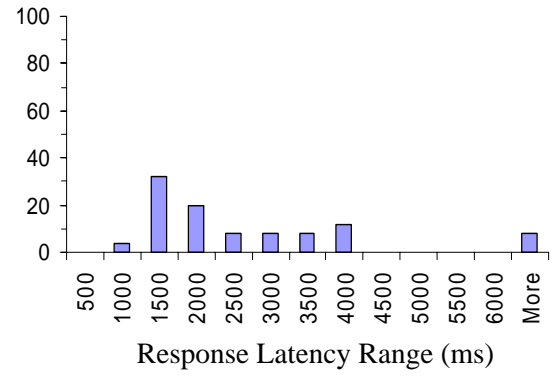
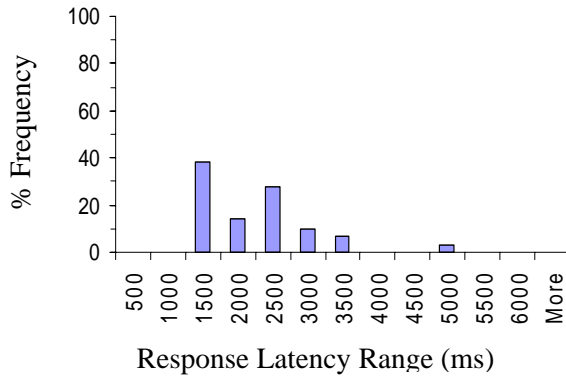
Histograms for each participant showing the percent frequencies of response latencies, for both Deck A and Deck B facts, over the four assessment stages (post-accuracy training, post-software training, 4-week retention assessment and 8-week retention assessment). Histograms are presented for each participant on a new page. Response latency, collected into 500-ms bins, is plotted on the x-axes. Percent frequency is plotted on the y-axes. Percent frequency was calculated by taking the number of response latencies (to both correct and incorrect responses) falling in each bin range, expressed as a percentage of the total facts answered in the 1-min timing.

P1

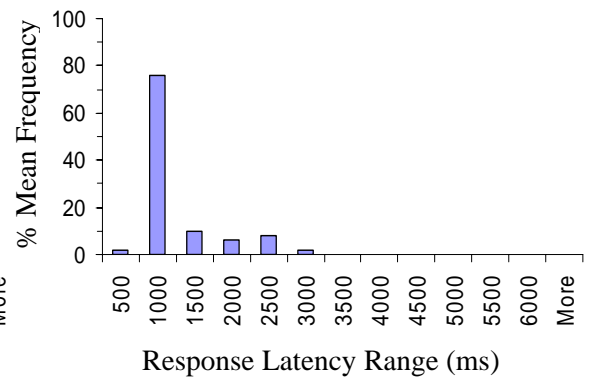
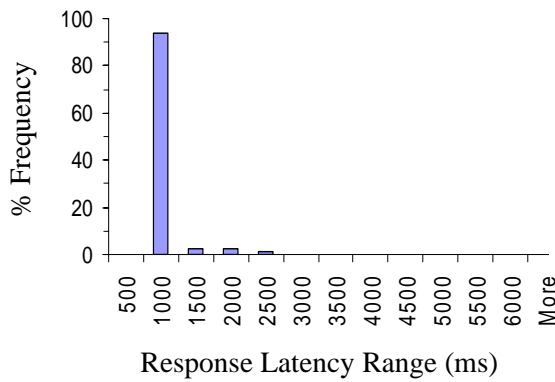
Deck A facts

Deck B facts

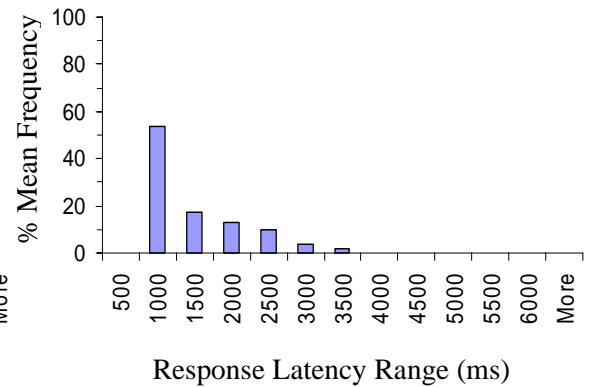
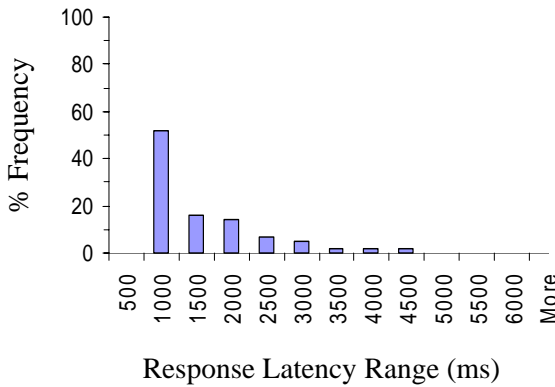
End of Accuracy Training



