An Investigation Into Rate-building and Cues on Conditional Discrimination Performance Using a Repeated Acquisition Procedure

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Department of Psychology

at

The University of Waikato

by

JOSHUA AARON LEVINE

2014

ABSTRACT

The present study consisted of eleven experiments divided between two series of studies. The first part of Series 1 aimed at replicating the findings of Porritt (2007) and Porritt et al. (2009). Findings from Series 1 showed that rate-building, when number of practices and reinforcement rate are controlled, enhance training accuracy. However, the greater response rates did not improve retention accuracy, a failure to replicate. Given the contrary outcomes, the studies in the second part of Series 1 attempted to fully replicate Porritt by using variables that have been shown to improve retention accuracy. These results replicated Porritt only when similar behaviours were trained under like conditions between the Training and Retention components. An interpretation of the Series 1 data suggests that, rather than response rate, response duration may contribute towards retention accuracy. The second series of studies investigated the role of stimuli in the repeated acquisition procedure. Findings show the use of colour cues generated the greatest accuracy while completing behaviour chains. However, both colour cues and position of last response were found to govern chain completion accuracy. Findings from Series 2 suggest attention should be paid to the use of cues when the repeated acquisition procedure is used in rate-building experiments. Overall, the present study found that focusing on duration-reduction, in an animal analogue study using a repeated acquisition procedure with no-colour cues, may reveal the prime contributor to greater retention in Precision Teaching.

ACKNOWLEDGEMENTS

I would first like to express my appreciation to Professor Mary Foster and Dr. James McEwan. I am grateful and thankful to Mary and James for their unwavering guidance and encouragement. Your superb academic instruction and guidance will surely stay with me throughout my career.

Thank you to the members of the Animal Behaviour and Welfare Research Centre for your help in running my experiments. Many thanks to Jenny Chandler and Rob Bakker who were instrumental in the maintenance of the hens and equipment.

The completion of this thesis would not have been possible without help from a very special group of individuals. A large portion of this thesis was written while I was living in Kuwait and India. I would like to express my gratitude to Sheikha Awrad Al-Sabah for her support while I lived in Kuwait. I would also like to express my appreciation to Bijesh Thakkar for providing me with accommodation in Mumbai during the writing of this thesis. Final revisions to this manuscript were completed while living back home in South Florida. Mom and Dad, thank you from the bottom of my heart. I am grateful to have parents that have showed me so much support.

This thesis would not have been completed without a little help from my friends. During the wonderful moments I’ve had to enjoy your company, the times you were available for great conversations and big hugs, and for just being there and helping me maintain balance, thank you! I am thankful and grateful to all the wonderful people I have had the pleasure to meet during this academic journey.

TABLE OF CONTENTS

PAGE

Experiment 1.1

Experiment 1.2

Experiment 1.3

Experiment 1.4

Experiment 1.5

Experiment 1.6

Experiment 1.7

Experiment 1.8

Experiment 2.1

Experiment 2.2

Experiment 2.3

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

One of the most important tasks of society is to ensure its people are properly educated. Nelson Mandela has stressed the importance of education when he said, “Education is the most powerful weapon which you can use to change the world.” To reflect their strong backing of education, the United States government spends over $500 billion dollars per year to fund over 98,000 public schools and 3.3 million teachers (NCES, 2012). But more important than money spent is the quality achieved.

The United States attempts to provide a high quality education for all Americans. The quality of education in the United States was first widely scrutinized in one of the largest educational report cards in U.S. history, the National Commission on Excellence in Education (NCEE, 1983). The NCEE reported that the government initiatives to provide quality education “seem to have lost sight of the basic purposes of schooling, and of the high expectations and disciplined effort needed to attain them”, and warned of a “rising tide of mediocrity that threatens our very future as a Nation and a people” (NCEE, 1983). Thus, despite the expenditures, there is a problem with the quality of education in the United States.

Attempts to improve the quality of education in America have aimed at standards- and outcomes- based education and spending more money on higher qualified teachers, merit based incentives for teachers, and increasing the number of school programs. To date, these reforms have not been successful at improving the quality of education (Burke, 2012; McNeir, 1993). As an alternative, Lindsley (1991) and Skinner (1984) suggested that focusing on the design of instruction could show promise at improving the quality of education, in addition to being cost-effective.

The significance of instructional design was first highlighted in one of the largest, and costliest, educational studies on “what works”, Project Follow Through (1968). In Project Follow Through, the effectiveness of several different instructional designs was tested on students’ learning across the United States over four years. It was found that Direct Instruction produced the highest scores in reading, arithmetic, spelling and language (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977).

The power of the Direct Instruction (DI) model rests on several essential teaching components (Becker, 1977). First, DI focuses on teaching towards generalization while providing as much rapid-paced instruction as possible. Second, DI is teacher-directed and ensures highly structured programs that are designed for small-groups. Third, DI uses reinforcement-based approaches to learning to ensure pre-requisite skills are being met and maintained throughout the learning process. Lastly, DI uses biweekly criterion referenced tests to help monitor student progress. One of the Direct Instructional approaches that is well researched and widely applied in schools is Precision Teaching.

Precision Teaching

Precision Teaching is an instructional design, rooted in the science of behaviour analysis, which describes the teaching, measuring, monitoring, and evaluating of educational pedagogy (Binder, 2003; Crawford & Olson, 1990; Kubina, 2005; Malabello, 1998). It has been used in sports (Keenan, 2002; McDowell), special education (Liberty & Paeth, 1990), and business (Binder & Bloom, 1989).

One of the most impressive and widely cited Precision Teaching developments was a 10 year research project named the "Great Falls Precision Teaching Project" (Beck & Clement, 1991). During the project, students who received Precision Teaching instruction for just 30 minutes a day outperformed their peers on state exams by 20% in reading and 40% in math; equivalent to outcomes observed from Direct Instruction during Project Follow Through. Precision Teaching proved to be so effective and reliable that a school using this model, Morningside Academy, promises only two months are required for a student to achieve a year's worth of learning (Binder, 1988; Johnson & Layng, 1992). What sets Precision Teaching apart from other instructional designs is its basis in a science that uses response rate as the standard metric (Lindsley, 1991).

The use of response rate in Precision Teaching began in the non-human operant lab of B. F. Skinner (Lindsley, 1972). One of B. F. Skinner's students, Ogden Lindsley, was the first to apply the response rate measure to human participants. Ogden Lindsley, along with colleagues, studied rate of task completion in school children (Haughton, 1972; Lindsley, 1964); techniques were then developed that characterize Precision Teaching to this day (Gallagher, 2006;

Kessissoglou & Farrell, 1995; Kubina, 2005; Kubina, Ward, & Mozzoni, 2000; Kubina & Morrison, 2000; Lindsley, 1972; White, 1986). An application of Precision Teaching is characterized by the following techniques:

1. The behaviour to improve (i.e., target behaviour) can be directly observed and a response rate is selected for the student to achieve.
2. Calculate how fast and accurate the student is currently performing the targeted behaviour during a 1-minute practice.
3. Immediately reinforce correct behaviour during practice and provide feedback.
4. Display behaviour on the Standard Celeration Chart (SCC).
5. Based on the charted data, make a decision on whether to continue with the current practice routine or make a change.

As a first step, Precision Teachers focus on behaviour that can be directly observed and measured (Kubina et al., 2000). For example, a student standing up from sitting in his chair would be a directly observable and measurable behaviour.

The behaviour targeted for increase is measured using counts of observable behaviour over time (i.e., response rate) and this is the standard measure of achievement for Precision Teachers, as opposed to the traditional percentage correct measure. Precision Teachers report count per minute (Graf & Lindsey, 2002); they may instead report count per second, hour, day, month, and year. For example, rate of behaviour could be reported as reading 50 words per minute.

Rate, rather than accuracy alone, provides more information about performance. For example, students A and B complete fifty math problems with 100% accuracy. Both students achieved similar accuracies; however, a rate measure would reveal student A completed the math problems in one minute while student B worked through the problems in 20 minutes. By using a rate measure, student A is shown to be most proficient at math. The response rate measure, the standard mark of achievement for Precision Teachers, provides more information about a student’s academic progress. Thus, rate alone is not a useful measure of academic behaviour or technical skills, whereas rate plus accuracy is useful.

Although the rate metric provides more information about performance, it reveals only how much behaviour was generated during one period of time. To obtain the *most* information about performance, Precision Teachers look for *change* in response rates across days, weeks, months, and years. Precision Teachers look for differences in response rates across time; showing a measure of learning (i.e., celeration). Response rates showing an upward trend over time are said to be accelerating, while rates showing a downward trend are said to be decelerating (Calkin, 2005). By adding a time dimension and using rate as the standard metric, Precision Teachers gain the most sensitive measure of learning (Binder, 1996).

Using operant conditioning procedures, precision teachers reinforce correct behaviour. For example, students may receive verbal praise after each correct response, or a token after every 20th correct response. The praise and tokens may serve as feedback for students correct responding.

Precision Teachers' chart response rates on a Standard Celeration Chart (SCC). In contrast to a linear scale, the SCC provides a logarithmic scale, an important feature to Precision Teachers. To illustrate, if Max improves his Arabic from one to two words, he doubled his word count; this is identical to growing from 50 to 100, not from 50 to 51. The SCC allows Precision Teachers to chart students' progress in multiplications, rather than in an arithmetic fashion. In addition, the SCC provides standard, appropriately sized graph for analysing the effectiveness of instruction (White, 1986).

Precision Teachers are committed to changing the type of instruction based upon student behaviour (i.e., "the child knows best"). The data on the SCC is used to guide the appropriate type of instruction. For example, if a student's rate of behaviour is accelerating, the type of instruction is considered appropriate. However, if the rate is not increasing or is slow, the instructional type is inappropriate and must be changed.

Overall, Precision Teachers record response rate data on the observable behaviour of their students. Further, Precision Teachers chart the response rate data on the SCC and monitor a student's progress based upon the charted data. These five characteristics of Precision Teaching help teachers guide students learning from the initial stages of acquisition and fluency-building, towards the final mastery stages of maintenance and generalization (Haring & Liberty, 1978; White & Haring, 1976).

Learning advances through predictable stages (Bryan & Harter, 1899), including acquisition, fluency, maintenance, and generalization (Alberto & Troutman, 2003; Haring & Liberty, 1978; White & Haring, 1976). During acquisition, behaviour is characterized as a mixture of accurate and inaccurate responses. The goal is to reduce errors and generate accurate responses. The fluency stage is characterized by fast and accurate responding (Haring & Liberty, 1978). Precision Teaching studies combine the acquisition and fluency stages by encouraging accurate and fast responding using rate-building procedures (Kubina & Wolfe, 2005). Rate-building describes any method that increases response rate (Binder, 1996) and is a necessary component in Precision Teaching studies.

The rate-building procedure

Precision Teaching studies set goals to build the rate of behaviour, goals differ across studies. Participants sometimes receive a relative goal (e.g., go as fast as you can, do your best, or go faster than yesterday; Binder, 1988; Chiesa & Robertson, 2000; Young, West, & Crawford, 1985) while some participants receive a quantitative goal (e.g., write your name 30 times per minute; Kessissoglou & Farrell, 1995; McDowell & Keenan, 2002; Shimamune & Jitsumori, 1999). Some studies combine relative and quantitative goals (Chapman, Ewing, & Mozzoni, 2005; Fox & Ghezzi, 2003; Hughes, Beverley, & Whitehead, 2007). There is a lack of research that compares whether a relative or quantitative goal impacts performance differently. However, goal setting researchers have shown that setting quantitative and relative goals both improve performance (Lock & Latham, 1990; Seijts & Latham, 2001). A lack of research comparing types of goals prevents any conclusions to be drawn on their effectiveness at encouraging high rates of behaviour.

The aim of the rate-building procedure is to encourage participants to respond as fast and accurately as possible until a targeted response rate is achieved (i.e., Performance Standards or Aims; Kubina & Wolfe, 2005). It is argued by Precision Teachers that achieving performance standards ensures maintenance and generalization (Binder, 1996, 2003; Lindsley, 1991). Maintenance is characterized by response rates that are the same after periods of no practice (i.e., retention; Berens, Boyce, Berens, Doney, & Kenzer, 1986) and that remain at the same rate for longer durations than used during training while in the face of distractions (i.e., endurance; Binder, Haughton, & Van Eyk, 1990). The generalization stage is characterized by responding that can quickly be incorporated into larger repertoires (i.e., application; Binder, 1996; Haughton, 1972). These learning outcomes are better known under the acronym REAPS, retention, endurance, application, and performance standards, respectively. Once responding shows retention, endurance, and application, behaviour is said to be *fluent* (e.g., at an expert skill level; Binder, 1996). A sizable amount of Precision Teaching literature has shown achieving performance standards leads towards improved retention, endurance, and application; however, some researchers have questioned these findings.

To illustrate, Hughes et al. (2007) used a group design to compare the effects of a rate-building and "teaching as usual" (TAU) on the retention, endurance, and application of vocabulary words. Participants in the rate-building group learned vocabulary words until a performance standard of 120-180 words per minute was achieved. Participants in the TAU group read from a preferred book with an assistant. Results showed the participants who achieved aims outperformed the control group in the areas of retention, endurance, and application. The authors pointed out participants achieving performance standards also required more practices; questioning whether the response rate or extra practice led towards retention, endurance, and application. Similar procedural confounds have been described by Precision Teachers (Bucklin, Dickinson, & Brethower, 2000; Chapman et al., 2005), further questioning whether achieving the performance standards or undertaking the extra practices ensures learning outcomes.

Problem controlling practice effects in rate-building studies

Extra practice past 100% mastery is called *overlearning* (Driskell, Willis, & Copper, 1992) and has been shown to enhance retention (Gillespie, 2002; Postman, 1962; Rohrer, Taylor, Pashler, Wixted, & Cepeda, 2005). For example, Postman (1962) required participants to recite several lists of 12 two-syllable nouns until 100% correct (i.e., mastery) was achieved. Participants were then placed into one of three groups, 0%, 50%, or 100%. The 0% group was required to reach mastery, no extra practice was required. The 50% group received extra practice equivalent to half the number of trials required to achieve the initial criterion. The 100% group practiced twice as many trials as it was required for them to reach the initial mastery criterion. Following a seven day retention interval, accuracy on reciting the lists improved across the 0%, 50%, and 100% overlearning groups. This finding demonstrates that retention was enhanced by extra practices past mastery (i.e., the overlearning effect).

The results by Postman (1962) and similar outcomes (Gillespie, 2002; Rohrer et al., 2005) suggest that number of practices must be controlled while investigating effects of rate-building in Precision Teaching studies. For example, a rate criterion of 60 responses per minute may show improved retention accuracy over an aim of 30 responses per minute. However, the former performance standard may necessitate more practices than the latter. This example shows how achieving performance standards in Precision Teaching studies may not be the critical component ensuring retention, but rather number of practices.

Problem controlling reinforcement effects in rate-building studies

In addition to practice, reinforcement rate has been shown to enhance the learning outcomes associated with rate-building procedures (Odum, Shahan, & Nevin, 2005). Odum et al. (2005) used a multiple Variable Interval-Delayed Matching-to-sample (VI-DMTS) task with pigeons to ask whether greater rates of reinforcement improved retention accuracy. Results demonstrated greater rates of reinforcement improved matching accuracy. Other studies have come to similar conclusions (Nevin & Grace, 2005), questioning whether reinforcement rate or response rate account for the outcomes of using rate-building procedures in Precision Teaching studies.

Taken together, Postman (1962) and Odum et al. (2005) showed that the number of practices and reinforcement rate enhance retention. These findings suggest that number of practices and reinforcement rate may account for the enhanced retention, endurance, and application observed in Precision Teaching studies, questioning the importance of achieving performance standards. This point was highlighted in a review of the Precision Teaching literature (Doughty, Chase, & O'Shields, 2004).

Doughty et al., (2004) reviewed 48 Precision Teaching studies that used rate-building procedures. They showed 45 studies lacked procedural controls for amount of practice and/or reinforcement. Doughty concluded there is insufficient empirical evidence to support Precision Teaching's claim that rate-building alone leads towards improved retention, endurance, and application. Of the 48 reviewed studies, however, three successfully controlled for reinforcement and number of practices (Evans, Merger, & Evans, 1983; Evans & Evans, 1985; Shirley & Pennypacker, 1994).

Evans et al. (1983) had three different groups practice consonant-vowel-consonant (CVC) trigrams towards performance standards of 80, 60, or 40 sounds per minute (SPM). After the 80 SPM was achieved, the 60 and 40 groups practiced slowly until achieving the same number of practices as the 80 group; this was done to control for number of practices. To control for reinforcement rate, praise was delivered on a fixed time schedule of 60 s (FT-60 s) across experimental groups. After each group achieved their individual aim and completed the same number of practices, post-tests were administered across five days. Results showed the 80 SPM group demonstrated the highest rate of corrects. There was little difference between groups during the post-tests. Evan et al. stated this finding was because greater response rates needed to be achieved by participants to show a difference in accuracy during post-tests. Thus, in a follow-up experiment using similar procedures, Evans and Evans (1985) encouraged participants to achieve performance standards of 60, 90, or 120 SPM. It was found participants from the 90 SPM group showed the highest rate of correct trigrams during post-tests rather than the 120 group. This finding suggests a relationship between response rate and later progress on CVC trigrams. These studies, however, did not test for all the purported fluency outcomes (e.g., retention and endurance).

Shirley and Pennypacker (1994) used a single-case experimental design to compare the effects of rate-building on retention of spelling words. Participants were required to write a list of 10 spelling words to 100% accuracy with or without the addition of a performance standard. Corrects and incorrect responses were held constant across groups. Results showed a small favourable result for rate-building from one participant.

Taken together, these three studies reviewed by Doughty et al., (2004) show that greater response rates did not always improve accuracy and that rate-building improved retention accuracy on some occasions (Evans et al., 1983; Evans & Evans, 1985; Shirley & Pennypacker, 1994). Three studies not reviewed by Doughty have also produced different conclusions on the effects of using the rate-building procedure while controlling for reinforcement rate and practice (Fox & Ghezzi, 2003; Holding, Bray, & Kehle, 2011; Porritt, 2007; Porritt, Wagner, & Poling, 2009)

Holding et al., (2011) used an alternating-treatment design to compare discrete trials and rate-building on noun labelling. During the training phase, participants were encouraged to "go fast" in order to achieve a performance standard during the rate-building condition or presented the noun and asked, "What is it?" during discrete trial conditions. A variable ratio schedule controlled for reinforcement across experimental conditions. The number of trials required to meet the performance standard during the rate-building condition was the same number of trials presented to the participant during the discrete trial condition. Following the training, participants received additional rate-building practice before post-tests were conducted. Holding et al. stated the purpose of the final phase was to "determine if the intervention was effective when it was administered on its own" (p.171). Post-test results showed Cohen's effect sizes ranged from medium (d = .57) to large (d = 1.9) across participants, suggesting a relatively large effect from the rate-building condition on retention accuracy. These results, however, should be taken with caution because of the extra rate-building practice participants received during the final phase condition.

Fox and Ghezzi (2003) used a group design to investigate the effects of rate-building and type of practice on identification of logical fallacies with 36 undergraduate students. Two groups practiced logical fallacies either using definitions (e.g., definition group) or examples (e.g., example group) aiming at 90% accuracy during the acquisition phase. During the next phase, the groups were further divided, producing four groups; participants practiced definitions or examples with either a performance standard (e.g., rate-building group) or no rate requirement (e.g., practice group). Participants in the rate-building groups were instructed to work as fast and as accurately as possible during their 1-minute timings. Participants received feedback on their response rate, accuracy, and new rate criterion to meet following each timing. Participants in the practice group were asked to respond as fast and as accurately as possible, but were not required to meet a performance standard. These participants did not practice in 1-minute timings, but rather completed the same number of trials as was required for the rate-building groups to achieve the rate criterion. Participants demonstrated greater percentage correct on generalization tests from training under the example practices, but did not show improvement from rate-building. Fox and Ghezzi did not collect response rate data but suggested the lack of effect from rate-building may be due to the response rate being similar across group. There was no specific mention of keeping the rate of reinforcement constant across rate-building and practice groups.

In his PhD thesis (Porritt, 2007) and published study (Porritt et al., 2009), Porritt used an alternating-treatments design to compare the effects of rate-building and rate-controlled conditions on the retention of spatial discrimination performance in pigeons while holding reinforcement rate and number of practices constant. Porritt developed response sequences in pigeons by training spatial discriminations with or without delays imposed between each response. Sequences were learned under two different conditions to generate fast (e.g., No-delay condition) and slow response rates (e.g., Within-chains delay condition) during training. In the no-delay condition, subjects were required to complete four consecutive occurrences of five response sequences within 45 s to meet the performance standard. To control for reinforcement rate, a variable-interval schedule was used across experimental conditions. To control for practices, the sum of correct and incorrect responses were held constant across No-delay and Within-chains delay conditions. Retention tests were conducted 23-hr following the no-delay and within-chain delay training. Porritt (2007) and Porritt et al. (2009) showed that pigeons receiving rate-building displayed enhanced retention accuracy over the rate-controlled group.

In summary, some rate-building studies controlling for reinforcement rate and number of practices have shown enhanced retention (Porritt, 2007; Porritt et al., 2009) and application (Evans & Evans, 1985). However, other studies have failed to show a strong support for the rate-building procedure (Evans et al., 1983, Fox & Ghezzi, 2003; Holding, 2011; Shirley & Pennypacker, 1994). Theses controlling for reinforcement rate and number of practices have not found that rate-building procedures improves retention, endurance, and application (Campbell, 2012; Cohen, 2008; Wheetley, 2005). Differences in methodology may account for the different findings in rate-building studies that control for reinforcement rate and number of practices.

The yoking procedure

One methodological difference which may account for the different findings in the rate-building studies is how extra practices are controlled across experimental conditions. Controlling for extra practice in rate-building studies requires a yoking procedure. In rate-building studies, yoking ensures participants are exposed to the same number of practices across rate-building and rate-controlled conditions. The number of practices required to meet a performance standard during a rate-building condition is the same as the number of practices participants are exposed to during a rate-controlled condition (e.g., a yoking procedure).

In one variation of the yoking procedure a practice is defined as a correct or incorrect response (i.e., all trials). Studies yoking trials have demonstrated rate-building enhanced retention accuracy (Porritt, 2007, Porritt et al., 2009) while others have not found a correlation between response rate and retention (Fox & Ghezzi, 2003; Shirley & Pennypacker, 1994). In another variation of the yoking procedure a practice is defined as only correct responses (i.e., corrects). Studies yoking corrects have shown that performance standards enhance retention accuracy (Evans & Evans, 1985) while others have not demonstrated improved retention, endurance, or application due to greater response rates (Campbell, 2012; Cohen, 2008; Evans et al., 1983; Wheetley, 2004).

There is no research that compares response accuracy based upon the type of yoking procedure employed. Thus, it is presently unclear whether the type of practice yoked (e.g., trials or corrects) accounts for the mixed outcomes in studies using the rate-building procedures while controlling for number of practices and reinforcement rate. One advantage of yoking corrects over trials is that fewer practices are generated, shortening a studies duration. However, the lack of evidence showing benefit to either method of yoking warrants further investigation.

Repeated acquisition procedure

A second methodological difference between studies that have used procedural controls for reinforcement rate and number of practices is the type of task. Some studies use flashcards while others have students write answers on worksheets (Fox & Ghezzi, 2003; Hughes et al., 2007; Shimmamune & Jitsumori, 1999). Flashcards and worksheets may not be well-suited to compare the effects of rate-building because participants can acquire the task very quickly; once the task is learned it cannot be learned again (Berens et al., 2003; Shimmamune & Jitsumori, 1999; Kubina, Young, & Kilwein, 2004), a threat to internal validity (e.g., testing). One procedure that removes this threat to internal validity is repeated acquisition (Baldwin, Chelonis, Prunty, & Paule, 2012).

Repeated acquisition procedure requires subjects or participants to learn a different series of spatially-defined responses (e.g., switches, lever press, nose poke, key pecks) each experimental session (Cohn & Paule, 1995). For example, Bickel, Higgens, and Hughes (1990) required participants to learn a new sequence of 10 responses using three touch-sensitive switches each session. During one session participants learned a sequence of left (L), centre (C), right (R), L, R, C, L, R, C, R, while the sequence for the following session might be R-C-L-C-R-L-C-L-R-C. Requiring participants to learn a new sequence every session allows for re-learning of the same task, removing the threats to internal validity (e.g., testing) discussed in some rate-building studies (Berens et al., 2003; Kubina et al., 2004; Shimmamune & Jitsumori, 1999).

The repeated acquisition procedure has been used successfully to demonstrate the effects of drugs on human (Bickel, Higgens, & Hughes, 1991; Higgins, Woodward, & Henningfield, 1989; Walker, 1981) and non-human (Galizio, McKinney, Cerutti, & Pitts, 2009; Picker & Poling, 1984; Turkkan & Hienz, 1992) responding. During each session, subjects are exposed to an acquisition and performance component. In the Acquisition component, subjects are exposed to a different sequence every session. The same sequence is used every session during the Performance component. Similar procedures have been used to study teaching methods (i.e., rate-building).

Sidman and Rosenberger (1967) stated that by using a repeated acquisition procedure, “one can follow progressive changes in the learning process in an individual as a function of such variables as lesions of the central nervous system, teaching methods, drugs...” (p.467). Investigations into teaching methods have used the repeated acquisition procedure successfully to show the effects of forward and backward chaining (Weiss, 1978) and sequence of instructions (Vaughan, 1985). In addition to these variables, the repeated acquisition procedure has proved to be useful in studying the effects of the rate-building procedure (Porritt et al., 2009). Thus, the present rate-building investigation used the repeated acquisition procedure.

Animal analogues of human responding

According to Porritt et al. (2009), training towards performance standards and testing human participants using the repeated acquisition procedure often require large number of practices and involves lengthy experimental sessions. Porritt suggests that repeatedly exposing participants to these lengthy experimental sessions while ensuring an effective positive reinforcer is available poses practical and sometimes ethical constraints. Others have suggested (Baron, Perone, & Galizio, 1991; Branch, 1991; Palmer & Donahoe, 1991) using non-humans is a sensible alternative to using human participants. Following these suggestions, non-humans were used as subjects in the present investigation.

Precision Teaching studies have demonstrated improvements in retention, endurance, and application (Berens et al., 2003; Binder, 1996; Binder et al., 1990; Haughton, 1972) by establishing performance standards using the rate-building procedure. Procedural confounds (e.g., practice, reinforcement) in studies using human participants have questioned whether performance standards are responsible for improved retention, endurance, and application (Doughty et al., 2004). For all but one thesis (Porritt, 2007) and a published study based on that thesis (Porritt et al., 2009), the outcomes from rate-building studies when procedural confounds have been controlled have been mixed (Campbell, 2012; Cohen, 2008; Evans et al., 1983, Fox & Ghezzi, 2003; Holding, 2011; Shirley & Pennypacker, 1994; Wheetley, 2005), warranting further investigation.

The present study used a repeated acquisition procedure analogous to Porritt (2007) to investigate the effects of performance standards on acquisition and retention. Reinforcement rate was held constant similar to Porritt (2007), however, the type of practice held constant across experimental conditions differed.

The present study defined a practice as a correct response because it is unclear whether a response that is not correct (e.g., an error) counts as a practice. For example, errors having similar topography to the previous response may be a repetition of the previous response, questioning whether the new response was to the stimuli or just repeated. Additionally, some incorrect responses may occur in the absence of stimuli. For example, a child says “blue” when shown a picture of a house and asked to point at the door. These examples show that a practice can be defined in different ways. This ambiguity makes it unclear what type of response is being yoked across experimental conditions in rate-building studies. The present study defined a practice as a correct response to ensure clarity of practice definition.

The performance standard differed between the present study and Porritt (2007). The performance standard in the present study required subjects to complete five consecutive occurrences of two chain completions, each within 45 s. This performance standard was selected to ensure all hens achieved the aim within a one hour session.

Method

Subjects

The 6 subjects, numbered 21 through 26, were Shaver-Starcross domestic hens (*Gallus gallus domesticus*). At the beginning of the experiment, the hens were two years old, and two of them, Hens 22 and 24, had some experience on ratio schedules of reinforcement; the rest were experimentally naive. The hens were housed individually in home cages (500-mm long × 510-mm wide × 420-mm high), in a ventilated room on a 12-hr light: 12-hr dark cycle. They had free access to water; grit and vitamins were provided weekly. Throughout the experiment all hens had red fleshy combs suggesting good health. Each hen was weighed every day an experimental session took place (approximately six days per week) and they were maintained at 80% (+/-5%) of their free-feeding body weights through feeding of commercial layer pellets.

Apparatus

The experimental chamber (400-mm long, 560-mm wide, 530-mm high) was made of white laminate encased particleboard (20-mm thick). The chamber floor was covered with a thick clear plastic that had black plastic matting on top (400-mm long x 560-mm wide). A food magazine was located on the right-hand wall of the chamber behind an opening (115-mm high × 70-mm wide) that was centered 90-mm above the floor. When operated, the magazine was lit with a clear bulb and raised; giving the subjects access to wheat. Three horizontally spaced (100-mm) keys (30-mm in diameter), which could be lit blue, red, or yellow with a 28 V multi-chip LED bulb were placed above the magazine opening (400-mm from the floor). Each key required a force of approximately 0.2 N to close a micro switch.

All experimental events were controlled and recorded by Med-PC® IV software run on a Dell Optiplex GX110. Summary data for each session were also manually written into a data book at the end of each session.

Procedures

Keypeck training. Experimental sessions were conducted daily at about the same time. Hens were initially trained to peck all three response keys using an autoshaping procedure similar to that of Brown and Jenkins (1968). At the start of each trial, at random, a left, centre, or right key was illuminated in blue, red, or yellow for 6 s. When either a keypeck occurred or 6 s had elapsed, the magazine was raised for 3.5 s. Keypecks to non-illuminated keys (i.e., dark-key pecks) did not produce any consequences, but were recorded. A 40-s intertrial interval (ITI) separated the lowering of the food magazine and the start of the next trial as suggested by Gibbon, Baldock, Locurto, Gold, and Terrace (1977). Experimental sessions ended following the 45th reinforcer. After subjects were responding to all three keys irrespective of colour illumination, Phase I began.

Phase I. Phase I procedures were similar to keypeck training with one exception. During Phase I, only a response to the lit key (i.e., correct key) raised the magazine for 3.5 s (i.e., an FR1 schedule of reinforcement). Dark-key pecks (e.g., errors) did not produce any consequences, but were recorded. A 5-s ITI separated the lowering of the magazine and the next keylight presentation. Each session ended following the 45th reinforcer. Phase II began after total errors for each session fell below three for all subjects. Hen 24 gradually stopped pecking all response keys and was no longer eating wheat from the magazine. She began pecking keys after supplemental feed was changed to wheat for two days.

Phase II. Conditions during Phase II were similar to the previous phase with one exception. A second keylight (i.e., distracter key) was illuminated at the same time and with the same colour as the correct key during each trial. Distracter key position was randomly selected between the two remaining key positions and pecks to it did not provide a consequence, but were recorded. Each session ended following the 45th reinforcer. Phase III began after total errors for each session fell below three for all subjects.

Phase III. During Phase III, two distracter keys were used in each trial. All other conditions during this phase of training remained the same as Phase II. Phase IV began after number of errors stabilized over 5 sessions as determined by visual inspection of graphed number of errors.

Phase IV. During Phase IV, subjects were required to complete three spatially-defined responses (i.e., a three-link behaviour chain) for magazine access (i.e., an FR 3) in each trial (see Figure 1.1.A for procedural outline). During the first link all three keys were illuminated in blue and a response to the designated correct key (e.g., left) immediately darkened all three keylights, advancing the chain to the next link. The keylights were then immediately illuminated in red and the key designated as correct changed (e.g., to the centre). A correct response again darkened the keys and advanced the schedule to the third and final link, in which all three keylights were illuminated in yellow and one of these (e.g., right) being designated as correct.

If a subject pecked an illuminated key not designated as correct during any link (i.e., error), all keylights were darkened for 1 s. During this blackout period, keypecks did not produce a consequence. After the 1 s, the three keylights were again illuminated with the same colours as before the blackout until a correct keypeck was made for that link.

A peck to the correct key in the third and final link of each chain immediately darkened the keylights and raised the magazine for 2 s. A 5-s ITI separated magazine access and re-presentation of the first link for the next trial. Magazine time was reduced to .9 s on three out of four trials selected at random to maintain body weight. The reduced magazine time resulted in a light flash and clicking sound, but did not allow subjects to consume any wheat. The ITI was removed after 30 sessions in order to make the procedures similar to those of Porritt (2007).

The position of the correct key for each link remained the same throughout each session. The keylight colour presented during each link of the chain schedule remained the same for every session throughout the study. Location of correct keypecks for each link during each session were chosen at random, except that no position could be designated as correct for two consecutive trials within a session and no position-colour combination was repeated across consecutive sessions. Twelve three-link chains were developed within these criteria. All subjects were exposed to a series of twelve chains which were repeated twelve times for a total of 144 training sessions; each lasting 45 minutes. Table 1.1.1 shows the correct key positions for each link of Chains 1-12. Chain numbers are listed in order of presentation for Series 1-12 during Phase IV.

During training, there were some sessions in which subjects made few keypecks or did not complete the chain, producing a very low response rate. A criterion was developed in which if a subject's overall response rate was less than .08 per s; the data were not used in the present analysis. This criterion was determined by calculating the mean response rate across all hens' first exposure to the 12 chain sequences (e.g., Exposure 1). Then, the 25th percentile for the distribution of mean rates was used as the criterion. After removing sessions not meeting the criterion, seven series of twelve chains, were used in the current analysis.

In summary, the auto-shaping procedure lasted for 38 sessions. The number of sessions to complete Phase I was 32. Phase II continued for 12 sessions and Phase III lasted a total of 44 sessions. Phase IV lasted for 144 sessions. Taken together, there were 270 training sessions. Following training, three behaviour chains in which the mean of the last three exposures was above 75% were selected at random to be used in Experiments 1.1-1.8. All three chains were used during each experimental session within each of the three experimental conditions.

Experimental sessions. Each experimental session was separated into three components. The first component of each session (e.g., Retention component) required subjects to complete 15 chains of the same sequence trained 23-hr prior. The function of this component was to provide a 23-hr measure of retention accuracy based upon the previous training condition. The order of components, chains used for each component, and the criterion to end each component is presented in Table 1.1.2.

The second component of each session required subjects to make 75 “distracter chain” completions (e.g., Distracter component). The first link of this distracter chain was the same as the chain sequence used during the Retention component; however, the correct key positions for the second and third links differed. The function of the distracter chains was to break up spatial discrimination performance between the sequence used during the Retention and Training component of each session.

The third component (e.g., Training component) exposed subjects to one of three experimental conditions using an alternating treatment, within-subject design (e.g., A/B/C). Each condition was in effect for three consecutive sessions, after which a new experimental condition began. The order of experimental conditions and chains used in each condition is presented in Table 1.1.2.

No-delay. During No-delay conditions, each hen was required to complete a performance standard for the session to end. The performance standard required subjects to complete two chains within 45 s, termed a bin. Five consecutive bin completions were required before the session ended.

Within-chains delay. During Within-chains delay conditions, a 5-s interval was imposed between a correct keypeck and the illumination of keylights for the next chain link. During this interval, all keylights darkened and responses to darkened keys did not produce any consequences. This condition ended once a subject completed the same number of correct responses for a chain as was needed to achieve the criterion used during the No-delay condition for the same chain (e.g., yoked correct practices). In this manner, the number of correct responses was held constant between experimental conditions for each chain sequence to control the number of practices. To control for reinforcement rate, wheat was available for 2 s after a variable interval of 50 s (i.e., a VI-50-s schedule) had elapsed following a chain completion.

Between-chains delay. During Between-chains delay conditions, a 15-s interval was imposed between each chain completion. During this interval, all keylights darkened and responses to darkened keys did not produce any consequences. This condition ended once a subject completed the same number of correct responses for a chain as was needed to achieve the criterion used during the No-delay condition for the same chain (e.g., yoked correct practices).

Summary data that were manually recorded in the data book at the end of each session included the total errors in each component, session time in seconds, and reinforcers delivered. Event data were recorded by Med-PC® using a system of 1's and 0's to represent events and responses that occurred within the chamber. These 1's and 0's were used to calculate percentage correct, latency to respond, response rate, reinforcement rate, and number of practices in each session. All statistical analyses were conducted using Statistical Package for the Social Sciences (SPSS) software. All raw data used in the following analysis, along with the programs used to analyse the data, can be found in the Appendix.

Figure 1.1.A.

Order of events during each link of a behaviour chain during the repeated acquisition procedure.

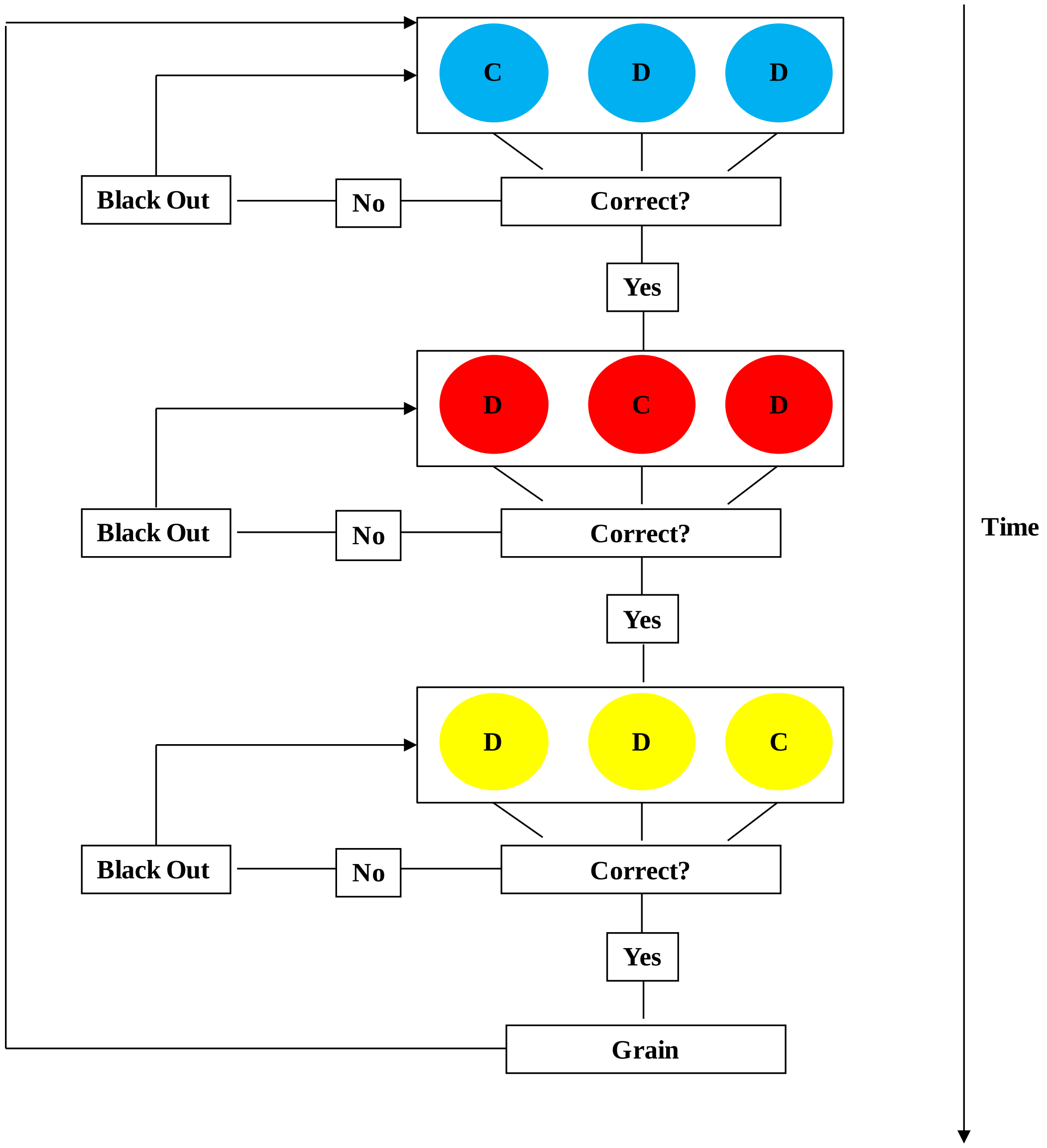


Table 1.1.1.

Correct key position for each link of every chain. Chain numbers are listed in order of presentation for each series during Phase IV.

Chain Number	Link 1	Link 2	Link 3
1	Right	Left	Centre
2	Left	Centre	Right
3	Centre	Right	Centre
4	Right	Centre	Left
5	Left	Right	Centre
6	Centre	Left	Right
7	Left	Right	Left
8	Right	Left	Right
9	Centre	Right	Left
10	Right	Centre	Right
11	Centre	Left	Centre
12	Left	Centre	Left

Table 1.1.2.

Order of components and experimental conditions for each session of Experiments 1.1-1.3, 1.5. Chain sequences used during each component and the criterion to end each component are given.

Component	Experimental Conditions	Chain	Criterion to End
	Session 1		
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	L-C-L	Performance Standard
	Session 2		
Retention		L-C-L	15chains
Distracter		L-R-C	75 chains
Training	No-delay	R-L-C	Performance Standard
	Session 3		
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	R-L-R	Performance Standard
	Session 4		
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	Within-chains delay	L-C-L	Yoked practices
	Session 5		
Retention		L-C-L	15 chains
Distracter		L-R-C	75 chains
Training	Within-chains delay	R-L-C	Yoked practices
	Session 6		
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	Within-chains delay	R-L-R	Yoked practices
	Session 7		
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	Between-chains delay	L-C-L	Yoked practices
	Session 8		
Retention		L-C-L	15 chains
Distracter		L-R-C	75 chains
Training	Between-chains delay	R-L-C	Yoked practices
	Session 9		
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	Between-chains delay	R-L-R	Yoked practices
	Session 10		
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	No delay	L-C-L	Yoked practices

Results

Figure 1.1.1 shows percentage of correct responses (+1 *SD*) in which the mean was calculated across the twelve chains for Exposures 1-7. A response is defined as a peck to only illuminated keys. Correct responses were defined as responses to illuminated keys designated as correct. Percentage correct was calculated by dividing the total number of correct responses in each chain link by the total number of responses in that session. The mean percentage correct data calculated across subjects is shown in the bottom left graph. Percentage of correct responses generally increased across Exposures 1-7 for all subjects.

Figure 1.1.2 shows the mean percentage of correct responses across Exposures 5, 6, and 7 (+1 *SD*) for each chain. Mean percentage correct was variable across chain sequences for all hens. The mean percentage correct data calculated across subjects is shown in the bottom left graph. The horizontal line in the bottom left graph represents the 75th percentile. The plus sign (+) represent the chains selected for the remainder of this study. The sequences in which the mean was above 75% were selected at random.

Group data for Figures 1.1.3-1.1.9 were analysed using a one-way repeated measures analysis of variance for all measures during the Training and Retention components for Experiment 1.1. The alpha level for all statistical comparisons in all situations was set at .05 and any results that reached this level were presented with an asterisk (*) in Table 1.1.3. Except where indicated with a hashtag (#) in Table 1.1.3, Mauchley's Test was not significant so sphericity was assumed. In these instances, and for Experiments 1.2-1.8, Greenhouse Geisser correction was used. Post-hoc tests were conducted using the Bonferroni correction, as recommended by Fields (2005).

Figure 1.1.3 shows mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. Percentage correct was calculated by dividing the total number of correct responses in each chain link during the Training or Retention components by the total number of responses in each chain link for that component. The mean percentage correct data calculated across subjects is shown in the bottom left graph. Generally, accuracy in the Training component was similar during the No-delay and Between-chains delay conditions, and lower under the Within-chains delay condition, for all subjects. Table 1.1.3 shows the overall effect was significant and effect size, partial eta squared, was moderate (Ferguson, 2009). None of the pairwise comparisons were significant. Retention accuracy was similar across the three experimental conditions for all subjects, Table 1.1.3 shows the data from these conditions were not significantly different.

Figure 1.1.4 shows the mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. Response rates were calculated by dividing the total number of responses emitted in each chain link for the Training or Retention component by that components duration. The mean response rate data calculated across subjects is shown in the bottom left graph. Training response rates were greatest during the No-delay condition, and lowest during the Within- and Between-chains delay conditions, for all subjects. Table 1.1.3 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention response rates were generally similar across the three experimental conditions for all subjects; showing no systematic effect from training conditions.

Figure 1.1.5 shows the mean correct response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. Correct response rates were calculated by dividing the total number of correct responses emitted in each chain link for the Training or Retention component by that components duration. The mean correct response rate data calculated across subjects is shown in the bottom left graph. Correct response rates during the Training component were greatest during the No-delay condition, and lowest during the Within- and Between-chains delay conditions, for all subjects. Correct response rates during the Retention component were generally similar across the three experimental conditions for all subjects; showing no systematic effect from training conditions. Table 1.1.3 shows the Training component findings were significantly different, whereas no significant differences were found between the Retention component conditions.

Figure 1.1.6 shows the mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and

Retention components for all subjects. Response latency represents the duration from the illumination of the keylights to the emission of a response. The mean response latency data calculated across subjects is shown in the bottom left graph. Response latency in the Training component was greatest during the Within-chains delay condition, and lowest during the Between-chains delay condition, for all subjects. Table 1.1.3 shows this finding was significant. Retention latencies were similar across all experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.1.3).

Figure 1.1.7 shows the mean correct response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. Correct response latency represents the duration from the illumination of the keylights for each chain link to the emission of a correct response for that link. The mean correct response latency data calculated across subjects is shown in the bottom left graph. Correct response latency in the Training component was greatest during the Within-chains delay condition, and lowest during the Between-chains delay condition, for all subjects. Retention latencies were similar across all experimental conditions for all subjects. Table 1.1.3 shows Training component finding were significantly different, whereas no significant differences were found between Retention component conditions (Table 1.1.3).

Figure 1.1.8 shows the mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition. Reinforcement rate was calculated by dividing the total number of grain presentations in each session by that sessions' duration. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph. A VI-50-s schedule would give a value of .02 reinforcers per second on the y-axis. In all cases, the means were less than this and were similar across each experimental condition for all subjects. Reinforcement rate was similar across each experimental condition for all subjects. Table 1.1.3 shows no significant differences across conditions.

Figure 1.1.9 shows the mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component. The left three bars on each graph shows the number of practices, when defined as corrects only. Correct practices were calculated by adding the number of correct keypecks during the Training component of the No-delay, Within- and Between-chains delay conditions. The right three bars on each graph show the number of trial practices. These trial practices were calculated by adding the total correct and incorrect responses during the Training component of the No-delay, Within- and Between-chains delay conditions. The mean number of practices data calculated across subjects is shown in the bottom left graph. The mean number of correct practices was similar across experimental conditions for all subjects. Table 1.1.3 shows no significant differences and no variance between conditions for all subjects. The mean number of trial practices was generally greatest during the Within-chains delay condition, and was lowest during the No-delay and Between-chains delay conditions, for all hens. Table 1.1.3 shows this finding was significant. Total number of practices for each experimental condition was generally greater when defined as trials.

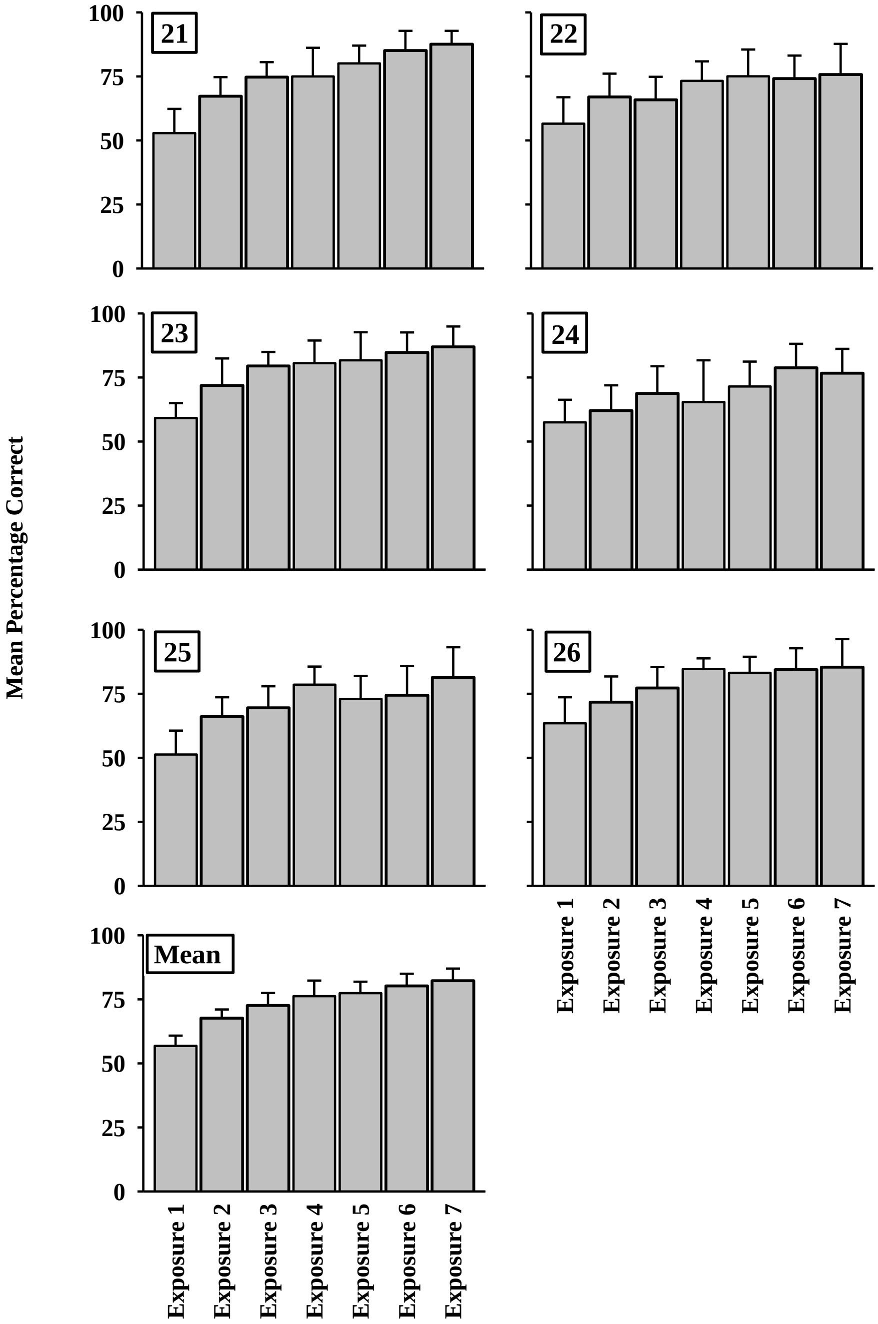


Figure 1.1.1. The mean percentage of correct responses (+1 *SD*) across the twelve chains for Exposures 1-7. The mean percent correct data calculated across subjects is shown in the bottom left graph.