End of Software Training



4-Week Retention Assessment



8-Week Retention Assessment

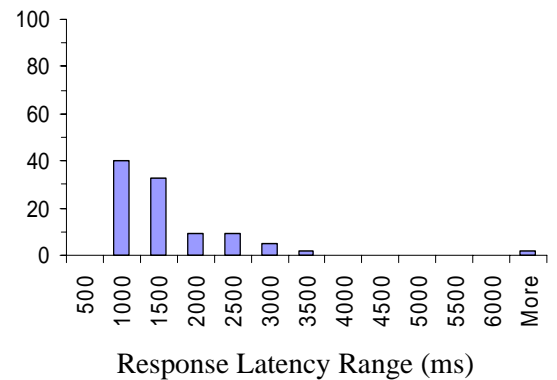
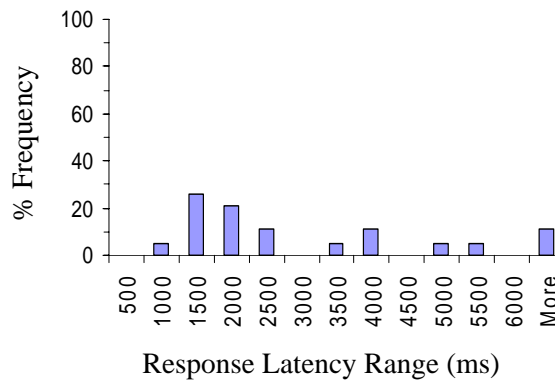


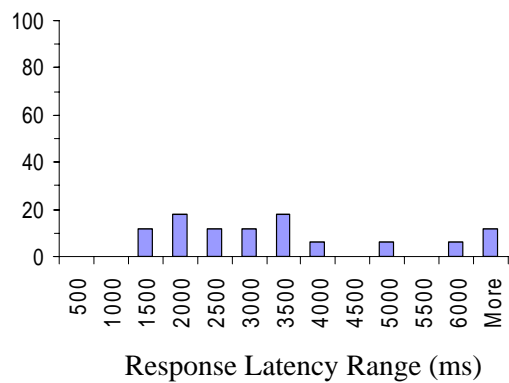
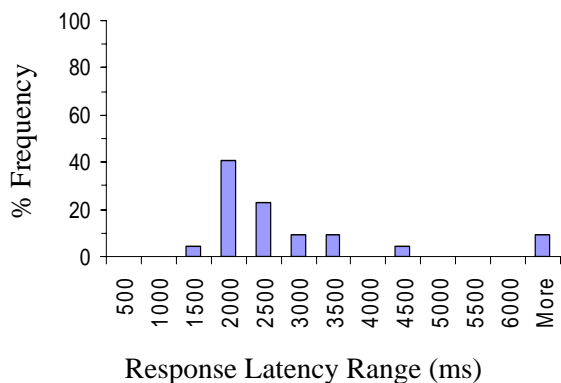
Figure H1. The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P1.

P2

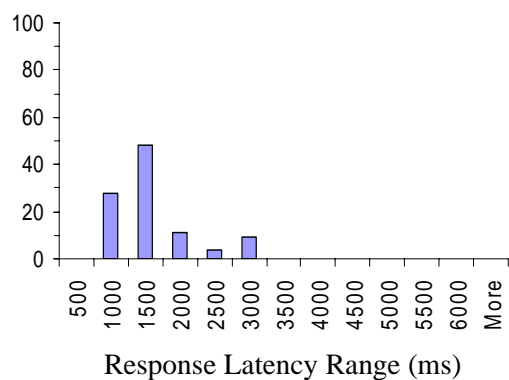
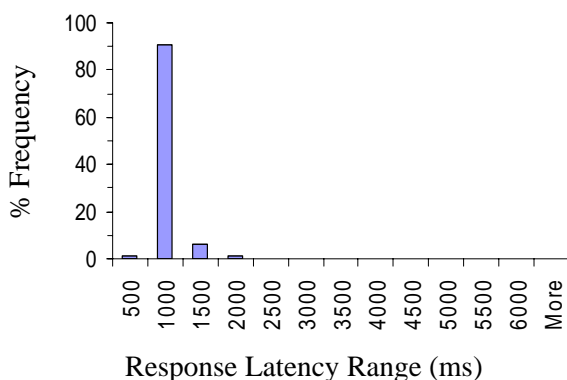
Deck A facts