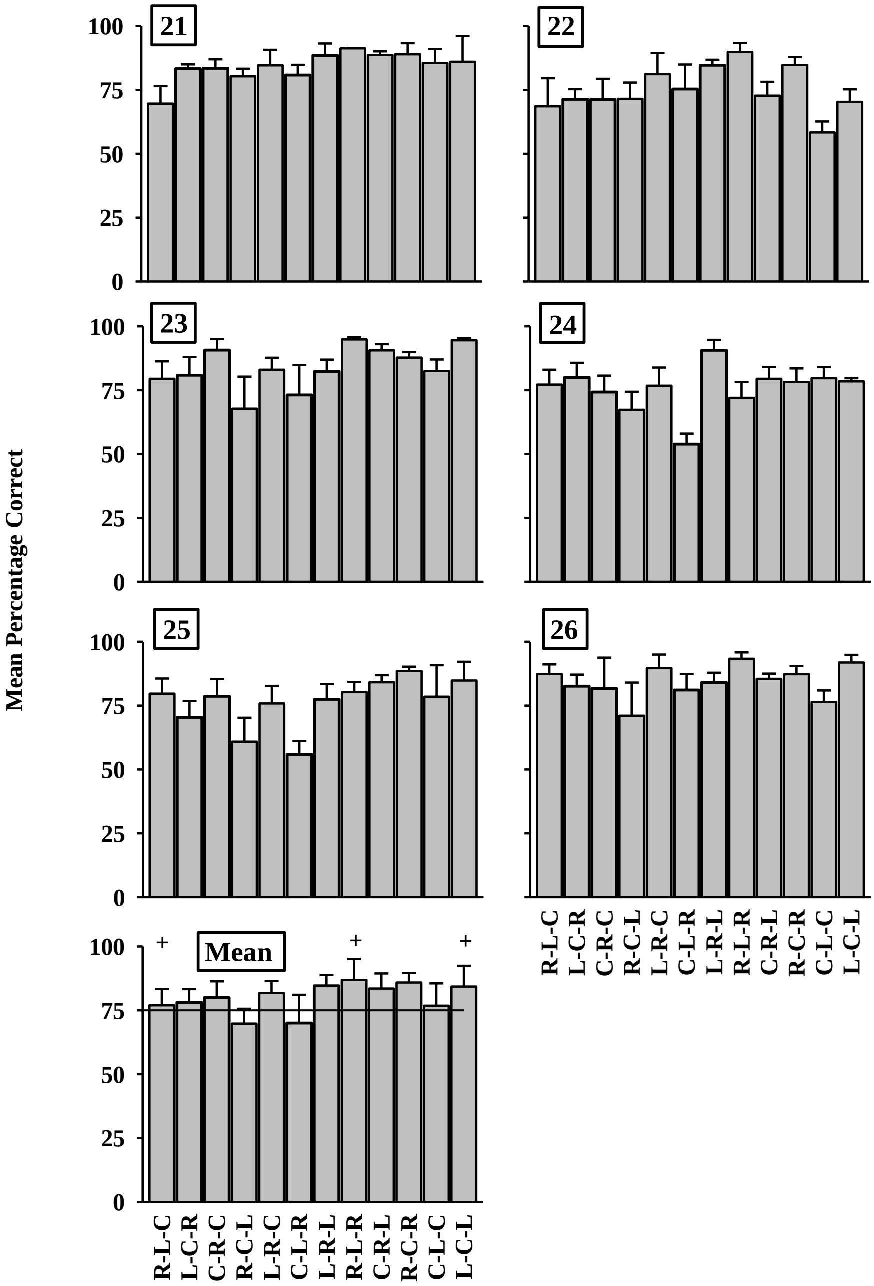


Figure 1.1.2. The mean percentage of correct responses across Exposures 5, 6, and 7 (+1 *SD*) for each chain. Mean percentage correct was variable across chain sequences for all hens. The mean percent correct data calculated across subjects is shown in the bottom left graph.

Table 1.1.3

Analysis of variance results during the Training and Retention components of Experiment 1.1.

Component	MS Treatment	MS Error	df	F	p	Partial Eta Squared
			Mean percent correct			
Training	256.34	59.28	2, 10	4.32	.04*	.46
Retention	45.63	29.66	2, 10	1.54	.26	.24
			Mean response rate			
Training	.32	.007	2, 10	48.22	.001*#	.91
Retention	.002	.001	2, 10	1.33	.31	.21
			Mean correct response rate			
Training	.16	.006	2, 10	27.46	.003*#	.85
Retention	.003	.001	2, 10	2.98	.10	.37
			Mean response latency			
Training	6.61	.30	2, 10	21.77	p<.0001*	.81
Retention	.04	.11	2, 10	0.33	.73	.06
			Mean correct response latency			
Training	4.08	.08	2, 10	53.12	p<.0001*	.91
Retention	.04	.08	2, 10	0.50	.62	.09
			Reinforcement rate			
Session	3.8E-7	3.4E-7	2, 10	1.10	.37	.18
			Number of practices			
Correct	--	--	2, 10	--	--	--
Trial	493.19	109.90	2, 10	4.49	.04*	.47

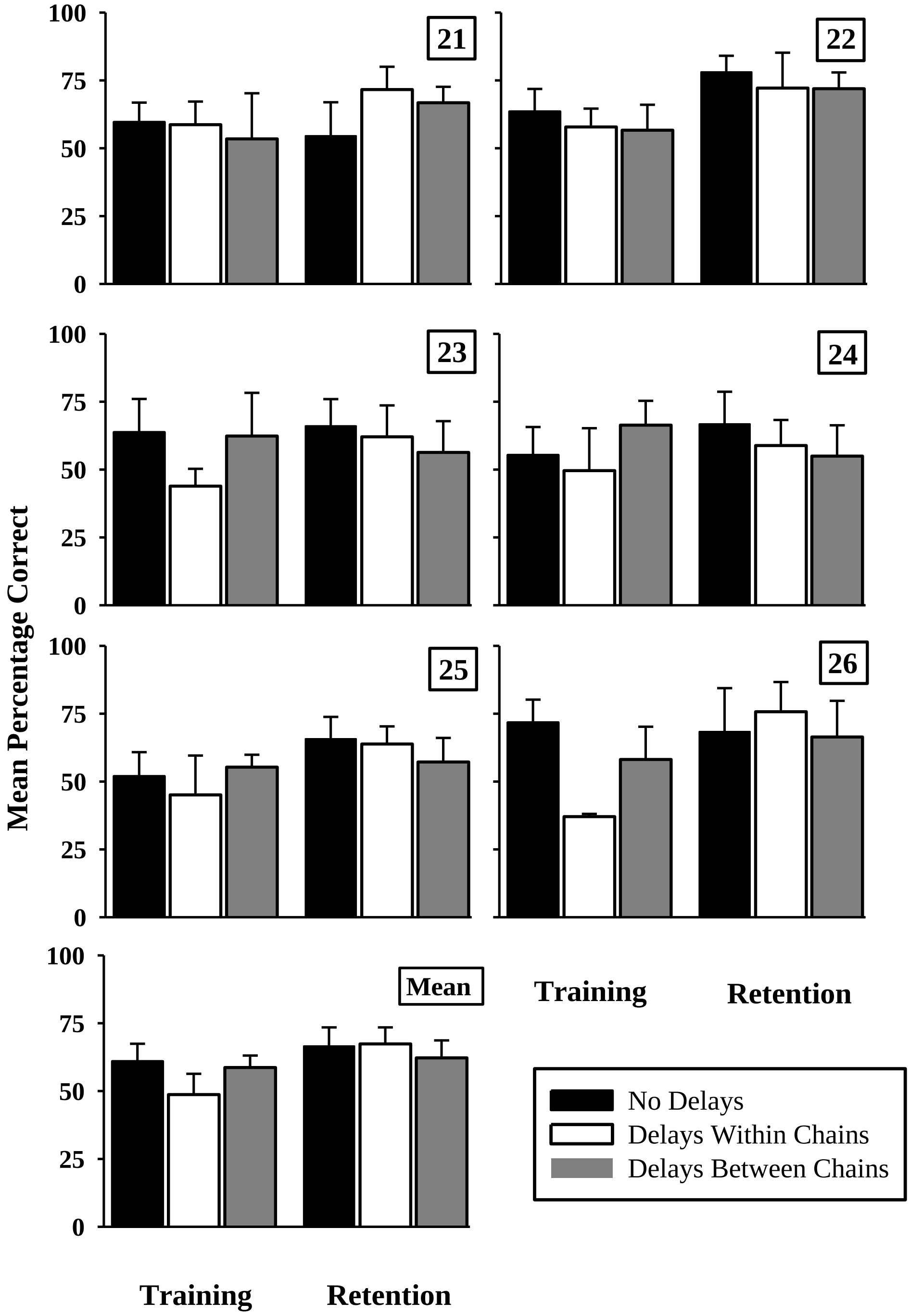


Figure 1.1.3. The mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean percent correct data calculated across subjects is shown in the bottom left graph.

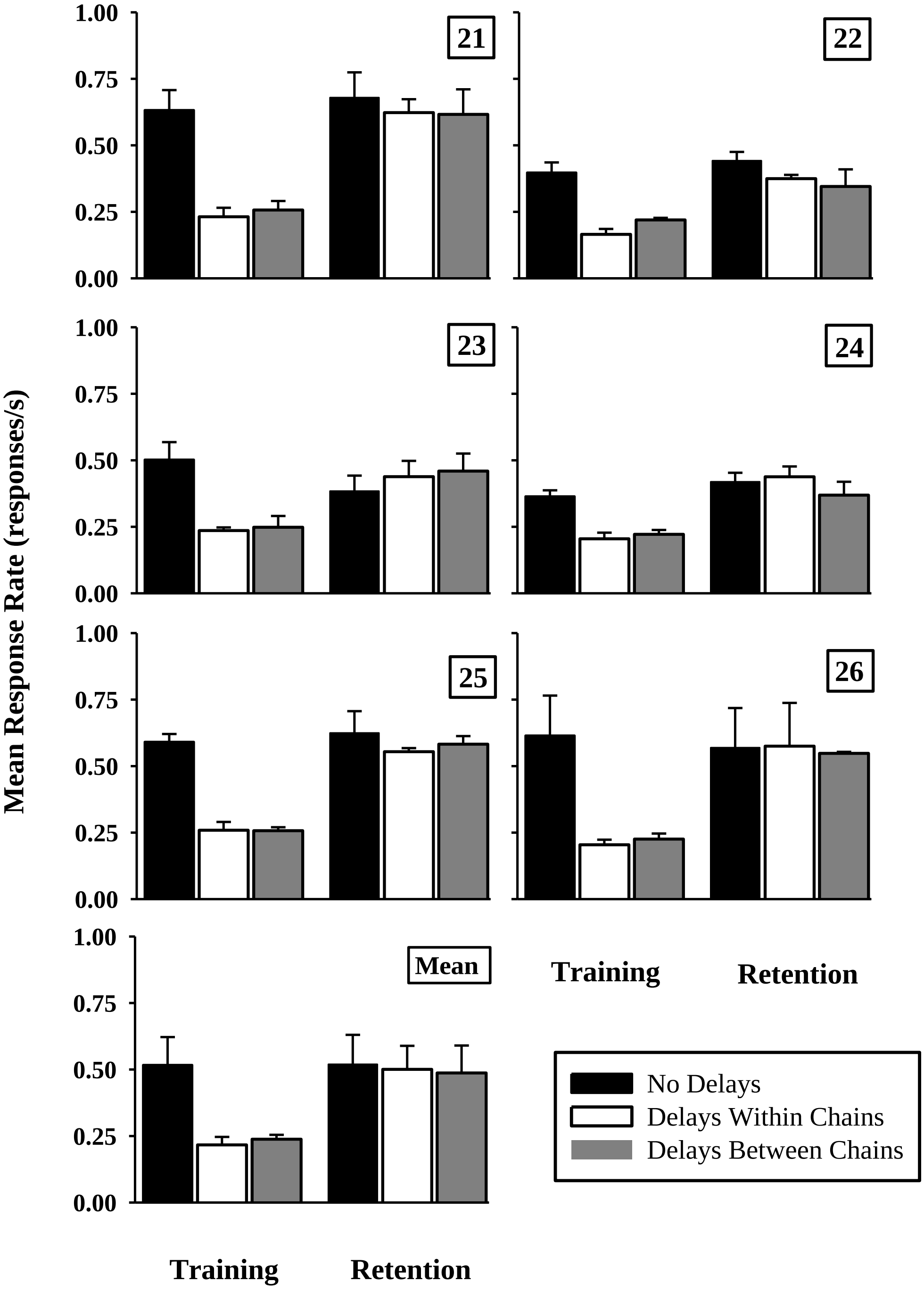


Figure 1.1.4. The mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response rate data calculated across subjects is shown in the bottom left graph.

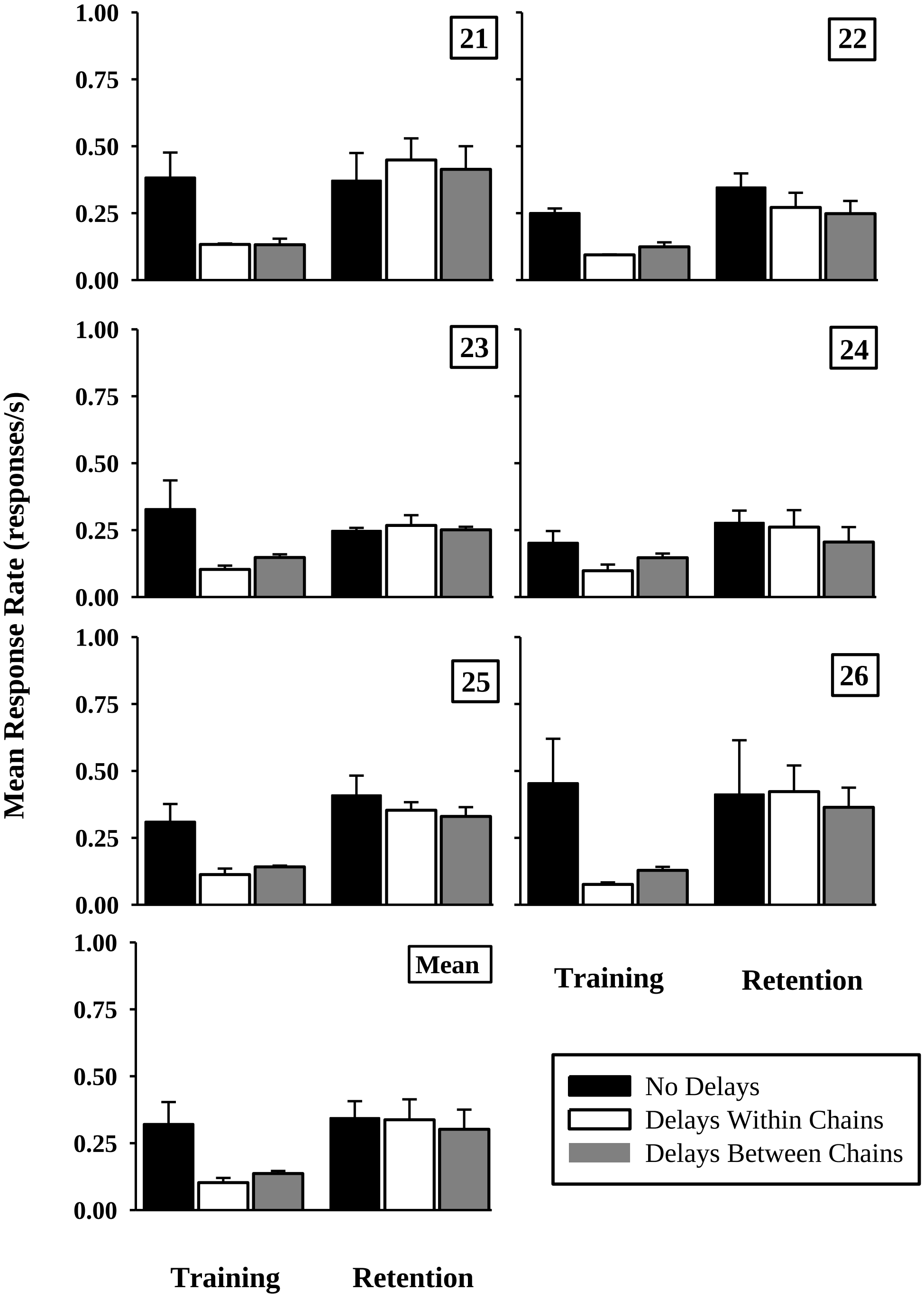


Figure 1.1.5. The mean correct response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean correct response rate data calculated across subjects is shown in the bottom left graph.

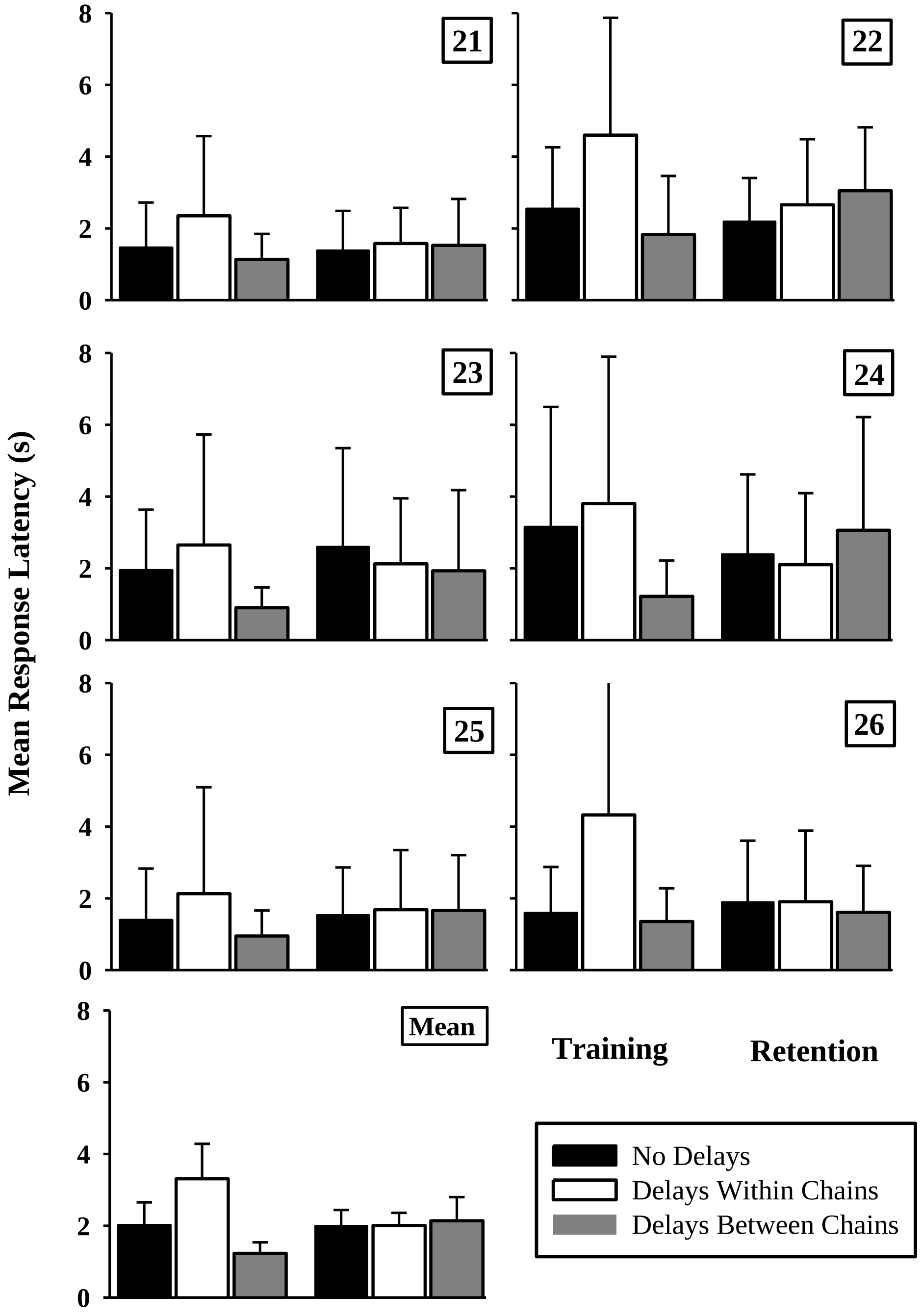


Figure 1.1.6. The mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph.

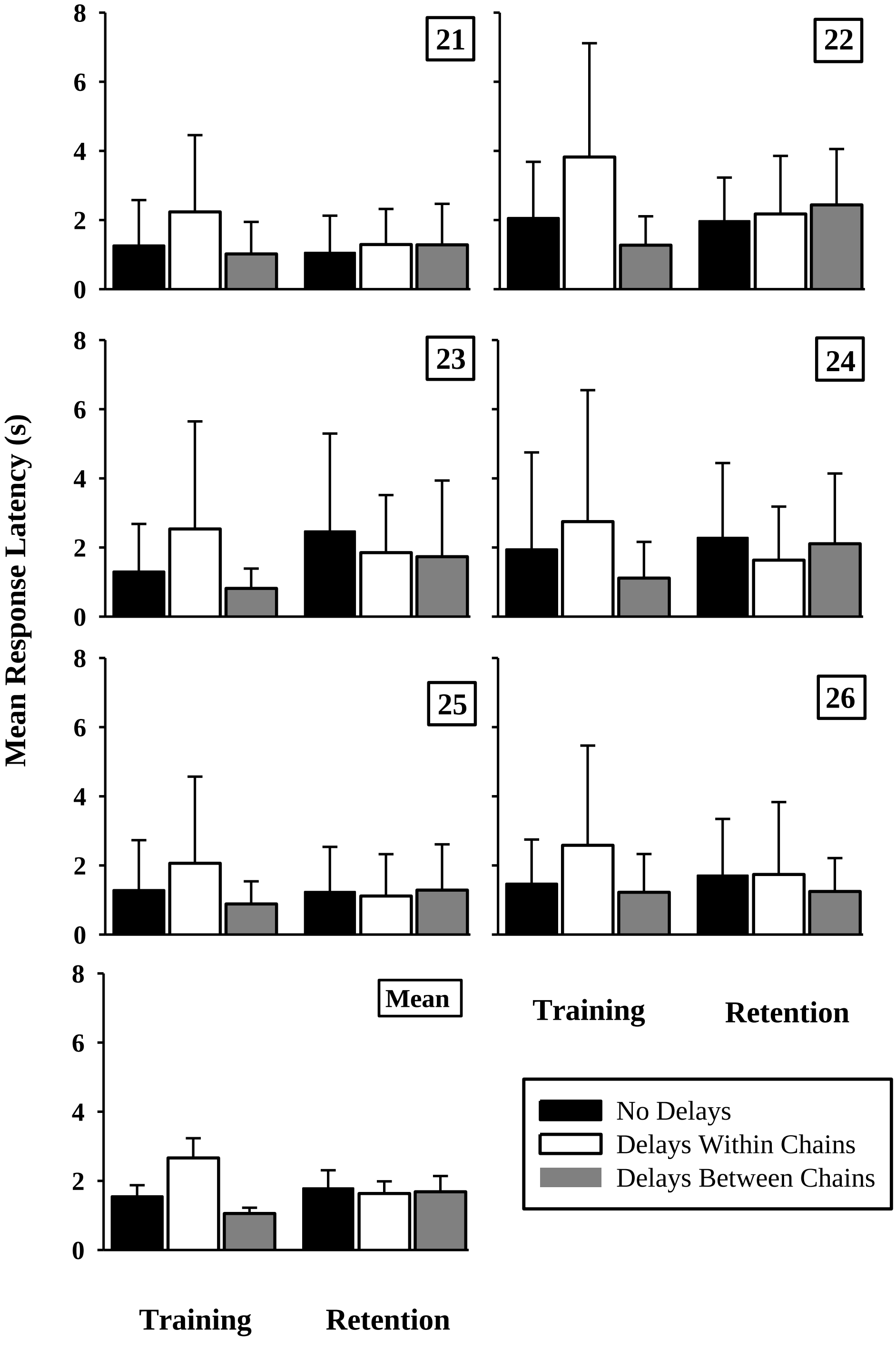


Figure 1.1.7. The mean correct response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean correct response latency data calculated across subjects is shown in the bottom left graph.

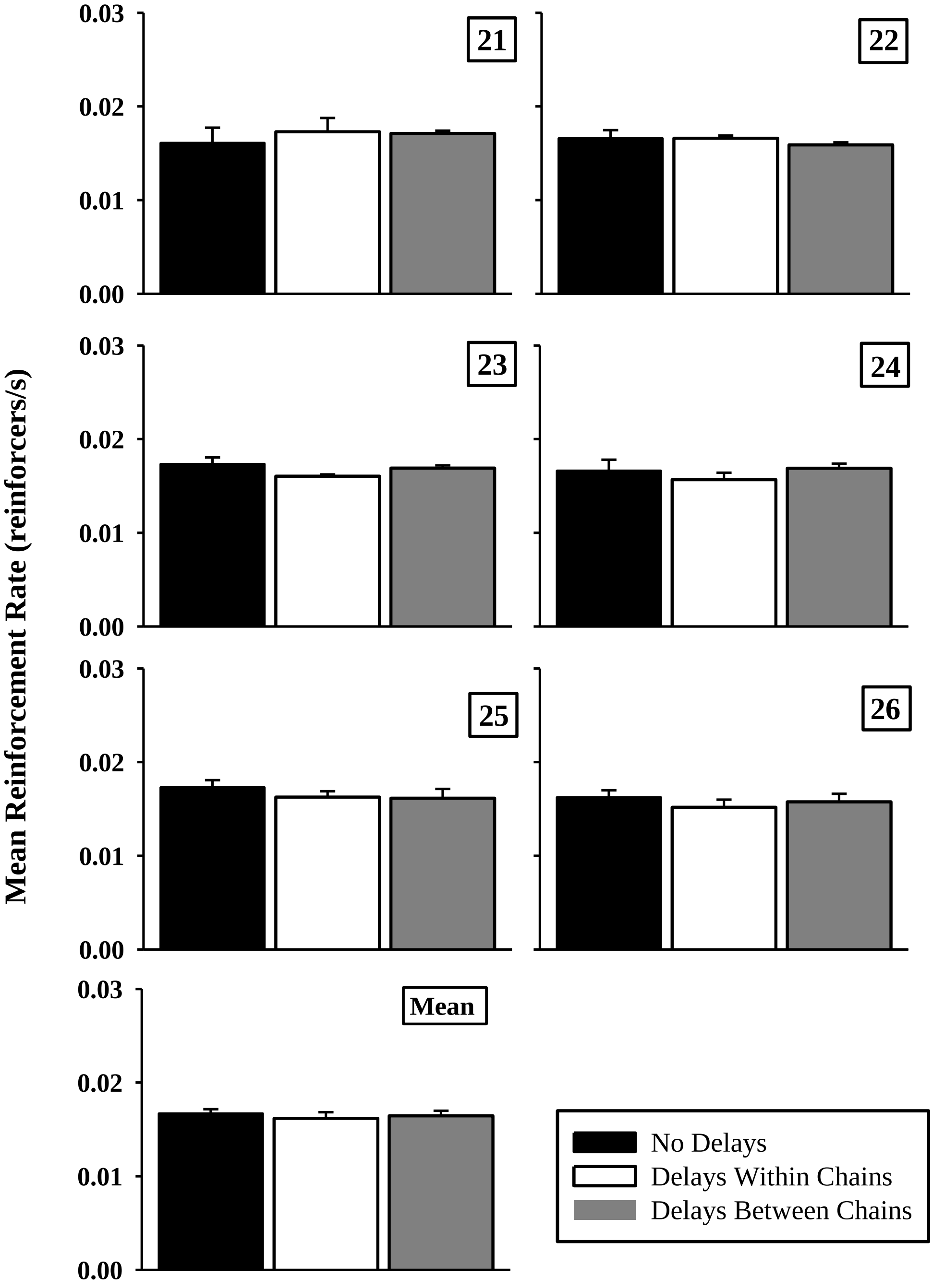


Figure 1.1.8. The mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

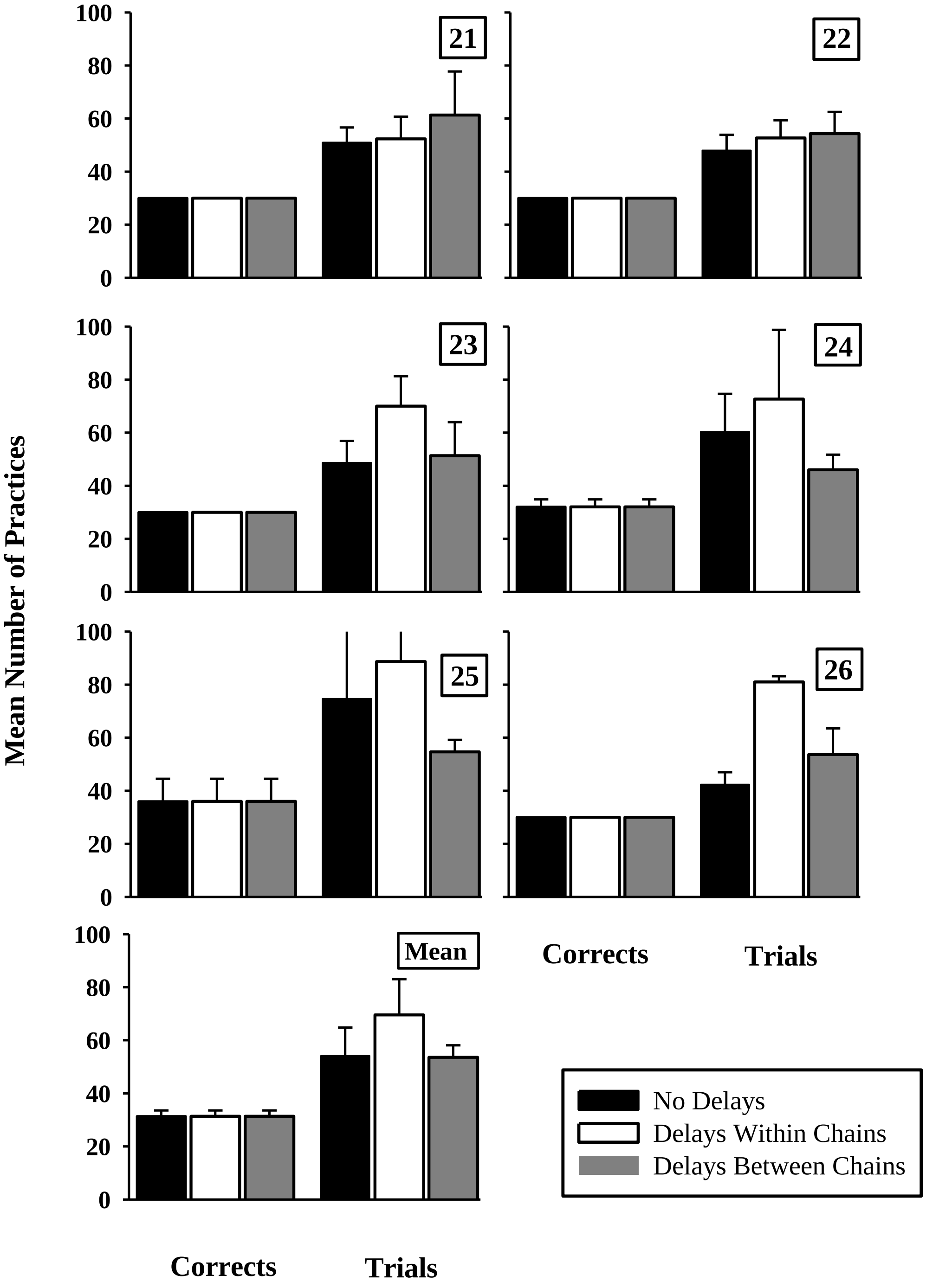


Figure 1.1.9. The mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component. The left three bars on each graph shows the number of practices when defined as corrects only. The right three bars on each graph show the number of practices when defined by trials. The mean number of practices data calculated across subjects is shown in the bottom left graph.

Discussion

The purpose of training was to ensure all subjects learned the 12 response sequences and to determine which sequences would be appropriate to use in the present study. Percentage correct increased as hens were repeatedly exposed to the chain sequences; showing that all subjects acquired the 12 chain sequences. The criterion ensured all hens had achieved 75% accuracy on the three chains used in the present study (i.e., R-L-R, L-C-L, and R-L-C).

Experiment 1.1 replicated the procedures from Porritt (2007) to compare results. The present study successfully held reinforcement rate and number of correct practices constant across experimental conditions. This finding suggests the methods used in the present experiment were similar to Porritt (2007). It was expected using similar methods to Porritt would yield similar results.

Similar to Porritt, mean percentage correct and mean response rates were greatest during the No-delay condition of the Training component when compared to the same means for the other two delay conditions in the same condition. In addition, mean response latencies were greatest during the Within-chains delay training condition when compared to the same means for the other two delay conditions in the same component. Similar results were found for the mean correct response rates and the mean correct response latencies. All three results from the Training component were similar to findings from Porritt, suggesting results during the Retention component of the present study should be similar.

Mean percentage correct, mean response rates, and mean latencies during the Retention component of the present study were similar across the three experimental conditions; a finding contrary to the results from Porritt (2007). Similar results were found for the mean correct response rates and the mean correct response latencies during the Retention component. The present study replicated only the Training component findings from Porritt, methodological differences may account for the partial lack of replication.

One difference in methodologies between the present study and Porritt was the performance standard. As previously mentioned [Experiment 1.1 Introduction], the present study required subjects to complete five consecutive bins of two chain completions, each within 45 s; this criterion was selected to ensure all hens achieved the performance standard within a one hour session.

Porritt (2007) required subjects to complete four consecutive bins of five chain completions, each within 45 s. Thus, the major difference in criterion is in the number of chain completions per bin.

The performance standard used by Porritt (2007) ensured more chain completions were made by subjects within each bin than the present study. Greater performance standards have been shown to improve accuracy during training and retention (Berens et al., 2003; Ivarie, 1986). For example, Ivarie (1986) conducted between group comparisons to study the effects of two different performance standards on accuracy and retention of writing Arabic numerals. Participants in the high rate group maintained 70 responses per minute with seven or less errors for three consecutive timings, participants in the low rate group were required to maintain 35 responses per minute with four or less errors for three consecutive timings. Results demonstrated that higher rates of writing numerals produced greater accuracy and retention; similar results were found by Berens et al. (2003).   This outcome suggests that increasing the performance standard in the present study should increase accuracy during the No-delay condition of the Training and Retention components, producing retention accuracies across experimental conditions similar to those of Porritt.

Another difference between the methods used in the present study and those of Porritt (2007) was how a practice was defined. As previously mentioned [Experiment 1.1 Introduction], the present investigation defined a practice as a correct response, while Porritt defined a practice as a correct or incorrect response (i.e., a trial).

The present data showed that yoking corrects, compared to yoking trials, generated less overall practice opportunities during the Within- and Between-chains delay conditions. On the one hand, as previously mentioned [Experiment 1.1 Introduction], yoking corrects removes the ambiguity of error definition. On the other hand, yoking trials across experimental conditions does not ensure for correct responses during the Within- and Between-chains delay condition, arranging the possibility for more incorrect responses during these yoked conditions. Thus, yoking trials may reduce accuracy during the Within- and Between-chains delay conditions, producing results similar to Porritt. The difference in how a practice is defined may be another reason for the failure to replicate the Retention component results of Porritt.

In the present study, the mean response latencies and the mean response rates were greater than the mean correct response latencies and the mean correct response rates in the Training component. These findings are not surprising because calculating rate using both correct and incorrect keypecks generates more responses over time than calculating corrects only. The manner in which rate and latency were calculated did not change the conclusions drawn from each experimental condition for the present study. Thus, the mean correct response rate and the mean correct response latency will not be calculated in Experiment 1.2.

Overall, Experiment 1.1 used similar methods to the ones used by Porritt (2007) and obtained similar findings during the Training component. The performance standard and type of practice yoked in the present experiment differed from those used by Porritt; either of these two differences in methodology may account for the failure to replicate Porritt's findings during the Retention component. To investigate these differences, it was first decided to examine the effect of performance standards. The performance standard was investigated before the yoking procedure because there is evidence to suggest that response rate is directly related to performance outcomes (Berens et al., 2003; Ivarie, 1986). Experiment 1.2 replicated the performance standard used by Porritt with the aim of comparing findings.

EXPERIMENT 1.2

Experiment 1.2 replicated Porritt's (2007) methods by increasing the performance standard with the aim of comparing results.

Methods

Subjects

The same 6 subjects from Experiment 1.1 participated in this study.

Apparatus

The apparatus was the same as that used in Experiment 1.1.

Procedures

Procedures were identical to those of the previous experiment, except that subjects did not complete the training procedures (e.g., Keypeck training, Phase I, II, III, IV). In addition, subjects were required to complete a new performance standard. The performance standard in Experiment 1.1 required subjects to complete five consecutive bins of two chain completions, each within 45 s. The new performance standard requires subjects to complete five consecutive bins of five chain completions, each within 45 s. After the performance standard was met, the number of correct practices was yoked onto a Within- and Between-chains delay condition [please see p. 16-19 for a full description of training procedures].

Results

Table 1.2.1 shows the completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens. Some hens did not complete the number of yoked practices within the session duration. These sessions were not used in the following analysis.

Group data for Figures 1.2.1-1.2.5 were analysed using a one-way repeated measures analysis of variance for all measures during the Training and Retention components for Experiment 1.2. The alpha level for all statistical comparisons in all situations was set at .05 and any results that reached this level were presented with an asterisk (*) in Table 1.2.2. Except where indicated with a hashtag (#) in Table 1.2.2, Mauchley's Test was not significant so sphericity was assumed. Post-hoc tests were conducted using the Bonferroni correction, as recommended by Fields (2005).

Figure 1.2.1shows mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean percent correct data calculated across subjects is shown in the bottom left graph. Accuracy in the Training component was similar during the No-delay and Between-chains delay conditions, and lower under the Within-chains delay condition, for all subjects. Table 1.2.2 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention accuracy was generally similar across the three experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.2.2).

Figure 1.2.2 shows the mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response rate data calculated across subjects is shown in the bottom left graph. Training response rates were greatest in the No-delay condition then decreased during the Within- and Between-chains delay conditions for all subjects. Table 1.2.2 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention response rates were generally similar across the three experimental conditions for all subjects; showing no systematic effect from training conditions.

Figure 1.2.3 shows the mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph. Response latency in the Training component was greatest during the Within-chains delay condition and generally similar during the No-delay and Between-chains delay conditions for all subjects. Table 1.2.2 shows this finding was significant. Retention latencies were similar across all experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.2.2).

Figure 1.2.4 shows the mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph. Reinforcement rate for the No-delay and Between-chains delay conditions were similar, and were lower during the Within-chains delay conditions. Table 1.2.2 shows the differences between conditions was significant.

Figure 1.2.5 shows the mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component. The left three bars on each graph shows the number of practices when defined as corrects only. The right three bars on each graph show the number of practices when defined by trials. The mean number of practices data calculated across subjects is shown in the bottom left graph. The mean number of correct practices was similar across experimental conditions for all subjects. Table 1.2.2 shows no significant differences across conditions. The mean number of trial practices was generally greatest during the Within-chains delay condition for all hens, Table 1.2.2 shows this finding was significant. Total number of practices for each experimental condition was generally greater when defined as trials. As mentioned, some experimental sessions were discarded and, when the mean was calculated for the three chains used in each experimental condition, produced different number of yoked correct practices across experimental sessions (e.g., Hen 21, Within- and Between-chains delay; Hen 24, Between-chains delay).

Table 1.2.1.

Completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens.

Condition	Chain	Hen 21	Hen 22	Hen 23	Hen 24	Hen 25	Hen 26
A	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨	▨	▨	▨
B	L-C-L		▨	▨	▨	▨	▨
	R-L-C		▨	▨	▨	▨	▨
	R-L-R		▨	▨	▨	▨	▨
C	L-C-L		▨	▨	▨	▨	▨
	R-L-C		▨	▨		▨	▨
	R-L-R		▨	▨		▨	▨

Table 1.2.2

Analysis of variance results during the Training and Retention components of Experiment 1.2.

Component	MS Treatment	MS Error	df	F	p	Partial Eta Squared
Mean percent correct						
Training	809.26	89.44	2, 8	9.05	.01*	.69
Retention	23.35	22.78	2, 6	1.03	.41	.26
Mean response rate						
Training	.17	.003	2, 8	67.85	p<.001*	.94
Retention	.003	.009	2, 6	0.06	0.94	.02
Mean response latency						
Training	8.76	.60	2, 8	14.59	.002*	.78
Retention	.02	.07	2, 6	0.24	.80	.07
Reinforcement rate						
Session	4.3E-6	3.6E-7	2, 8	11.60	.004*	.74
Number of practices						
Correct	454.74	450.59	2, 8	1.00	.37#	.20
Trial	7989.03	623.261	2, 8	12.82	.003*	.76

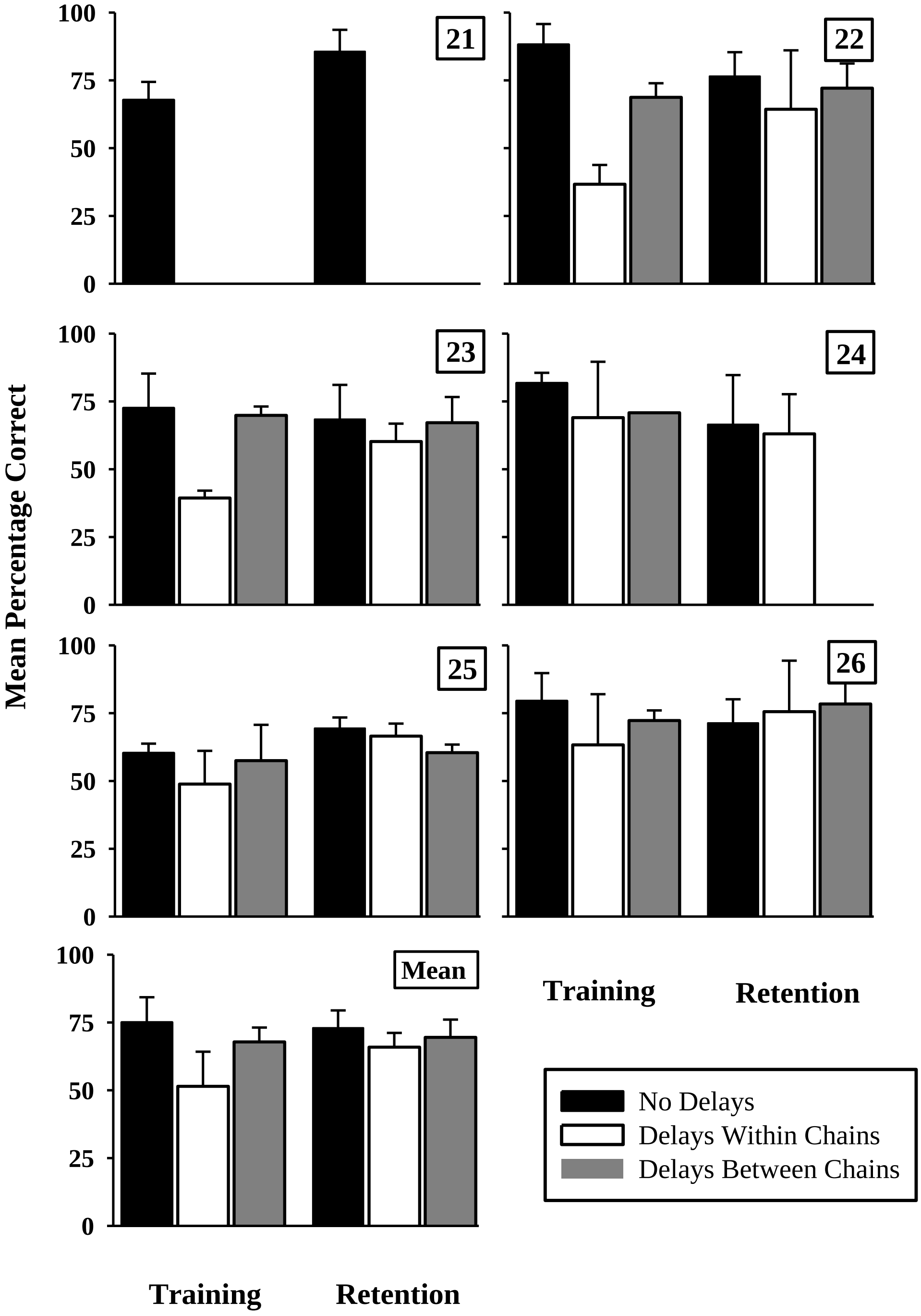


Figure 1.2.1. The mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean percent correct data calculated across subjects is shown in the bottom left graph.

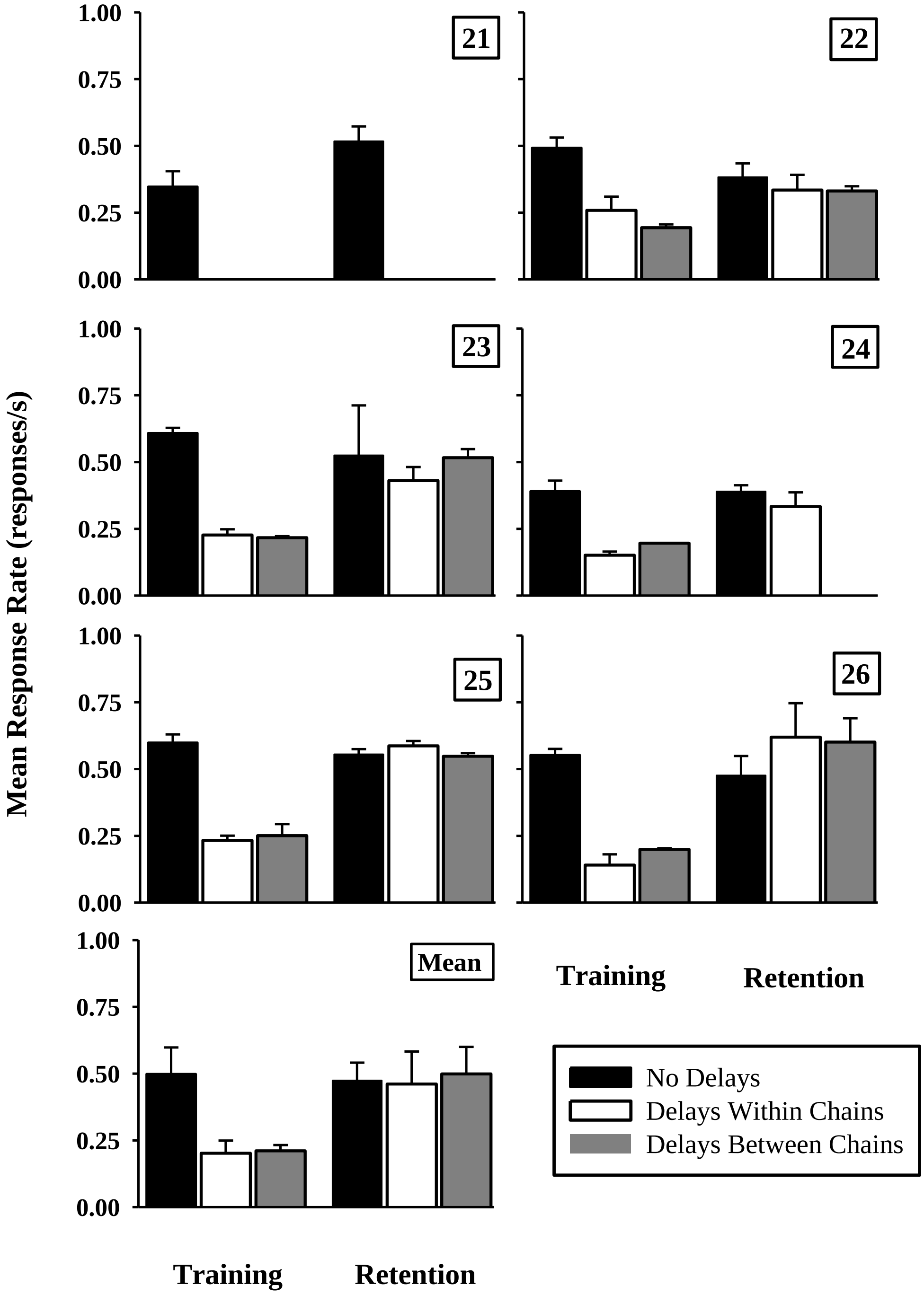


Figure 1.2.2. The mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response rate data calculated across subjects is shown in the bottom left graph.

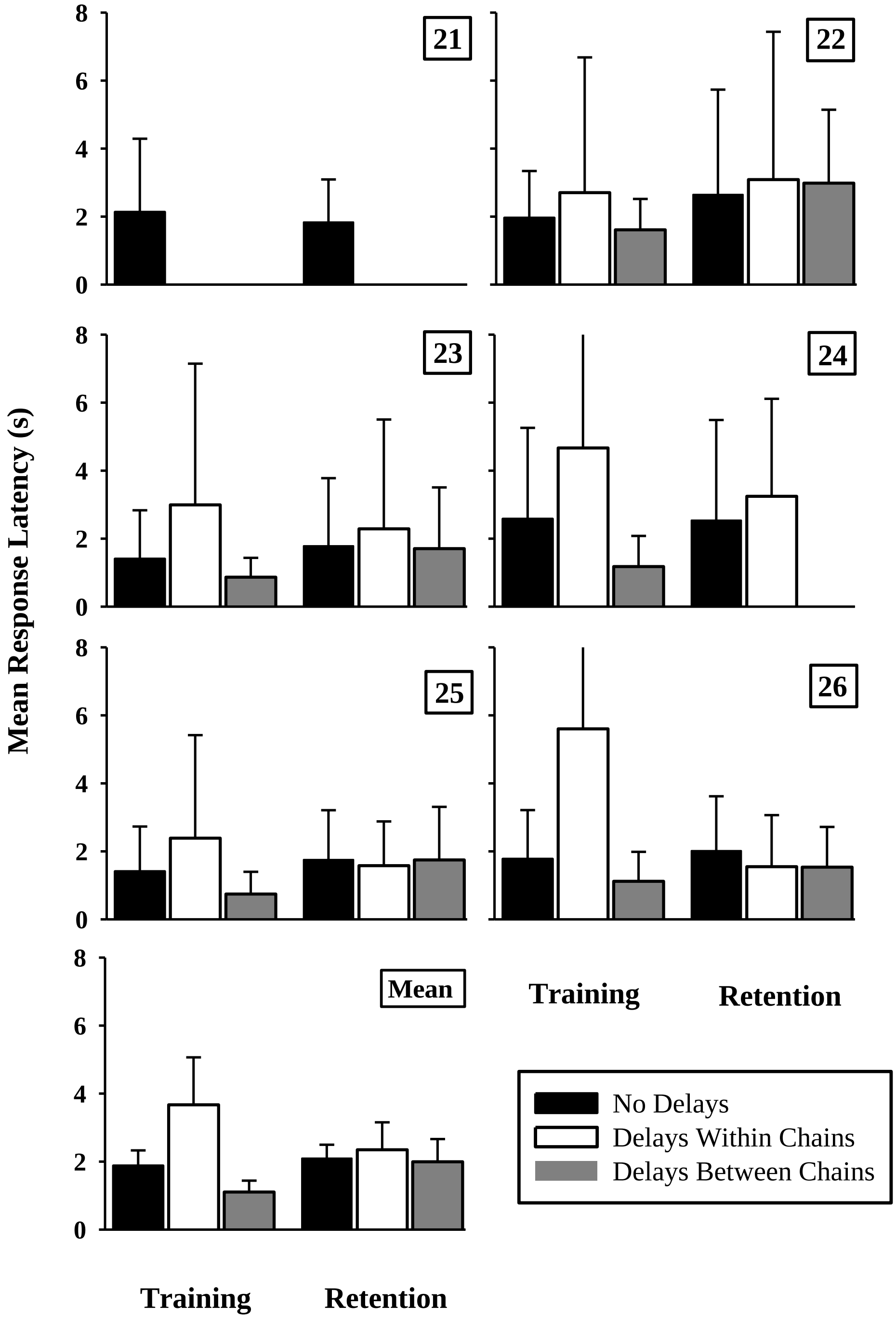


Figure 1.2.3. The mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph.

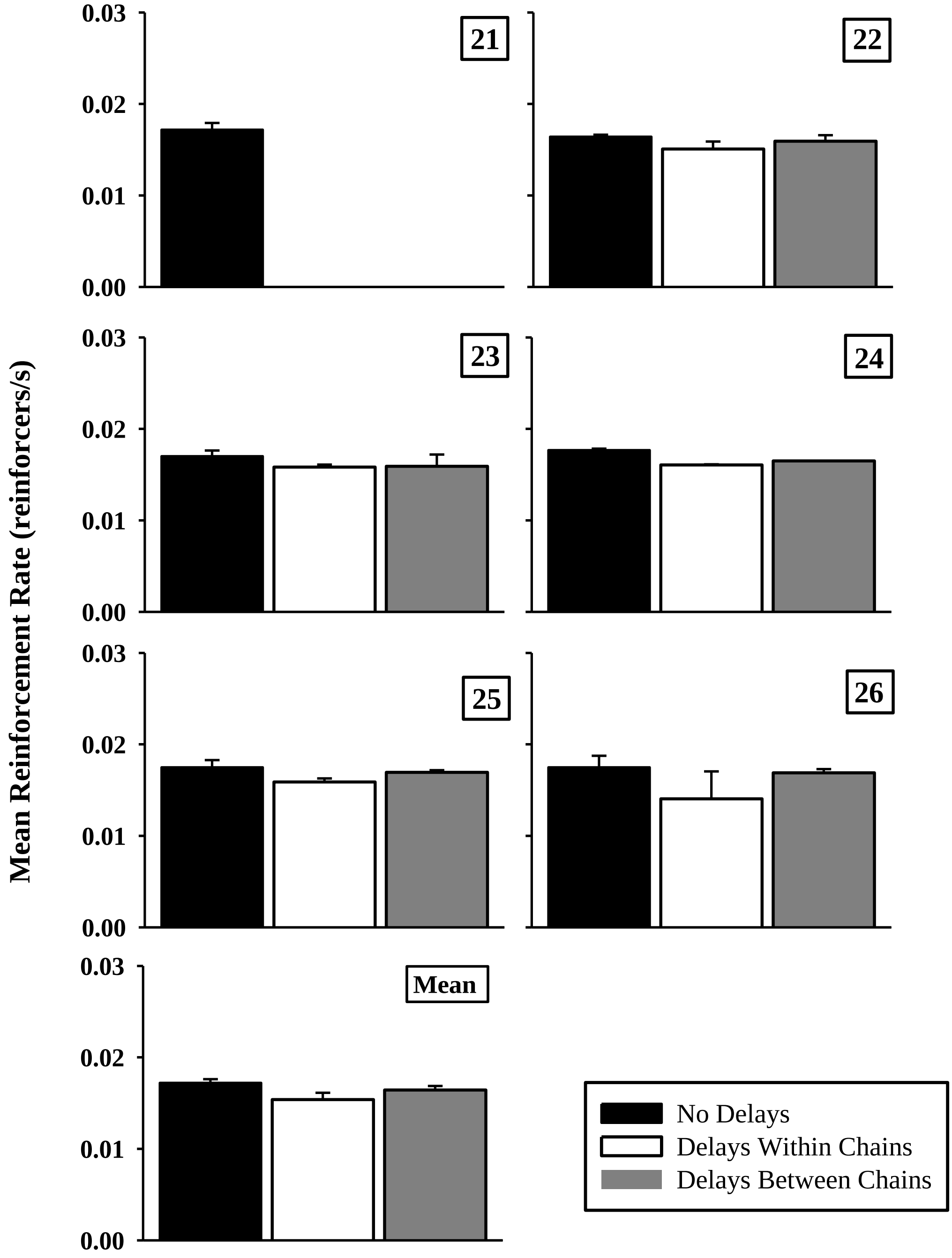


Figure 1.2.4. The mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

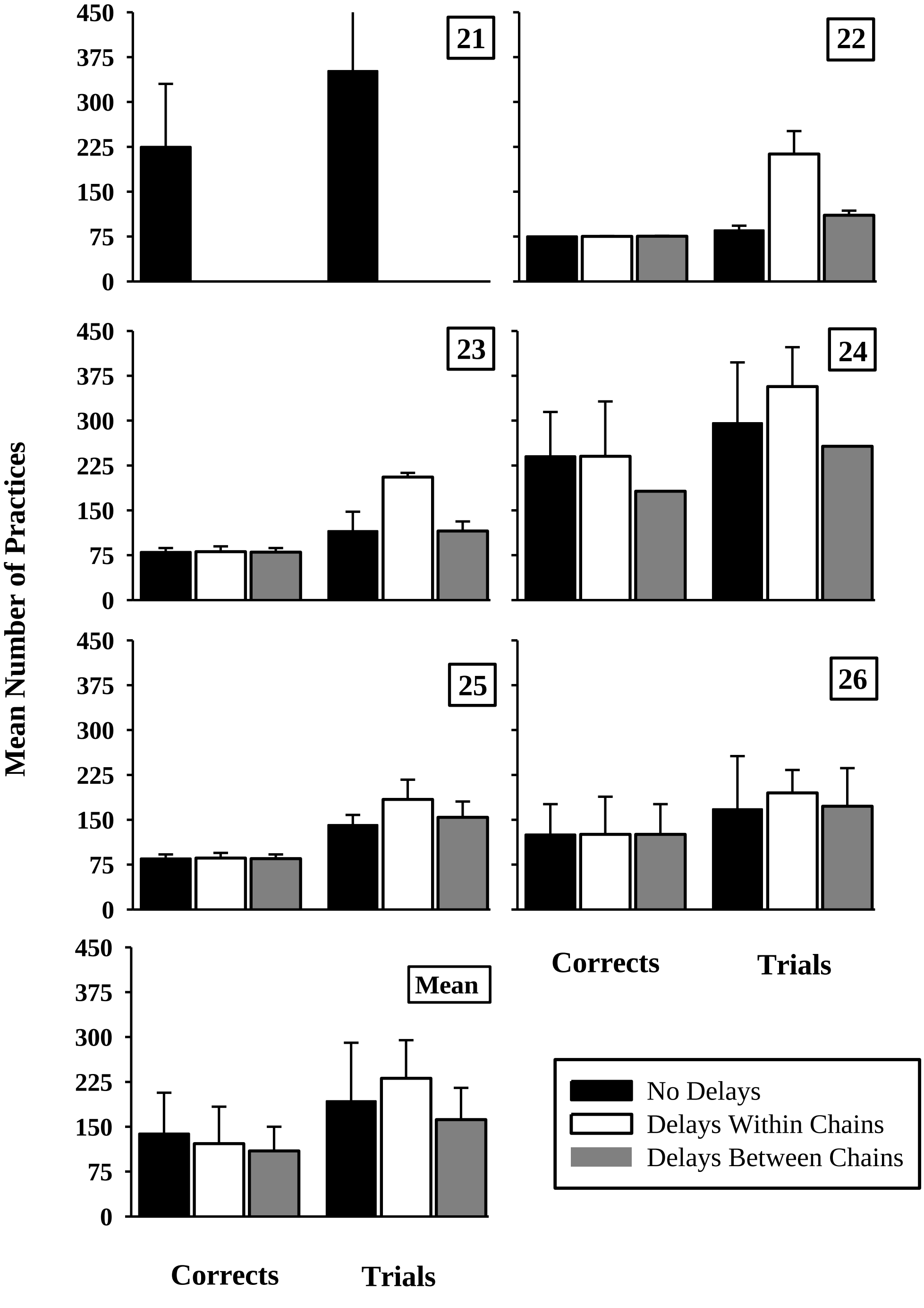


Figure 1.2.5. The mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component. The left three bars on each graph shows the number of practices when defined as corrects only. The right three bars on each graph show the number of practices when defined by trials. The mean number of practices data calculated across subjects is shown in the bottom left graph.

Discussion

The present study replicated Porritt's (2007) methods by increasing the performance standard with the aim of comparing results. It was expected that increasing the performance standard would increase percentage correct during the No-delay condition of the Training and Retention component, replicating findings from Porritt.

As in Porritt (2007), the mean percentage correct and mean response rate was greatest during the No-delay training condition when compared to the other two delay conditions in the same component. In addition, mean response latency was greatest during the Within-chains delay training condition when compared to the other two delay conditions in the same component. Increasing the performance standard in the present experiment generated greater overall mean percentage correct during the No-delay condition of the Training component than that of the mean percentage correct of the same component during Experiment 1.1, supporting the finding that greater response rates improve accuracy (Berens et al., 2003; Ivarie, 1986). This finding should be taken with caution because a greater number of trial and correct practices were required to achieve the performance standard in the present study than in that of Experiment 1.1.

Contrary to Porritt (2007), there were no systematic differences across experimental conditions in mean response rate, mean percentage correct, or mean latency during the Retention component. This finding suggests that increasing the performance standard did not increase mean percentage correct during the No-delay condition of the Retention component. Mean reinforcement rates were significantly different between the No-delay and Within-chains delay conditions (.002) and the No-delay and Between-chains delay conditions (.001). While these differences were statistically significant, they are similar to the difference in mean between the Within- and Between-chains delay conditions (.001).

Overall, using a similar performance standard to that of Porritt (2007) during the No-delay training condition in the present study improved mean percentage correct when compared to the mean percentage correct during the No-delay condition of the Training component in Experiment 1.1. However, this increased criterion did not produce differences in mean percentage correct across the three experimental conditions in the Retention component. Similar to

Experiment 1.1, results from the present investigation partially replicated the findings from Porritt (2007). One reason for this lack of replication may be that the hens used in the present experiment require a greater performance standard than pigeons.

As previously mentioned [Experiment 1.1 Introduction], greater performance standards have been shown to improve retention accuracy (Berens et al., 2003; Ivarie, 1986), it may that the hens used in the present study require a greater performance standard than pigeons to replicate findings from Porritt (2007).

Before investigating the effects from using an even greater performance standard, other methodological differences were investigated. As previously mentioned [Experiment 1.1 Discussion], the type of practice yoked may account for the differences in findings between the present study and those of Porritt. Experiment 1.3 yoked trial practices with the aim of comparing findings to Porritt (2007).

EXPERIMENT 1.3

Experiment 1.3 yoked trial practices, similar to Porritt (2007), with the aim of comparing results.

Method

Subjects

Only five subjects (22-26) from Experiment 1.1 and 1.2 participated in this study. Hen 21 became ill and could not participate.

Apparatus

The apparatus was the same as that used in Experiment 1.1 and 1.2.

Procedures

Procedures were identical to those of Experiment 1.2 with one exception. Instead of yoking number of correct practices, the present study held the number of trial practices (e.g., correct and incorrect responses) constant across experimental conditions [please see p. 16-19 for a full description of training procedures].

Results

Table 1.3.1 shows the completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens. Hen 24 did not complete the number of yoked practices within the programmed duration for some sessions. These sessions were not used in the following analysis.

Group data for Figures 1.3.1-1.3.7 were analysed using a one-way repeated measures analysis of variance for all measures during the Training and Retention components for Experiment 1.3. The alpha level for all statistical comparisons in all situations was set at .05 and any results that reached this level were presented with an asterisk (*) in Table 1.3.2. Except where indicated with a hashtag (#) in Table 1.3.2, Mauchley's Test was not significant so sphericity was assumed. Post-hoc tests were conducted using the Bonferroni correction, as recommended by Fields (2005).

Figure 1.3.1 shows mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean percent correct data calculated across subjects is shown in the bottom left graph. Generally, accuracy in the Training component was similar during the No-delay and Between-chains delay conditions, and lower during the Within-chains delay condition, for all subjects. Table 1.3.2 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention accuracy was generally similar across the three experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.3.2).

Figure 1.3.2 shows the mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response rate data calculated across subjects is shown in the bottom left graph. Training response rates were greatest in the No-delay condition then decreased during the Within- and Between-chains delay conditions for all subjects. Table 1.3.2 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009).

Retention response rates were generally similar across the three experimental conditions for all subjects; showing no systematic effect from training conditions.

Figure 1.3.3 shows the mean correct response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean correct response rate data calculated across subjects is shown in the bottom left graph. Correct response rate was greatest during the No-delay condition and lowest during the Within-chains delay condition for all subjects. Correct response rates during the Retention component were generally similar across the three experimental conditions for all subjects; showing no systematic effect from training conditions. Table 1.3.2 shows the Training component findings were significantly different, whereas no significant differences were found between the Retention component conditions.

Figure 1.3.4 shows the mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph. Response latency in the Training component was greatest during the Within-chains delay condition, and generally similar during the No-delay and Between-chains delay conditions, for all subjects. Table 1.3.2 shows this finding was significant. Retention latencies were similar across all experimental conditions for all subjects; the data from these conditions were not significantly different.

Figure 1.3.5 shows the mean correct response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean correct response latency data calculated across subjects is shown in the bottom left graph. Correct response latency in the Training component was greatest during the Within-chains delay condition, and generally similar during the No-delay and Between-chains delay conditions, for all subjects. Retention latencies were similar across all experimental conditions for all subjects. Table 1.3.2 shows Training component finding were significantly different, whereas no significant differences were found between Retention component conditions.

Figure 1.3.6 shows the mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph. Reinforcement rates were similar for the No-delay and Between-chains delay conditions, and lower during the Within-chains delay condition. The differences between conditions were not significant (Table 1.3.2).

Figure 1.3.7 shows the mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component. The left three bars on each graph shows the number of practices when defined as corrects only. The right three bars on each graph show the number of practices when defined by trials. The mean number of trial practices was generally similar during the Within- and Between-chains delay conditions, and largest during the No-delay condition. Table 1.3.2 shows the differences between conditions was significant. The mean number of correct practices was variable across experimental conditions for all subjects, this finding was significant (Table 1.3.2). The mean number of trial practices differs across experimental conditions for Hen 24 because she never completed Chain 3 of the Within-chains delay condition. Total number of practices for each experimental condition was generally greater when defined as trials.

Table 1.3.1.

Completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens.

Condition	Chain	Hen 21	Hen 22	Hen 23	Hen 24	Hen 25	Hen 26
A	L-C-L		▨	▨	▨	▨	▨
	R-L-C		▨	▨	▨	▨	▨
	R-L-R		▨	▨	▨	▨	▨
B	L-C-L		▨	▨	▨	▨	▨
	R-L-C		▨	▨	▨	▨	▨
	R-L-R		▨	▨		▨	▨
C	L-C-L		▨	▨		▨	▨
	R-L-C		▨	▨		▨	▨
	R-L-R		▨	▨		▨	▨

Table 1.3.2

Analysis of variance results during the Training and Retention components of Experiment 1.3.

Component	MS Treatment	MS Error	df	F	p	Partial Eta Squared
		Mean percent correct				
Training	1046.94	87.30	2, 6	12.00	.01*	.80
Retention	47.65	43.39	2, 6	1.10	.39	.27
		Mean response rate				
Training	.19	.003	2, 6	72.04	p<.0001*	.96
Retention	.002	.001	2, 6	2.50	.17	.45
		Mean correct response rate				
Training	.133	.003	2, 6	52.80	p<.0001*	.95
Retention	0	.001	2, 6	0.18	.84	.06
		Mean response latency				
Training	13.29	.802	2, 6	16.57	.004*	.85
Retention	.180	.069	2, 6	1.58	.28	.35
		Mean correct response latency				
Training	17.09	4.99	2, 6	3.45	.16#	.53
Retention	.081	.052	2, 6	1.56	.29	.34
		Reinforcement rate				
Session	2E-5	2.4E-6	2, 6	8.58	.06	.74
		Number of practices				
Correct	1872.29	141.47	2, 6	13.23	.01*	.82
Trial	3.95	.67	2, 6	5.90	.04*	.66

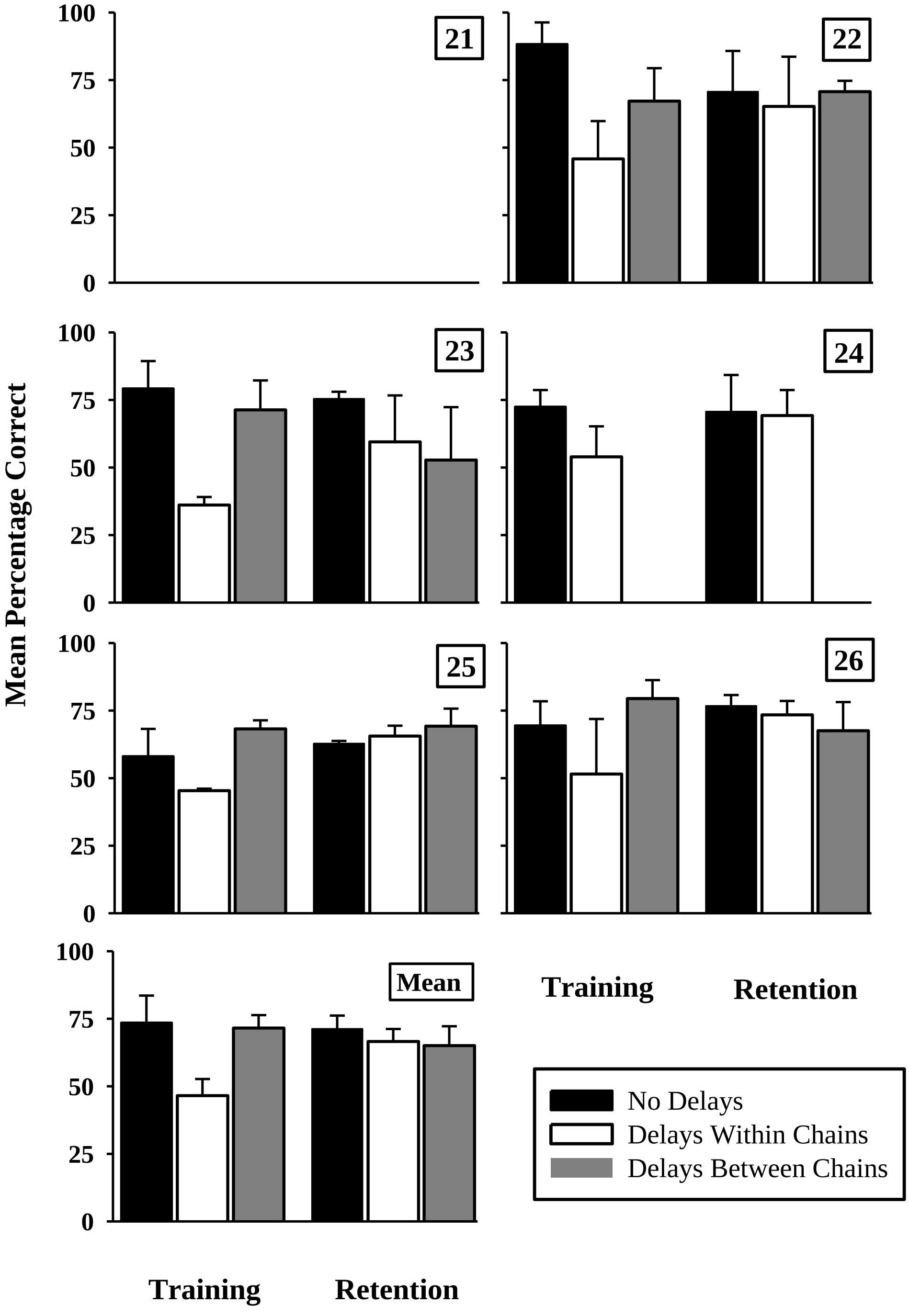


Figure 1.3.1. The mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean percent correct data calculated across subjects is shown in the bottom left graph.

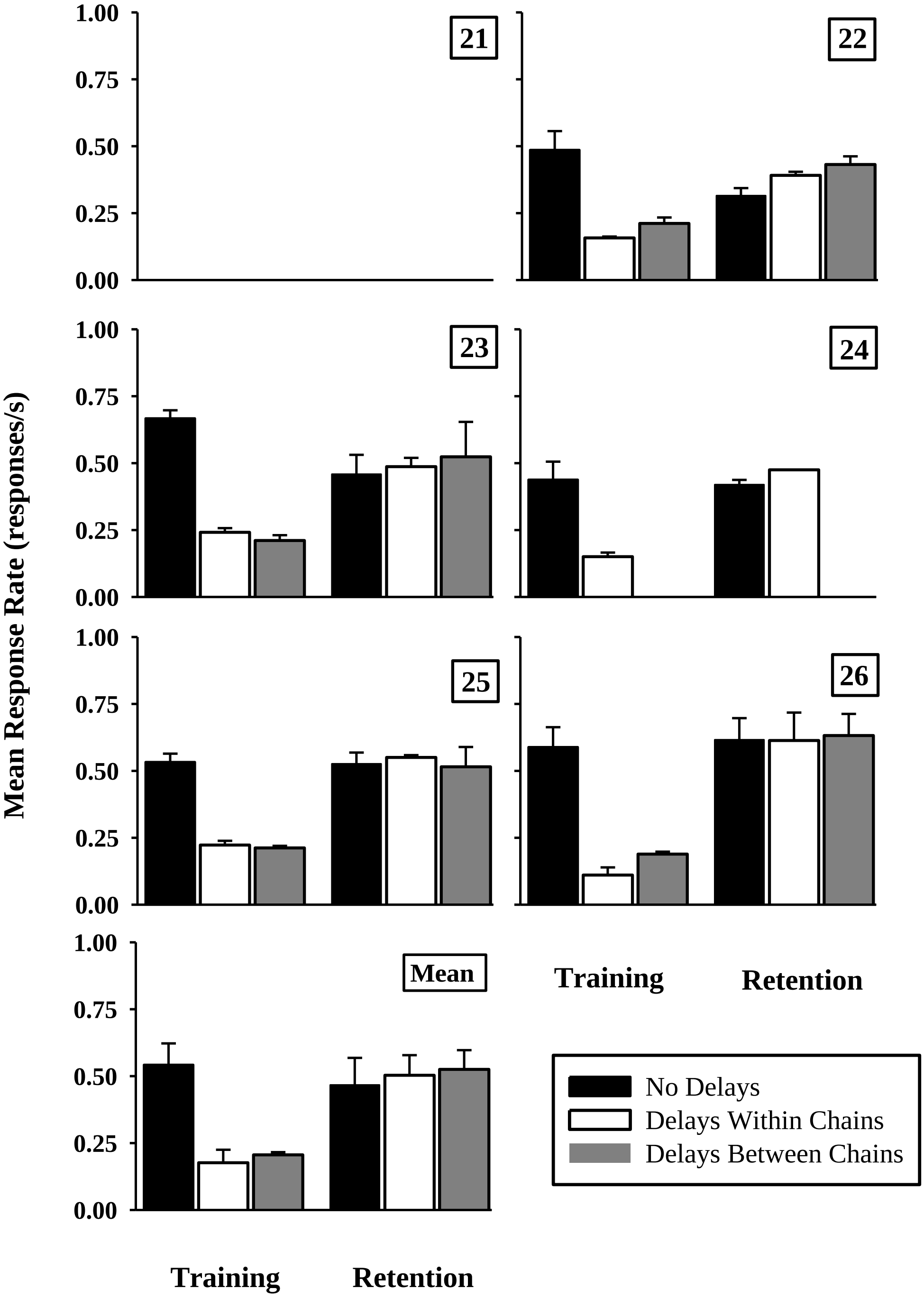


Figure 1.3.2. The mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response rate data calculated across subjects is shown in the bottom left graph.

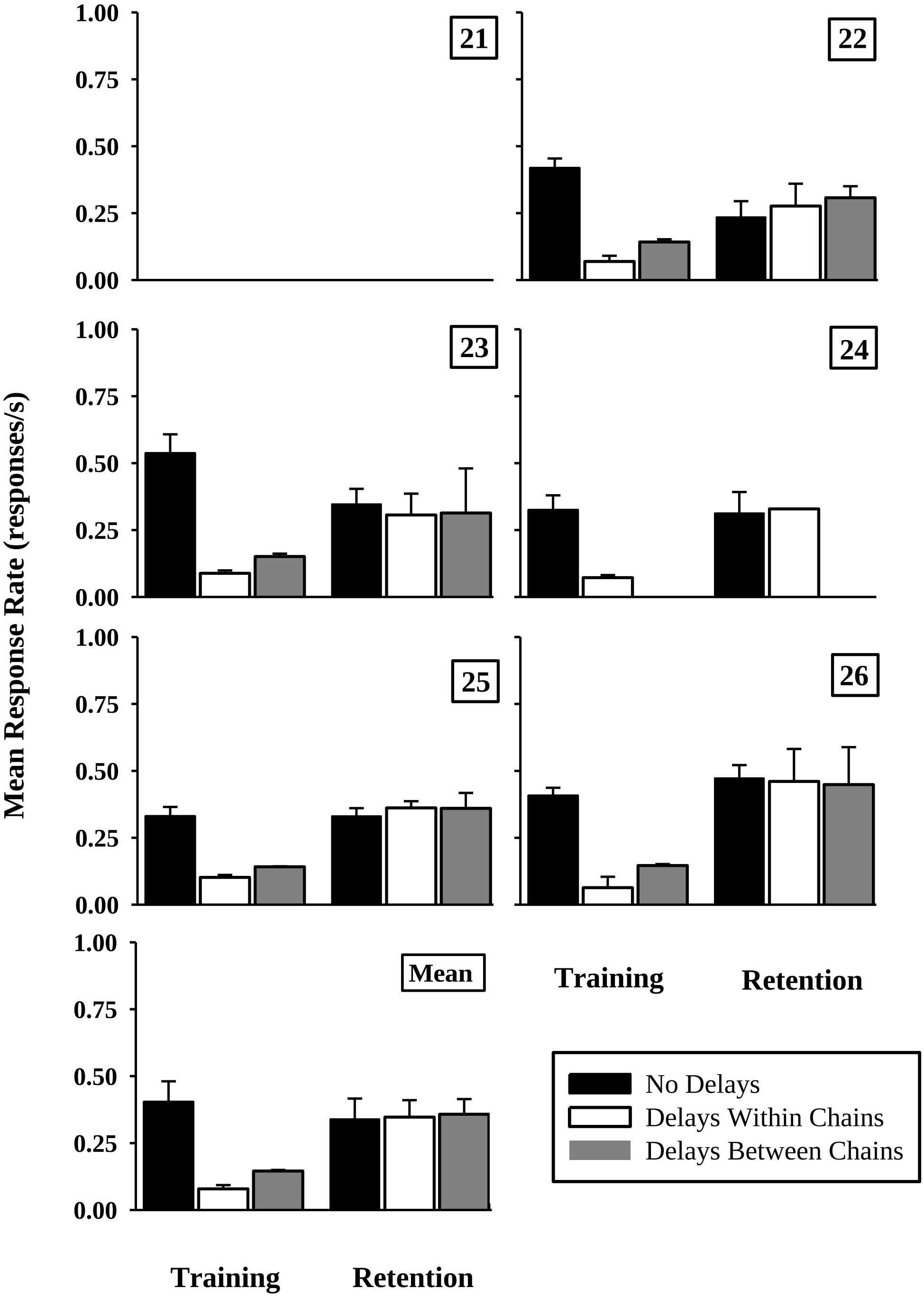


Figure 1.3.3. The mean correct response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean correct response rate data calculated across subjects is shown in the bottom left graph.

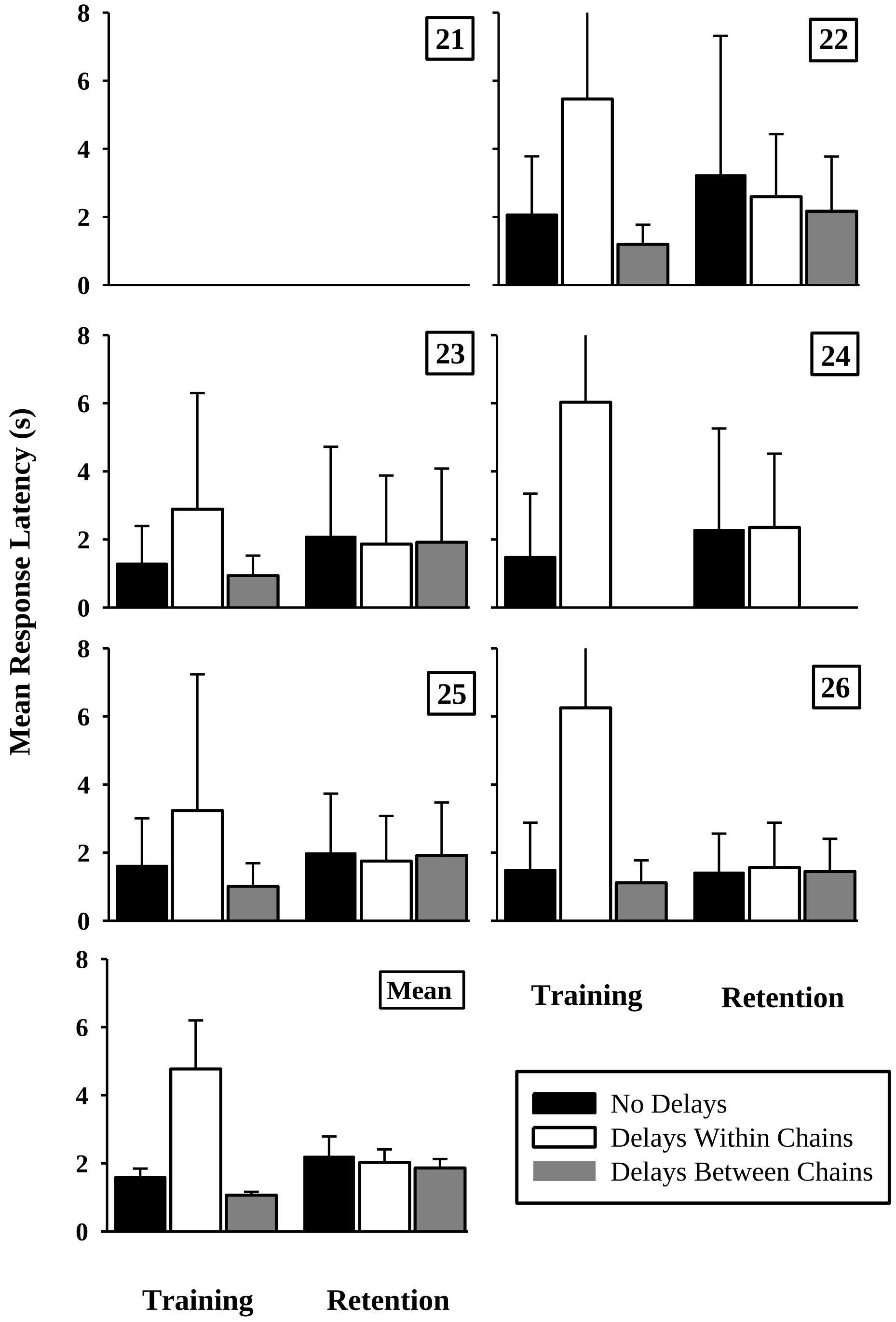


Figure 1.3.4. The mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph.

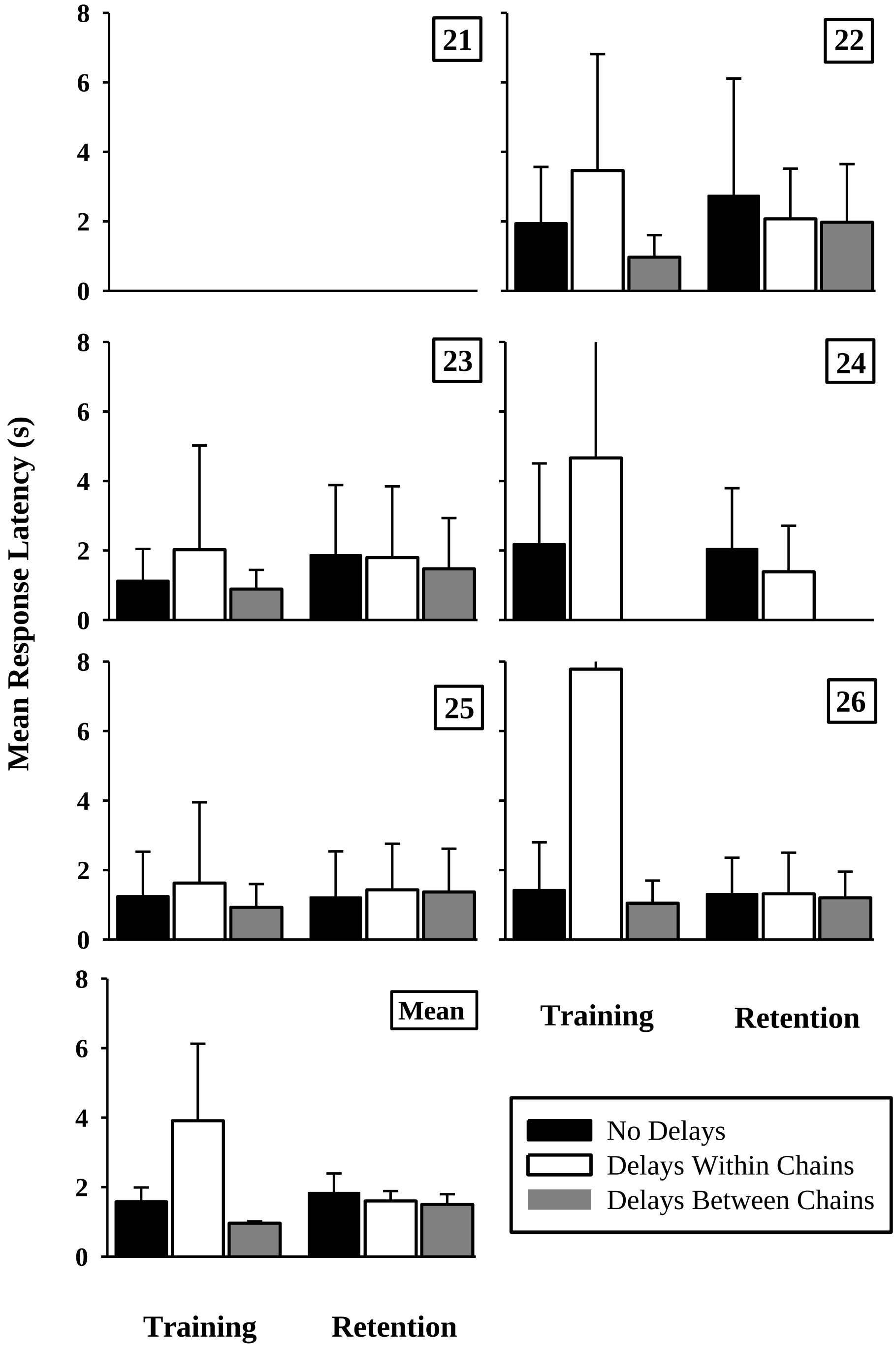


Figure 1.3.5. The mean correct response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean correct response latency data calculated across subjects is shown in the bottom left graph.

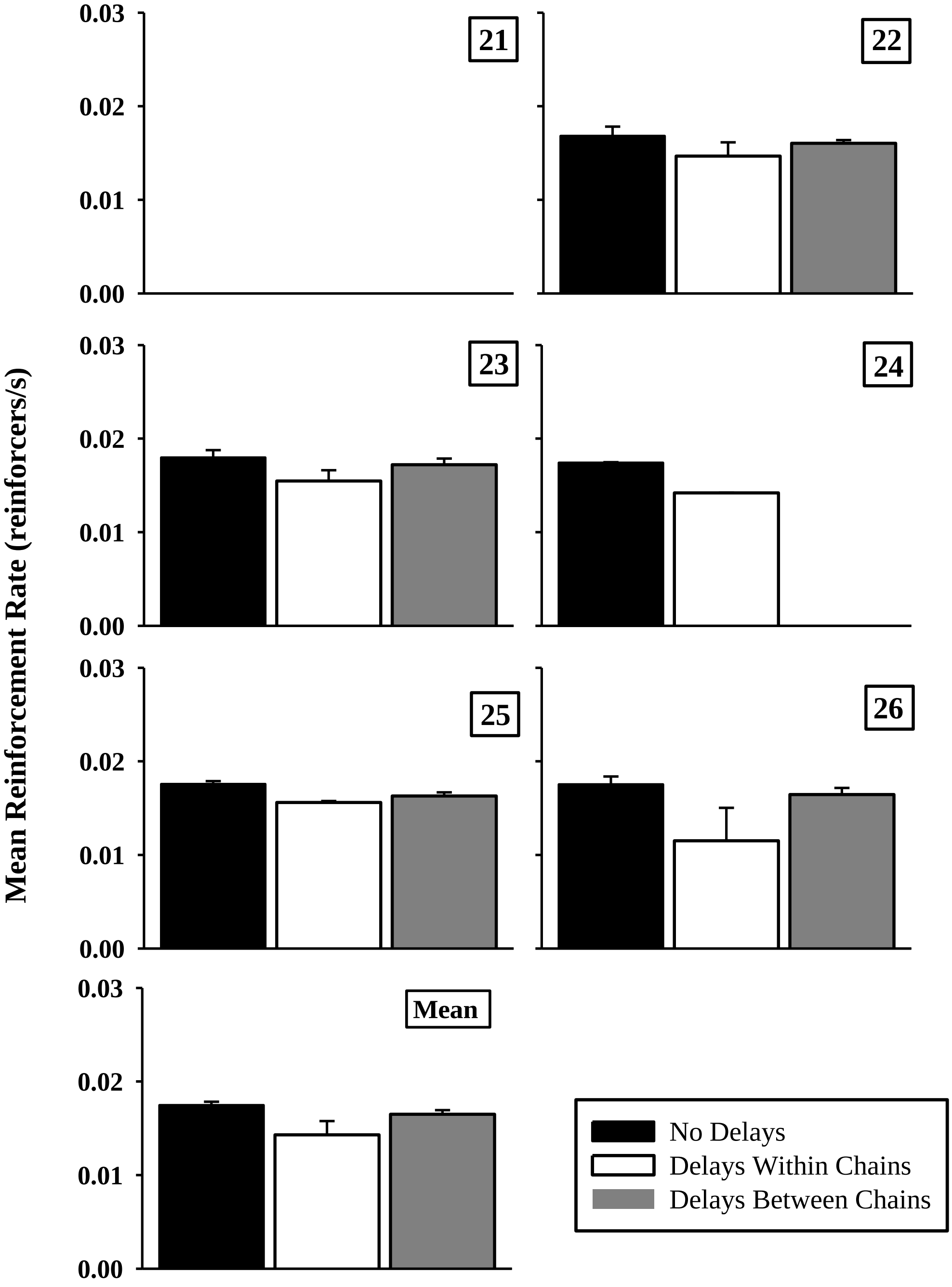


Figure 1.3.6. The mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

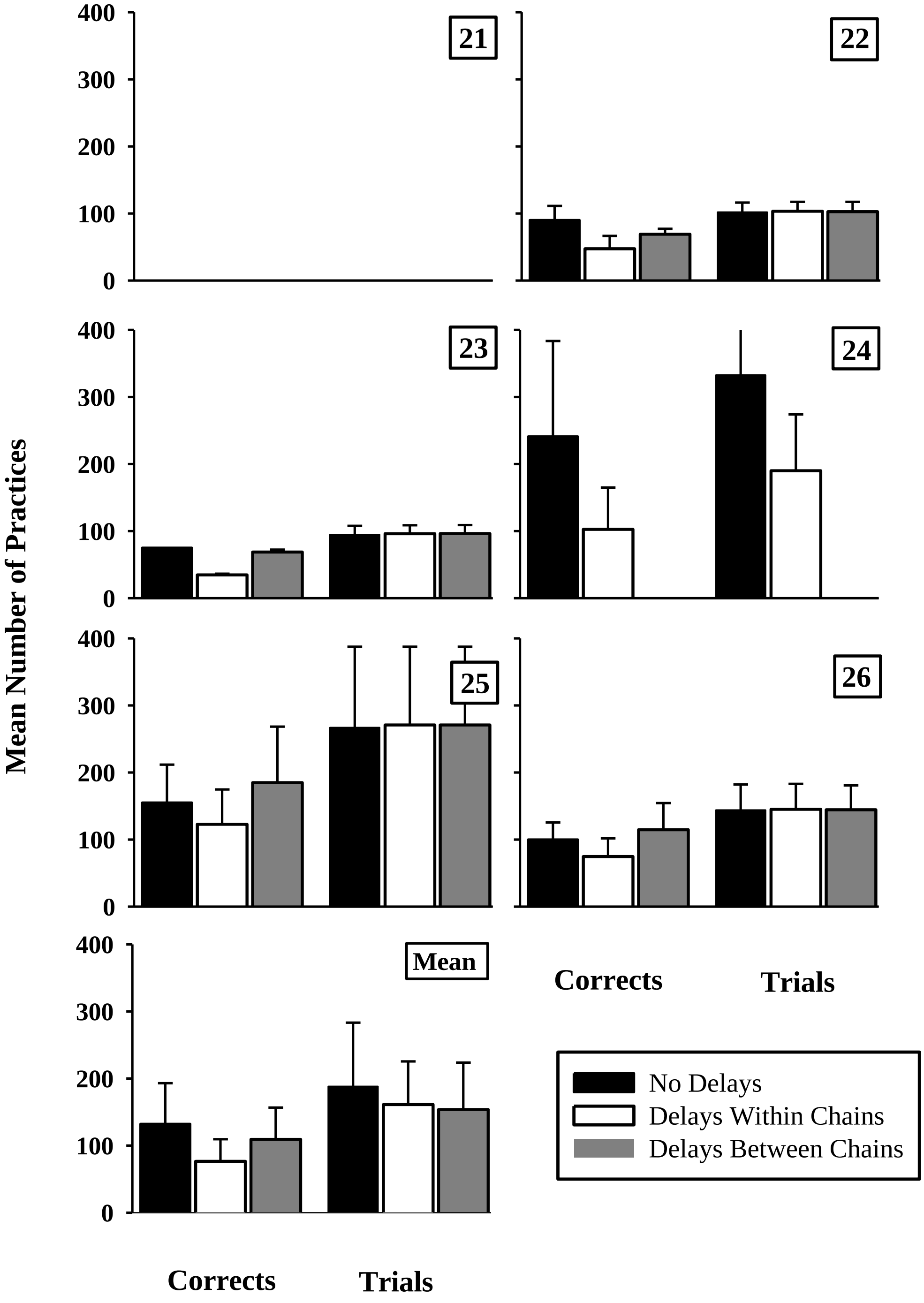


Figure 1.3.7. The mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component. The left three bars on each graph shows the number of practices when defined as corrects only. The right three bars on each graph show the number of practices when defined by trials. The mean number of practices data calculated across subjects is shown in the bottom left graph.

Discussion

The present study yoked trial practices, similar to Porritt (2007), with the aim of comparing results. It was expected that yoking practices, when defined by trials, would generate differences in accuracy between experimental conditions during the Retention component, replicating findings from Porritt.

As in the No-delay condition during the Training component of Experiment 1.2 and Porritt (2007), mean percentage correct and mean response rates were greatest during the No-delay training condition in the present study. In addition, mean response latency was greatest during the Within-chains delay condition of the Training component when compared to the mean response latency of the other two delay conditions during the same component in the present study. Similar results were found for correct response rates and correct response latencies. Similar to findings in Experiment 1.1, calculating latencies and rates as trials (e.g., all responses) produced greater values than calculating correct response latencies and correct response rates. These differences in calculation, however, did not produce different results on the effects of each experimental condition when trials were yoked. Thus, correct response rate and correct response latency will not be calculated in the remaining studies.

The number of trial practices, when the mean was calculated across hen data, was significantly different across experimental conditions (Figure 1.3.1). This finding was not expected given the yoking procedure used. However, individual data in Figure 1.3.1 shows the findings for all but Hen 24 were similar across conditions, suggesting the yoking procedure worked for these hens. The observed difference across conditions for Hen 24 appears to be accounted for by the failure to complete all experimental conditions, as shown in Table 1.3.1. Thus, the significant difference was a product of Hen 24 failing to complete all conditions, and not a failure of the yoking procedure. Perhaps a more relevant comparison of trial practices is one that leaves out Hen 24. Further analysis of this result, removing data from Hen 24, shows there is no significant difference between experimental conditions, F (2, 6) = 5.9, p > .05, η_p^2 = .66.