Deck B facts

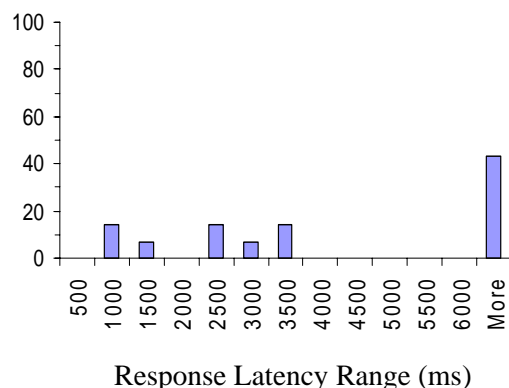
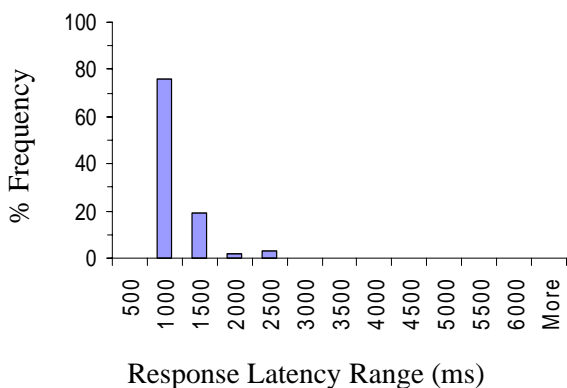
End of Accuracy Training



End of Software Training



4-Week Retention Assessment



8-Week Retention Assessment

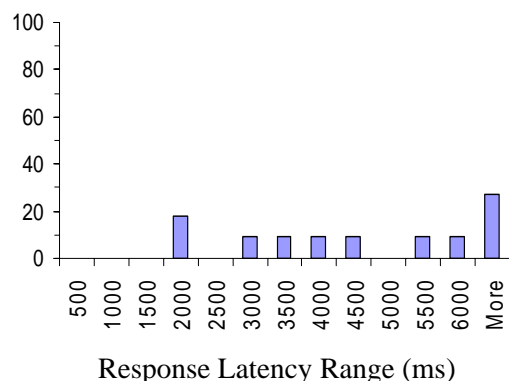
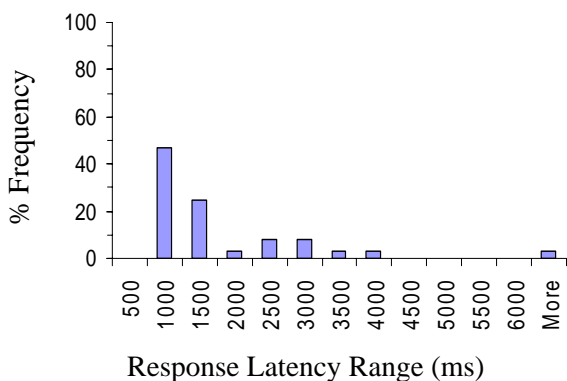


Figure H2. The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P2.