Contrary to Porritt (2007), there were no systematic differences across experimental conditions in response rate, percentage correct, or latency during the Retention component. This finding suggests that yoking trials, rather than corrects, did not produce differences in percentage correct across the experimental conditions during the Retention component. The mean reinforcement rates for the Within-chains delay condition were not significantly different from the means for the other two delay conditions.

Yoking trials in the present experiment generated more practice opportunities during the Within- and Between-chains delay conditions than yoking corrects in Experiment 1.2. The extra practice opportunities, however, did not create outcomes similar to Porritt (2007), suggesting that how a practice is defined does not change performance. As previously mentioned [Experiment 1.1 Introduction], yoking corrects has advantages over yoking trials. Corrects were yoked during the remaining experiments of this study and will now be referred to as practices.

In addition to Porritt (2007), Porritt et al., (2009) was the only other study to show strong support for the use of a rate-building procedure while controlling for reinforcement rate and number of practices. There was one major difference between the two studies; Porritt et al. (2009) used an experimental design that repeatedly exposed subjects to the No-delay, Within- and Between-chains delay conditions. Repeatedly exposing subjects to experimental conditions helps reduce multiple-treatment interference (Johnston & Pennypacker, 1993, p. 341). It may be that using an experimental design similar to Porritt et al. (2009) in the present study produces similar results to Porritt (2007).

Similar to Experiment 1.2, the present study partially replicated findings from Porritt. As previously mentioned [Experiment 1.2 Discussion], it may that the hens used in the present study require a greater performance standard than pigeons to replicate findings from Porritt (2007). Before investigating any species differences by using a greater performance standard, differences in experimental design were investigated. Experiment 1.4 used an experimental design similar to Porritt et al. (2009) with the aim of comparing results.

EXPERIMENT 1.4

Experiment 1.4 replicated the experimental design from Porritt et al. (2009) by repeatedly exposing subjects to experimental conditions with the aim of comparing results.

Methods

Subjects

The same 6 subjects from Experiments 1.1-1.3 participated in this study. Hen 21 became well and participated in the current experiment.

Apparatus

The apparatus was the same as that used in Experiments 1.1-1.3.

Procedures

Procedures were identical to those of Experiment 1.2 with one exception. Using repeated measures, within-subject design (e.g., A/B/A/B/A/C/A/C/A), subjects were repeatedly exposed to the No-delay (Termed A), Within-chains delay (Termed B), and Between-chains delay (Termed C) conditions. The number of practices to achieve the performance standard during Chains 1, 2, and 3 of each No-delay condition was yoked onto the chain sequences for the next condition, either a Within-chains delay or Between-chains delay condition. The order of components, experimental conditions, chains used for each component, and the criterion to end each component is presented in Table 1.4.1 [please see p. 16-19 for a full description of training procedures].

Results

Table 1.4.2 shows the completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens. Some hens did not complete the number of yoked practices within the session duration. These sessions were not used in the following analysis.

Group data for Figures 1.4.1-1.4.5 were analysed using a one-way repeated measures analysis of variance for all measures during the Training and Retention components for Experiment 1.4. The alpha level for all statistical comparisons in all situations was set at .05 and any results that reached this level were presented with an asterisk (*) in Table 1.4.3. Except where indicated with a hashtag (#) in Table 1.4.3, Mauchley's Test was not significant so sphericity was assumed. Post-hoc tests were conducted using the Bonferroni correction, as recommended by Fields (2005).

Figure 1.4.1 shows mean percentage of correct responses (+1 *SD*) obtained from the six sessions of the Within- and Between-chains delay condition and 12 sessions of the No-delay condition during the Training and Retention components for all subjects [calculated by dividing the total number of responses in each chain link during the Training or Retention components by the total number of correct responses in each chain link for that component]. The mean percent correct data calculated across subjects is shown in the bottom left graph. Generally, accuracy in the Training component was similar during the No-delay and Between-chains delay conditions, and lower during the Within-chains delay condition, for all subjects. Table 1.4.3 shows these differences were significant and effect size, partial eta squared, was moderate (Ferguson, 2009). Retention accuracy was generally similar across the three experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.4.3).

Figure 1.4.2 shows the mean response rates (+1 *SD*) obtained from the six sessions of the Within- and Between-chains delay condition and 12 sessions of the No-delay condition during the Training and Retention components for all subjects [calculated by dividing the total number of responses emitted in each chain link for the Training or Retention component by that components duration]. The mean response rate data calculated across subjects is shown in the bottom left graph.

Training response rates were greatest in the No-delay condition then decreased during the Within- and Between-chains delay conditions for all subjects. Table 1.4.3 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention response rates were generally similar across the three experimental conditions for all subjects; showing no systematic effect from training conditions.

Figure 1.4.3 shows the mean response latency [the duration from the illumination of the keylights to the emission of a response] (+1 *SD*) obtained from the six sessions of the Within- and Between-chains delay condition and 12 sessions of the No-delay condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph. Response latency in the Training component was greatest during the Within-chains delay condition, and generally similar during the No-delay and Between-chains delay conditions for all subjects. Table 1.4.3 shows this finding was significant. Retention latencies were similar across all experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.4.3).

Figure 1.4.4 shows the mean reinforcement rate (+1 *SD*) obtained from the six sessions of the Within- and Between-chains delay condition and 12 sessions of the No-delay condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph. Reinforcement rates were similar for the No-delay and Between-chains delay conditions, and lowest during the Within-chains delay condition. The differences between conditions were significant (Table 1.4.3).

Figure 1.4.5 shows the mean number of practices (+1 *SD*) obtained from the three sessions of each condition for the first (e.g., A/B, Sessions 1-6; A/C, Sessions 13-18) and second (e.g., A/B, Sessions 1-6; A/C, Sessions 13-18) set of condition changes. The left four bars on each graph show the number of practices from the first set of condition changes. The right four bars on each graph show the number of practices from the second set of condition changes. The mean number of practices data calculated across subjects is shown in the bottom left graph. The mean number of practices was generally similar across experimental conditions for all subjects. The mean number of practices was similar between the first and second set of condition changes. Table 1.4.3 shows no significant differences across conditions.

Table 1.4.1.

Order of components and experimental conditions for each session of Experiments 1.4. Chain sequences used during each component and the criterion to end each component are given.

Component	Experimental Conditions	Chain	Criterion to End
Session 1			
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	L-C-L	Performance Standard
Session 2			
Retention		L-C-L	15chains
Distracter		L-R-C	75 chains
Training	No-delay	R-L-C	Performance Standard
Session 3			
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	R-L-R	Performance Standard
Session 4			
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	Within-chains delay	L-C-L	Yoked practices
Session 5			
Retention		L-C-L	15 chains
Distracter		L-R-C	75 chains
Training	Within-chains delay	R-L-C	Yoked practices
Session 6			
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	Within-chains delay	R-L-R	Yoked practices
Session 7			
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	L-C-L	Performance Standard
Session 8			
Retention		L-C-L	15chains
Distracter		L-R-C	75 chains
Training	No-delay	R-L-C	Performance Standard
Session 9			
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	R-L-R	Performance Standard
Session 10			
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	Within-chains delay	L-C-L	Yoked practices
Session 11			
Retention		L-C-L	15 chains
Distracter		L-R-C	75 chains
Training	Within-chains delay	R-L-C	Yoked practices
Session 12			
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	Within-chains delay	R-L-R	Yoked practices
Session 13			
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	L-C-L	Performance Standard
Session 14			
Retention		L-C-L	15chains
Distracter		L-R-C	75 chains
Training	No-delay	R-L-C	Performance Standard
Session 15			
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	R-L-R	Performance Standard
Session 16			
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	Between-chains delay	L-C-L	Yoked practices
Session 17			
Retention		L-C-L	15 chains
Distracter		L-R-C	75 chains
Training	Between-chains delay	R-L-C	Yoked practices
Session 18			
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	Between-chains delay	R-L-R	Yoked practices
Session 19			
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	L-C-L	Performance Standard
Session 20			
Retention		L-C-L	15chains
Distracter		L-R-C	75 chains
Training	No-delay	R-L-C	Performance Standard
Session 21			
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	R-L-R	Performance Standard
Session 22			
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	Between-chains delay	L-C-L	Yoked practices
Session 23			
Retention		L-C-L	15 chains
Distracter		L-R-C	75 chains
Training	Between-chains delay	R-L-C	Yoked practices
Session 24			
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	Between-chains delay	R-L-R	Yoked practices
Session 25			
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	L-C-L	Performance Standard
Session 26			
Retention		L-C-L	15chains
Distracter		L-R-C	75 chains
Training	No-delay	R-L-C	Performance Standard
Session 27			
Retention		R-L-C	15 chains
Distracter		R-C-L	75 chains
Training	No-delay	R-L-R	Performance Standard
Session 28			
Retention		R-L-R	15 chains
Distracter		R-C-L	75 chains
Training	Within-chains delay	L-C-L	Yoked practices

Table 1.4.2.

Completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens.

Condition	Chain	Hen 21	Hen 22	Hen 23	Hen 24	Hen 25	Hen 26
A	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨		▨	▨
	R-L-R	▨	▨	▨		▨	▨
B	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨	▨	▨	▨
A	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨	▨	▨	▨
B	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨		▨	▨
A	L-C-L	▨	▨	▨		▨	▨
	R-L-C	▨	▨	▨		▨	▨
	R-L-R	▨	▨	▨		▨	▨
C	L-C-L	▨	▨	▨			▨
	R-L-C	▨		▨			▨
	R-L-R	▨		▨			▨
A	L-C-L	▨		▨	▨	▨	▨
	R-L-C	▨		▨	▨	▨	▨
	R-L-R	▨		▨	▨	▨	▨
C	L-C-L	▨		▨	▨	▨	▨
	R-L-C	▨		▨		▨	▨
	R-L-R	▨		▨		▨	▨
A	L-C-L	▨		▨	▨	▨	▨
	R-L-C	▨		▨	▨	▨	▨
	R-L-R	▨		▨	▨	▨	▨

Table 1.4.3

Analysis of variance results during the Training and Retention components of Experiment 1.4.

Component	MS Treatment	MS Error	df	F	p	Partial Eta Squared
			Mean percent correct			
Training	191.57	92.83	2, 10	2.10	.18	.29
Retention	1.62	15.24	2, 10	0.11	.90	.02
			Mean response rate			
Training	.29	.006	2, 10	47.35	.001*#	.90
Retention	0	.001	2, 10	0.21	.81	.04
			Mean response latency			
Training	9.54	.45	2, 10	20.92	p<.0001*	.81
Retention	.44	.35	2, 10	1.25	.33	.20
			Reinforcement rate			
Session	3.6E-6	8.2E-8	2, 10	43.38	p<.0001*	.90
			Number of practices			
First Set (A/B)	2720.04	2720.04	1, 5	1.00	.36#	.17
First Set (A/C)	120.13	120.13	1, 5	1.00	.36#	.17
Second Set (A/B)	252.08	252.08	1, 5	1.00	.39#	.25
Second Set (A/C)	1587	1587	1, 5	1.00	.36#	.17

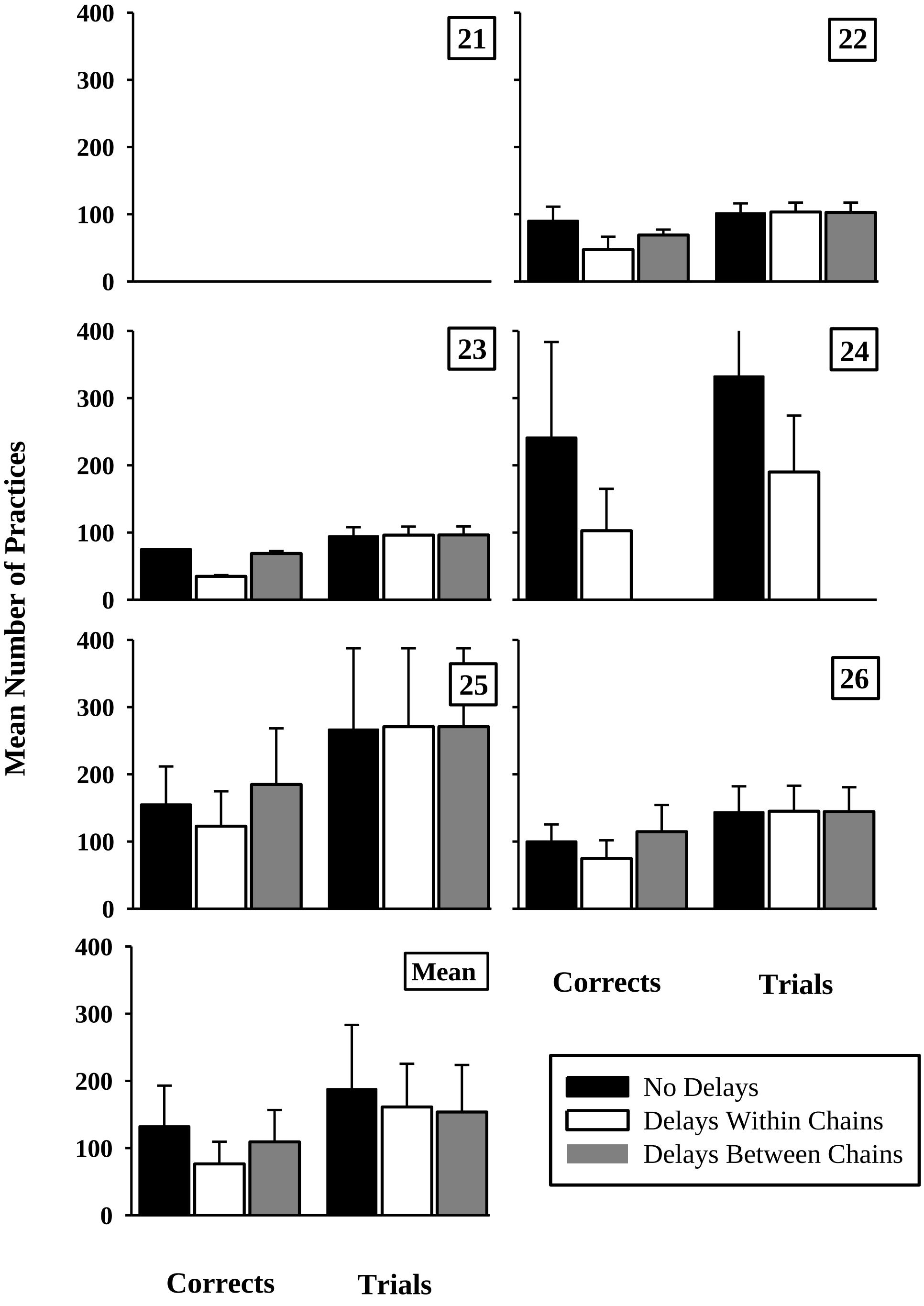


Figure 1.4.1. The mean percentage of correct responses (+1 *SD*) obtained from the six sessions of the Within- and Between-chains delay condition and 12 sessions of the No-delay condition during the Training and Retention components for all subjects. The mean percent correct data calculated across subjects is shown in the bottom left graph.

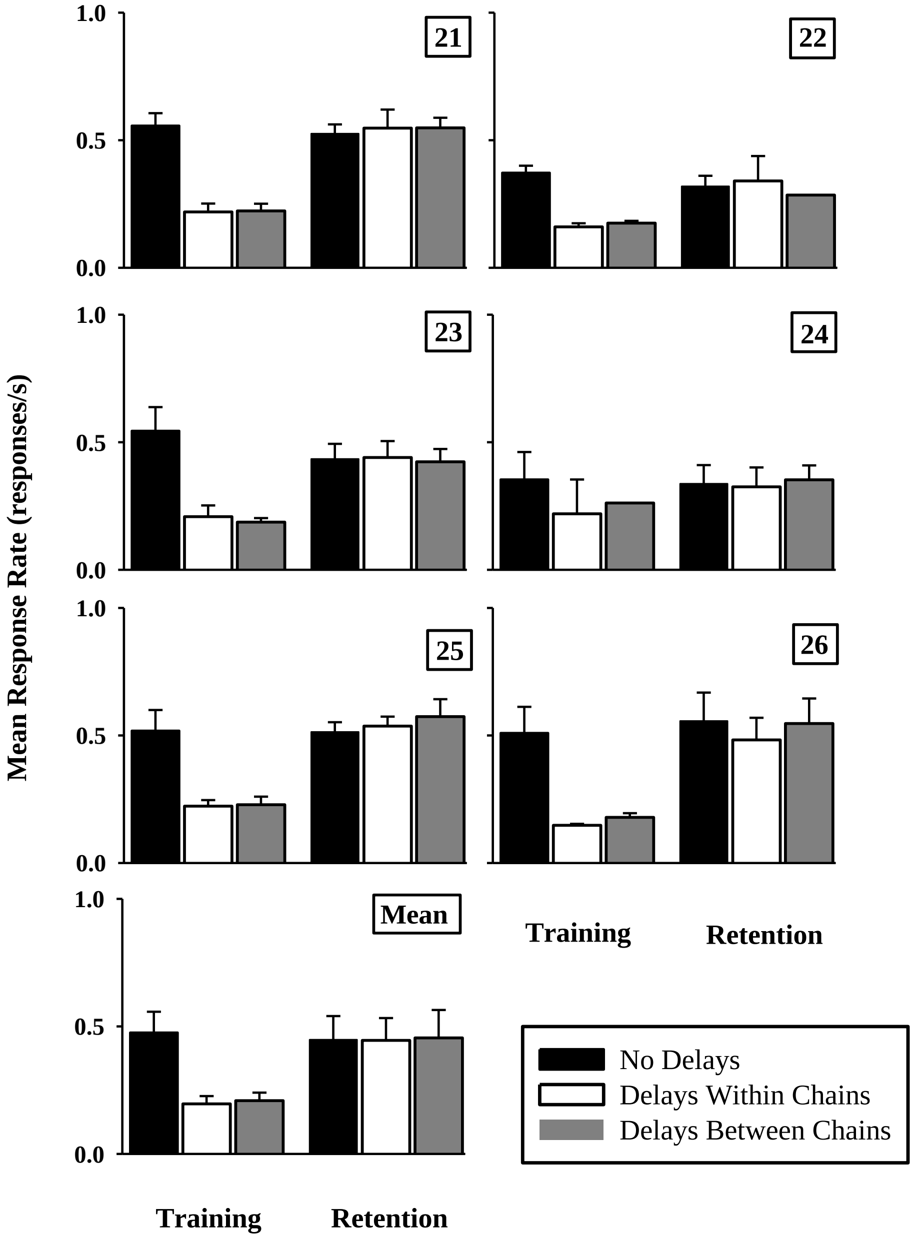


Figure 1.4.2. The mean response rates (+1 *SD*) obtained from the six sessions of the Within- and Between-chains delay condition and 12 sessions of the No-delay condition during the Training and Retention components for all subjects. The mean response rate data calculated across subjects is shown in the bottom left graph.

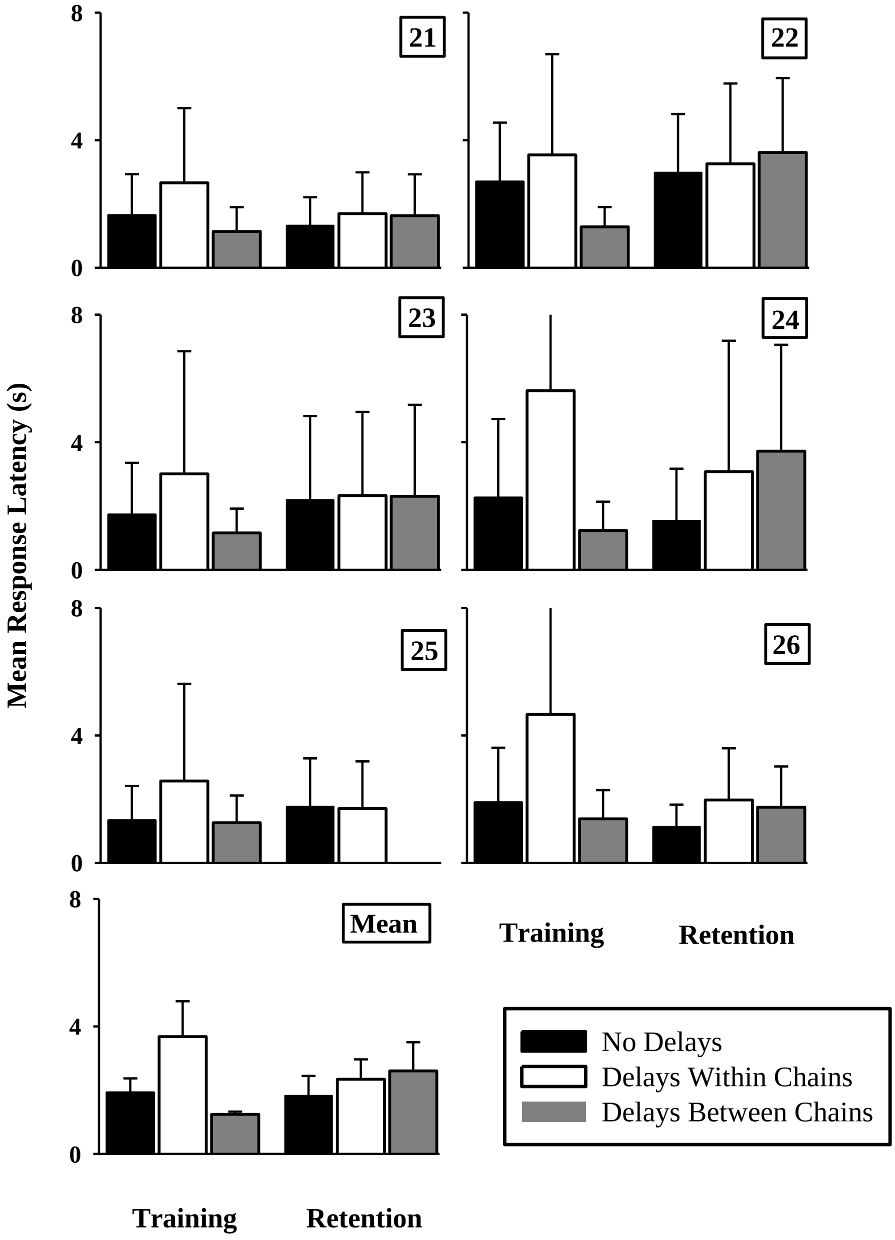


Figure 1.4.3. The mean response latency (with +1 *SD*) obtained from the six sessions of the Within- and Between-chains delay condition and 12 sessions of the No-delay condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph.

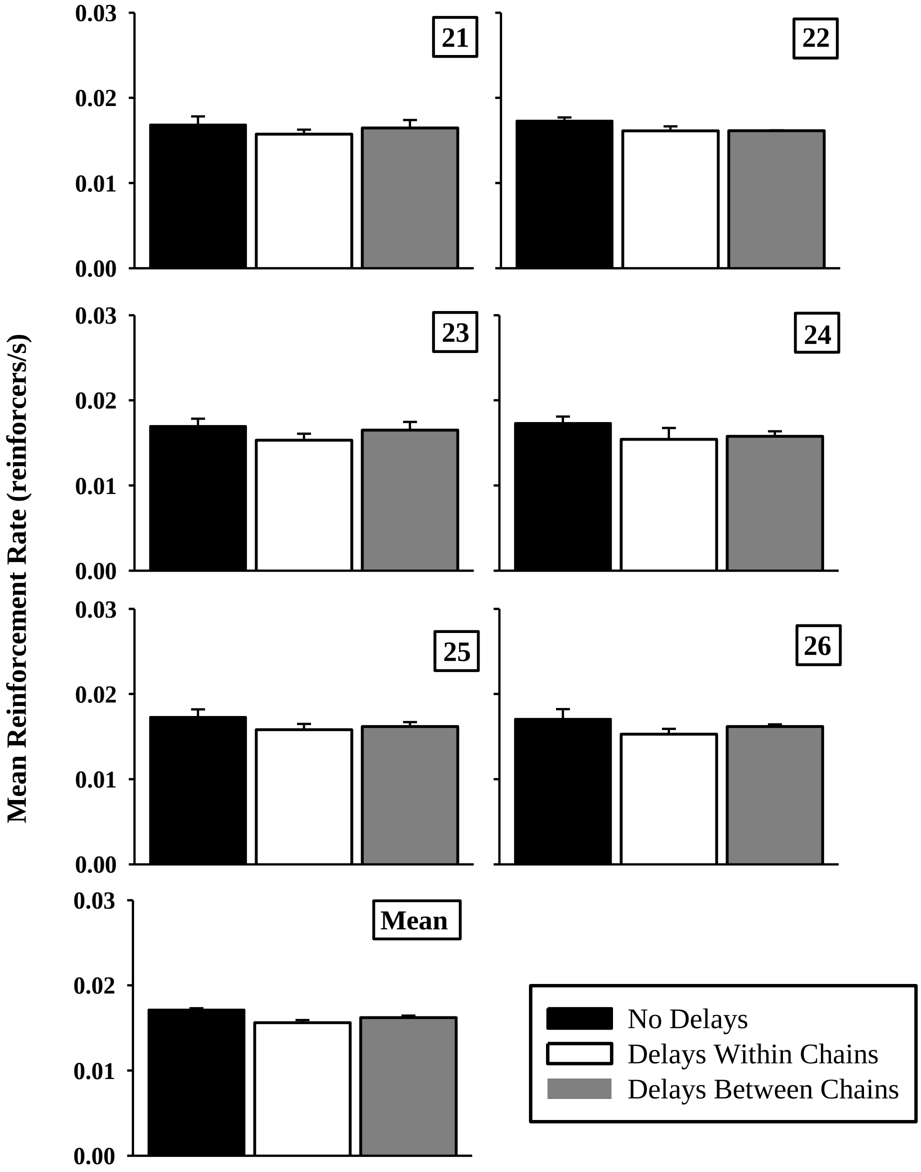


Figure 1.4.4. The mean reinforcement rate (+1 *SD*) obtained from the six sessions of the Within- and Between-chains delay condition and 12 sessions of the No-delay condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

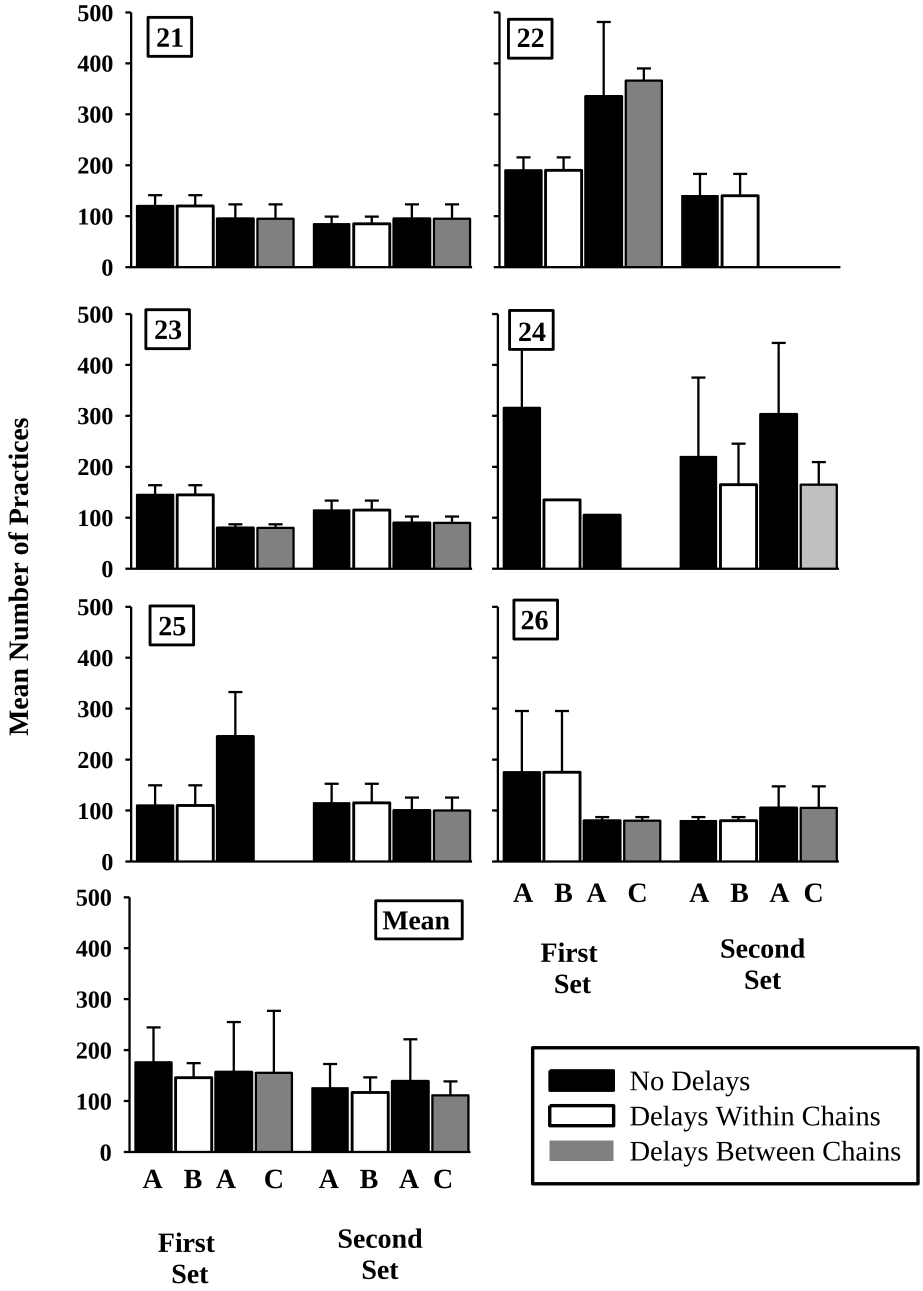


Figure 1.4.5. The mean number of practices (+1 *SD*) obtained from the three sessions of each condition for the first (e.g., A/B, Sessions 1-6; A/C, Sessions 13-18) and second (e.g., A/B, Sessions 1-6; A/C, Sessions 13-18) set of condition changes. The left four bars on each graph show the number of practices from the first set of condition changes. The right four bars on each graph show the number of practices from the second set of condition changes. The mean number of practices data calculated across subjects is shown in the bottom left graph.

Discussion

Experiment 1.4 replicated the experimental design from Porritt et al. (2009) with the aim of comparing results. It was expected that repeatedly exposing subjects to experimental conditions would produce results similar to those of Porritt (2007).

Similar to Experiment 1.2 and Porritt et al. (2009), percentage correct and response rate was found to be greatest during the No-delay condition. In addition, response latency was greatest during the Within-chains delay condition. There were no systematic differences in number of practices across conditions. Mean reinforcement rates were significantly different between the No-delay and Within-chains delay conditions (.002) and the No-delay and Between-chains delay conditions (.001). While these differences were statistically significant, they are not too different from the difference in means between the Within- and Between-chains delay conditions ($M < .0001$).

Contrary to Porritt et al. (2009), there were no systematic differences across experimental conditions in response rate, percentage correct, or latency during the Retention component. Thus repeatedly exposing subjects to experimental conditions did not produce Retention component results similar to those of Porritt (2007).

The findings of the present experiment show significant differences between training conditions, but all three conditions produced similar accuracy during the Retention component. Similar findings were obtained from Experiment 1.2, in which an A/B/C design was used. Given the finding from the present experiment and that using an A/B/C design results in similar outcomes and requires fewer sessions, the A/B/C design was used in the remaining experiments.

SUMMARY

Experiments 1.1-1.4 attempted to replicate procedures used by Porritt (2007) and Porritt et al. (2009) to compare results. Findings from Experiments 1.1-1.4 confirm the results from Porritt and Porritt et al., showing that rate-building improved training accuracy when number of practices and reinforcement rate were controlled. These findings support the use of rate-building procedures to improve accuracy during training in the Precision Teaching literature (Bucklin et al., 2000; Kubina, Aho, Mozzoni, & Malanga, 1998; McDowell & Keenan, 2001).

Experiment 1.1 replicated the procedures of Porritt (2007), with exception to yoking correct practices, to compare results. The No-delay condition produced the greatest accuracy during the Training component, but there were no differences in accuracy between experimental conditions during the Retention component, partially replicating the findings from Porritt (2007). The findings from Experiment 1.1 support the use of rate-building procedures to improve accuracy during training in the Precision Teaching literature (Bucklin et al., 2000; Kubina et al., 1998; McDowell & Keenan, 2001). As previously mentioned [Experiment 1.1 Discussion], the lack of replication may be due to either the differences in performance standard or type of practice yoked from those used by Porritt.

Experiment 1.2 replicated Porritt's (2007) methods by increasing the performance standard with the aim of comparing results. The No-delay condition produced the greatest accuracy during the Training component, but there were no differences in accuracy between experimental conditions during the Retention component, partially replicating the findings from Porritt. This finding shows that using a similar performance standard to that of Porritt (2007) did not generate similar outcomes. As previously mentioned [Experiment 1.1 Discussion], the lack of replication may be due to yoking correct practices, whereas Porritt yoked trials.

Experiment 1.3 yoked trial practices, similar to Porritt (2007), with the aim of comparing results. The No-delay condition produced the greatest accuracy during the Training component, but there were no differences in accuracy between experimental conditions during the Retention component, partially replicating the findings from Porritt. Yoking trials generated similar training and retention accuracy to that of Experiment 1.2. This finding suggests that yoking trials or correct responses does not impact percentage correct. Being that yoking corrects requires less practices, this finding is informative for future studies that investigate the effects of rate-building.

Experiment 1.4 used a stronger experimental design, identical to the one used by Porritt et al. (2009), in an attempt to replicate the Retention component results from Porritt (2007). The no-delay condition produced the greatest accuracy during the Training component, but there were no differences in accuracy between experimental conditions during the Retention component, partially replicating the findings from Porritt (2007) and Porritt et al. (2009). Using a stronger experimental design generated similar training and retention accuracy to that of Experiment 1.2, suggesting that using an experimental design that repeatedly exposes subjects to experimental conditions does not impact percentage correct during the Training or Retention component differently than using an alternating treatment design. Given these finding, and that using an A/B/C design requires fewer sessions, this outcome is informative for future studies that investigate the effects of rate-building.

Contrary to findings by Porritt (2007) and Porritt et al. (2009), results from Experiments 1.1-1.4 show that the No-delay conditions, arrangements that produced greater response rates, did not lead towards greater retention accuracy. Variations in methodology and experimental design were tested during Experiments 1.2-1.4 to compare findings to Porritt. These attempts failed to replicate Porritt’s findings and there are no obvious theoretical reasons for this lack of replication.

As previously mentioned [Experiment 1.3 Discussion], increasing the performance standard has been shown to improve retention accuracy (Berens et al., 2003; Ivarie, 1986). It may be that a greater performance standard is required for hens to replicate the Retention component findings from Porritt (2007). Experiment 1.5 increased the performance standard in the hope to compare findings to Porritt.

EXPERIMENT 1.5

As mentioned previously [Experiment 1.3 Discussion], Experiment 1.5 increased the performance standard to determine if hens require a greater performance standard to replicate the Retention component findings from Porritt (2007). Given results of Experiment 1.4, a design similar to Experiments 1.1 and 1.2 was used.

Method

Subjects

The same 6 subjects from Experiments 1.1-1.4 participated in this study.

Apparatus

The apparatus was the same as that used in Experiments 1.1-1.4.

Procedures

Procedures were identical to those of Experiments 1.1 and 1.2 with exception to the performance standard. The performance standard in Experiment 1.2 required subjects to complete five consecutive bins of five chain completions, each within 45 s. The new performance standard during the first series of conditions required subjects to complete five consecutive bins of seven chain completions, each within 45 s (i.e., Series 1). After hens were exposed to each experimental condition, the performance standard was increased to five consecutive bins of nine chain completions, each within 45 s, for a second series of the same conditions (i.e., Series 2) [please see p. 16-19 for a full description of training procedures].

Results

Table 1.5.1 shows the completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens in Series 1. Some hens did not complete the number of yoked practices within the session duration. These sessions were not used in the following analysis.

Table 1.5.2 shows the completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens in Series 2. There was very little data collected for this second series of conditions because the hens either could not complete the performance standard (e.g., Hens 22, 23, 24, 25) or could not complete the number of practices during the Within-chains delay condition within the session duration (e.g., Hens 21 & 26). In spite of the limited data, results for the Training and Retention components could still be described.

Group data for Figures 1.5.1-1.5.10 were analysed using a one-way repeated measures analysis of variance for all measures during the Training and Retention components for Experiment 1.5. The alpha level for all statistical comparisons in all situations was set at .05 and any results that reached this level were presented with an asterisk (*) in Table 1.5.3. Except where indicated with a hashtag (#) in Table 1.5.3, Mauchley's Test was not significant so sphericity was assumed. Post-hoc tests were conducted using the Bonferroni correction, as recommended by Fields (2005). Statistical analysis for group data in Figures 1.5.6-1.5.10 was not completed due to lack of completed sessions.

Figure 1.5.1 shows mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean percent correct data calculated across subjects is shown in the bottom left graph. Generally, accuracy in the Training component was similar during the No-delay and Between-chains delay conditions, and lower during the Within-chains delay condition, for all subjects. Table 1.5.3 shows these differences were significant and effect size, partial eta squared, was moderate (Ferguson, 2009). Retention accuracy was similar across the three experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.5.3).

Figure 1.5.2 shows the mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean response rate data calculated across subjects is shown in the bottom left graph. Training response rates were greatest in the No-delay condition then decreased during the Within- and Between-chains delay conditions for all subjects. Table 1.5.3 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention response rates were generally similar across the three experimental conditions for all subjects; showing no systematic effect from training conditions.

Figure 1.5.3 shows the mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean response latency data calculated across subjects is shown in the bottom left graph. Response latency in the Training component was greatest during the Within-chains delay condition, and was generally similar during the No-delay and Between-chains delay conditions, for all subjects. Table 1.5.3 shows this finding was not significant. Retention latencies were similar across all experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.5.3).

Figure 1.5.4 shows the mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition for all subjects during Series 1. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph. Reinforcement rates were greatest during the No-delay condition, and generally similar during the Within- and Between-chains delay conditions. The differences between conditions were significant (Table 1.5.3).

Figure 1.5.5 shows the mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component for all subjects in Series 1. The mean number of practices data calculated across subjects is shown in the bottom left graph. The number of practices to achieve the performance standard was similar across each experimental condition for all subjects. Table 1.5.3 shows no significant differences across conditions.

Figure 1.5.6 shows mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean percent correct data calculated across subjects is shown in the bottom left graph. Generally, accuracy during the No-delay condition for the Training and Retention components were similar to the first series data for all hens.

Figure 1.5.7 shows the mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean response rate data calculated across subjects is shown in the bottom left graph. Response rates during the No-delay condition for the Training and Retention components were similar to the first series data for all hens.

Figure 1.5.8 shows the mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean response latency data calculated across subjects is shown in the bottom left graph. Response latency during the No-delay condition for the Training and Retention components were similar to the first series data for all hens.

Figure 1.5.9 shows the mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition for all subjects during Series 2. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph. Reinforcement rate during the No-delay condition was similar to the first series data for all hens.

Figure 1.5.10 shows the mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component for all subjects in Series 2. The mean number of practices data calculated across subjects is shown in the bottom left graph. Generally, the number of practices to achieve the performance standard during the No-delay condition of the Training component was similar to the first series data for all hens.

Table 1.5.1.

Completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens during Series 1.

Condition	Chain	Hen 21	Hen 22	Hen 23	Hen 24	Hen 25	Hen 26
A	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨	▨	▨	▨
B	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨			▨
	R-L-R	▨	▨	▨			▨
C	L-C-L	▨	▨	▨			▨
	R-L-C		▨	▨			▨
	R-L-R		▨	▨			▨

Table 1.5.2.

Completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens during Series 2.

Condition	Chain	Hen 21	Hen 22	Hen 23	Hen 24	Hen 25	Hen 26
A	L-C-L	▨		▨	▨	▨	▨
	R-L-C	▨		▨			▨
	R-L-R	▨					▨
B	L-C-L						
	R-L-C						
	R-L-R						
C	L-C-L						
	R-L-C						
	R-L-R						

Table 1.5.3

Analysis of variance results during the Training and Retention components of Experiment 1.5.

Component	MS Treatment	MS Error	df	F	p	Partial Eta Squared
			Mean percent correct			
Training	155.07	78.36	2, 6	1.98	.22	0.40
Retention	.033	4.52	2, 6	0.01	1.00	0.00
			Mean response rate			
Training	.192	.004	2, 6	44.57	p<.0001*	0.94
Retention	.004	.003	2, 6	1.37	.32	0.31
			Mean response latency			
Training	9.78	1.58	2, 6	6.20	.09#	0.67
Retention	.04	.04	2, 6	0.83	.48	0.22
			Reinforcement rate			
Session	2.4E-6	1.9E-7	2, 6	12.40	.01*	0.81
			Number of practices			
Correct	197.83	199.06	2, 6	0.99	.39#	0.25

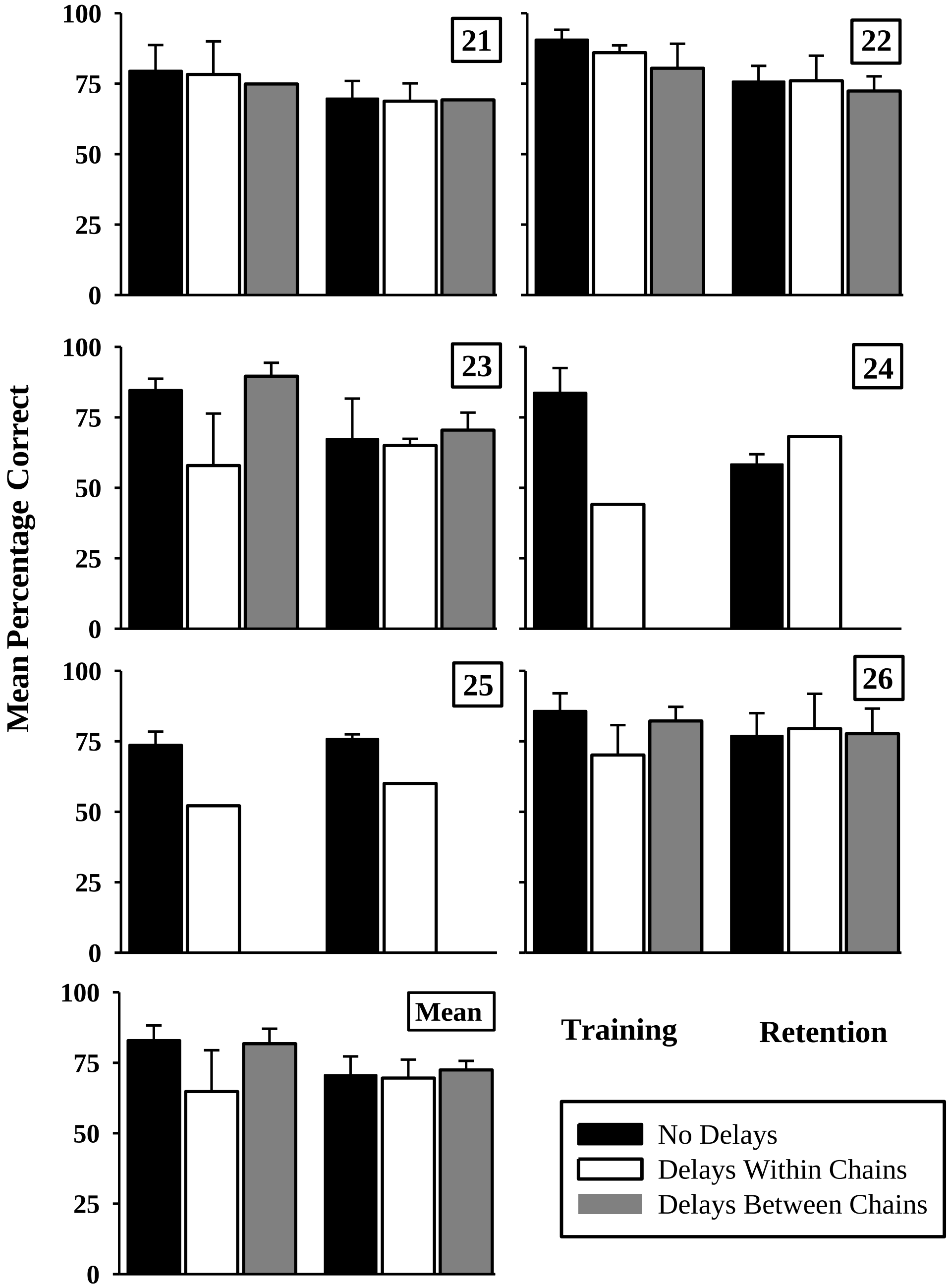


Figure 1.5.1. The mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean percent correct data calculated across subjects is shown in the bottom left graph.

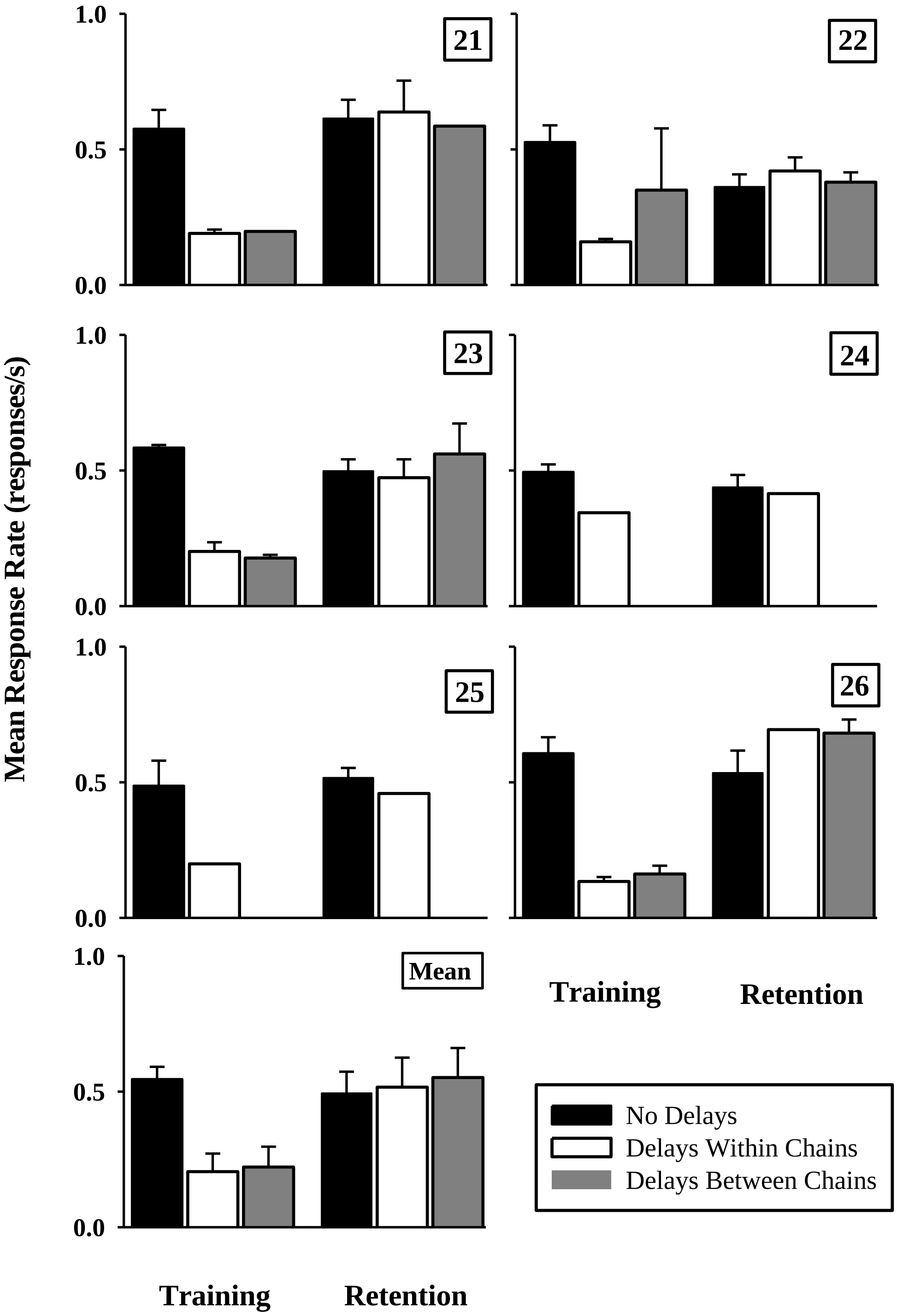


Figure 1.5.2. The mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean response rate data calculated across subjects is shown in the bottom left graph.

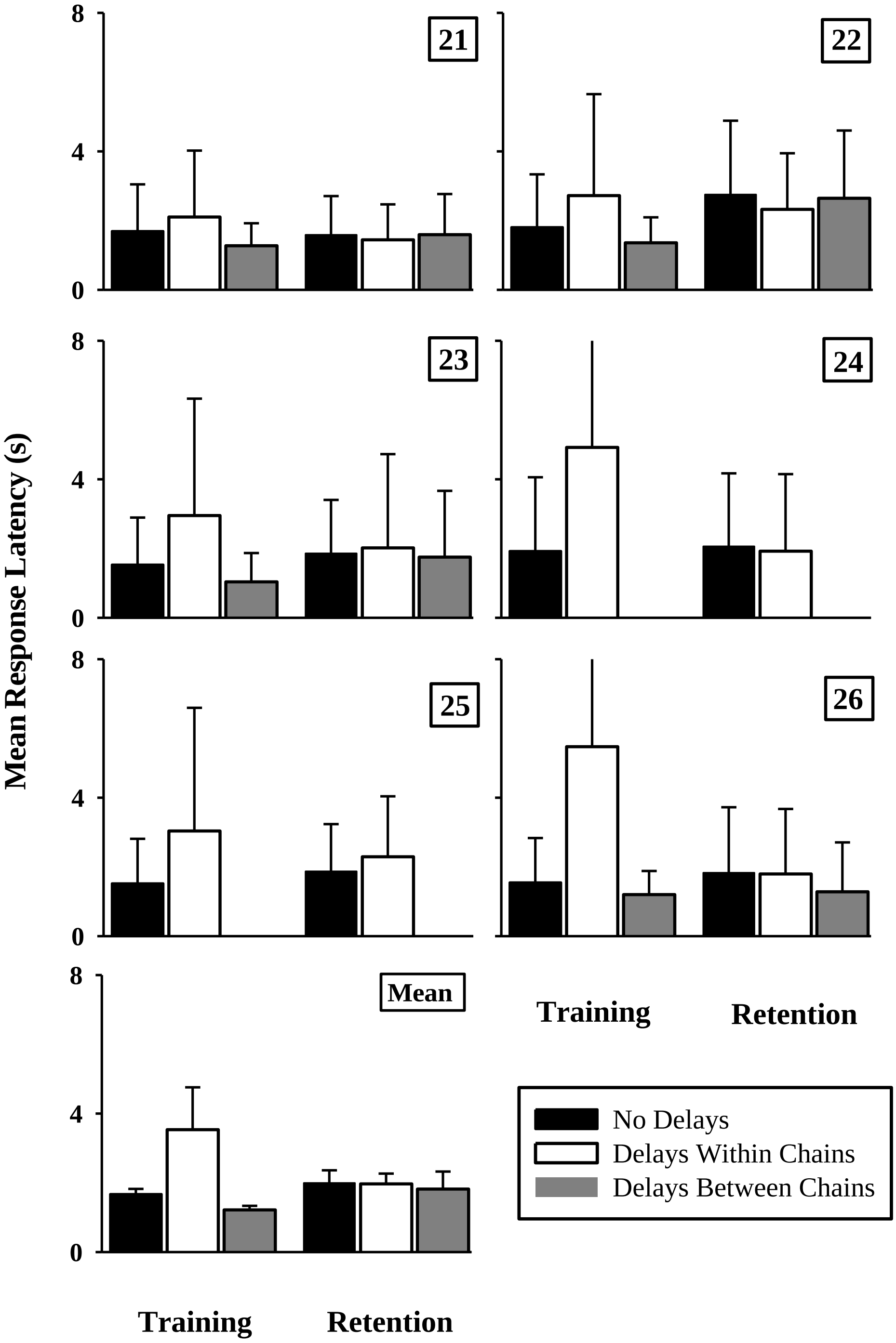


Figure 1.5.3. The mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean response latency data calculated across subjects is shown in the bottom left graph.

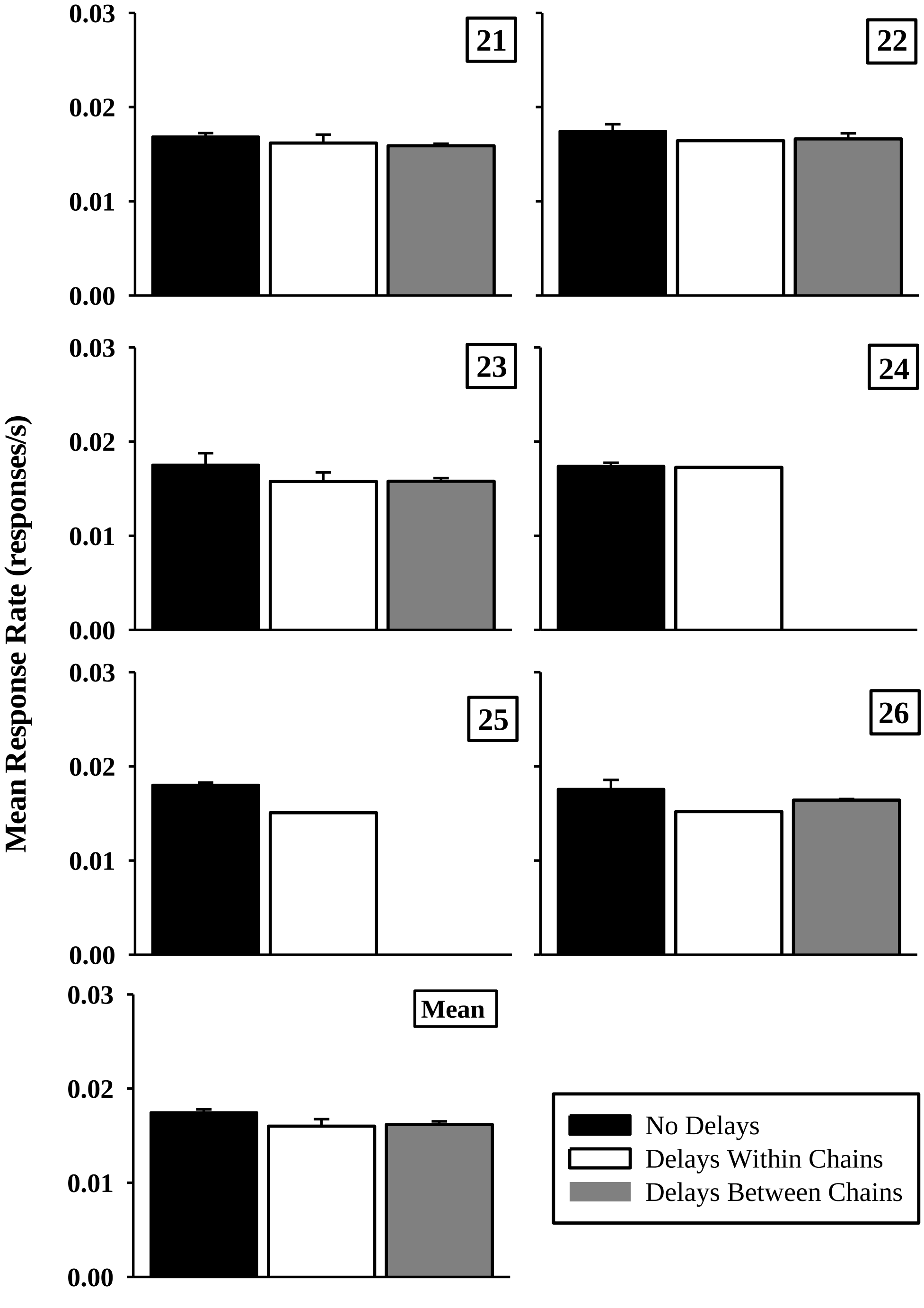


Figure 1.5.4. The mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition for all subjects during Series 1. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

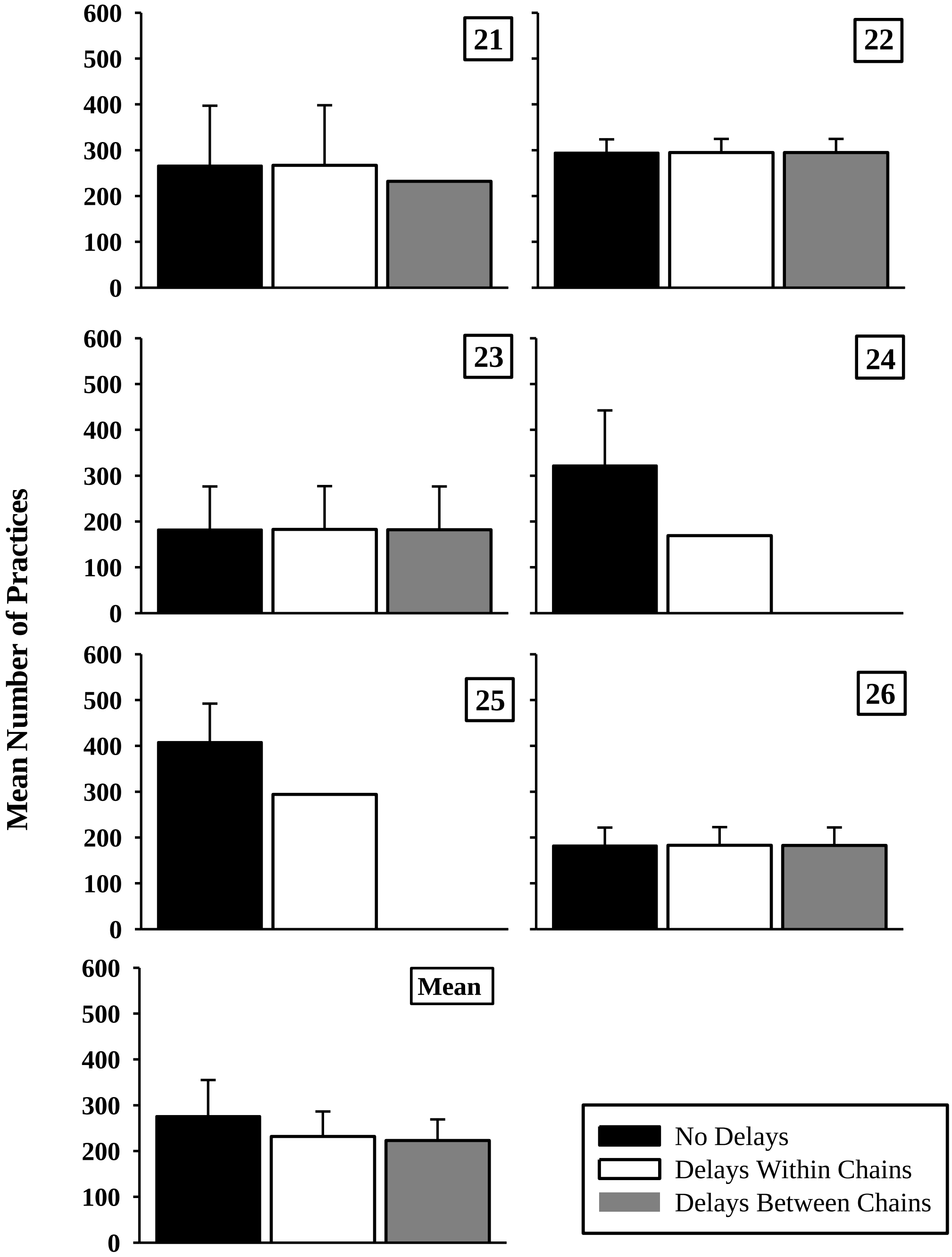


Figure 1.5.5. The mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component for all subjects in Series 1. The mean number of practices data calculated across subjects is shown in the bottom left graph.

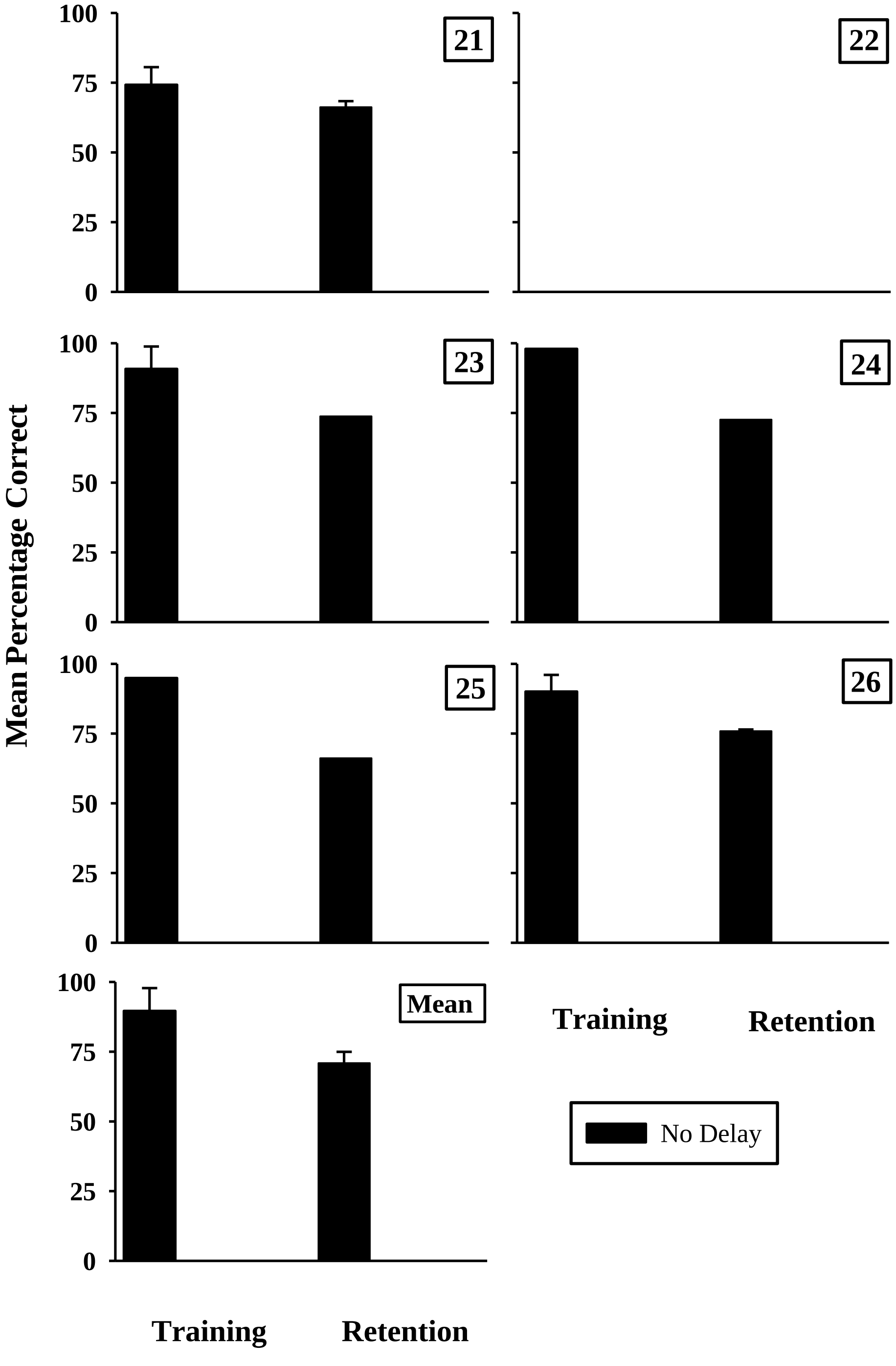


Figure 1.5.6. The mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

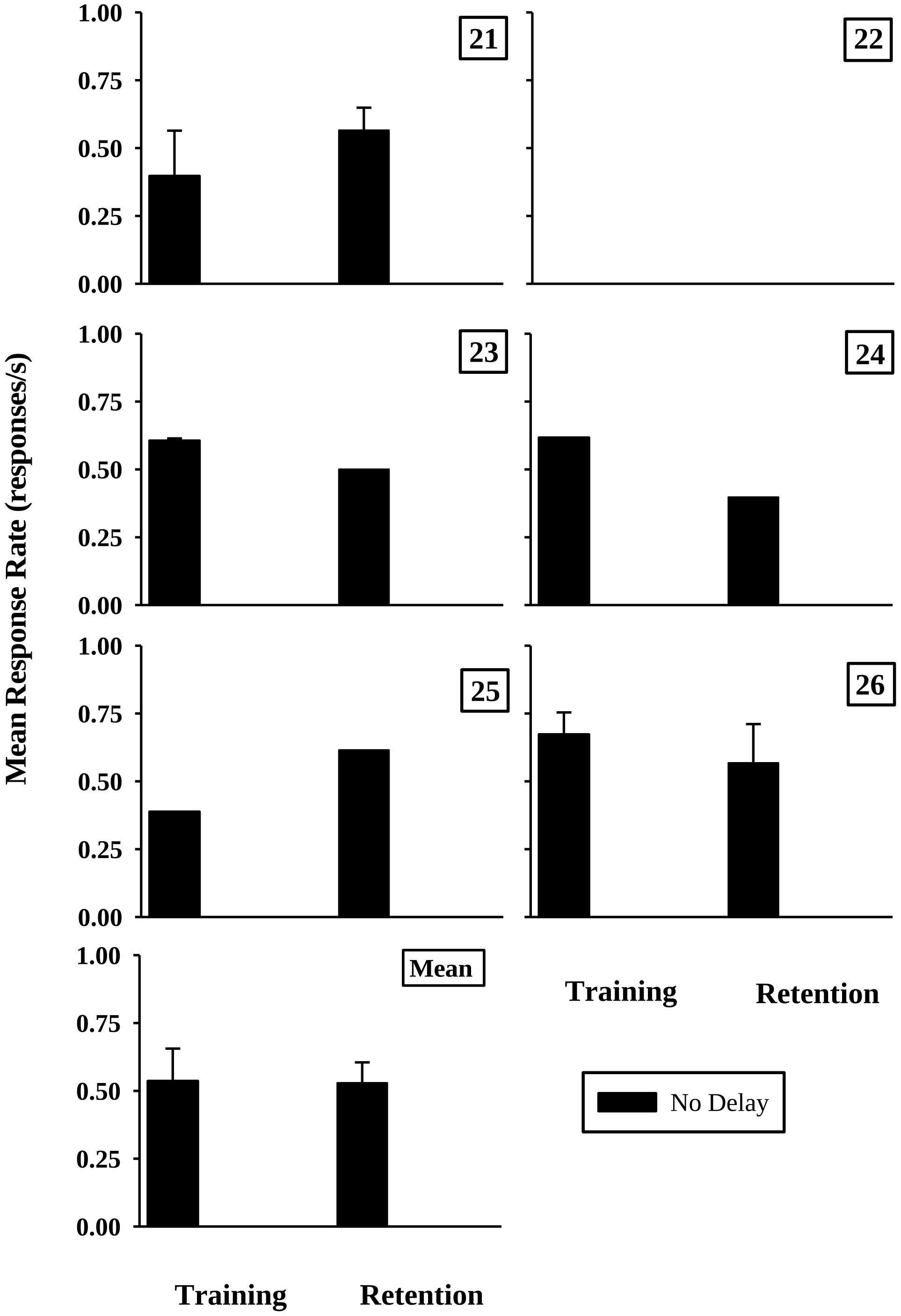


Figure 1.5.7. The mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean response rate data calculated across subjects is shown in the bottom left graph.

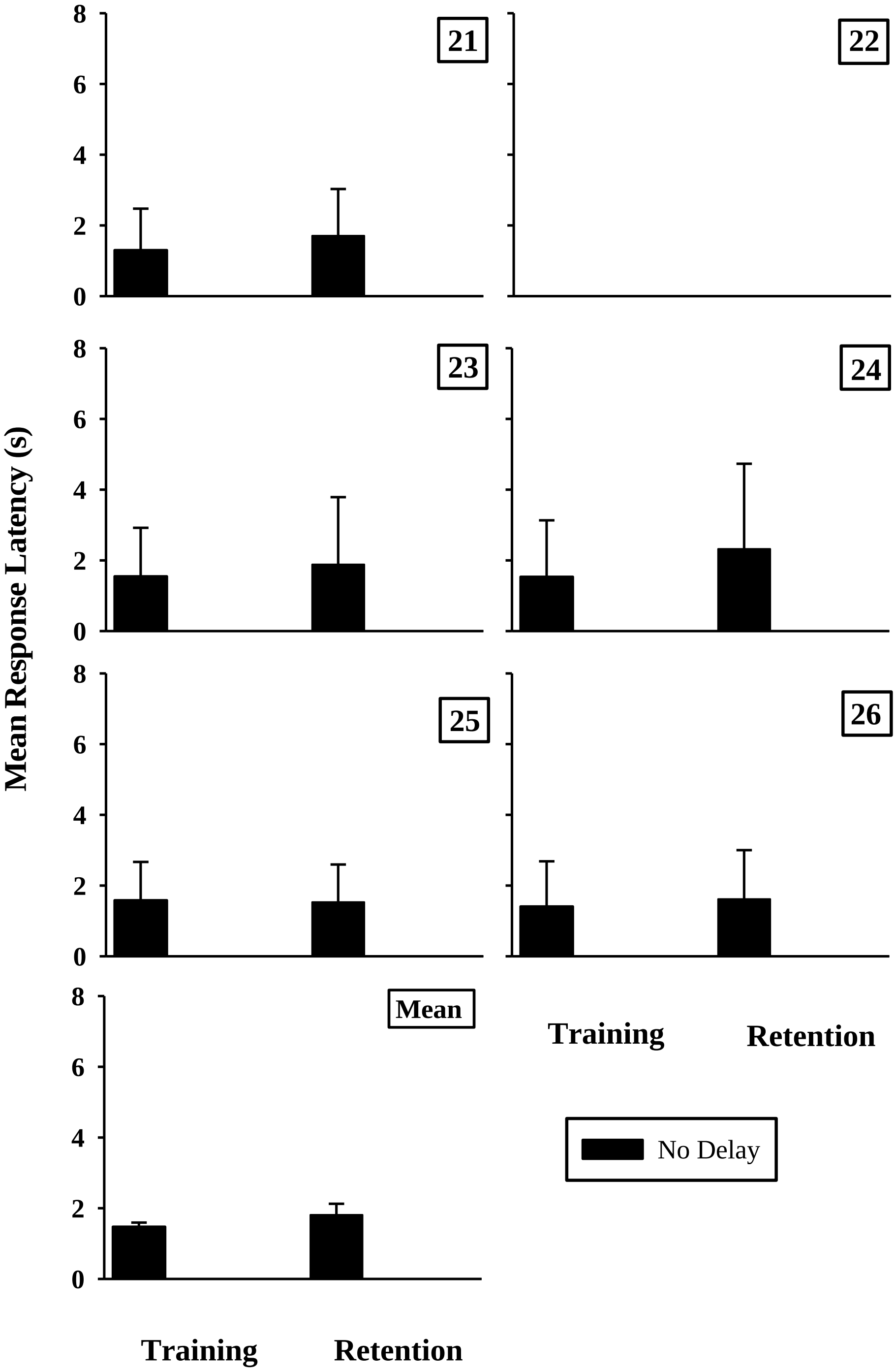


Figure 1.5.8. The mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean response latency data calculated across subjects is shown in the bottom left graph.

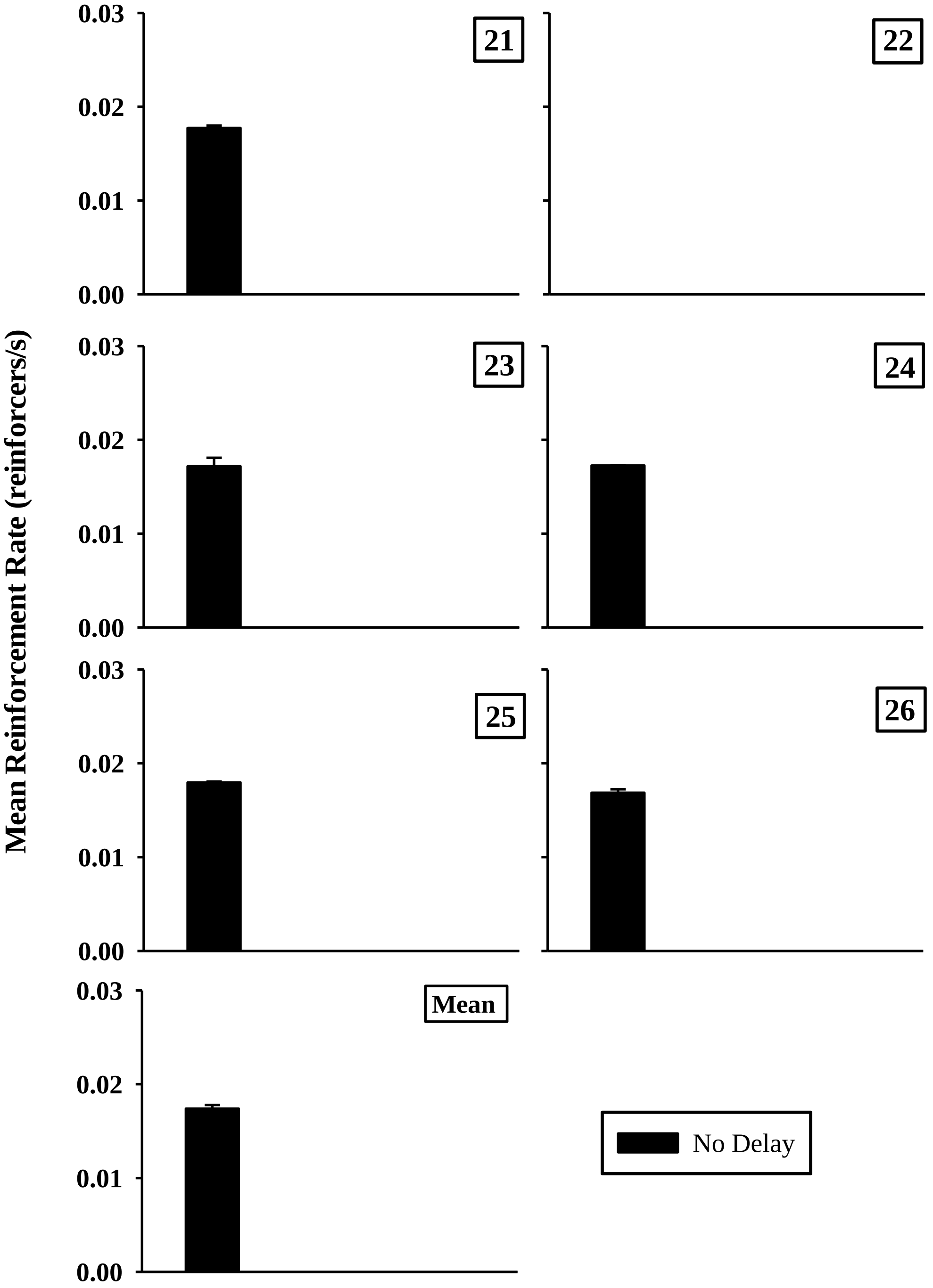


Figure 1.5.9. The mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition for all subjects during Series 2. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

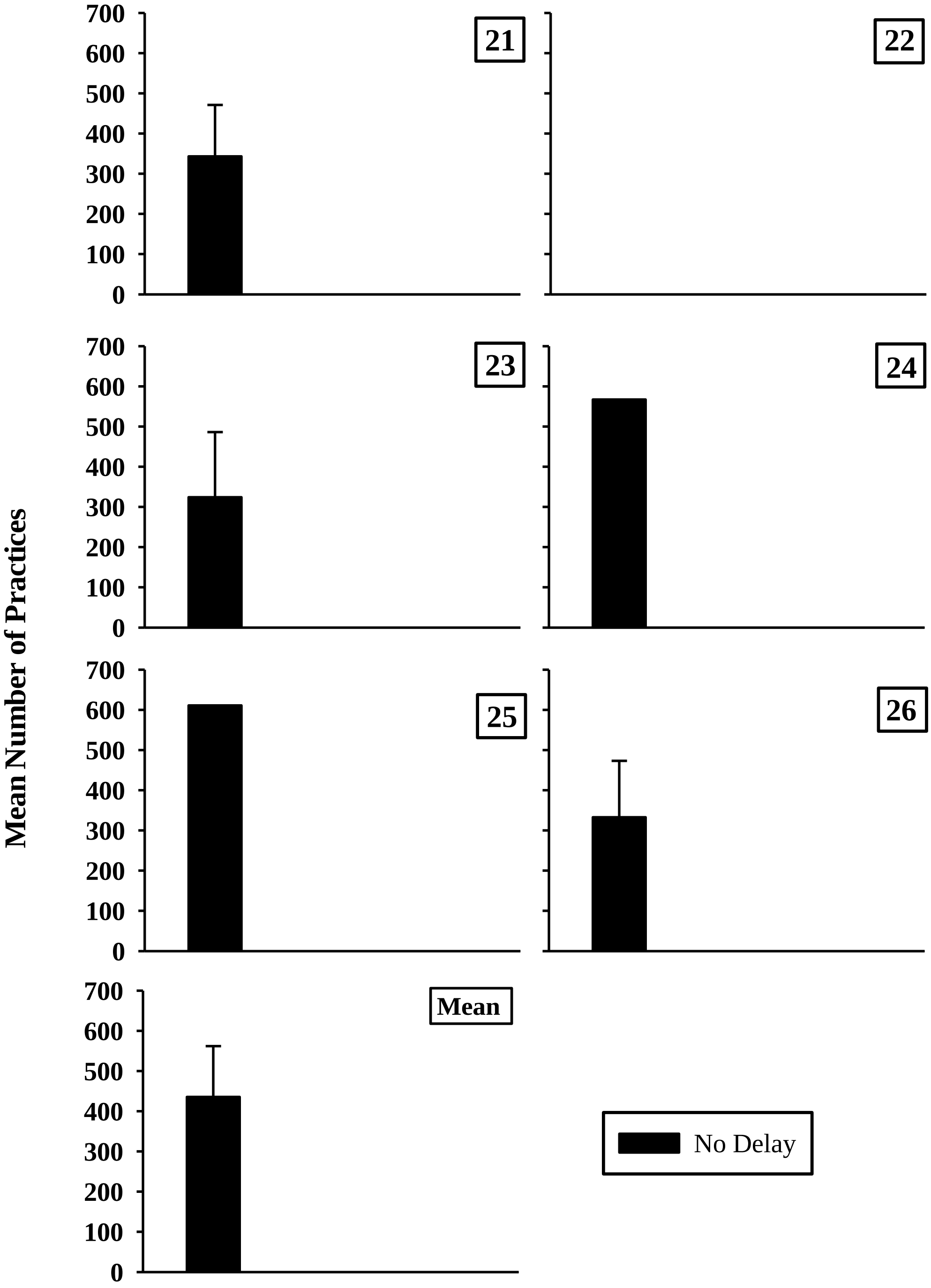


Figure 1.5.10. The mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component for all subjects in Series 2. The mean number of practices data calculated across subjects is shown in the bottom left graph.

Discussion

Experiment 1.5 increased the performance standard to determine if hens require a greater performance standard to replicate the Retention component findings from Porritt (2007).

As in Experiment 1.2 and Porritt (2007), percentage correct and response rate for Series 1 in the present study were greatest during the No-delay condition of the Training component. In addition, response latency was greatest during the Within-chains delay condition. There were no systematic differences in number of practices across conditions; however, mean reinforcement rates were significantly different between the No-delay and Between-chains delay conditions (.001). While this difference was statistically significant, it is similar to the differences in means for the No-delay and Within-chains delay conditions (.001) and the Within- and Between chains delay conditions ($M < .0001$).

It was expected that increasing the performance standard during Series 1 of the present study would produce Retention component findings similar to Porritt (2007). However, contrary to Porritt, there were no systematic differences across experimental conditions in response rate, percentage correct, or latency during the Retention component. This finding is similar to the Retention component results for Experiment 1.2.

Series 2 of the present investigation used a greater performance standard than Series 1 to replicate the Retention component findings of Porritt (2007). Subjects were unable to complete most experimental conditions during Series 2 given the present session parameters, as shown in Table 1.5.2. This lack of data for each experimental condition prevented comparisons to be made across conditions. This finding suggests, given the session parameters, there is a maximum performance standard that can be used with hens.

Overall, a performance standard of nine chains produced greater percentage correct during the Training component than when seven chains within 45 s was used. This finding should be taken with caution because more practices were required to achieve the greater performance standard. The nine chains performance standard did not generate greater percentage correct during the Retention component than the seven chains standard. This finding is contrary to

Precision Teaching research that shows greater performance standards lead towards enhanced retention accuracy (Berens et al., 2003; Ivarie, 1986).

Retention has been shown to improve as the time between practices is distributed further apart (e.g., the spacing effect, Baddeley & Longman, 1978, Ebbinghaus, 1913). Using himself as the subject, Ebbinghaus repeated a 12-syllable series and found 68 immediately successive repetitions made an errorless recital possible. However, he achieved the same result by distributing 38 practices over three days. Ebbinghaus concluded that "with any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time" (1913, p.89), more recent findings have drawn similar conclusions (Baddeley & Longman, 1978; Bloom & Shuell, 1981). Based upon Ebbinghaus's conclusion and recent research on distributed practice (Baddeley & Longman, 1978; Bloom & Shuell, 1981) retention accuracy should improve by increasing the duration between each practice. The aim of Experiment 1.6 was to examine the effects of distributed practice on retention accuracy as this has been shown to affect retention accuracy.

EXPERIMENT 1.6

Experiment 1.6 examined the effects of distributed practice on retention accuracy.

Method

Subjects

The same 6 subjects from Experiments 1.1-1.5 participated in this study.

Apparatus

The apparatus was the same as that used in Experiments 1.1-1.5.

Procedures

Procedures were identical to those of Experiment 1.2 with one exception. Subjects were exposed to a Within-chains delay condition that imposed a 10-s delay between each practice (e.g., correct response) [please see p. 16-19 for a full description of training procedures].

Results

Table 1.6.1 shows the completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens. Some hens did not complete the number of yoked practices within the session duration. These sessions were not used in the following analysis.

Group data for Figures 1.6.1-1.6.5 were analysed using a one-way repeated measures analysis of variance for all measures during the Training and Retention components for Experiment 1.6. The alpha level for all statistical comparisons in all situations was set at .05 and any results that reached this level were presented with an asterisk (*) in Table 1.6.2. Except where indicated with a hashtag (#) in Table 1.6.2, Mauchley's Test was not significant so sphericity was assumed. Post-hoc tests were conducted using the Bonferroni correction, as recommended by Fields (2005).

Figure 1.6.1 shows mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean percent correct data calculated across subjects is shown in the bottom left graph. Generally, accuracy in the Training component was similar during the No-delay and Between-chains delay conditions, and lower under the Within-chains delay condition, for all subjects. Table 1.6.2 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention accuracy was similar across the three experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.6.2).

Figure 1.6.2 shows the mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response rate data calculated across subjects is shown in the bottom left graph. Training response rates were greatest during the No-delay condition, and lowest during the Within- and Between-chains delay conditions, for all subjects. Table 1.6.2 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention response rates were similar during the No-delay and Between-chains delay conditions, and lowest during the Within-chains delay condition. The data from these conditions were significantly different (Table 1.6.2) and effect size, partial eta squared, was large (Ferguson, 2009).

Figure 1.6.3 shows the mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph. Response latency in the Training and Retention components were largest during the Within-chains delay condition, and generally similar during the No-delay and Between-chains delay conditions. Both Training and Retention component findings were significant (Table 1.6.2).

Figure 1.6.4 shows the mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph. Reinforcement rates were similar for the No-delay and Between-chains delay conditions, and lowest during the Within-chains delay condition. Table 1.6.2 shows significant differences across conditions.

Figure 1.6.5 shows the mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component. The mean number of practices data calculated across subjects is shown in the bottom left graph. The number of practices to achieve the performance standard was similar across each experimental condition for all subjects. Table 1.1.3 shows no significant differences across conditions.

Table 1.6.1.

Completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens.

Condition	Chain	Hen 21	Hen 22	Hen 23	Hen 24	Hen 25	Hen 26
A	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨	▨	▨	▨
B	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨		▨
	R-L-R	▨	▨	▨	▨		
C	L-C-L	▨	▨	▨	▨		
	R-L-C	▨	▨	▨	▨		
	R-L-R	▨	▨	▨	▨		

Table 1.6.2

Analysis of variance results during the Training and Retention components of Experiment 1.6.

Component	MS Treatment	MS Error	df	F	p	Partial Eta Squared
			Mean percent correct			
Training	860.33	23.01	2, 6	37.39	p<.0001*	.93
Retention	.41	17.82	2, 6	0.02	.98	.01
			Mean response rate			
Training	.40	.005	2, 6	80.65	.003*#	.97
Retention	.008	.001	2, 6	6.32	.03*	.68
			Mean response latency			
Training	3.05	.09	2, 6	32.38	.001*	.92
Retention	.07	.01	2, 6	8.20	.02*	.73
			Reinforcement rate			
Session	3.5E-6	6.1E-7	2, 10	5.71	.02*	.53
			Number of practices			
Correct	8729.15	3598.15	2, 10	2.43	.18#	.38

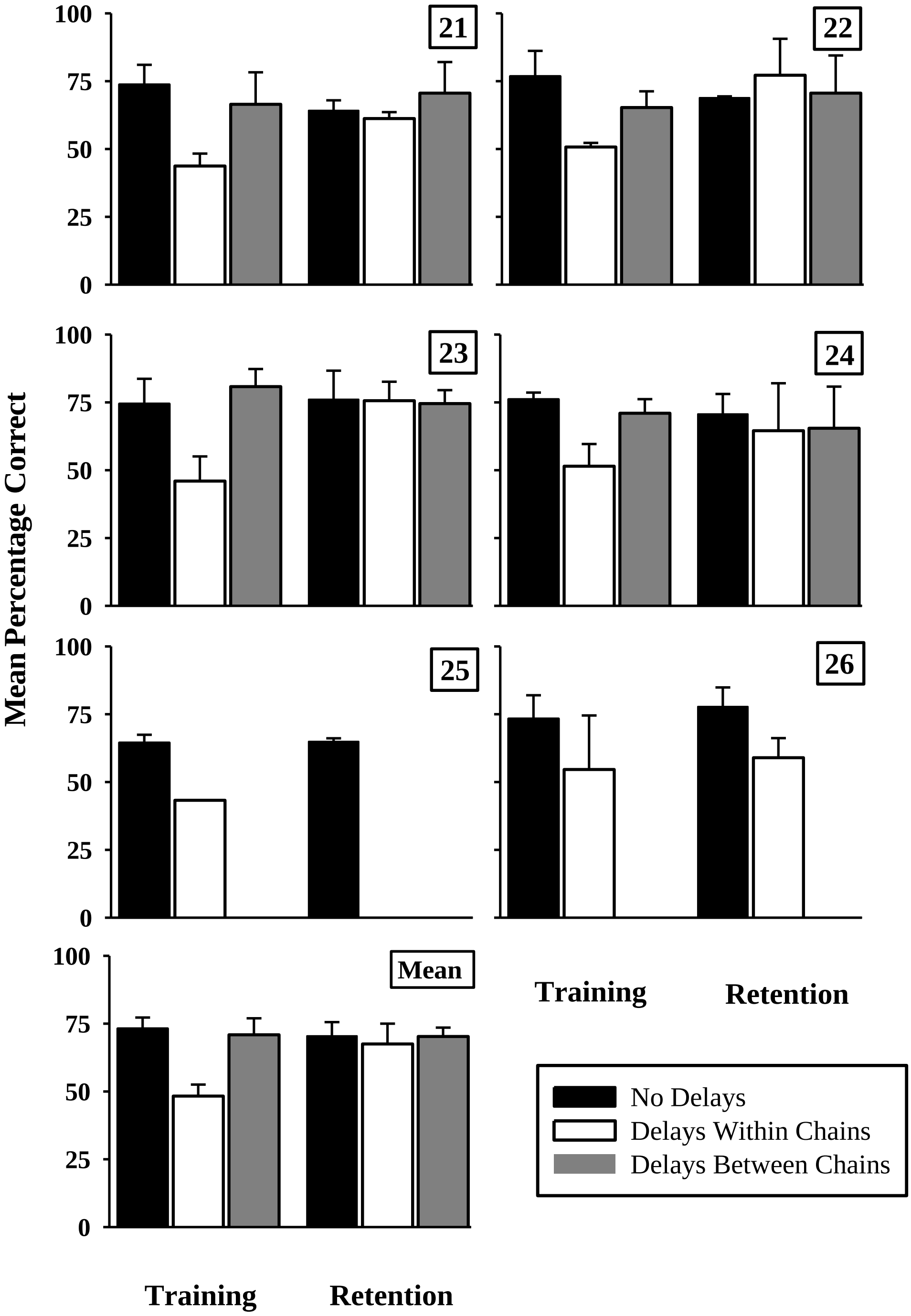


Figure 1.6.1. The mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean percent correct data calculated across subjects is shown in the bottom left graph.

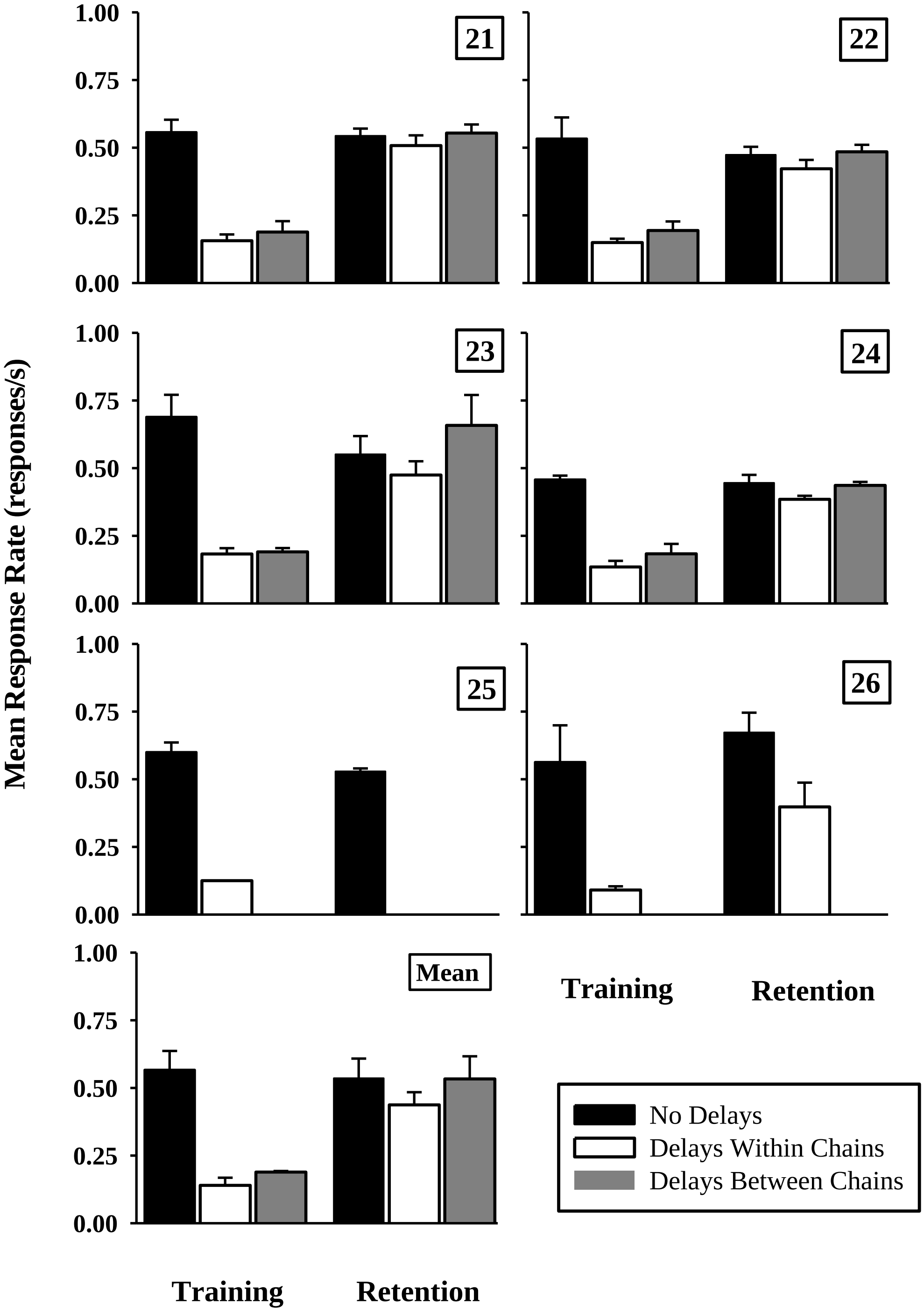


Figure 1.6.2. The mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response rate data calculated across subjects is shown in the bottom left graph.

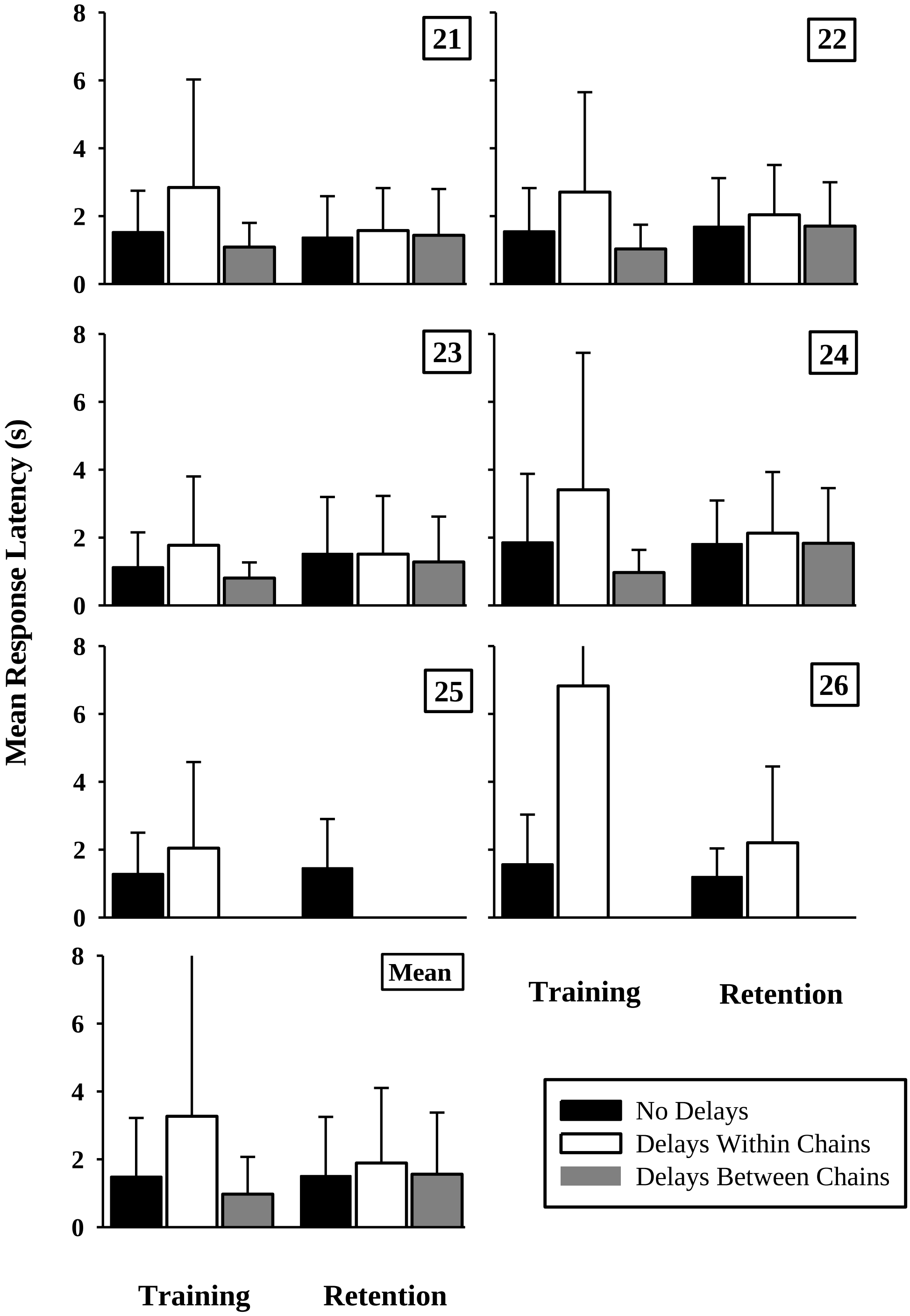


Figure 1.6.3. The mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph.