P3

Deck A facts

Deck B facts

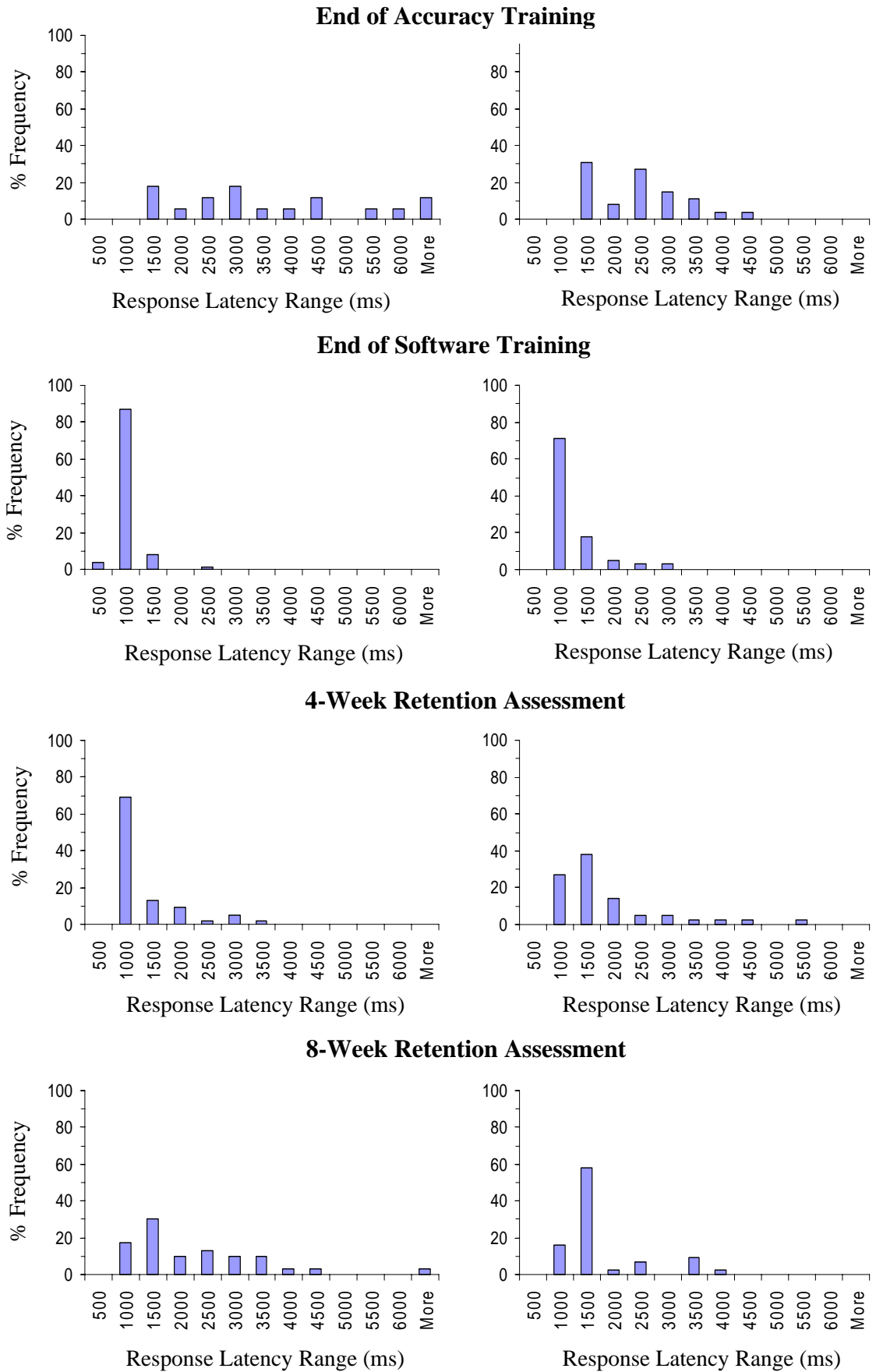


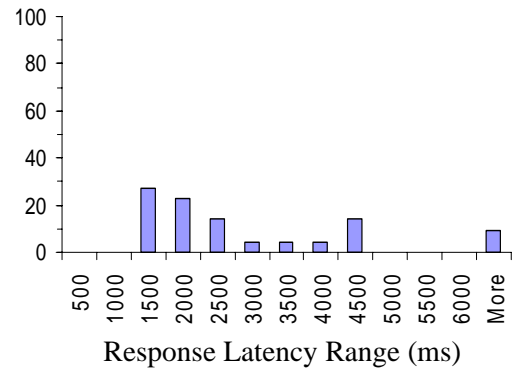
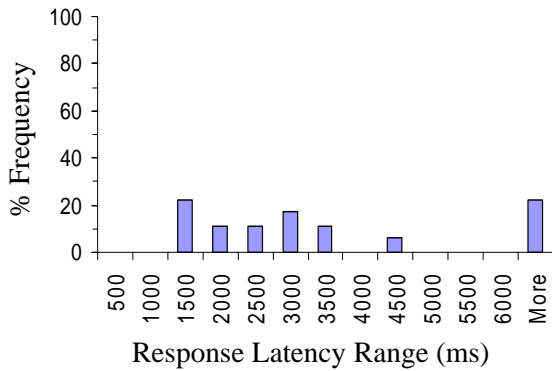
Figure H3. The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P3.

P4

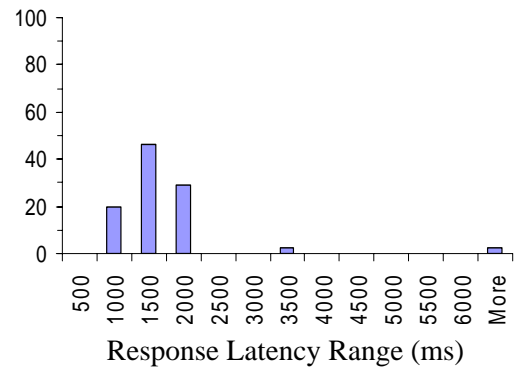
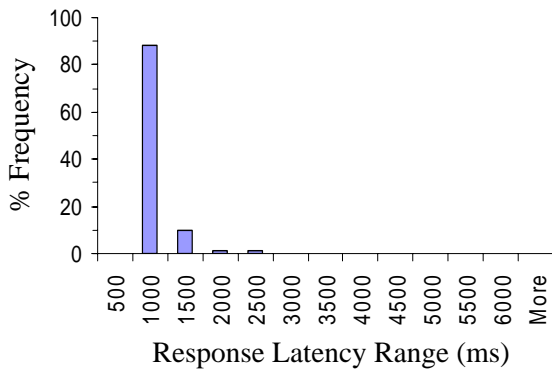
Deck A facts