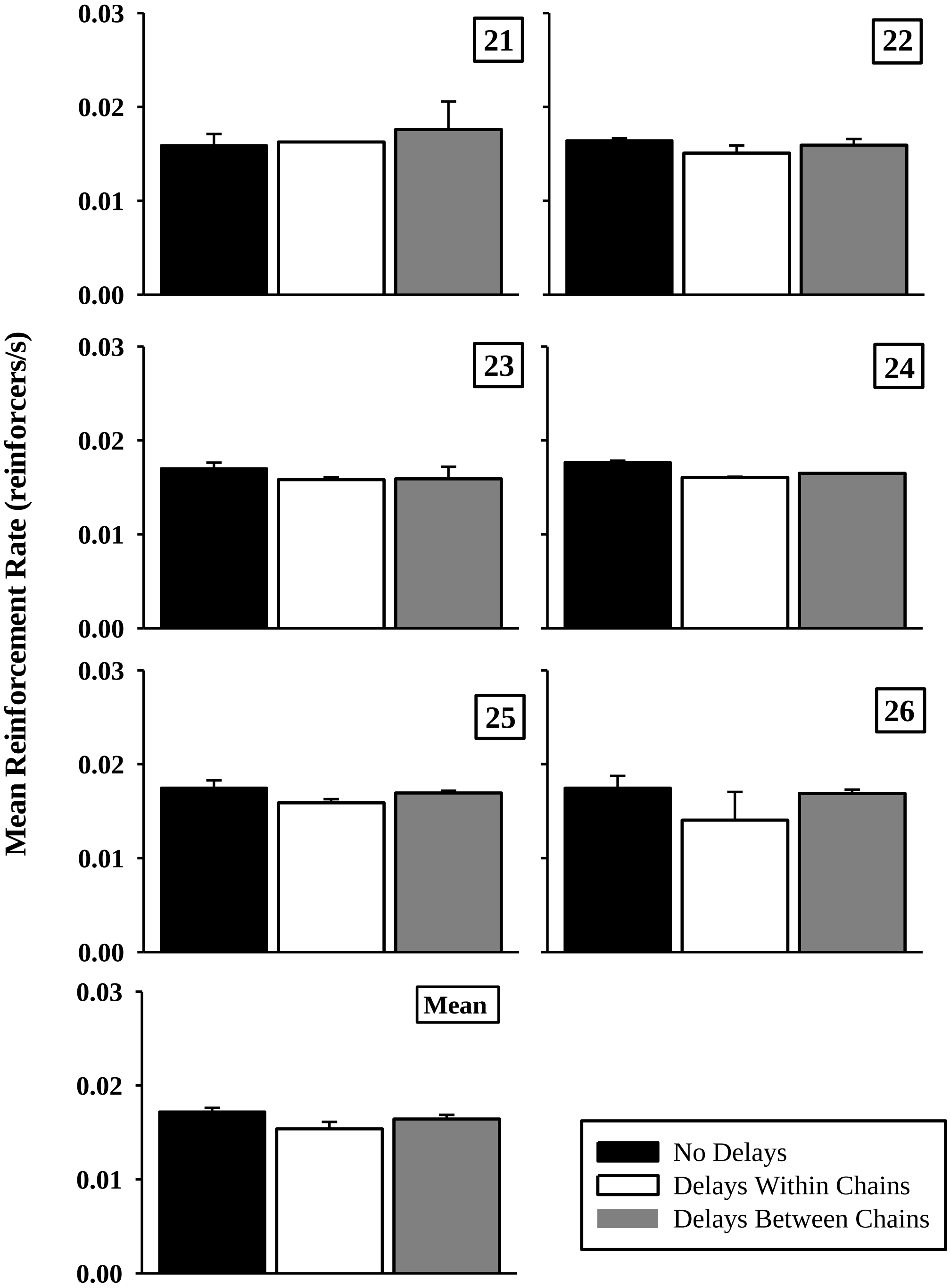


Figure 1.6.4. The mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

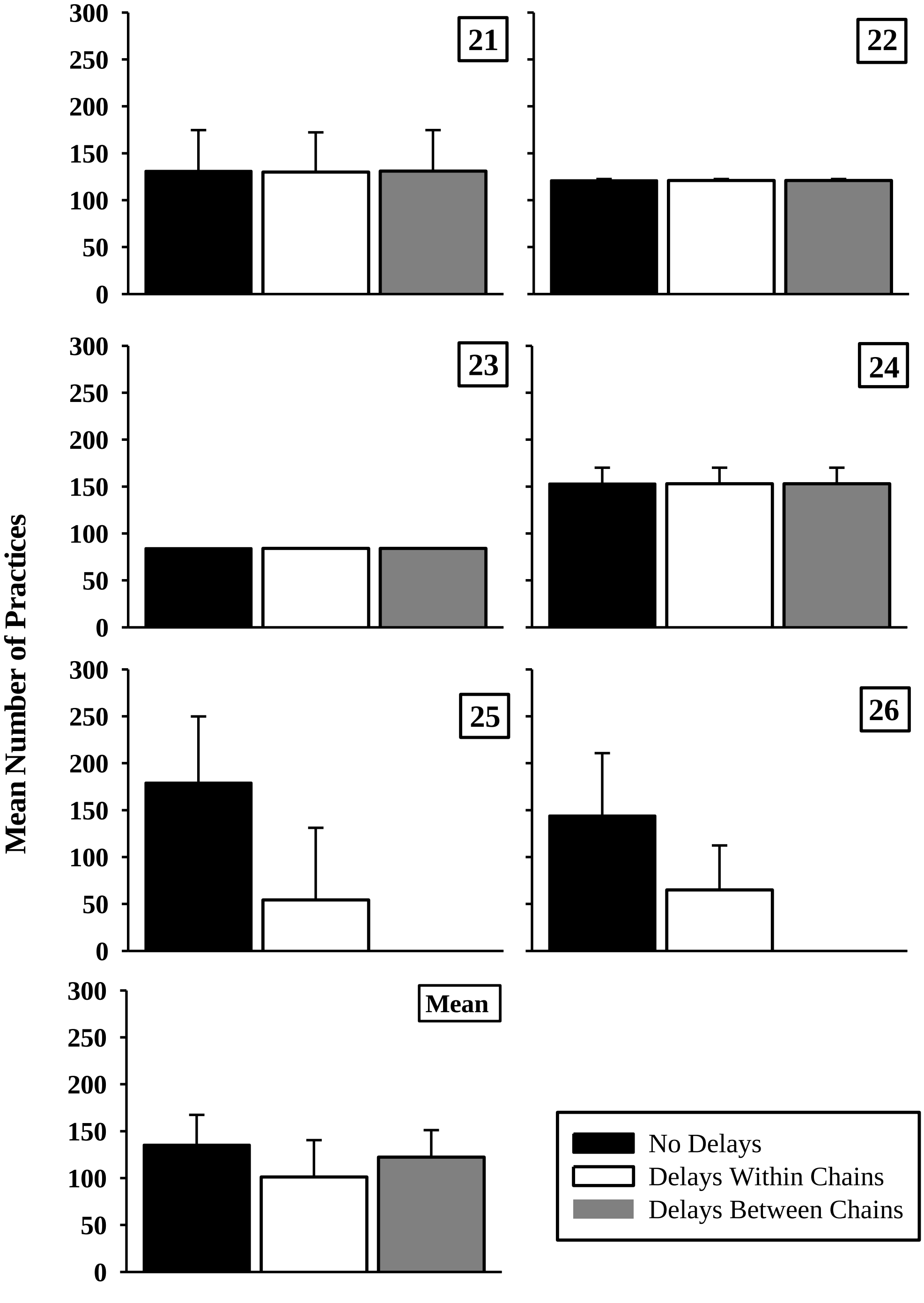


Figure 1.6.5. The mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component. The mean number of practices data calculated across subjects is shown in the bottom left graph.

Discussion

Experiment 1.6 examined the effects of distributed practice on retention accuracy. As in Experiment 1.2, percentage correct and response rate were greatest during the No-delay condition of the Training component. In addition, response latency was greatest during the Within-chains delay condition. There were no systematic differences in number of practices across conditions. However, reinforcement rates were significantly different between the No-delay and Within-chains delay conditions and the No-delay and Between-chains delay conditions. Post-hoc tests suggest this effect was not due to any specific pair of mean differences.

It was expected that distributing practices would enhance retention accuracy. Latency and response rates differed across experimental conditions during the Retention component; suggesting conditions during the Training component had some effect over responding in the Retention component. However, percentage correct during the Retention component was similar across experimental conditions. Thus, no conclusions can be drawn about whether the Training component conditions affected responding during the Retention component. The percentage correct findings are similar to Experiment 1.2, suggesting that distributing practices did not enhance retention accuracy.

The Retention component results of the present experiment are contrary to studies showing that distributed practices increase retention accuracy (Baddeley & Longman, 1978; Bloom & Shuell, 1981). One difference between the present study and those from the distributed practice literature, however, is that reinforcement rate and number of practices was controlled in the present investigation. This difference could account for the difference in findings.

Another variable that has been shown to effect retention accuracy is the retention interval. White (1985, 2001) points out that in studies of memory [where no occasioning stimuli is available at the point of recall], retention accuracy decreases as the time from the to-be-remembered stimulus increases. Typically, this research uses retention intervals from 1s-60s (Roberts & Grant, 1976). Stimulus control is often lost at the longer intervals. Given this, reliable recall after 23 hours seems unlikely. The first series of conditions during

Experiment 1.7 examined the effects of a shorter retention interval on retention accuracy under the No-delay, Within- and Between-chains delay conditions.

The results of Experiment 1.6 suggest that distributing practices during the Within-chains delay condition using a 23-h retention interval did not improve accuracy during the Retention component. Given that stimulus control from the to-be-remembered stimulus weakens as time passes (White, 1985, 2001), distributing practices within-chains at a shorter retention interval may enhance retention accuracy more so than at the longer interval. The second series of conditions in the present experiment replicated the procedures of Experiment 1.6 using a shorter retention interval to compare findings.

EXPERIMENT 1.7

The first series of Experiment 1.7 examined the effects of a shorter retention interval on accuracy. The second series of conditions replicated the procedures of Experiment 1.6 using a shorter retention interval to compare findings.

Method

Subjects

The same 6 subjects from Experiments 1.1-1.6 participated in this study.

Apparatus

The apparatus was the same as that used in Experiments 1.1-1.6.

Procedures

Procedures for the first series of conditions were identical to those of Experiment 1.2 with two exceptions. First, the retention interval was reduced from 23-hr to 10min. Second, the Retention component followed the Training component during each session. Third, the Distracter component was removed from each session. Thus, during each session, a subject was first exposed to the Training component followed by the retention interval and then the Retention component. The second series of conditions used these procedures with one exception. The within-chains delay was increased from 5 s to 10 s, similar to the delay used in Experiment 1.6 [please see p. 16-19 for a full description of training procedures].

Results

Tables 1.7.1 and 1.7.2 show the completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens during Series 1 and Series 2. Some hens did not complete the number of yoked practices within the session duration. These sessions were not used in the following analysis.

Group data for Figures 1.7.1-1.7.10 were analysed using a one-way repeated measures analysis of variance for all measures during the Training and Retention components for Experiment 1.7. The alpha level for all statistical comparisons in all situations was set at .05 and any results that reached this level were presented with an asterisk (*) in Table 1.7.3 and 1.7.4. Except where indicated with a hashtag (#) in Table 1.7.3 and 1.7.4, Mauchley's Test was not significant so sphericity was assumed. Post-hoc tests were conducted using the Bonferroni correction, as recommended by Fields (2005).

Figure 1.7.1 shows mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean percent correct data calculated across subjects is shown in the bottom left graph. Generally, accuracy in the Training component was similar during the No-delay and Between-chains delay conditions, and lower under the Within-chains delay condition, for all subjects. Table 1.7.3 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention accuracy was largest during the Between-chains delay condition, and generally similar during the No-delay and Between-chains delay conditions. The differences between conditions were significantly different (Table 1.7.3) and effect size, partial eta squared, was moderate (Ferguson, 2009).

Figure 1.7.2 shows the mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean response rate data calculated across subjects is shown in the bottom left graph. Training response rates were greatest during the No-delay condition, and similar during the Within- and Between-chains delay conditions, for all subjects. Table 1.7.3 shows these differences were significant and effect size, partial eta squared, was large

(Ferguson, 2009). Retention response rates were generally similar across the three experimental conditions for all subjects; showing no systematic effect from training conditions.

Figure 1.7.3 shows the mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean response latency data calculated across subjects is shown in the bottom left graph. Response latency in the Training component was greatest during the Within-chains delay condition, and generally similar during the No-delay and Between-chains delay conditions, for all subjects. Table 1.7.3 shows this finding was significant. Retention latencies were similar across all experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.7.3).

Figure 1.7.4 shows the mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition for all subjects during Series 1. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph. Reinforcement rates were similar for the No-delay and Between-chains delay conditions, and lowest during the Within-chains delay condition. Table 1.7.3 shows significant differences across conditions.

Figure 1.7.5 shows the mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component for all subjects in Series 1. The mean number of practices data calculated across subjects is shown in the bottom left graph. Number of practices during the No-delay, Within- and Between-chains delay conditions for the Training and Retention were similar across conditions for all hens. Table 1.7.3 shows no significant differences across conditions.

Figure 1.7.6 shows mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean percent correct data calculated across subjects is shown in the bottom left graph. Generally, accuracy in the Training component was similar during the No-delay and Between-chains delay conditions, and lower under the Within-chains delay condition, for all subjects. Table 1.7.4 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention accuracy was similar across the three experimental conditions for all subjects; the data from these conditions were not significantly different (Table 1.7.4).

Figure 1.7.7 shows the mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean response rate data calculated across subjects is shown in the bottom left graph. Training response rates were largest during the No-delay condition, and generally similar during the Within- and Between-chains delay conditions, for all subjects. Table 1.7.4 shows these differences were significant and effect size, partial eta squared, was large (Ferguson, 2009). Retention response rates were generally greatest during the Between-chains delay condition, and similar during the No-delay and Within-chains delay condition. The data from these conditions were significantly different (Table 1.7.4) and effect size, partial eta squared, was moderate (Ferguson, 2009).

Figure 1.7.8 shows the mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean response latency data calculated across subjects is shown in the bottom left graph. Response latency in the Training component was greatest during the Within-chains delay condition, and generally similar during the No-delay and Between-chains delay conditions. Retention latency was lowest under the Between-chains delay condition, and generally similar during the No-delay and Within-chains delay conditions. Table 1.7.4 shows data across conditions for both the Training and Retention component were significant.

Figure 1.7.9 shows the mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition for all subjects during Series 2. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph. Reinforcement rate was variable across each experimental condition for all subjects. Table 1.7.4 shows significant differences across conditions.

Figure 1.7.10 shows the mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component for all subjects in Series 2. The mean number of practices data calculated across subjects is shown in the bottom left graph. The number of practices to achieve the performance standard was similar across each experimental condition for all subjects. Table 1.7.4 shows no significant differences across conditions.

Table 1.7.1.

Completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens during Series 1.

Condition	Chain	Hen 21	Hen 22	Hen 23	Hen 24	Hen 25	Hen 26
A	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨	▨	▨	▨
B	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨			
C	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨	▨	▨	▨

Table 1.7.2.

Completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens during Series 2.

Condition	Chain	Hen 21	Hen 22	Hen 23	Hen 24	Hen 25	Hen 26
A	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨		▨	▨
B	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨		▨	
C	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨		▨	▨

Table 1.7.3

Analysis of variance results during the Training and Retention components of Series 1 for Experiment 1.7.

Component	MS Treatment	MS Error	df	F	p	Partial Eta Squared
Mean percent correct						
Training	966.04	30.50	2, 10	31.68	p<.0001*	.86
Retention	74.59	16.37	2, 10	4.56	.04*	.48
Mean response rate						
Training	.47	.003	2, 10	139.32	p<.0001*#	.97
Retention	.004	.004	2, 10	0.99	.41	.17
Mean response latency						
Training	7.61	.108	2, 10	70.30	p<.0001*	.93
Retention	.10	.07	2, 10	1.41	.29	.22
Reinforcement rate						
Session	2.1E-5	2.8E-6	2, 10	7.39	.01*	.60
Number of practices						
Correct	166.12	51.48	2, 10	3.22	.13#	.39

Table 1.7.4

Analysis of variance results during the Training and Retention components of Series 2 for Experiment 1.7.

Component	MS Treatment	MS Error	df	F	p	Partial Eta Squared
Mean percent correct						
Training	1079.35	16.72	2, 10	64.54	p<.0001*	.93
Retention	163.77	65.22	2, 10	2.50	.13	.33
Mean response rate						
Training	.44	.005	2, 10	97.14	p<.0001*#	.95
Retention	.02	.002	2, 10	8.44	.01*	.63
Mean response latency						
Training	46.98	4.57	2, 10	10.28	.02*#	.67
Retention	181.23	36.69	2, 10	4.94	.03*	.50
Reinforcement rate						
Session	1E-5	6.5E-7	2, 10	15.87	.001*	.76
Number of practices						
Correct	87.11	87.11	2, 10	1.00	.36#	.17

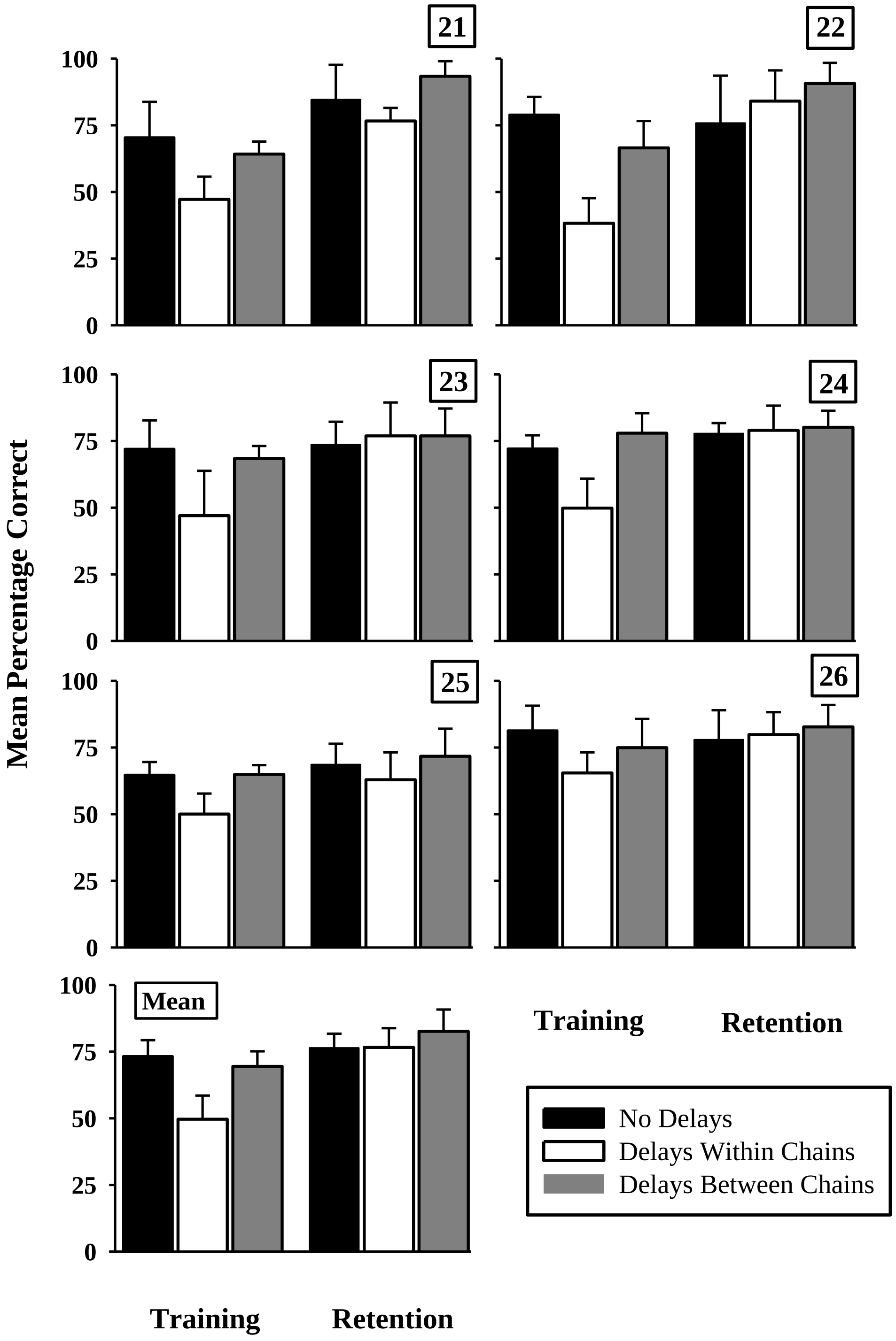


Figure 1.7.1. The mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean percent correct data calculated across subjects is shown in the bottom left graph.

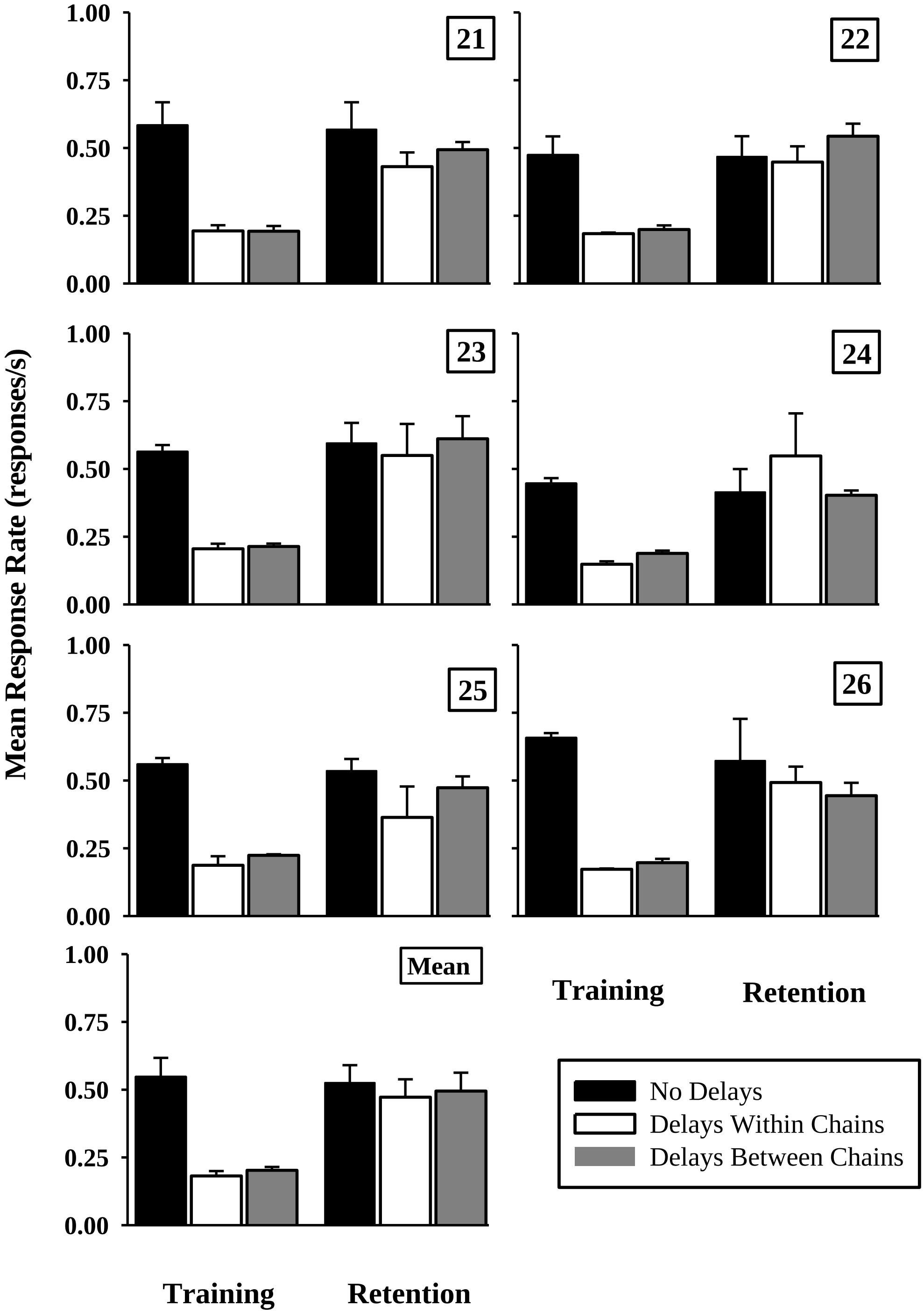


Figure 1.7.2. The mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean response rate data calculated across subjects is shown in the bottom left graph.

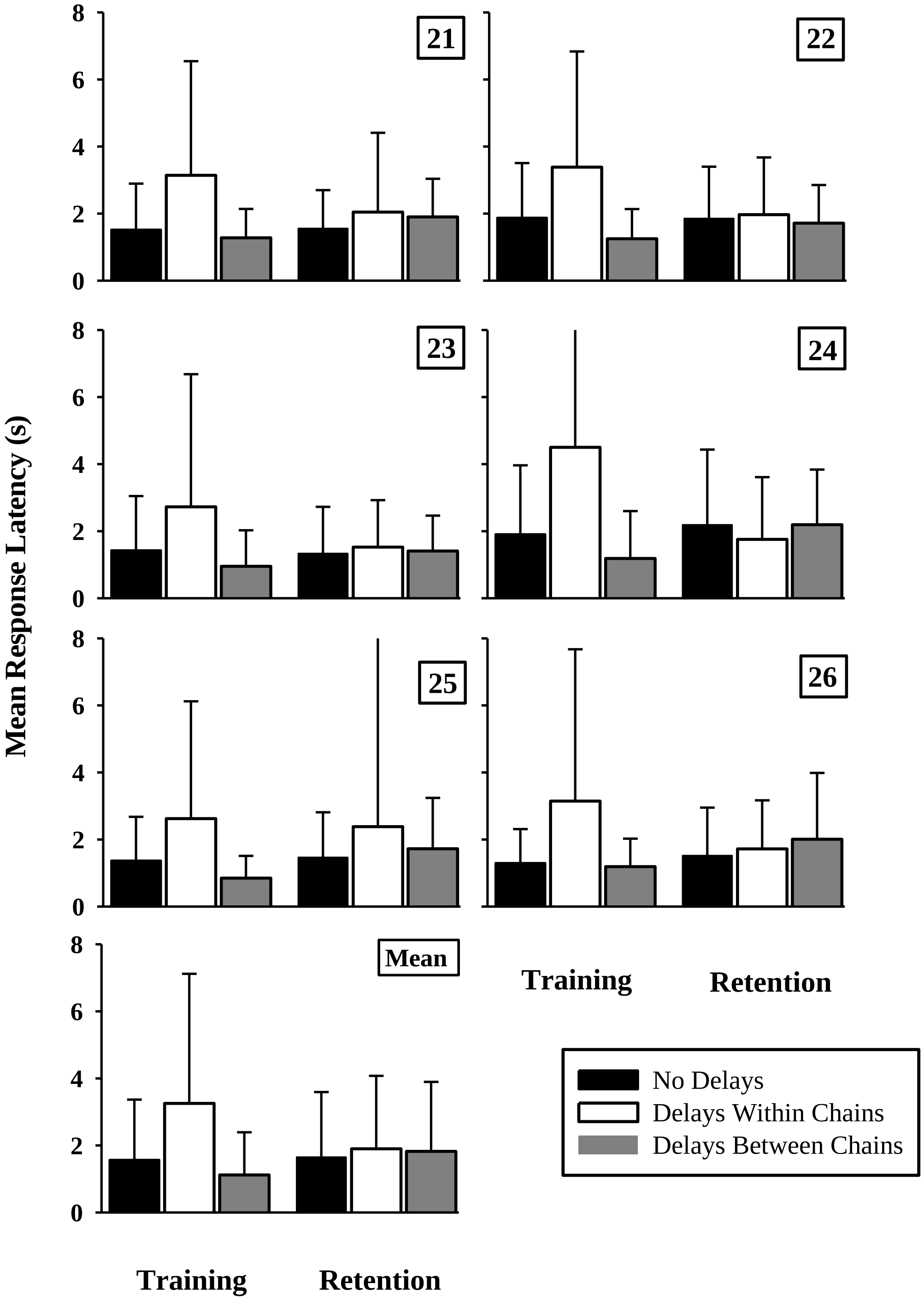


Figure 1.7.3. The mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 1. The mean response latency data calculated across subjects is shown in the bottom left graph.

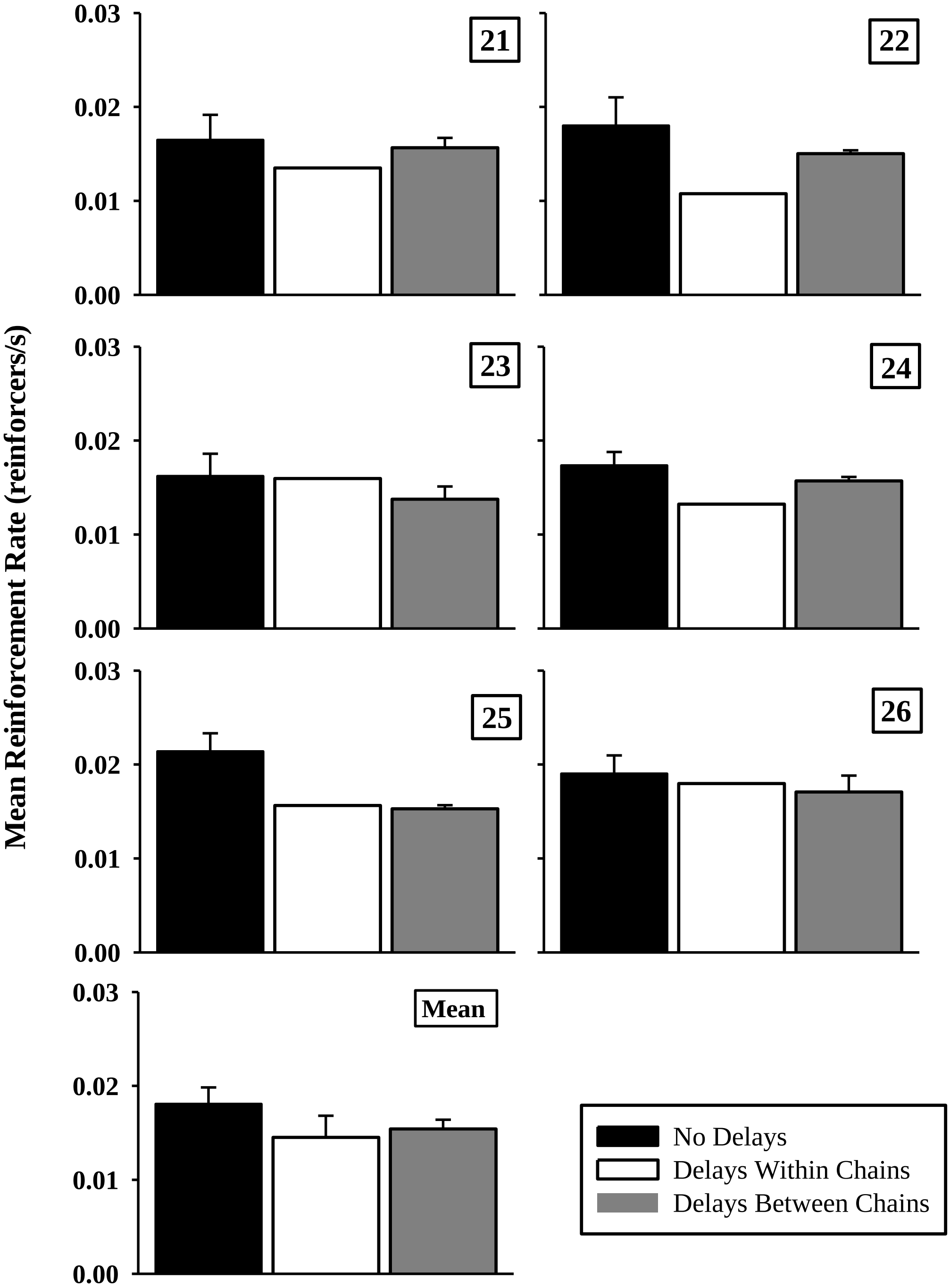


Figure 1.7.4. The mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition for all subjects during Series 1. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

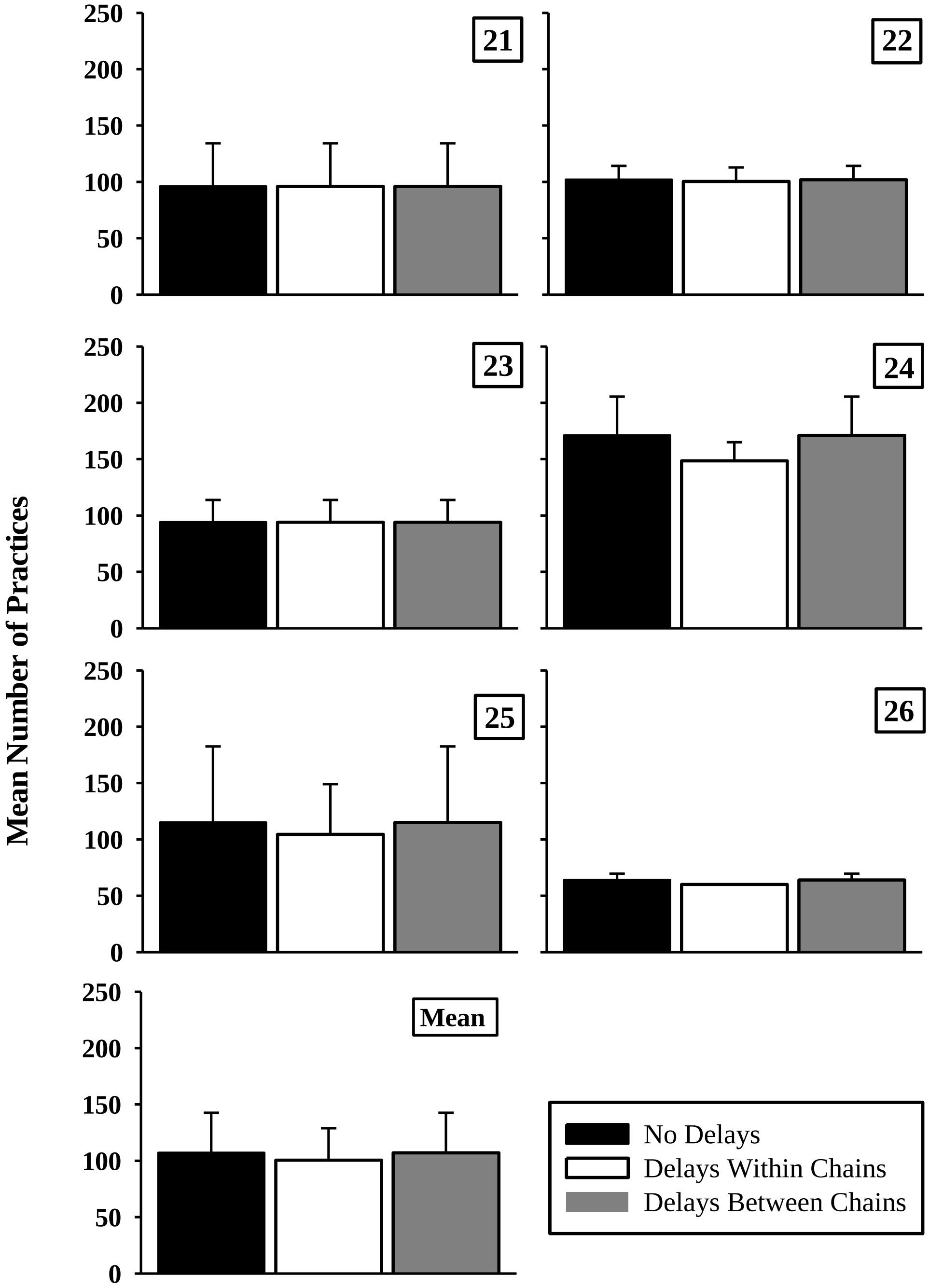


Figure 1.7.5. The mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component for all subjects in Series 1. The mean number of practices data calculated across subjects is shown in the bottom left graph.

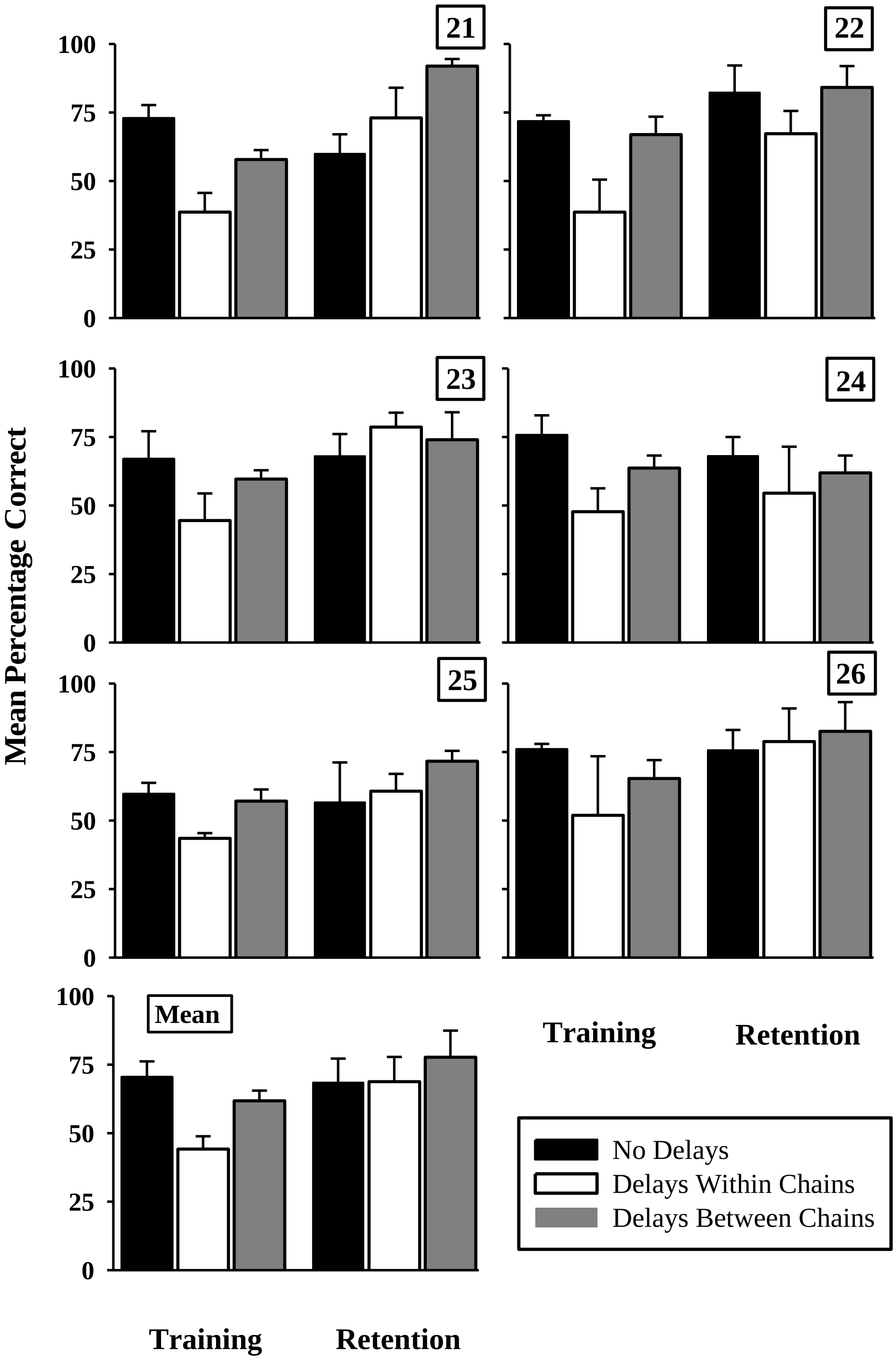


Figure 1.7.6. The mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean percent correct data calculated across subjects is shown in the bottom left graph.

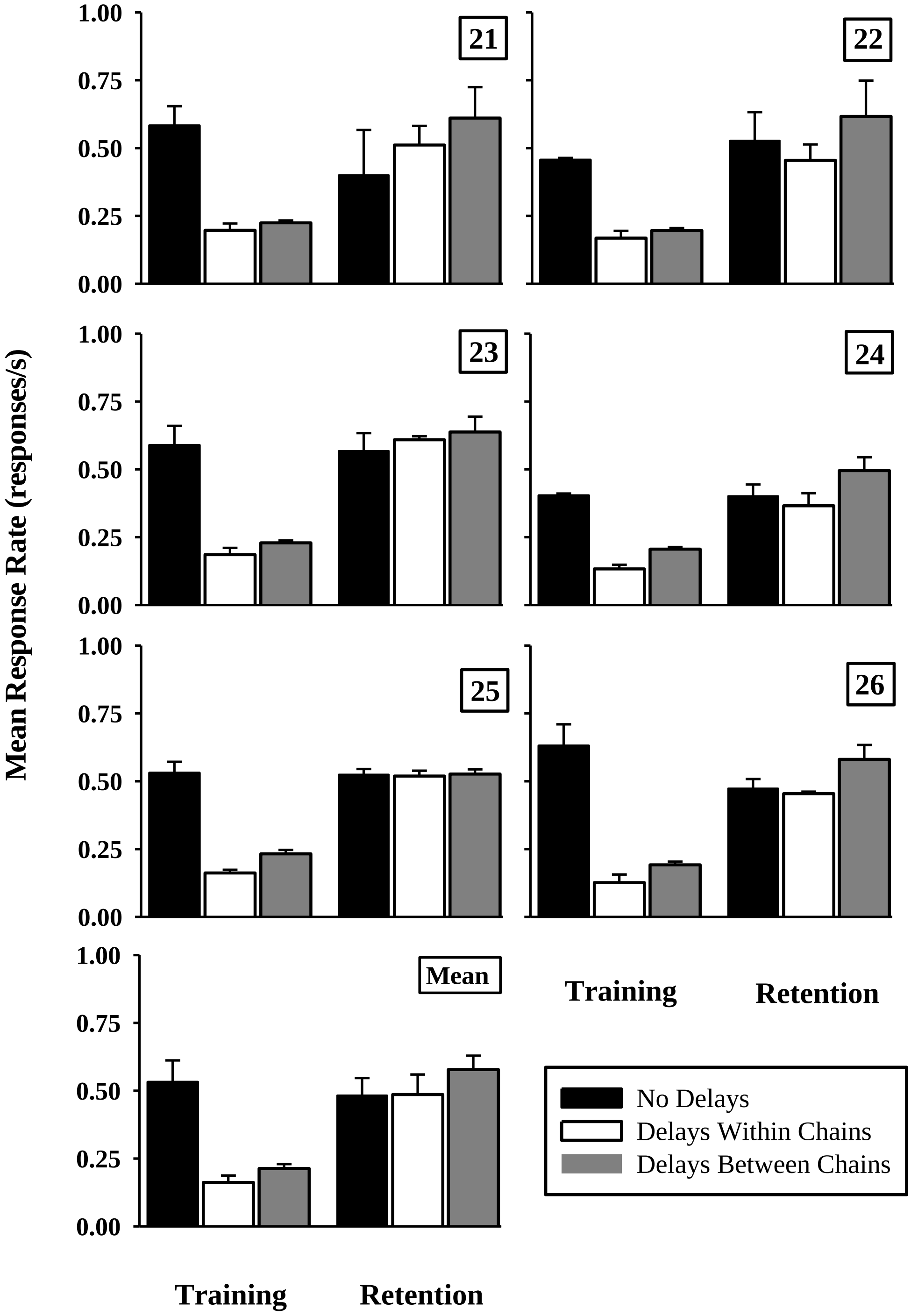


Figure 1.7.7. The mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean response rate data calculated across subjects is shown in the bottom left graph.

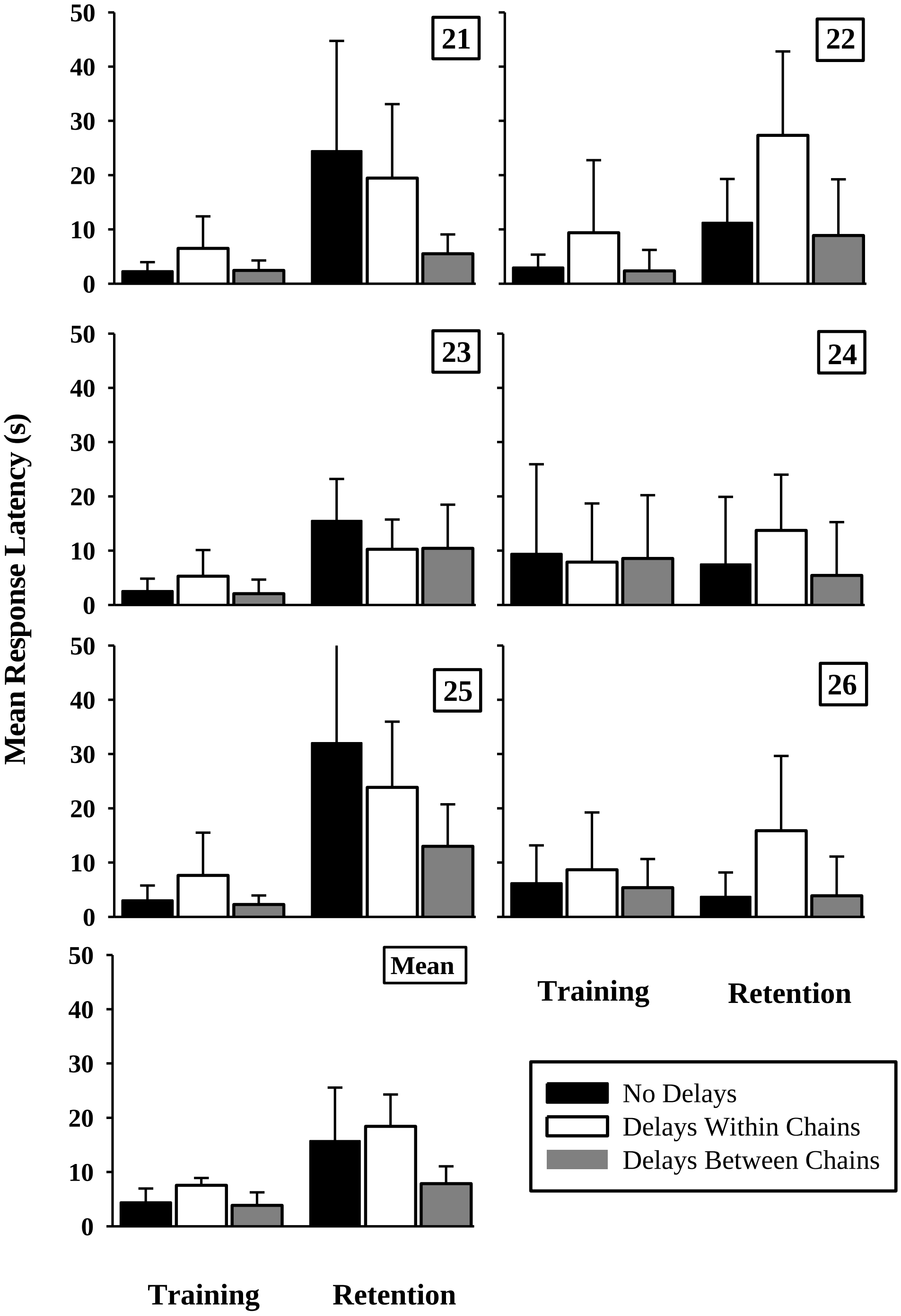


Figure 1.7.8. The mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects during Series 2. The mean response latency data calculated across subjects is shown in the bottom left graph.