Deck B facts

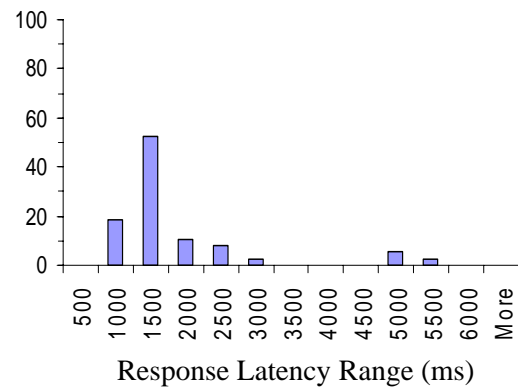
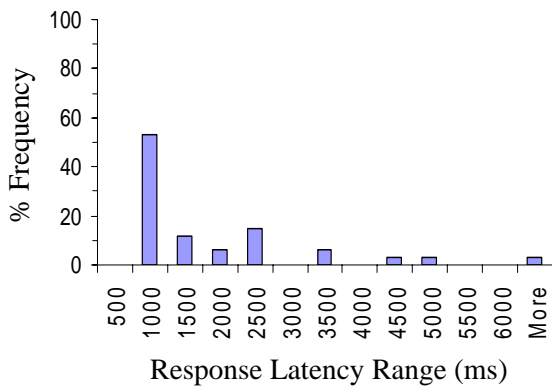
End of Accuracy Training



End of Software Training



4-Week Retention Assessment



8-Week Retention Assessment

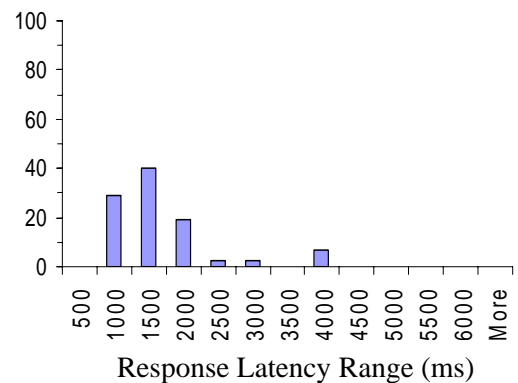
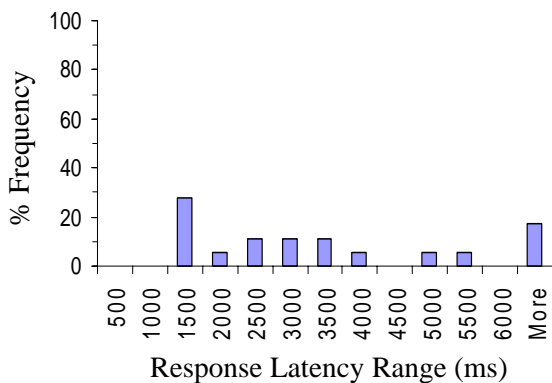


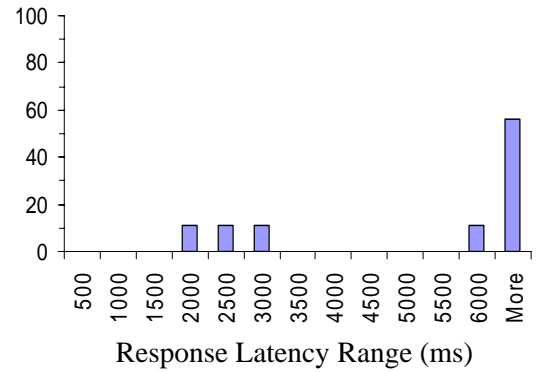
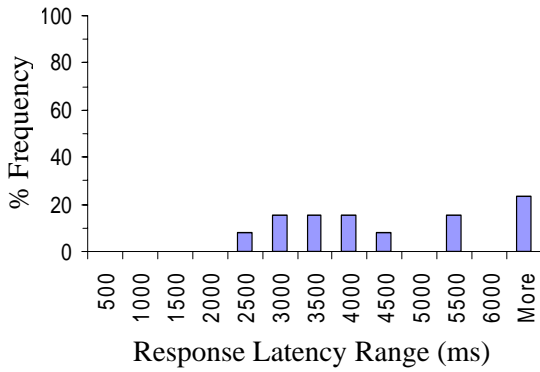
Figure H4. The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P4.

P5

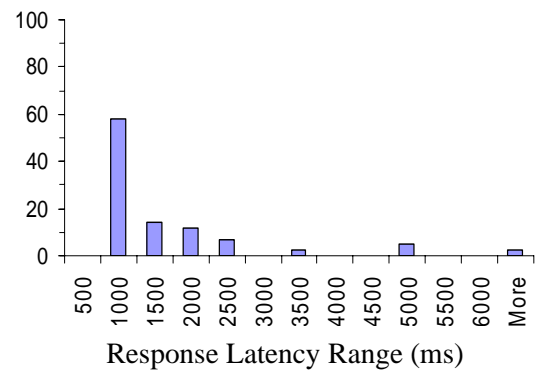
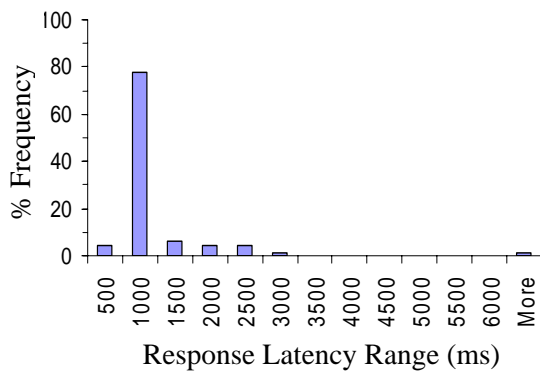
Deck A facts