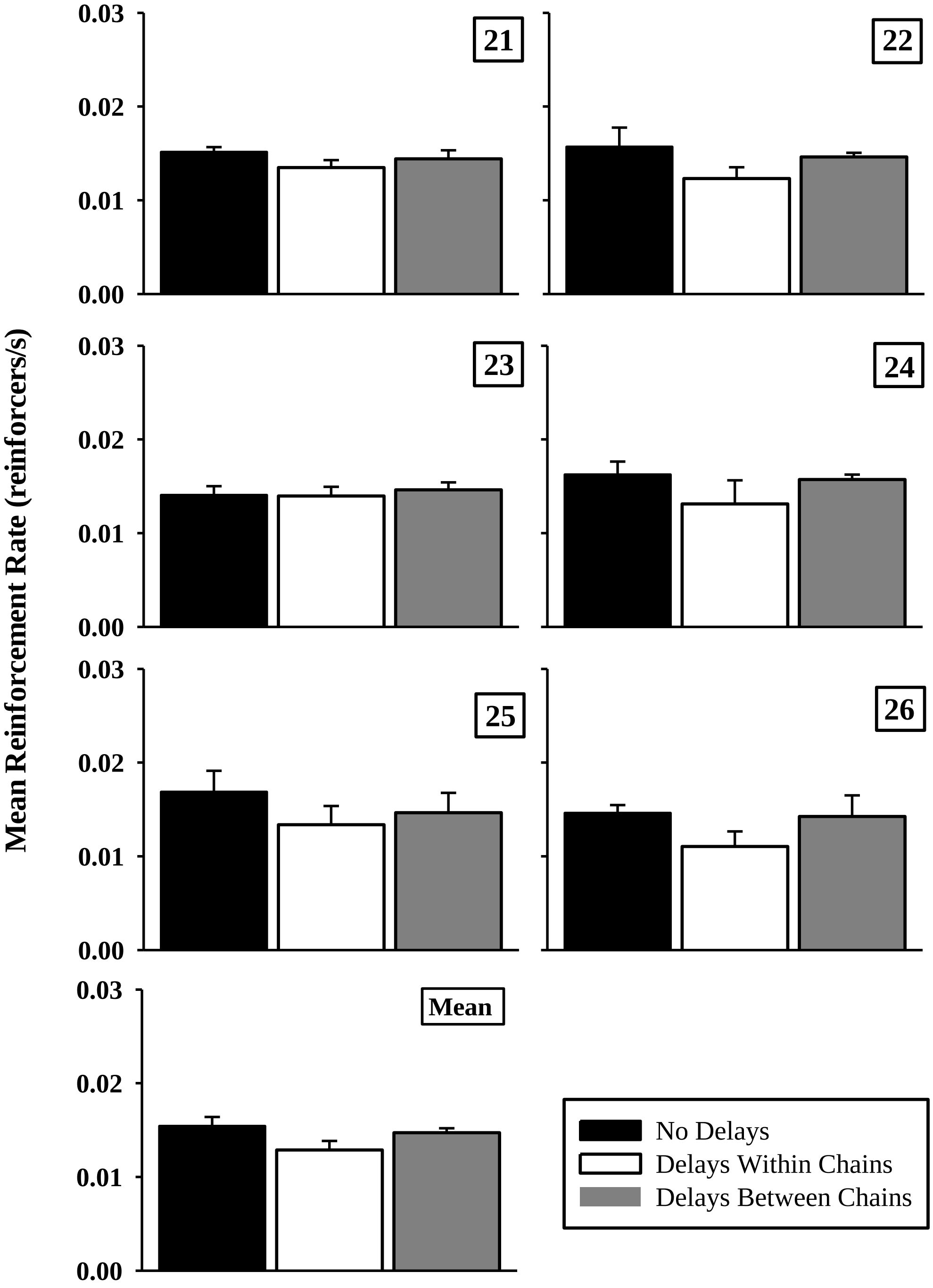


Figure 1.7.9. The mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition for all subjects during Series 2. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

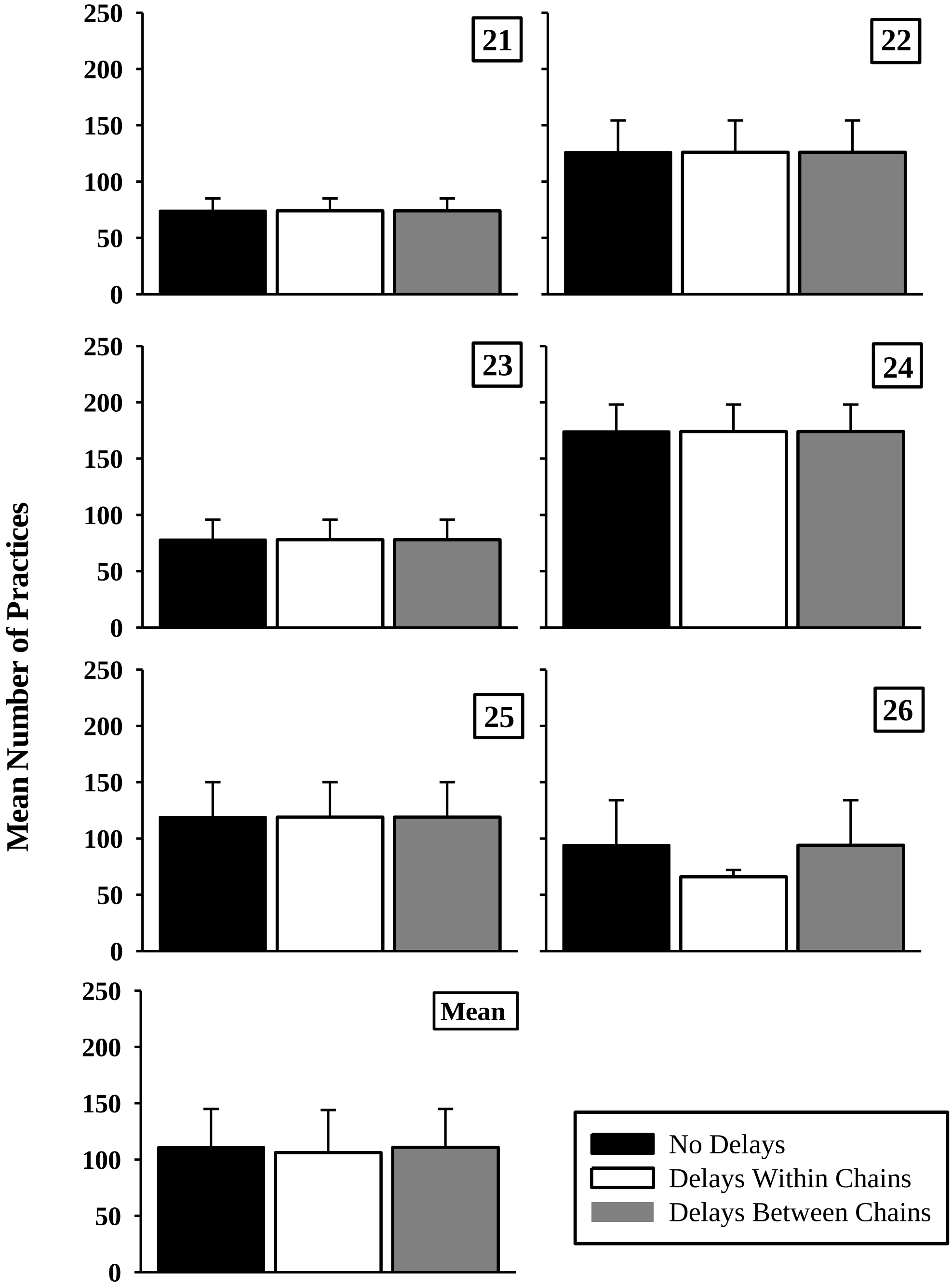


Figure 1.7.10. The mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component for all subjects in Series 2. The mean number of practices data calculated across subjects is shown in the bottom left graph.

Discussion

The first series of Experiment 1.7 examined the effects of decreasing the retention interval on accuracy. As in Experiment 1.2, percentage correct and response rate were greatest during the No-delay condition. In addition, response latency was greatest during the Within-chains delay condition. There were no systematic differences in number of practices across conditions. However, mean reinforcement rates were significantly different between the No-delay and Between-chains delay conditions (.003). While this difference was statistically significant, it is similar to the difference in means for the No-delay and Within-chains delay conditions (.004) and the Within- and Between-chains delay conditions (M < .0001).

It was expected that retention accuracy would improve by shortening the duration between training and testing. Although response rates and latency findings during the Retention component were similar across experimental conditions, retention accuracy did slightly improve during the Between-chains delay condition. This finding suggests the conditions during the Training component had some effect on response accuracy during the Retention component. However, the effect size for this finding was moderate; suggesting that decreasing the retention interval had little effect on the Retention component, a similar finding to Experiment 1.2.

The present study shortened the retention interval to 10 min, whereas the longest duration studied in much of the memory literature was 60 s (Roberts & Grant, 1976). It is possible that stimulus control established during the Training component was absent during the Retention component due to the long retention interval (White, 1985, 2001). It may be that an even shorter retention interval is needed to find any reliable recall.

The second series of Experiment 1.7 examined the effects of distributing practices at the shorter retention interval. As in Experiment 1.6, percentage correct and response rate were greatest during the No-delay condition. In addition, response latency was greatest during the Within-chains delay condition. There were no systematic differences in number of practices across conditions. However, mean reinforcement rates were significantly between the No-delay and Within-chains delay conditions (.003) and the Within- and Between-chains delay conditions (.002). While these differences were statistically significant, they are similar to the difference in mean between the No-delay and Between-chains delay conditions (M < .0001).

It was expected that distributing practices at the shorter retention interval would increase retention accuracy. Latency and response rates during the Between-chains delay condition significantly differed from other conditions during the Retention component. This finding suggests conditions during the Training component had some effect on responding during the Retention component. However, retention accuracy was similar across experimental conditions. This finding suggests that distributing practices at the shorter retention interval did not enhance retention accuracy, a finding similar to Experiment 1.6. The Retention component results of the second series of Experiment 1.7, along with Experiment 1.6, are contrary to studies showing that distributing practices enhances retention accuracy (Baddeley & Longman, 1978; Bloom & Shuell, 1981). One difference between the present study and those from the distributed practice literature, however, is that reinforcement rate and number of practices was controlled in the present investigation. This difference could account for the difference in findings.

Findings from Experiments 1.1-1.7 show the No-delay, Within- and Between-chains delay conditions during the Training component each produced different percentage of corrects, response rates, and latencies. This finding was contrary to the results from the Retention component, in which percentage of corrects, response rates, and latencies were generally similar across all three experimental conditions. It can be argued the reason for this finding was due to the differences in experimental arrangements for the Training and Retention components.

During the Training component, three different behaviours were learned under three experimental conditions. Responding without delays, responding with delays imposed after each correct response, and responding with delays imposed following three correct responses were learned under the No-delay, Within- and Between-chains delay conditions, respectively. Under this arrangement, three different behaviours were learned. For example, the behaviours of pecking without delays, peck-wait-peck-wait-peck, and peck-peck-peck-wait were learned under the No-delay, Within- and Between-chains delay, respectively.

During the Retention component different conditions were arranged, thus different behaviours were tested. Delays were not imposed during the No-delay, Within- and Between-chains delay conditions. With this type of arrangement, the behaviours tested across each experimental condition of the Retention component were peck-peck-peck. Thus, the reason the outcomes for each experimental condition of the Retention component during Experiments 1.1-1.7 were similar may be because similar behaviours were being tested.

Arranging the Retention component conditions similarly to the ones in the Training component, thus training and testing similar behaviours, may produce outcomes similar to Porritt (2007). To accomplish this, similar conditions were arranged between the Training and Retention components of Experiment 1.8.

EXPERIMENT 1.8

Experiment 1.8 examined whether the behaviour learned under the three experimental conditions of the Training component would be retained under similar conditions during the Retention component.

Method

Subjects

The same 6 subjects from Experiments 1.1-1.7 participated in this study.

Apparatus

The apparatus was the same as that used in Experiments 1.1-1.7

Procedures

Procedures were identical to those of Experiment 1.7 (Series 2) with one exception. A 10-s delay was imposed following each correct response during the Within-chains delay condition of the Retention component. In addition, a 15-s delay was imposed followed every third correct response during the Between-chains delay condition of the Retention component. Delays were not imposed following correct responding during the No-delay condition of the Retention component [please see p. 16-19 for a full description of training procedures].

Results

Table 1.8.1 shows the completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens. Some hens did not complete the number of yoked practices within the session duration. These sessions were not used in the following analysis.

Group data for Figures 1.8.1-1.8.5 were analysed using a one-way repeated measures analysis of variance for all measures during the Training and Retention components for Experiment 1.8. The alpha level for all statistical comparisons in all situations was set at .05 and any results that reached this level were presented with an asterisk (*) in Table 1.8.2. Except where indicated with a hashtag (#) in Table 1.8.2, Mauchley's Test was not significant so sphericity was assumed. Post-hoc tests were conducted using the Bonferroni correction, as recommended by Fields (2005).

Figure 1.8.1 shows mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean percent correct data calculated across subjects is shown in the bottom left graph. Generally, accuracy in the Training component was similar during the No-delay and Between-chains delay conditions, and lower during the Within-chains delay condition, for all subjects. Retention accuracy was generally similar during the No-delay and Between-chains delay conditions, and lower during the Within-chains delay, for all subjects. Table 1.8.2 shows the Training and Retention component findings were significant and effect size, partial eta squared, was large for both components (Ferguson, 2009).

Figure 1.8.2 shows the mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response rate data calculated across subjects is shown in the bottom left graph. Response rates were greatest under the No-delay condition then decreased during the Within- and Between-chains delay conditions during the Training and Retention components, showing systematic differences from experimental conditions. Table 1.8.2 shows the Training and Retention component findings were significant and effect size, partial eta squared, was large for both components (Ferguson, 2009).

Figure 1.8.3 shows the mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph. Response latency was greatest during the Within-chains delay condition during the Training and Retention components. Table 1.8.2 shows the Training and Retention component findings were significant and effect size, partial eta squared, was large for both components (Ferguson, 2009).

Figure 1.8.4 shows the mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph. Reinforcement rate was variable across experimental condition for all subjects. Table 1.8.2 shows significant differences across experimental conditions.

Figure 1.8.5 shows the mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component. The mean number of practices data calculated across subjects is shown in the bottom left graph. The number of practices to achieve the performance standard was similar across each experimental condition for all subjects. Table 1.8.2 shows no significant differences across experimental conditions.

Table 1.8.1.

Completed (filled boxes) and non-completed (open boxes) sessions for every chain of each experimental condition for all hens.

Condition	Chain	Hen 21	Hen 22	Hen 23	Hen 24	Hen 25	Hen 26
A	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨	▨	▨	▨
B	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨		▨	▨	▨	▨
	R-L-R	▨	▨	▨		▨	
C	L-C-L	▨	▨	▨	▨	▨	▨
	R-L-C	▨	▨	▨	▨	▨	▨
	R-L-R	▨	▨	▨	▨	▨	▨

Table 1.8.2

Analysis of variance results during the Training and Retention components of Experiment 1.8.

Component	MS Treatment	MS Error	df	F	p	Partial Eta Squared
			Mean percent correct			
Training	688.56	33.55	2, 10	20.52	p<.0001*	.80
Retention	963.14	52.75	2, 10	18.26	p<.0001*	.79
			Mean response rate			
Training	.634	.003	2, 10	229.99	p<.0001*#	.98
Retention	.29	.002	2, 10	139.40	p<.0001*	.97
			Mean response latency			
Training	139.04	10.81	2, 10	12.87	.02*#	.72
Retention	169.30	7.59	2, 10	22.31	.005*#	.82
			Reinforcement rate			
Session	1.55E-5	8.84E-7	2, 10	17.59	.001*	.78
			Number of practices			
Correct	1482.25	640.52	2, 10	2.3	.19#	.32

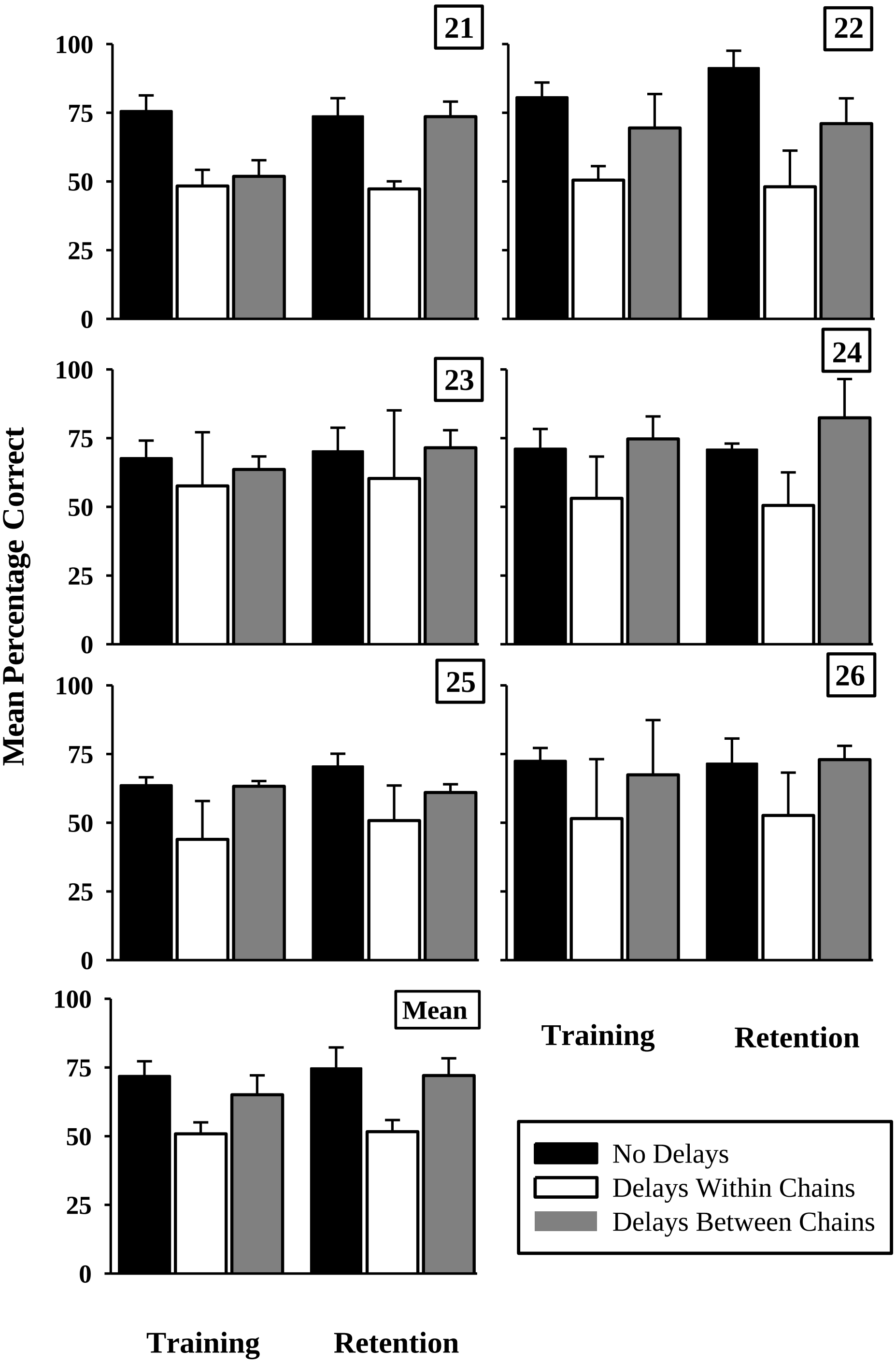


Figure 1.8.1. The mean percentage of correct responses (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean percent correct data calculated across subjects is shown in the bottom left graph.

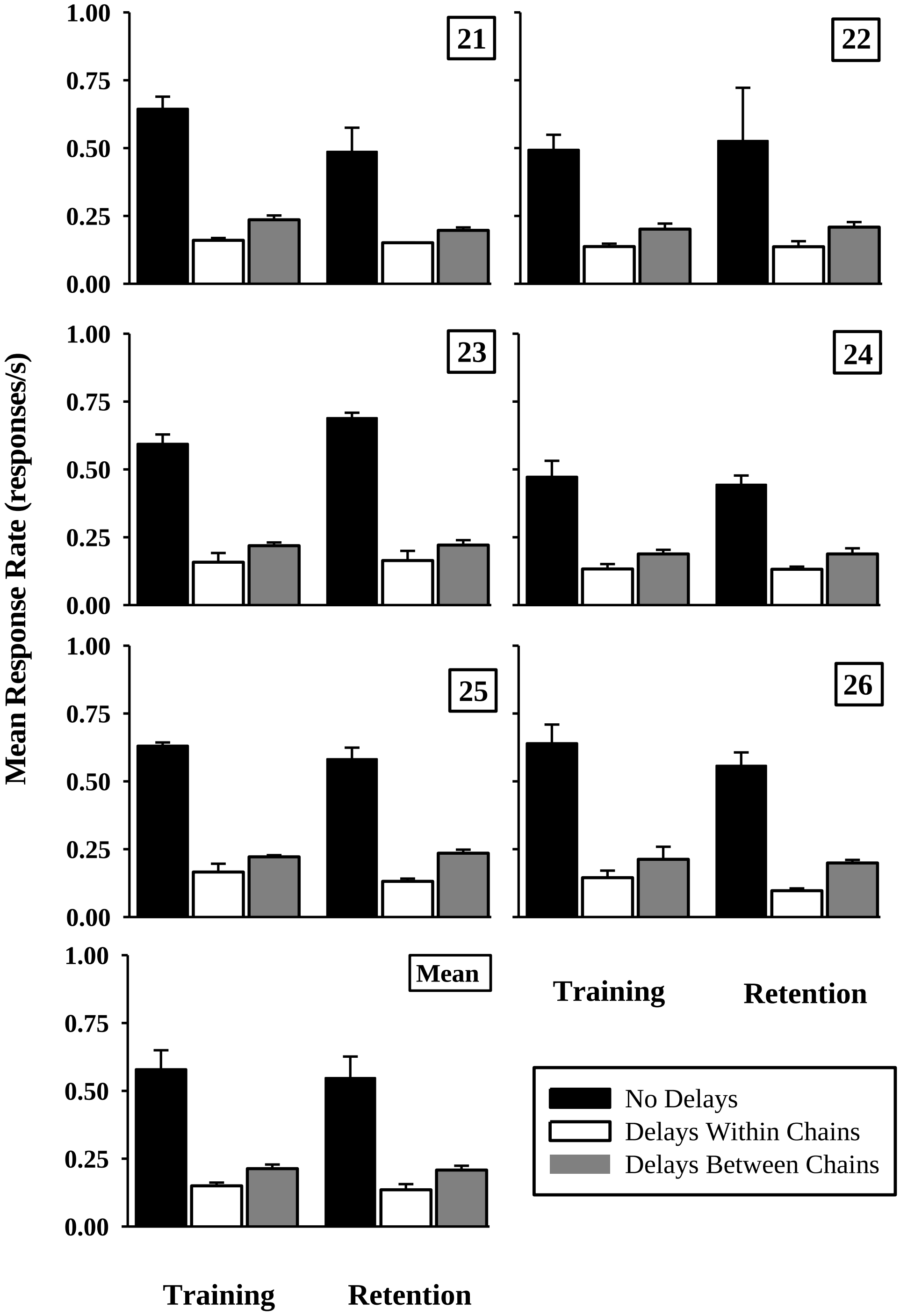


Figure 1.8.2. The mean response rates (+1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response rate data calculated across subjects is shown in the bottom left graph.

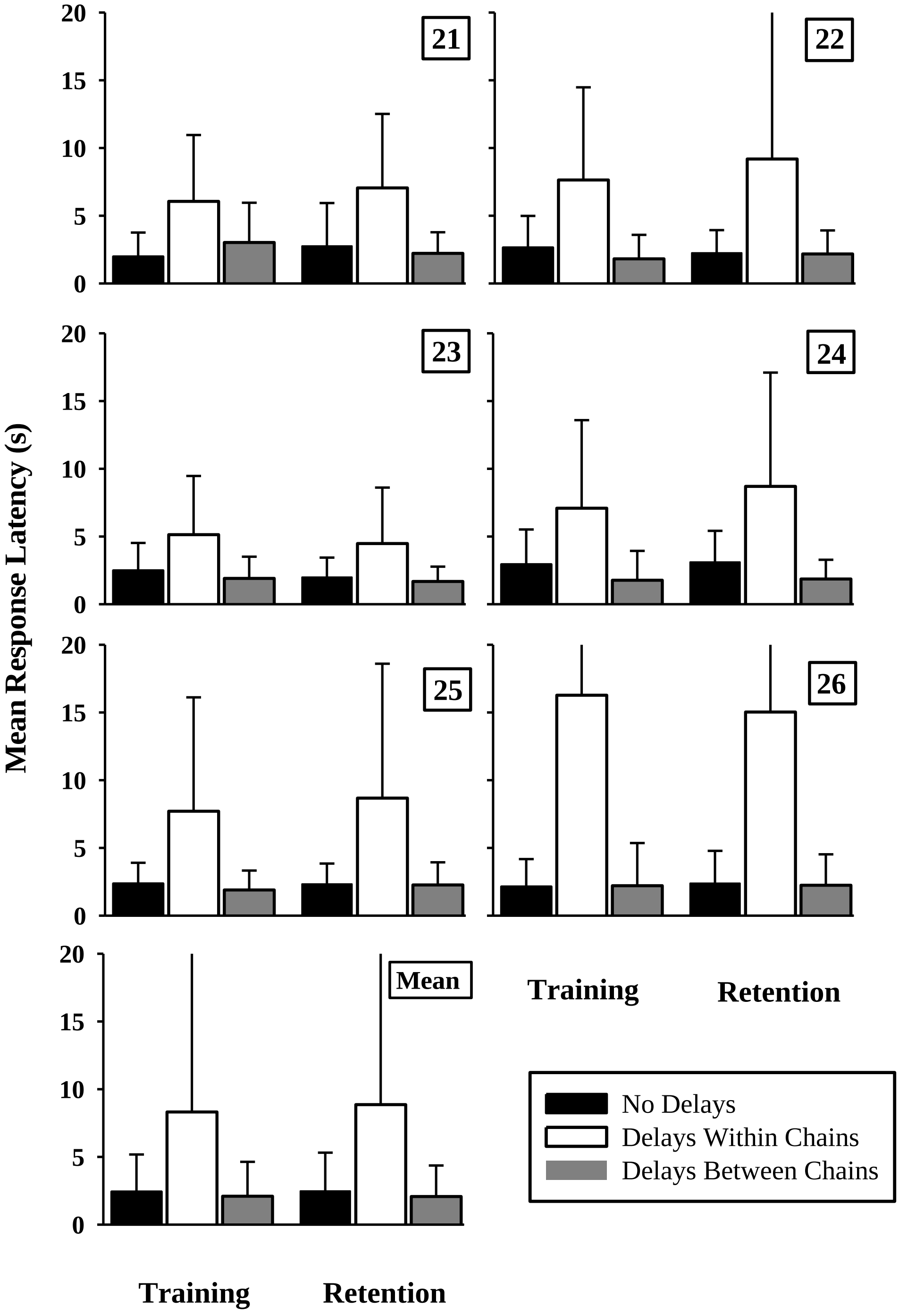


Figure 1.8.3. The mean response latency (with +1 *SD*) obtained from the three sessions of each experimental condition during the Training and Retention components for all subjects. The mean response latency data calculated across subjects is shown in the bottom left graph.

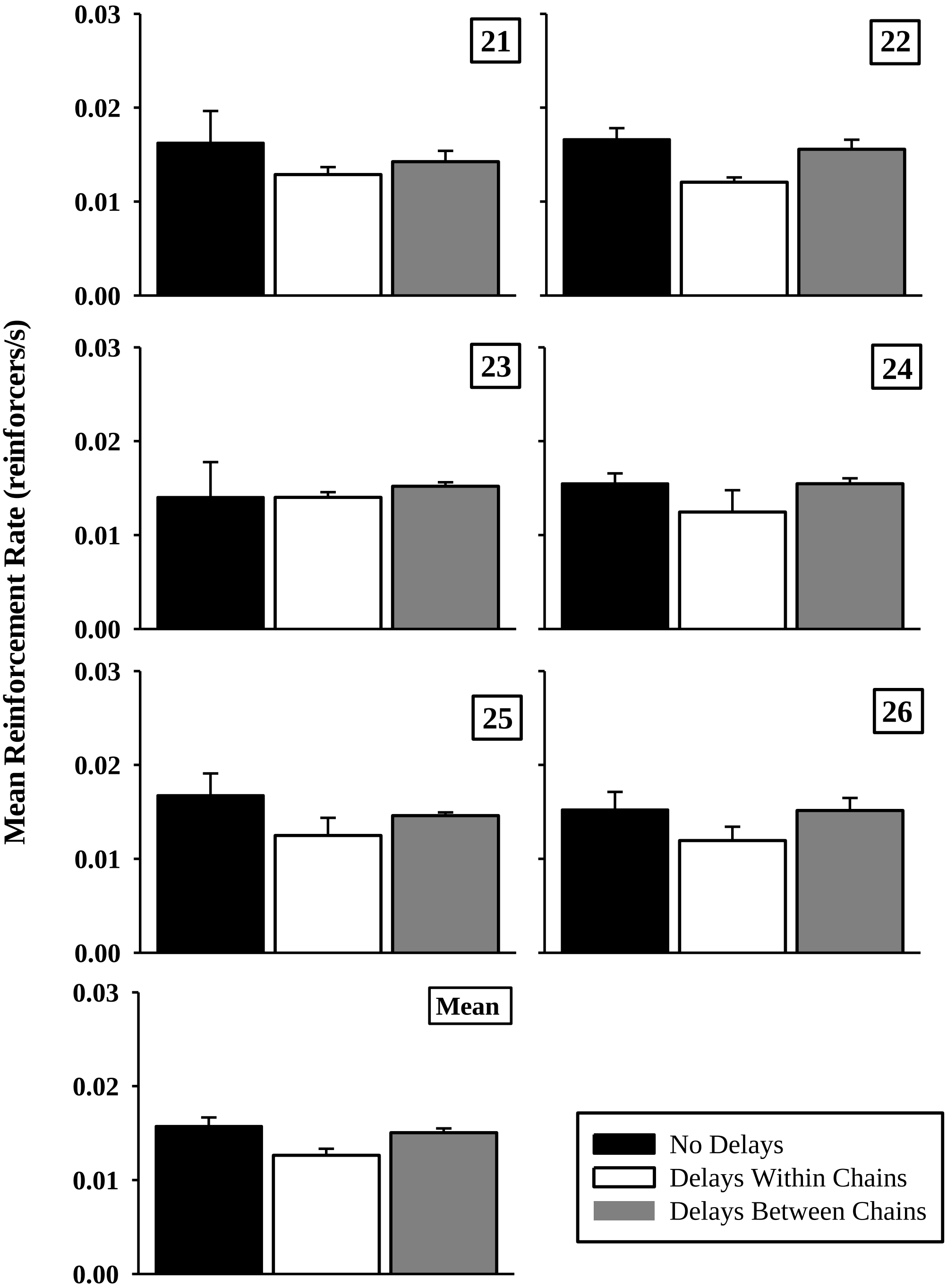


Figure 1.8.4. The mean reinforcement rate (+1 *SD*) obtained from the three sessions of each experimental condition. The mean reinforcement rate data calculated across subjects is shown in the bottom left graph.

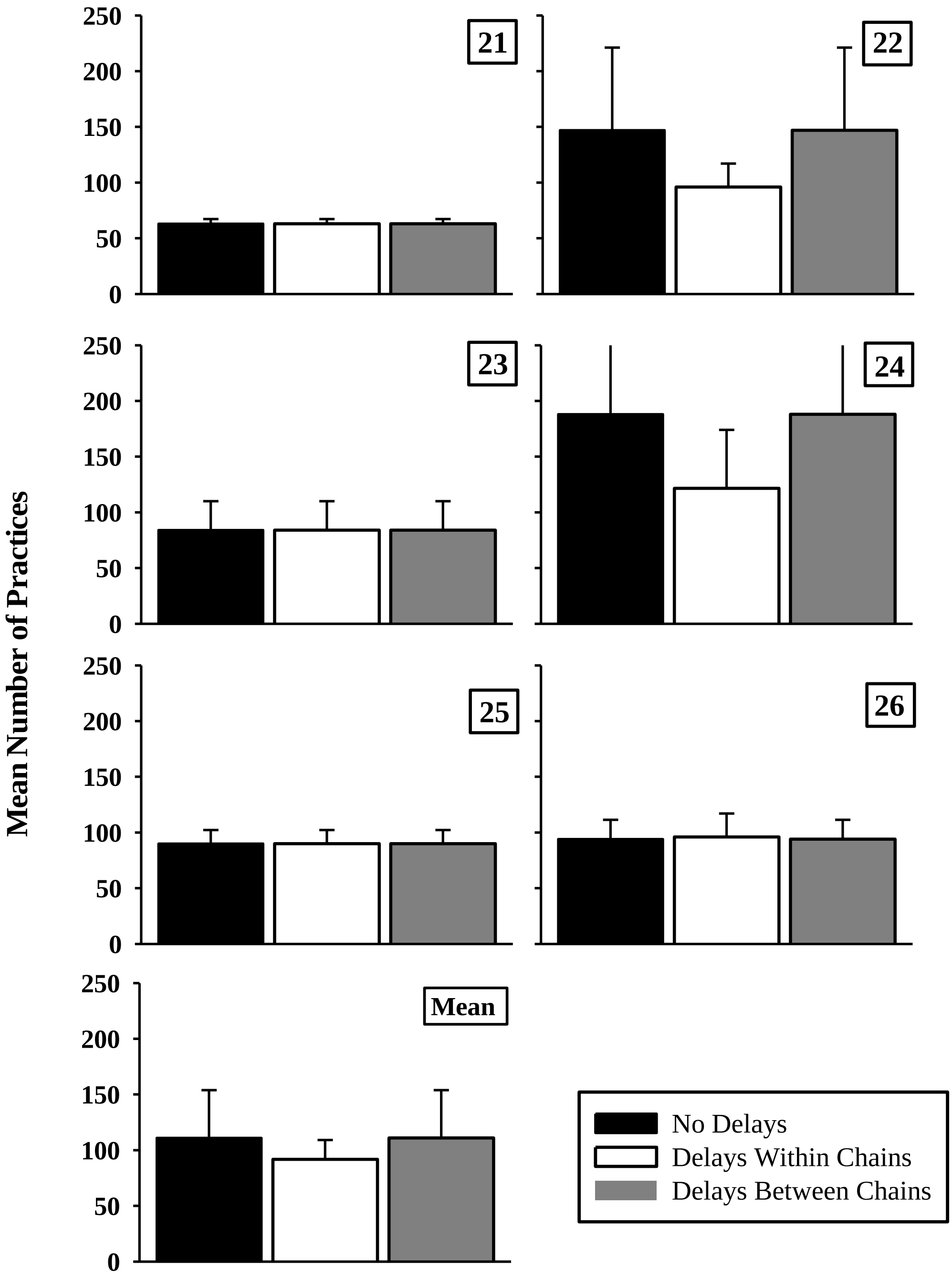


Figure 1.8.5. The mean number of practices (+1 *SD*) obtained from the three sessions of each experimental condition during the Training component. The mean number of practices data calculated across subjects is shown in the bottom left graph.

Discussion

Experiment 1.8 examined whether the behaviour learned under the three experimental conditions of the Training component would be retained under similar conditions during the Retention component. As in Experiment 1.2, percentage correct and response rate were greatest during the No-delay condition. In addition, response latency was greatest during the Within-chains delay condition. There were no systematic differences in number of practices across conditions. However, mean reinforcement rate was significantly different between the No-delay and Within-chains delay conditions (.003) and the Within- and Between-chains delay conditions (.002). While these differences were statistically significant, they are similar to the difference in mean between the No-delay and Between-chains delay conditions (M < .0001).

Contrary to Experiment 1.2, accuracy and response rates were greatest under the No-delay condition during the Retention component. In addition, response latency was greatest during the Within-chains delay condition. These findings suggest that greater response rates improve retention accuracy when similar behaviours are trained and tested under like conditions.

Similar to Porritt (2007) and Porritt et al. (2009), the No-delay condition in the present experiment produced the greatest accuracy during the Training and Retention components. The results of the present investigation were obtained, however, using different procedures from those of Porritt and Porritt et al. Had Porritt and Porritt et al. trained similar behaviours under like conditions during the Training and Retention Components in their studies, they would arguably have come up with similar results to the present study.

SUMMARY

Experiments 1.5-1.8 used experimental arrangements that have been shown to effect retention accuracy in an attempt to replicate the Retention component findings from Porritt (2007).

The performance standard was increased across two series of conditions (e.g., seven and nine chains) during Experiment 1.5 to determine if hens require a greater performance standard than pigeons to replicate the Retention component results from Porritt (2007). The No-delay condition produced the greatest accuracy during the Training component, but there were no differences in accuracy between experimental conditions during the Retention component, partially replicating the findings from Porritt. This finding suggests that the hens used in the present experiment did not require a greater performance standard to replicate the Retention component findings from Porritt (2007).

Experiment 1.6 examined the effects of distributed practice on retention accuracy. Increasing the duration between practices did not increase retention accuracy. This finding suggests that distributed practice does not impact retention accuracy. This finding is contrary to studies showing that distributed practice increases retention accuracy (Baddeley & Longman, 1978; Bloom & Shuell, 1981). One difference between the present study and those from the distributed practice literature is that reinforcement rate and number of practices was controlled in the present investigation. This difference could account for the difference in findings; thus, no conclusions can be drawn on the effects of distributed practice in the present study as it relates to the literature. As mentioned [Experiment 1.6 Discussion], it is possible that stimulus control established during the Training component is absent during the Retention component due to the long retention interval (White, 1985, 2001).

Series 1 of Experiment 1.7 examined the effects of a shorter retention interval on retention accuracy under three different experimental conditions. Results showed that shortening the duration between training and testing components did not improve retention accuracy. Being that the longest duration of retention interval studied was 60s (Roberts & Grant, 1976), it is possible that an even shorter retention interval is needed to find any reliable recall. Thus, no conclusions can be drawn about the effects of shortening the retention interval in the present study as it related to the memory literature.

Series 2 of Experiment 1.7 examined the effects of distributed practice at the shorter retention interval. Similar to Experiment 1.6, the results from the second series of Experiment 1.7 showed that retention accuracy did not improve by distributing practices, even at the shorter retention interval. It was argued that the differences in the experimental conditions between the Training and Retention components in the present study accounted for the lack of replicating the Retention component findings of Porritt (2007).

Experiment 1.8 trained and tested similar behaviours by arranging the experimental conditions alike during the Training and Retention components. Results showed that the No-delay condition led towards enhanced accuracy during the Training and Retention components. These findings suggest that the behaviour learned under the three experimental conditions of the Training component were retained under similar conditions during the Retention component. Similar to Porritt (2007), the present results suggest that greater response rates lead toward greater training and retention accuracy. The present results, however, were obtained under different experimental arrangements than Porritt (2007). Thus, it is still unclear how Porritt (2007) obtained his results using his arrangements.

GENERAL DISCUSSION OF EXPERIMENT 1

Controlling procedural confounds in Precision Teaching studies is a necessary and important step for researchers; because, it helps identify the critical components that lead towards improved retention, endurance and application. Porritt (2007) and Porritt et al., (2009) have been the only studies to show a result in strong support for the rate-building procedure when reinforcement rate and number of practices are controlled. It was hypothesized that differences in methodology and experimental design may account for the results obtained by Porritt.

The aim of Experiments 1.1-1.4 was to replicate the procedures used by Porritt (2007) and Porritt et al., (2009). These experiments tested the following research questions:

1. Do greater response rates improve accuracy during training and retention when number of practices and reinforcement rate are held constant?
2. Do alternative methodology and experimental designs account for the present findings?

Both experimental questions were addressed across four studies and led towards similar findings. The data collected across the four studies showed that greater response rates improved training accuracy, but not retention accuracy. These data support Porritt's results and the findings of the Precision Teaching research that suggests rate-building improves training accuracy. In addition, the present studies show support for yoking correct practices, as opposed to trials, and using an A/B/C experimental design. Thus, future rate-building studies interested in controlling for number of practices across experimental conditions could yoke correct practices under an A/B/C design. This arrangement would generate similar outcomes to an A/B/A/B/A/C/A/C design, thus shortening the duration of a study.

As previously mentioned [Experiment 1.1 Introduction], future rate-building studies can focus on the effects of instructions (e.g., goal setting) used by Precision Teachers to generate response rates, as the two methods have both been shown to produce beneficial outcomes (Lock & Latham, 1990; Seijts & Latham, 2001). A clearer understanding of how goal setting affects responding during the rate-building procedure may shed light on the mixed outcomes within the Precision Teaching literature. Future research could focus on the effects of rate-building on other purported outcomes of fluency (e.g., endurance, stability, application) while controlling for practice and reinforcement rate.

Given the failure to show any strong effect of the training conditions on retention accuracy, Experiments 1.5-1.8 attempted to replicate the Retention component findings of Porritt by using experimental arrangements that have been shown to enhance retention accuracy. These experiments tested the following research questions:

1. Would greater performance standards replicate Porritt's Retention component findings?
2. Would distributing practices replicate Porritt's Retention component findings?
3. Would decreasing the duration between training and testing replicate Porritt's Retention component findings?
4. Would training and testing similar behaviours under similar conditions replicate Porritt's Retention component findings?

With respect to the first research question, the data did not show improved retention accuracy when increasing the performance standard. These data are contrary to research showing that increasing the performance standard also improves retention accuracy (Berens et al., 2003; Ivarie, 1986). The contrary findings, however, may be due to methodological differences. Number of practices and reinforcement rate were held constant in the present study. Studies showing that greater performance standards improve retention accuracy did not use these procedural controls. Thus, as previously stated [Summary #2], no conclusions can be drawn about the present data as it relates to the applied literature.

With respect to the second research question, the data did not show improved retention accuracy when practices were distributed during the training component, both at the 23-hr and 10min retention intervals. These data are contrary to research showing that distributing practices further apart improves retention accuracy (Baddeley & Longman, 1978; Bloom & Shuell, 1981). The contrary findings may be due to the duration of the practice distribution in the present study. In the distributed practice literature, practices are typically distributed across days (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). In the present investigation, practices were distributed by imposing a 10-s delay between each correct response. It is possible that longer delays between practices are required to observe an affect from distributed practices on retention accuracy. As mentioned [Experiment 1.7 Discussion], the difference in findings between the present study and distributed practice literature may be due to the procedural controls used in the present study. Thus, no conclusions can be drawn about how the present results compare to the distributed practice literature.

With respect to the third research question, the data did not show improved retention accuracy when the retention interval was shortened. These data are contrary to research showing that increasing the retention interval decreases retention accuracy (Roberts, 1972; White, 1985; White & Wixted, 1999). The contrary findings may be due to the duration of the delay interval used in the present study. The present study shortened the retention interval to 10 min, whereas the longest duration studied in much of the memory literature was 60 s (Roberts & Grant, 1976). It is possible that stimulus control established during the Training component was absent during the Retention component due to the long retention interval (White, 1985, 2001). It may be that an even shorter retention interval is needed to find any reliable recall.

With respect to the fourth research question, the rate-building condition showed the greatest retention accuracy when similar behaviours were trained and tested under similar conditions. Thus, the most important factor in retention accuracy, it seems, was the response rate during the Retention component. Comparing response rates during the Retention component in the present study with those of Porritt would help shed light on the contrary findings. Unfortunately, these response rate data were not reported by Porritt (2007) nor Porritt et al. (2009), preventing any conclusions to be drawn about the failure to replicate.

Effect sizes in the rate-building literature

The results from Porritt (2007) suggest that the rate-building procedure used during the training component increased retention accuracy when reinforcement rate and number of practices were held constant. The present study used similar procedures and produced contrary results; findings from both studies were statistically significant. However, according to Kirk (1996), statistical significance testing may provide a misleading representation of the effect from independent variables.

Statistical significance provides a measure of the likelihood that the change between conditions (or difference between groups) was not due to chance (Sterne & Smith, 2001). Kirk (1996) argues this measure can be misleading because the probability value of .05 is arbitrary, leading researchers towards different conclusions about similar treatment effects. For example, a one percentage difference in accuracy scores using a sample of 100 participants can become statistically significant if 100 participants were added, even though the treatment effects remained similar. Effect size provides another measure on the relationship between variables.

The various effect size measures provide estimates of the magnitude of the effect between two or more variables (Ferguson, 2009) and they are recommended for use in addition to statistical tests (Wilkinson & Task Force, 1999, p. 599). Although the present findings and those of Porritt (2007) were statistically significant, investigating the size of the effect seen may provide clarity on the different conclusions.

Partial eta squared was the effect size calculated in the present study. Partial eta squared is generally a good index of the relationship between experimental conditions (Ferguson, 2009). To determine effect size, Cohen (1992) suggests $r = .1$ as a cut-off for small effect sizes. However, Ferguson (2009) argues that Cohen did not anchor this cut-off across effect sizes, presenting unequal values (e.g., Cohen's r and d). As an alternative, Ferguson suggests effect sizes are anchored to a minimum of $r = .2$ as a measure of practical significance. Ferguson notes this is the suggested minimum, not a guarantee of practical significance and that it should not, as Cohen (1992) suggests, be applied rigidly. Ferguson's suggestion creates partial eta effect sizes of .04, .25, and .64 for minimum, moderate, and strong effect sizes, respectively. Ferguson's suggestions were not only used in the present study, but also used to determine effect size in the rate-building literature for the present discussion.

Effect sizes for the present data were calculated by dividing the sum of squares effect by the sum of the sum of squares effect and the sum of squares error. The effect sizes for the present data for percentage correct in the present study during the Training and Retention components were .69 and .26, respectively. The effect sizes for the Training and Retention components are strong and moderate, respectively (Ferguson, 2009). The effect size for percentage correct in Porritt's (2007) study, based on the data presented in his analysis of variance tables, during the Training and Retention components were .94 and .93, respectively. The experimental conditions showed a strong effect during the Training and Retention components (Ferguson). Although both studies showed strong effect sizes during the Training component, the effect size for the Retention component differed. This difference raises the question about effect sizes in related research.

As mentioned previously [Experiment 1.1 Introduction], results from studies that have controlled for reinforcement rate and number of practices have either not shown favourable results for using the rate-building procedure (Campbell, 2012; Cohen, 2008; Evans et al., 1983; Fox & Ghezzi, 2003; Wheetley, 2005), have been mixed (Shirley & Pennypacker, 1994) or have been shown to produce a moderate effect size (Holding et al., 2011). Results from these studies, along with the present findings and a recent review of the literature (Johnson & Street, 2014), questions the retention enhancing effects of rate-building when procedural controls are used. Some well-cited Precision Teaching studies not controlling for reinforcement rate or number of practices, however, report that rate-building improves retention accuracy (Berens et al., 2003; Bucklin et al., 2000; Hughes et al., 2007; Ivarie, 1986; Young et al., 1985). Taken the unfavourable outcomes from rate-building studies using procedural controls, closer inspection of the effect size from studies not controlling for reinforcement rate and number of practices is warranted.

Initial inspection of the data from rate-building studies not using procedural controls showed the effect size from only two could be calculated (Bucklin et al., 2000, Ivarie, 1986); data needed to calculate effect sizes in other studies was not presented (Berens et al., 2003; Hughes et al., 2007; Young et al., 1985).

As mentioned previously [Experiment 1.1 Discussion], Ivarie (1986) investigated the effects of two different performance standards on accuracy and retention of identifying Arabic numerals across three achievement category groups. Students were placed into achievement category groups of above mean, mean, and below mean, based upon a pre-test. Ivarie concluded the faster response rates led towards greater retention accuracy. Closer inspection of the data revealed this effect was not statistically significant for the mean and above mean groups, suggesting outcomes for using the rate-building procedure were favourable only for the below mean group. The effect size for the below mean group (.87) is considered to be strong (Ferguson, 2009), suggesting a significant and strong effect of the rate-building procedure on retention accuracy with participants in the below mean group.

Bucklin et al. (2000) compared the effects of rate-building and accuracy training on the retention and application of Hebrew symbols, nonsense symbols, and Arabic numerals using a stimulus equivalence task. Following initial training, participants either continued at 100% correct or were required to meet a performance standard. Tests for retention and application were administered four weeks post training. Retention tests were then administered every two weeks to half of each group, while the other half of each group received retention tests every four weeks. The final retention test was administered sixteen weeks post training. Application tests revealed no significant differences between the two groups. Bucklin et al. tested for percentage correct and corrects per minute during the retention tests. At four weeks post training, the percentage correct difference between groups was 2.5 and is considered to be strong (Ferguson, 2009). Percent correct was also significantly different between groups during follow-up retention tests. At four weeks post training, number of corrects per minute was not significant between the two groups. Group differences in fluency scores revealed a minimum effect (.58). Corrects per minute were also not significantly different between the two groups during the follow-up retention tests. The lack of differences in corrects per minute is particularly important because greater response rates, generated by the rate-building procedure, are argued to maintain at similar rates before and after periods of no practice (Doughty et al., 2004); this effect was not found in the data presented by Bucklin et al. Closer inspection of studies in which data needed to calculate effect sizes was not presented revealed mixed outcomes and procedural confounds (Berens et al., 2003; Hughes et al., 2007; Young et al., 1985).

Young et al. (1985) used a within-subject design to compare the outcomes of teaching two behaviours at two different performance standards using the rate-building procedure. Young et al. suggested greater performance standards maintained at similar rates following a four month retention interval, showing favourable outcomes for generating greater response rates using the rate-building procedure. However, closer inspection of the data revealed the lower performance standard generated similar response rates before and after the retention interval, suggesting both performance standards ensured retention. Mixed outcomes were also found in the data presented by Berens et al. (2003).

Berens et al. (2003) investigated the effects of rate-building on saying computation flashcards. Participants were encouraged to reach a performance standard and were then tested for retention, endurance, and application. The authors concluded that greater performance standards improved retention accuracy, however, closer inspection of the data revealed this finding rarely occurred. For example, there were several occasions when a greater performance standard did not enhance retention accuracy or if it did, the change in retention accuracy was minimal. This outcome, along with the findings from Ivarie (1986), Bucklin et al. (2000), and Young et al. (1985) suggest the effects of rate-building on retention accuracy are mixed.

As mentioned previously [Experiment 1.1 Introduction], Hughes et al. (2007) compared the effects of a rate-building procedure (RBP) and a teaching as usual (TAU) condition on reading retention accuracy. The authors concluded the group receiving rate-building showed improved accuracy during retention tests. Closer inspection of procedures revealed that children in the rate-building condition were exposed to an error correction procedure, an additional variable not used during the TAU conditions. This finding is important because the favourable outcomes for the rate-building group may have also been a product of the error correction procedure. Thus, it is possible that error correction confounded the rate-building procedure for the RBP group.

A closer inspection of the results from Ivarie (1986), Young et al. (1985), Bucklin et al. (2000), Berens (2003), and Hughes et al. (2007) revealed mixed outcomes and problems with internal validity. Young's data showed outcomes were similar during both performance standards, while other studies showed the rate-building procedure failed to produce similar response rates before and after the retention interval (Berens et al., 2003; Bucklin et al., 2000; Ivarie, 1986). In one study, it was unclear whether the rate-building or other differences between conditions produce the stated outcomes (Hughes et al., 2007), generating problems with internal validity (e.g., history).