Deck B facts

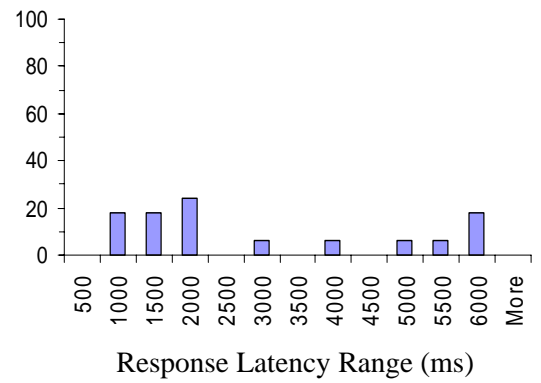
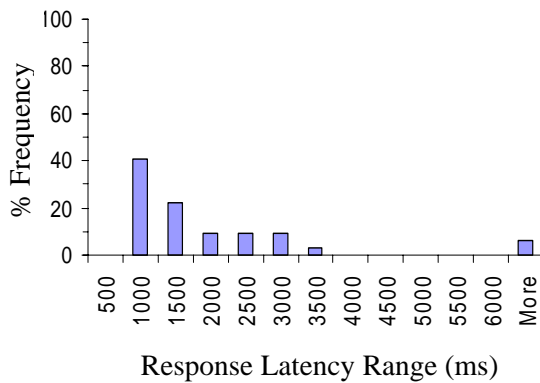
End of Accuracy Training



End of Software Training



4-Week Retention Assessment



8-Week Retention Assessment

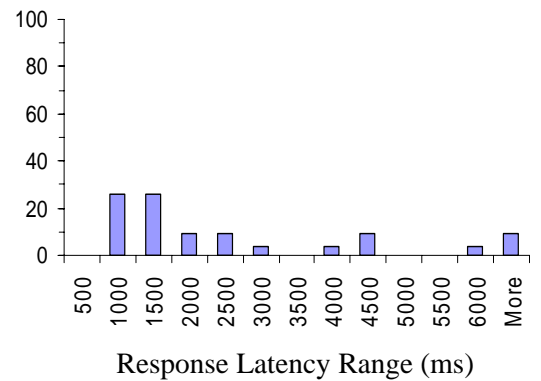
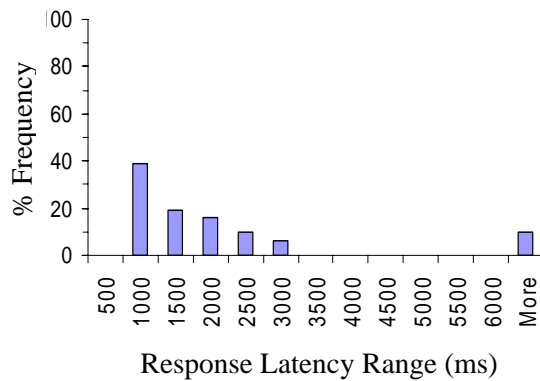


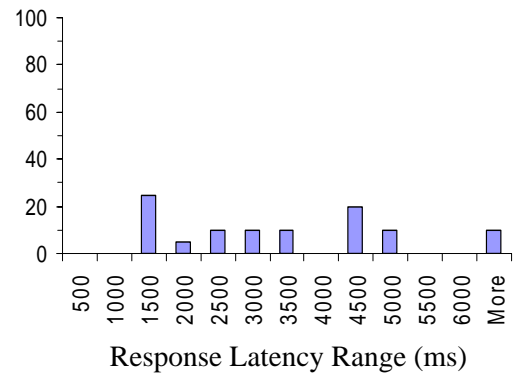
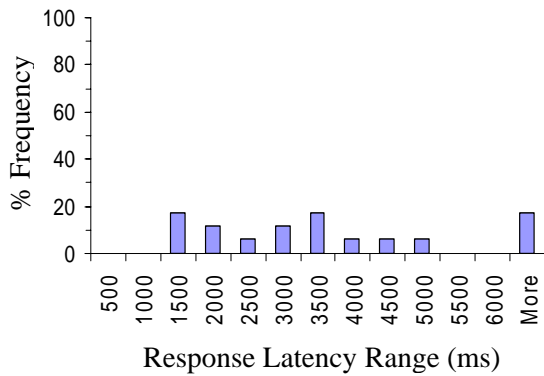
Figure H5. The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P5.

P6

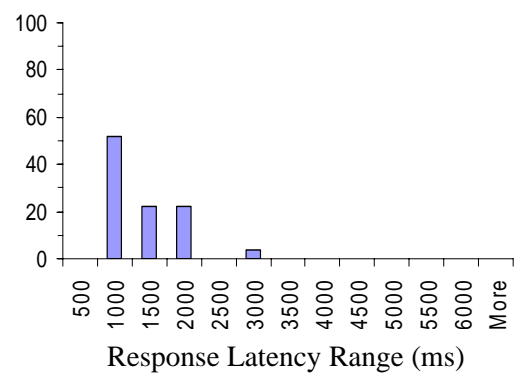
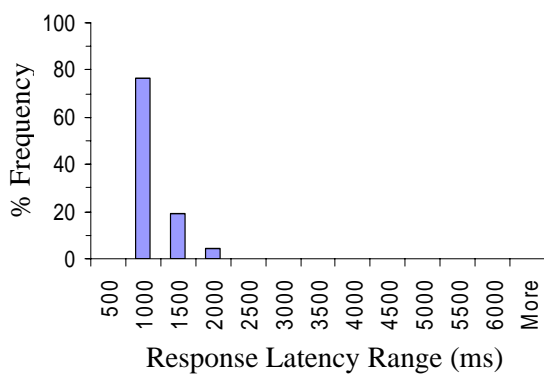
Deck A facts

Deck B facts

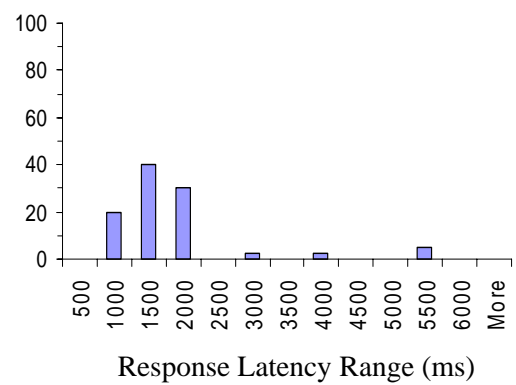
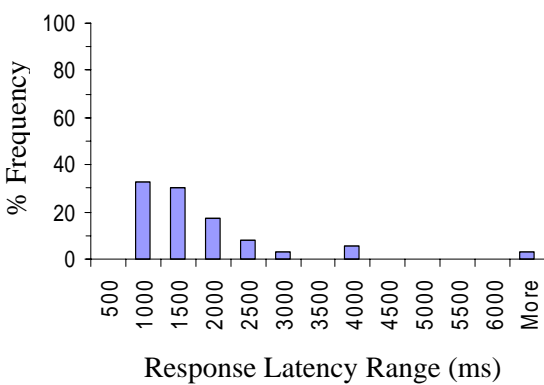
End of Accuracy Training



End of Software Training



4-Week Retention Assessment



8-Week Retention Assessment

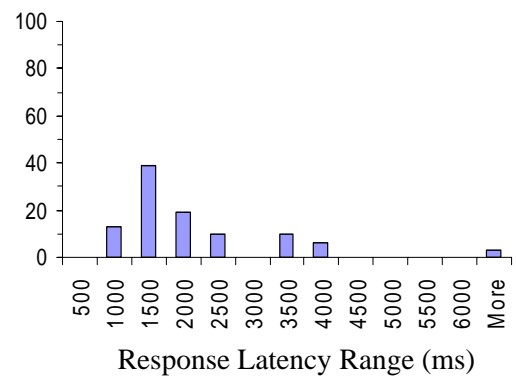
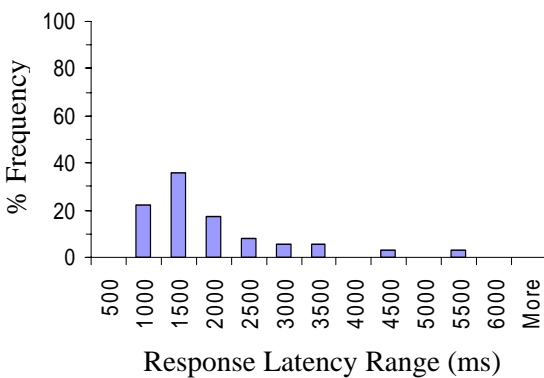


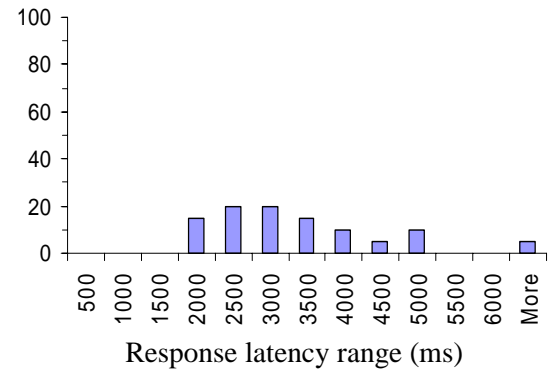
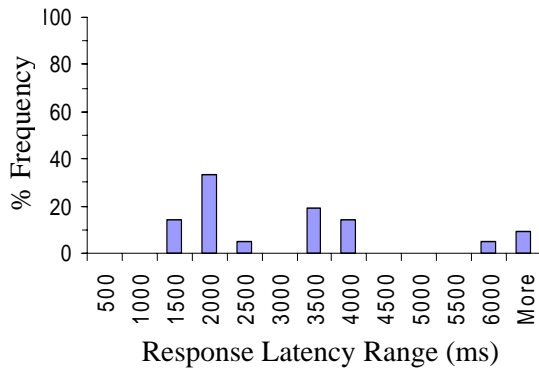
Figure H6. The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P6.

P7

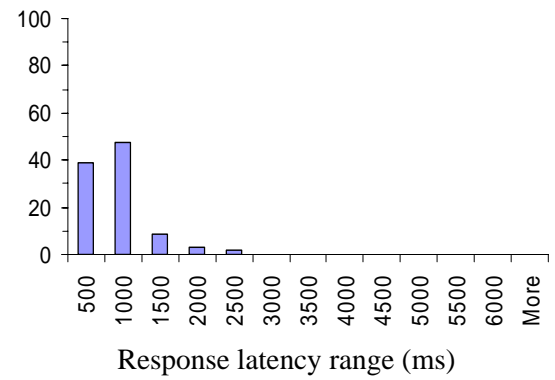
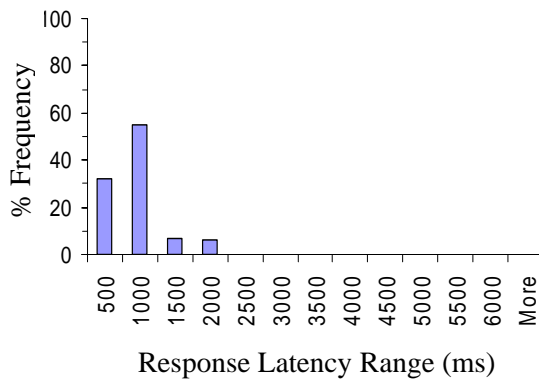
Deck A facts

Deck B facts

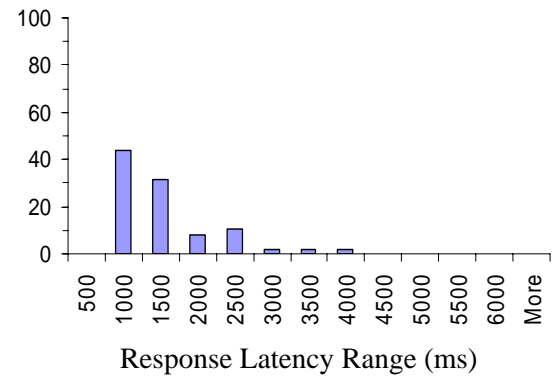
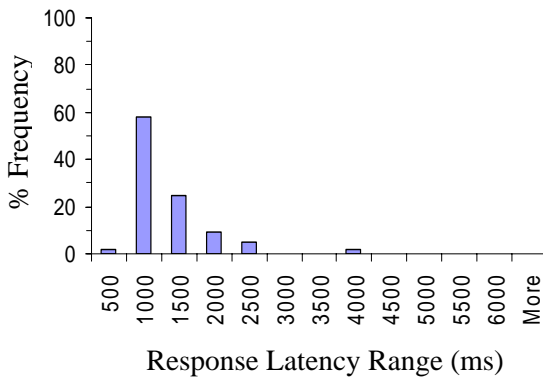
End of Accuracy Training



End of Software Training



4-Week Retention Assessment



8-Week Retention Assessment

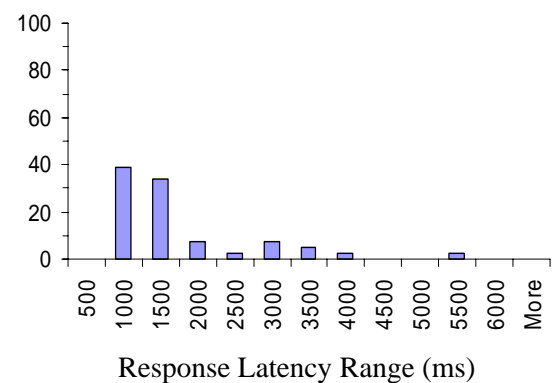
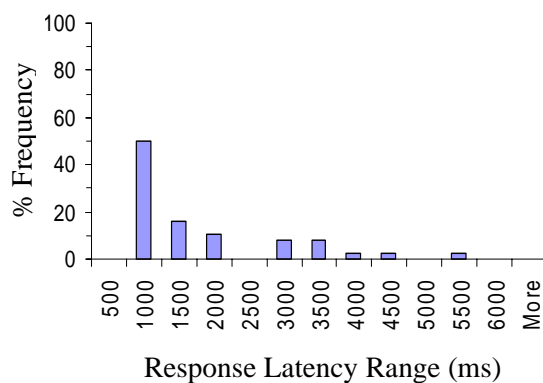


Figure H7. The percent frequencies of response latencies for both Deck A and Deck B facts over the four assessment stages for P7.