A number of Precision Teaching studies have stated benefit from using the rate-building procedure without incorporating procedural controls (Berens et al., 2003; Bucklin et al., 2000; Hughes et al., 2007; Ivarie, 1986; Young et al., 1985). A closer inspection of these studies revealed mixed outcomes from using the rate-building procedure and, in one case, procedural confounds. Data needed to calculate effect sizes was not presented in most of these studies. In the two studies in which effect sizes could be calculated (Bucklin et al., 2000, Ivarie, 1986), the effect size was determined to be large (Ferguson, 2009), suggesting a significant and strong effect from the rate-building procedure on retention accuracy. However, as Doughty et al. (2004) points out, number of practices and/or reinforcement rate may account for the large effect size.

Investigating effect sizes in studies not incorporating procedural controls revealed there is little support that the rate-building procedure has a large effect on retention accuracy. Less support was found for the effects of rate-building on retention fluency. Rate-building studies that incorporate procedural controls have shown rate-building does not improve retention accuracy (Cohen, 2008; Fox & Ghezzi, 2003; Wheetley, 2005; Experiments 1.1-1.8), with exception to Porritt (2007) and Porritt et al. (2009). Therefore, it is possible the reason the present investigation did not replicate the Retention component finding from Porritt (2007) is because the rate-building procedure, alone, has a minimum effect on retention accuracy.

EXPERIMENT 2.1

As mentioned earlier [Experiment 1.1 Introduction], the goal for Precision Teachers is to establish fluent performances in their students. Fluency ensures a set of outcomes referred to as REAPS. It is argued that to achieve fluency requires rate-building (Berens et al., 2003; Binder et al., 1990; Haughton, 1972; Binder, 1996). As discussed earlier, the majority of rate-building research has had procedural confounds (Doughty et al., 2004). However, Porritt et al. (2009) used a repeated acquisition procedure to overcome these confounds and concluded from their results that rate-building produces greater retention accuracy.

An important aspect of the repeated acquisition procedure is that it provides a repeatable, within-subject measure of learning. In the work by Porritt (2009), subjects learned a new chain sequence of responses during each experimental session. As mentioned [Experiment 1.1 Introduction], relearning the same task each session removes threats to internal validity (e.g., testing) discussed in some rate-building studies (Berens et al., 2003; Kubina et al., 2004; Shimmamune & Jitsumori, 1999). Despite its utility in rate-building studies, some dimensions of the repeated acquisition procedure are not well understood.

The behaviour chains developed when using repeated acquisition are said to be governed by environmental stimulus changes (e.g., colour cues) and from the location of the previous response (Keller & Schoenfeld, 1950, p.200). For example, the source of stimulus control for an L-C-R behaviour chain could come from the key colour designated for each link, the previous response position, or both (Kelleher, 1966). These different sources of stimulus control present two different explanations for how chains are developed.

One explanation is the chaining hypothesis (Skinner, 1938, p. 32). The critical aspect of this chaining hypothesis is the role assigned to response produced stimulus changes. For example, a correct response during Link 1 produces a stimulus that signals the correct response for Link 2. Thus, one explanation for how behaviour chains are formed is that each stimulus serves a discriminative function for the next correct response.

A second explanation for how behaviour chains are formed is the unitary-response hypothesis (Hull, 1952). The critical aspect of this hypothesis is that once a behaviour chain has been established, it may function as a single, unitary response (Kelleher, 1966) in the absence of discriminative stimulus changes. For example, the movement to complete a correct response during Link 1 governs the correct response during Link 2; behaviour chains are learned without additional environmental cues.

The chain and unitary-response hypotheses pose two conflicting explanations of the same behaviour phenomenon. The former states that behaviour chains are learned from response produced stimulus changes in the environment, whereas the latter states behaviour chains can be developed without discriminative stimuli. There are two studies that have attempted to clarify the role of stimuli governing behaviour chain formation in the repeated acquisition procedure (Snodgrass & McMillan, 1989; Thompson, 1970).

Thompson (1970) required pigeons to learn a new four link behaviour chain across three consecutive sessions. Using a reversal design (e.g., ABA) within each session, a new chain was learned using colour cues during the first and last part of each session. During this second part of each session, coloured cues were removed so that responses did not produce any discriminative stimuli. Keylights were dimmed following each correct response during the second part of each session. Thompson’s results showed a greater decrease in errors when responding produce discriminative stimuli, suggesting colour cues facilitated learning.

Similar to Thompson (1970), Snodgrass and McMillan (1989) examined whether responding during a repeated acquisition procedure is facilitated by discriminative stimuli. Using a reversal design, subjects learned a new behaviour chain during a training session. A test session followed each training session. During the test session, the same behaviour chain was trained, but the final 15 trials were a “test block”. During this test block, the colours presented for each link changed to allow examination of whether subjects’ behaviour was under the control of colour cues.

The test blocks were under one of two conditions, either a sequence- or conditional discrimination-appropriate condition. In the sequence-appropriate condition, the next chain link was presented only if the pigeons responded based upon the correct position, not colour. In the conditional discrimination-appropriate condition, the next chain link was presented only if the pigeons responded based upon colour, not position. Results showed more errors were made when chain links advanced during the sequence-appropriate condition. Results suggest that colour cues facilitated learning of behaviour chains. Findings from Experiment 1 may be brought to bear on the results of Snodgrass and McMillan.

Experiment 1 found that greater response rates improved accuracy. It may be that the different conditions used by Thompson (1970) and Snodgrass and McMillan (1989) generated different rates of reinforcement, thus response rates differed across conditions. This difference in response rates may have produced the greater accuracy in their experiments, not the use of colour cues.

Thompson (1970) and Snodgrass and McMillan (1989) have shown that discriminative stimuli facilitate behaviour chains. Conducting a microanalysis of response patterns during each link of the repeated acquisition procedure may reveal additional effects from independent variables on behaviour chain formation that were not revealed in original analysis (Cohn, MacPhail, & Paule, 1996; Cohn & Paule, 1993)

Experiment 2.1 replicated the procedures from Snodgrass and McMillan (1989) to compare findings. A microanalysis was conducted by examining the response rate and response latency during each experimental condition (Cohn et al., 1996; Cohn & Paule, 1993). Being that some chain sequences have been shown to be more difficult than others (Wright & Paule, 2007), the chains selected for Experiment 2 were identical to the ones used in Experiment 1. Using similar chains across experiments ensures that comparisons made between outcomes of experiments were not due to differences in task difficulty.

Method

Subjects

Six subjects, numbered 11 through 16, were experimentally naïve Shaver-Starcross domestic hens (*Gallus gallus domesticus*). The hens were housed individually in home cages (500-mm long × 510-mm wide × 420-mm high), in a ventilated room on a 12-hr light: 12-hr dark cycle. They had free access to water; grit and vitamins were provided weekly. Throughout the experiment all hens had red fleshy combs suggesting good health. Each hen was weighed every day an experimental session took place (approximately six days per week) and they were maintained at 80% (+/-5%) of their free-feeding body weights through feeding of commercial layer pellets.

Apparatus

The apparatus was a particleboard experimental chamber (530-mm long, 470-mm wide, 530-mm high). The chamber floor was covered with a thick clear plastic that had black plastic matting on top (300-mm long x 300-mm wide). A food magazine was located on the right-hand wall of the chamber behind an opening (115-mm high × 70-mm wide) that was centered 100-mm above the floor. When operated, the magazine was lit with a clear bulb and raised; giving the subjects access to wheat. Three horizontally spaced (100-mm) keys (30-mm in diameter), which could be lit blue, red, or yellow with a 28 V multi-chip LED bulb were placed above the magazine opening (400-mm from the floor). Each key required a force of approximately 0.2 N to close a micro switch.

All experimental events were controlled and recorded by Med-PC® IV software run on a Dell Optiplex GX110. Summary data for each session were also manually written into a data book at the end of each session.

Procedures

The procedures for keypeck training, Phase I, Phase II, and Phase III were identical to those used during Experiment I. The auto-shaping procedure lasted for 13 sessions. Five sessions were required to complete Phase I. Phase II continued for 10 sessions and Phase III lasted for 20 sessions.

The procedures for the repeated acquisition procedure for Phase IV were similar to those used during Phase IV of Experiment I except there was no ITI separating chain completions. All hens were exposed to six series of the same 12 chains sequences used during Phase IV of Experiment 1. Phase IV lasted for 48 sessions.

Experimental Sessions. During each experimental session, the hopper was raised for 2 s access to wheat following completion of five chains. The hopper light was illuminated following each chain completion. Experimental sessions were either training or testing.

During the training sessions, each hen learned a new chain sequence. Each training session was divided into six blocks, with each block requiring subjects to complete 25 chain completions. A training session ended once subjects completed all six blocks, 150 total chain completions. Test sessions occurred the day following a training session. The number of responses, chain completions, and reinforcers delivered during each block of the Training and Test session is presented in Table 2.1.1.

The chain sequence trained during the previous training session was used during the Test session. Each test session was divided into six blocks. Blocks 1-5 each required subjects to complete 25 chains. Block 6 was a test block. During the Test block, the colours presented for each link changed. The colour illuminating each key for Link 1, 2 and 3 were now red, yellow, and blue, respectively. Two experimental conditions were used during the Test block.

Sequence-appropriate condition. During sequence-appropriate conditions, correct responses were defined by the sequence position used during the training session and Blocks 1-5 of the Test session. For example, if an L-C-R sequence was used during the training session and Blocks 1-5 of the testing session, a left keypeck during Link 1 of the Test block was considered correct and the chain sequence advanced to the next link. After subjects completed 15 chain completions, a new chain was trained during the next training session.

Conditional discrimination-appropriate condition. During conditional discrimination-appropriate conditions, correct responses were defined by the colour presented during each chain link for the training session and Blocks 1-5 of the Test session. For example, if an L-C-R sequence was used during the Training session and Blocks 1-5 of the Test session, a centre keypeck during Link 1 of the Test block would be considered correct and the chain sequence advanced to the next link. After subjects completed 15 chain completions, a new chain was trained during the Training session.

If a subject pecked an illuminated key not designated as correct during any link (i.e., error), all keylights were darkened for 1 s. During this blackout period, keypecks did not produce a consequence. After the 1 s, the three keylights were again illuminated with the same colours as before the blackout until a correct keypeck was made for that link.

After three chains were trained (termed Phase A) and testing under the Sequence-appropriate condition (termed Phase B), the Conditional discrimination-appropriate condition began (termed Phase C). Each experimental condition was in effect for six consecutive sessions, after which a different condition began (i.e., an A/B/A/B/A/B/A/C/A/C/A/C experimental design). The session type, order of sessions, chain used in each session, and the experimental conditions are presented in Table 2.1.2.

Summary data that were manually recorded in the data book at the end of each session included the total errors in each component, session time in seconds, and reinforcers delivered. Event data were recorded by Med-PC® using a system of 1's and 0's to represent events and responses that occurred within the chamber. These 1's and 0's were used to calculate percentage correct, response rate, and latency. All raw data used in the following analysis, along with the programs used to analyse the data, can be found in the Appendix.

Table 2.1.1

Number of responses, chain completions, and reinforcers delivered during each block of the Training and Test session.

Training session		Test session		
75 responses 25 chain completions 5 reinforcers (FR5)	Block 1	75 responses 25 chain completions 5 reinforcers (FR5)	Block 1	
75 responses 25 chain completions 5 reinforcers (FR5)	Block 2	75 responses 25 chain completions 5 reinforcers (FR5)	Block 2	
75 responses 25 chain completions 5 reinforcers (FR5)	Block 3	75 responses 25 chain completions 5 reinforcers (FR5)	Block 3	
75 responses 25 chain completions 5 reinforcers (FR5)	Block 4	75 responses 25 chain completions 5 reinforcers (FR5)	Block 4	
75 responses 25 chain completions 5 reinforcers (FR5)	Block 5	75 responses 25 chain completions 5 reinforcers (FR5)	Block 5	
75 responses 25 chain completions 5 reinforcers (FR5)	Block 6	15 responses 5 chain completions 1 reinforcers (FR1)	Block 6	(Test block)

Table 2.1.2

Session type, order of sessions, chain used in each session, and the experimental conditions are given.

Session Type	Chain	Experimental Condition
Session 1		
Training	R-L-R	None
Session 2		
Test	R-L-R	Sequence-appropriate
Session 3		
Training	R-C-L	None
Session 4		
Test	R-C-L	Sequence-appropriate
Session 5		
Training	L-C-L	None
Session 6		
Test	L-C-L	Sequence-appropriate
Session 7		
Training	R-L-R	None
Session 8		
Test	R-L-R	Conditional discrimination-appropriate
Session 9		
Training	R-C-L	None
Session 10		
Test	R-C-L	Conditional discrimination-appropriate
Session 11		
Training	L-C-L	None
Session 12		
Test	L-C-L	Conditional discrimination-appropriate

Results

Single-subject data for Table 2.1.4 and Figures 2.1.1-2.1.4 were analysed using a one-way repeated measures analysis of variance for all measures during the Sequence- and Conditional discrimination-appropriate conditions for Experiment 2.1. The alpha level for all statistical comparisons in all situations was set at .05 and any results that reached this level were presented with an asterisk (*) in Table 2.1.3. Except where indicated with a hashtag (#) in Table 2.1.4, Mauchley's Test was not significant so sphericity was assumed. In these instances, and for Experiments 2.1-2.3, Greenhouse Geisser correction was used. Post-hoc tests were conducted using the Bonferroni correction, as recommended by Fields (2005).

Table 2.1.4 shows the percentage correct, percentage errors, and number of errors during the Sequence- and Conditional discrimination-appropriate conditions. On the left- and right-hand panels of the table, the numbers from left to right indicate the percentage of correct during the block preceding the Test block (bracketed numbers), percentage correct during the Test block, percentage of total errors on the conditional discrimination or sequence key during the Test block, percentage of total errors on the "other" key during the Test block, and the total number of errors during the Test block. The following outlines the measures used in this study, how they were calculated, and what was found. Table 2.1.5 shows, for comparison, the same data from the study by Snodgrass and McMillan (1989).

A correct response represents a key peck to an illuminated key designated as correct. The percentage of correct during the block preceding the Test block was calculated by dividing the total number of responses to an illuminated key during Block 5 of the Test session by the total number of correct responses for that block. Percentage correct was similar between the two conditions for all subjects and the data from these conditions were not significantly different (Table 2.1.3)

The percentage of correct during the Test block was calculated by dividing the total number of responses to an illuminated key during Block 6 of the Test session by the total number of correct responses for that block. Percentage correct was greatest during the Conditional discrimination-appropriate condition for all subjects. The data from these conditions were significantly different (Table 2.1.3) and effect size, partial eta squared, was large (Ferguson, 2009).

The percentage of total errors on the conditional discrimination or sequence key represents a response to an illuminated key during the Test block that was not correct and was either a sequence- or conditional discrimination-appropriate response, based upon the experimental condition. For example, a sequence-appropriate keypeck under the Conditional discrimination-appropriate condition would be considered a sequence-appropriate error. The percentage was calculated by dividing the total number of responses to illuminated keys during Block 6 of the Test session by the total number of errors for either the sequence- or conditional discrimination-appropriate key, based upon condition, for that block. The percentage of conditional discrimination-appropriate errors was greater than the sequence-appropriate percentage of errors. The data from these conditions were significantly different (Table 2.1.3) and effect size, partial eta squared, was moderate (Ferguson, 2009).

The percentage of total errors on the “other” key represents a response to an illuminated key during the Test block that was neither a sequence- nor conditional discrimination-appropriate response. This percentage was calculated by dividing the total number of incorrect responses on the “other” key during the Test block by the total number of responses to an illuminated key during Block 6 of the Test session. The percentage of errors on the “other” key was generally similar and the data from these conditions were not significantly different (Table 2.1.3)

The total number of errors represents the number of responses to illuminated keys during the Test block that were not designated as correct. The total number of errors was greatest during the Sequence-appropriate condition for all subjects. The data from these conditions were significantly different (Table 2.1.3) and effect size, partial eta squared, was large (Ferguson, 2009).

Fig. 2.1.1 shows the mean response rates (+ 1 *SD*) across chains for Blocks 1-6 of the Test sessions under the Sequence-appropriate and Conditional discrimination-appropriate conditions for all subjects. Response rates were greatest during the Sequence-appropriate condition during Blocks 1-6 for all subjects. The data from these conditions were significantly different (Table 2.1.3) and effect size, partial eta squared, was moderate (Ferguson, 2009).

Fig. 2.1.2 shows the mean response latency (+/- 1 *SD*) across chains for Blocks 1-6 during the Test session under the Sequence-appropriate and Conditional discrimination-appropriate conditions for all subjects. For each group of data points, the first, second, and third plot represents Links 1, 2, and 3, respectively. Response latencies were greatest during Link 1, regardless of experimental condition. This finding was significant (Table 2.1.3) and effect size, partial eta squared, was large (Ferguson, 2009).

Fig. 2.1.4 shows the mean response latency (+/- 1 *SD*) across chains for Link 1 during Blocks 1-6 under the Sequence-appropriate and Conditional discrimination-appropriate conditions for all subjects. Response latency was greatest following the fifth chain completion during each block, regardless of experimental condition. This finding was significant (Table 2.1.3) and effect size, partial eta squared, was large (Ferguson, 2009).

Table 2.1.3

Analysis of variance results during the Training and Test Blocks of Experiment 2.1

Comparison	MS Treatment	MS Error	df	F	p	Partial Eta Squared
	Percentage correct prior to Test block					
Sequence X Conditional Discrimination	132.25	32.96	1, 17	4.01	.06	.19
	Percentage correct during Test block					
	37960.03	146.56	1, 17	259.00	p<.0001*	.94
	Percentage of total errors during Test block					
	11844.69	938.69	1, 17	12.62	.002*	.43
	Percentage of total errors on the "other" key during Test block					
	641.78	736.54	1, 17	.87	.36	.05
	Total number of errors during Test block					
	30276.00	7248.00	1, 17	71.01	p<.0001*	.81
	Mean response rates for Blocks 1-6 during Test sessions					
	.083	.002	1, 35	37.78	p<.0001*	.52
	Mean response latency for Blocks 1-6 during Test sessions					
Link 1 X Link 2 X Link 3	12.74	.04	2, 20	289.02	p<.0001*	.96
	Mean response latency for Link 1 during Blocks 1-6 of Test sessions					
Trial 1 X Trail 2 X Trial 3 X Trial 4 X Trial 5	165.6	.83	4, 40	200.71	p<.0001*	.95

Table 2.1.4 and Table 2.1.5

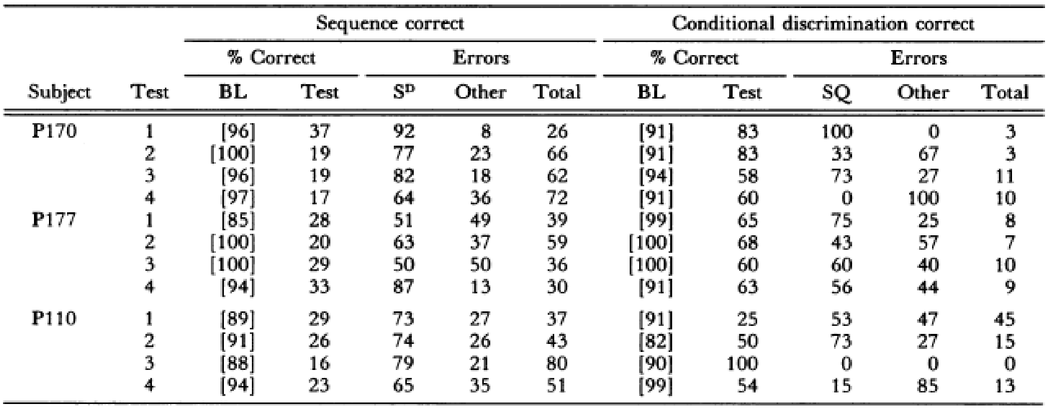
Percentage correct, percentage errors, and number of errors during the sequence- and conditional discrimination-appropriate conditions. On the left- and right-hand panels of the table, the numbers from left to right indicate the percentage of correct during the block preceding the Test block (bracketed numbers), percentage correct during the Test block, percentage of total errors on the conditional discrimination or sequence key during the Test block, percentage of total errors on the "other" key during the Test block, and the total number of *errors during the Test block.*

Table 2.1.4

		Sequence correct					Conditional discrimination correct				
		% Correct		Errors			% Correct		Errors		
Subject	Test	BL	Test	% S^D	% Other	Total	BL	Test	% SQ	% Other	Total
11	1	[97]	36	78	22	27	[94]	83	33	67	3
	2	[99]	36	85	15	27	[82]	88	100	0	2
	3	[95]	16	87	13	79	[90]	94	100	0	1
12	1	[95]	22	69	31	54	[95]	79	25	75	4
	2	[82]	31	53	47	34	[93]	83	100	0	3
	3	[93]	9	73	27	144	[84]	88	0	100	2
13	1	[92]	21	69	31	55	[87]	83	0	100	3
	2	[80]	21	79	21	56	[75]	100	0	0	0
	3	[91]	13	82	18	98	[87]	100	0	0	0
14	1	[91]	22	72	28	54	[94]	88	0	100	2
	2	[77]	25	78	22	46	[86]	71	83	17	6
	3	[92]	17	81	19	72	[89]	83	67	33	3
15	1	[92]	25	70	30	44	[68]	47	71	29	17
	2	[92]	27	73	28	40	[84]	83	33	67	3
	3	[92]	21	67	33	55	[87]	100	0	0	0
16	1	[94]	19	81	19	63	[96]	100	0	0	0
	2	[97]	18	87	13	69	[92]	100	0	0	0
	3	[89]	16	81	19	77	[88]	94	100	0	1

Table 2.1.5

Reproduced with permission from John Wiley and Sons



		Sequence correct					Conditional discrimination correct				
		% Correct		Errors			% Correct		Errors		
Subject	Test	BL	Test	S^D	Other	Total	BL	Test	SQ	Other	Total
P170	1	[96]	37	92	8	26	[91]	83	100	0	3
	2	[100]	19	77	23	66	[91]	83	33	67	3
	3	[96]	19	82	18	62	[94]	58	73	27	11
	4	[97]	17	64	36	72	[91]	60	0	100	10
P177	1	[85]	28	51	49	39	[99]	65	75	25	8
	2	[100]	20	63	37	59	[100]	68	43	57	7
	3	[100]	29	50	50	36	[100]	60	60	40	10
	4	[94]	33	87	13	30	[91]	63	56	44	9
P110	1	[89]	29	73	27	37	[91]	25	53	47	45
	2	[91]	26	74	26	43	[82]	50	73	27	15
	3	[88]	16	79	21	80	[90]	100	0	0	0
	4	[94]	23	65	35	51	[99]	54	15	85	13

Snodgrass and McMillan (1989)

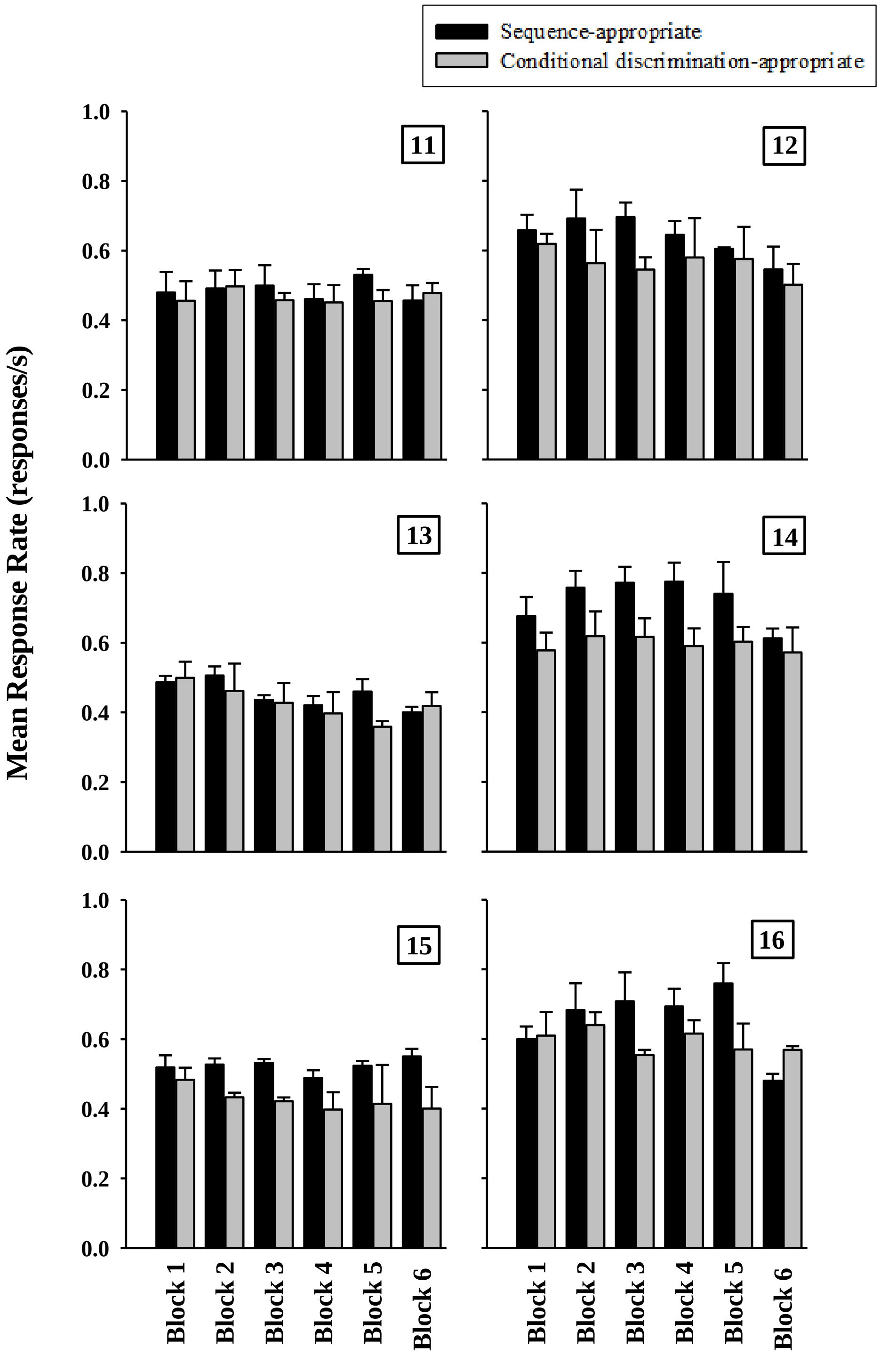


Fig. 2.1.1. The mean response rates (+ 1 *SD*) across chains for Blocks 1-6 of the Test sessions under the Sequence-appropriate and Conditional discrimination-appropriate conditions for all subjects.

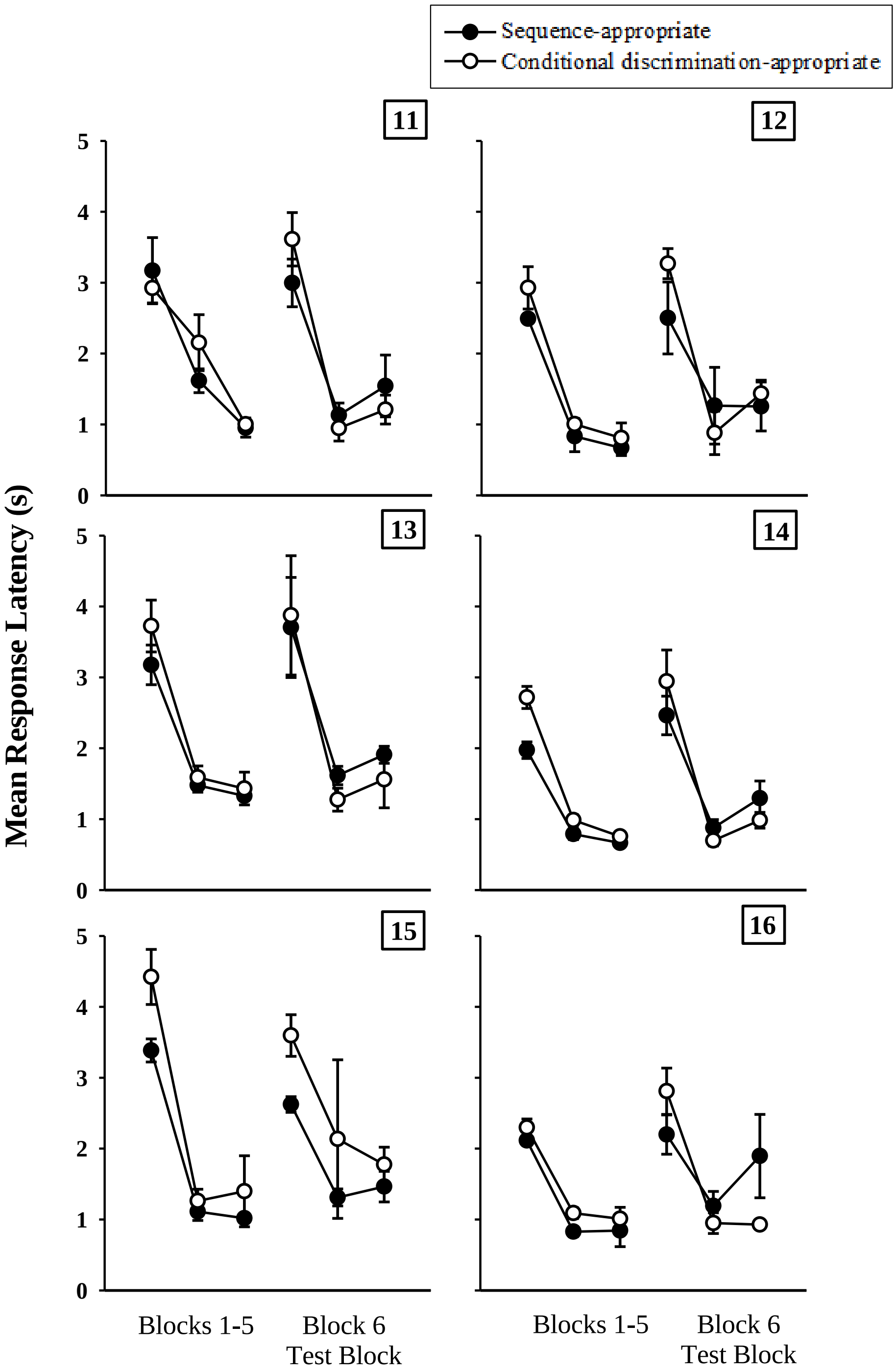


Fig. 2.1.2. The mean response latency (+/- 1 *SD*) across chains for Blocks 1-6 during the Test session under the Sequence-appropriate and Conditional discrimination-appropriate conditions for all subjects. For each group of data points, the first, second, and third plot represents Links 1, 2, and 3, respectively.

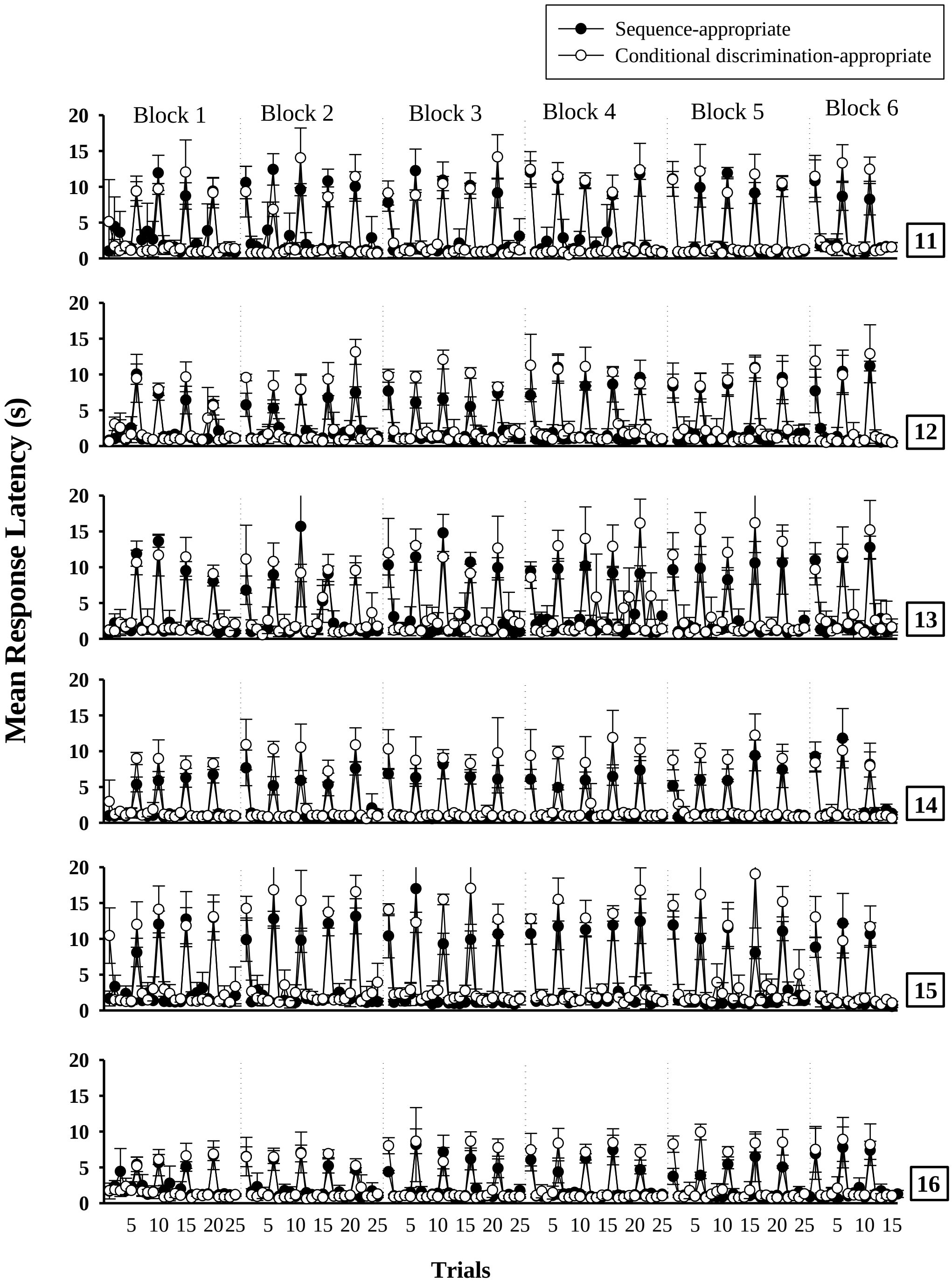


Fig. 2.1.4. The mean response latency across chains for Link 1 during Blocks 1-6 under the Sequence-appropriate and Conditional discrimination-appropriate conditions for all subjects.

Discussion

Experiment 2.1 replicated the procedures from Snodgrass and McMillan (1989) with hens to compare results. Similar to Snodgrass and McMillan, percentage correct during Blocks 1-5 was generally similar between experimental conditions, suggesting all subjects were responding with equal accuracy before each Test block. Percentage correct during the Test block was greatest during the Conditional discrimination-appropriate condition. This finding suggests hens were responding to key colour more than position of last response, supporting the chaining hypothesis (Skinner, 1938, p. 32). Given the case here, and given the similarity in procedures (e.g., repeated acquisition), it is likely subjects responding in Experiment 1 was governed by key colour.

Similar to Snodgrass and McMillan (1989), the percentage of errors on a conditional discrimination-appropriate key was greater than errors made on a sequence-appropriate key, supporting the finding that hens were responding to key colour more so than the position of last response. The percentage of errors on the “other” key was generally similar during the Conditional discrimination- and Sequence- appropriate conditions.

As previously mentioned [Experiment 2.1 Introduction], conducting a microanalysis of response patterns during each link of the repeated acquisition procedure may reveal additional effects from independent variables on behaviour chain formation that were not revealed in original analysis (Cohn et al., 1996; Cohn & Paule, 1993). Response rates were assessed because Experiment 1 found that greater rates enhanced accuracy. Response rates were greatest during the Sequence-appropriate condition for all hens. This finding suggests that colour cues, not response rates, accounts for the greater percentage correct during the conditional discrimination-appropriate condition. This result supports the conclusion that colour cues facilitate responding during the repeated acquisition procedure (Snodgrass & McMillan, 1989).

A microanalysis of response patterns revealed latency to respond was greatest during Link 1, regardless of experimental condition; showing that subjects responded faster as the chain links progressed. This finding suggests that control of responding is coming from previous responding, regardless of condition. Closer inspection of Link 1 latencies revealed the greater Link 1 latency durations occurred after reinforcement was delivered. This finding suggests that these Link 1 latency durations were due to post-reinforcement pausing. This finding suggests that the latency to respond to Links 1-3 were generally similar, supporting the conclusion that colour cues facilitate responding during the repeated acquisition procedure (Snodgrass & McMillan, 1989).

In summary, Experiment 2.1 replicated the procedures of Snodgrass and McMillan (1989) and found similar results. The results of the present study suggest subject's responding during chain completions in Experiment 1 were governed by the coloured cues presented during each link. Microanalysis of response rates and latencies provided additional information and further supported the same conclusion as Snodgrass and McMillan, supporting the chaining hypothesis (Skinner, 1938, p. 32)

As previously mentioned [Experiment 2.1 Introduction], Thompson's (1970) data showed a greater decrease in errors when correct responses produced discriminative stimuli (e.g., colour cues). Conducting a microanalysis of response patterns during each link of the repeated acquisition procedure may reveal additional effects from independent variables on behaviour chain formation that were not revealed in original analysis (Cohn et al., 1996; Cohn & Paule, 1993). Experiment 2.2 replicated the procedures from Thompson (1970) to compare findings and extended the analysis by examining response rates.

EXPERIMENT 2.2

Introduction

The goal of Experiment 2.2 was to replicate the procedures from Thompson (1970) and compare findings.

Method

Subjects

The same six subjects from Experiment 2.1 participated in this study.

Apparatus

The apparatus was the same as that used in Experiment 2.1.

Procedures

Subjects did not require pre-training due to previous exposure to the repeated acquisition procedure. The same three chain sequences used in Experiment 2.1 were used in the present investigation. A reversal design was used in each session (e.g., ABA) in which subjects were required to complete sixty chain sequences (i.e., 180 trials) across three experimental conditions. Table 2.2.1 shows the number of responses, chain completions, reinforcement schedule and experimental conditions during each session.

Colour cue condition. During the Colour cue condition (i.e., Condition A), subjects made chain completions using the repeated acquisition procedure in a similar fashion to Experiment 2.1.

No-colour cue condition. During the No-colour cue condition (i.e., Condition B) subjects completed chains using the repeated acquisition procedure in a similar fashion to Experiment 2.1 with one exception. The three keylights for each link were illuminated in white and were dimmed for .9 s following each correct keypeck.

If a subject pecked an illuminated key not designated as correct during any link (i.e., error) in either condition, all keylights were darkened for 1 s. During this blackout period, keypecks did not produce a consequence. After the 1 s, the three keylights were again illuminated with the same colours as before the blackout until a correct keypeck was made for that link.

Each experimental condition ended following twenty chain completions (i.e., sixty trials) and each condition was broken into 2 blocks of ten chains. Reinforcement was presented for 2 s following every fifth chain completions.

Table 2.2.1.

The number of responses, chain completions, reinforcement schedule and experimental conditions during each session.

30 responses 10 chain completions 5 reinforcers (FR5)	Block 1	Condition A Colour cues
30 responses 10 chain completions 5 reinforcers (FR5)	Block 2	
30 responses 10 chain completions 5 reinforcers (FR5)	Block 3	Condition B No-colour cues
30 responses 10 chain completions 5 reinforcers (FR5)	Block 4	
30 responses 10 chain completions 5 reinforcers (FR5)	Block 5	Condition A Colour cues
30 responses 10 chain completions 5 reinforcers (FR5)	Block 6	

Results

Figure 2.2.1 shows the mean number of errors (+/- 1 *SD*) across chain type for each block of ten trials under the Colour cue or No-colour cue experimental conditions. The data show greater number of errors during the No-colour cue conditions for all subjects. This finding was significant (F (2, 10) = 144.67, p < .0001, η_p^2 = .82.) and effect size, partial eta squared, was large (Ferguson, 2009).

Figure 2.2.2 shows the mean response rate (+1 *SD*) across chain type for each block of ten trials under the Colour cue and No-colour cue experimental conditions. Response rates were greatest during the No-colour cue condition for all subjects. This finding was significant (F (2, 10) = 38.87, p < .0001, η_p^2 = .73.) and effect size, partial eta squared, was large (Ferguson, 2009).

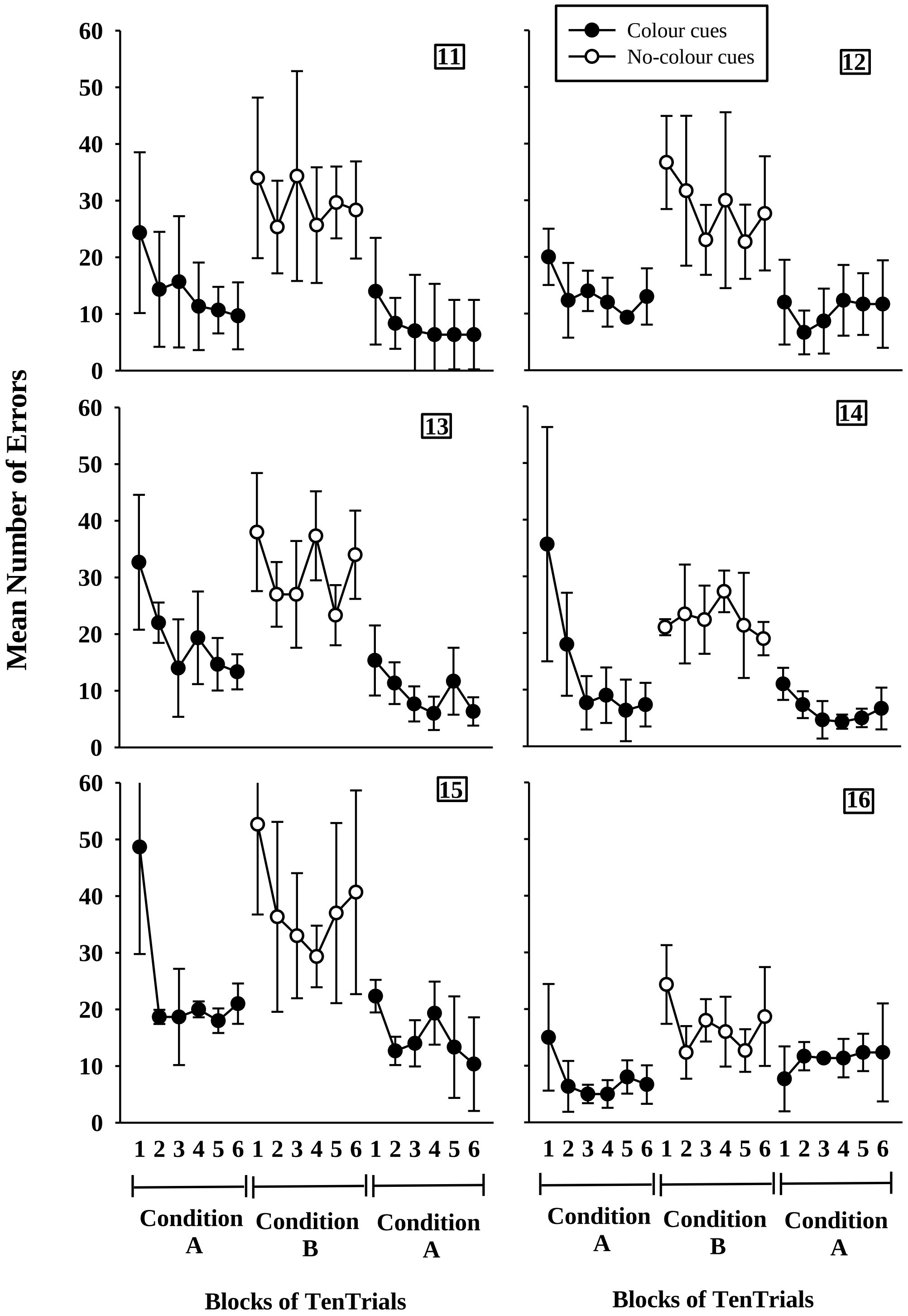


Figure 2.2.1. The mean number of errors (+1 *SD*) across chain type for each block of ten trials under the Colour and No-colour experimental conditions.

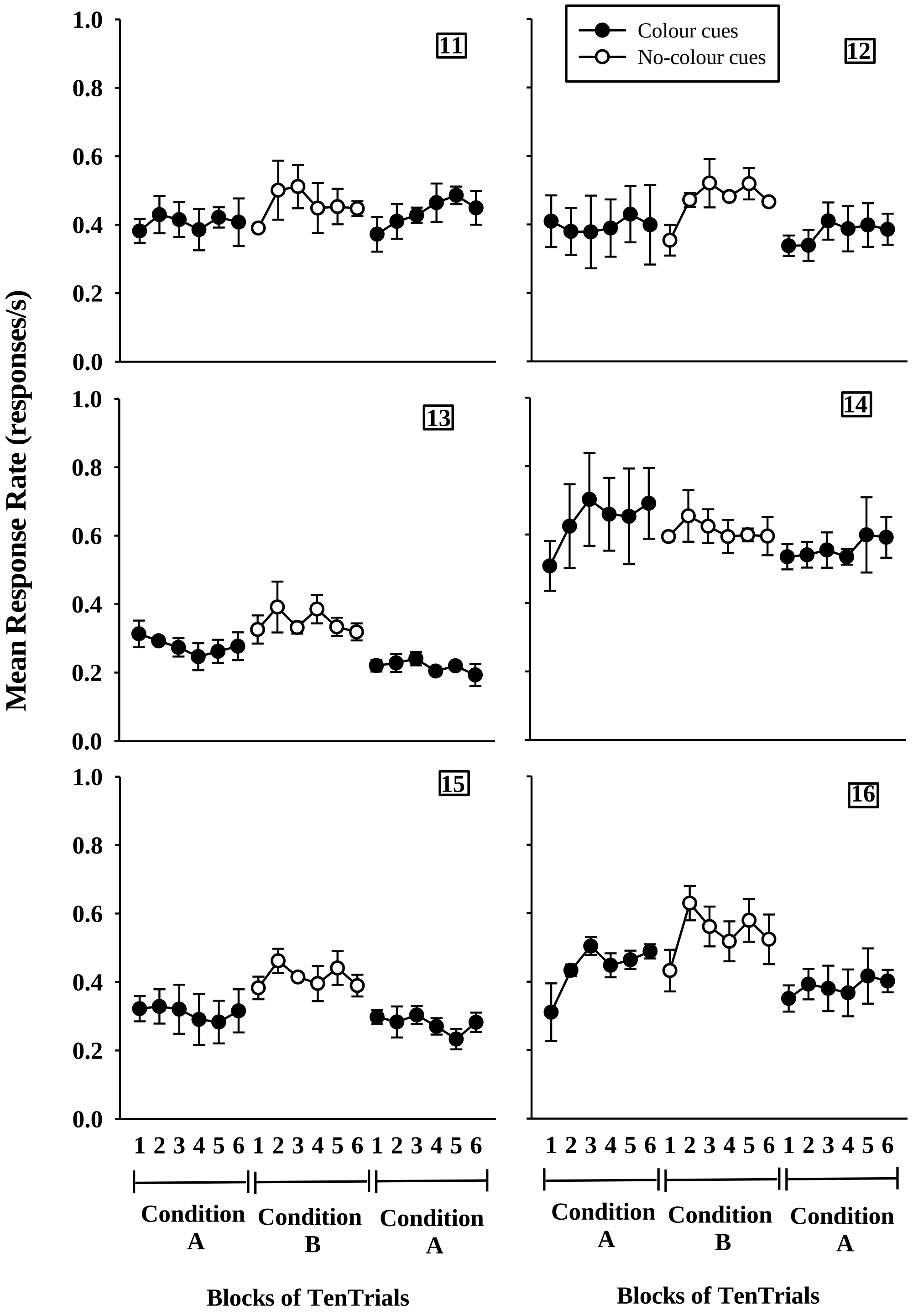


Figure 2.2.2. The mean response rate (+1 *SD*) across chain type for each block of ten trials under the Colour and No-colour experimental conditions.

Discussion

Experiment 2.2 replicated the procedures from Thompson (1970) to compare findings. Similar to Thompson, there was a greater decrease in errors during the Colour cue condition. This finding suggests that subjects learned the chain sequence more rapidly with the colour cues.

The greater percentage of corrects during the conditional discrimination condition of Experiments 2.1 and 2.2 suggests that response produced stimulus changes (e.g., colour change from Link 1 to Link 2) governed subjects' responding during the repeated acquisition procedure, supporting the chaining hypothesis (Skinner, 1938, p. 32). Given the case here, and given the similarity in procedures (e.g., repeated acquisition), it is likely subjects responding in Experiment 1 was governed by key colour.

As previously mentioned [Experiment 2.1 Introduction], response rates were assessed because Experiment 1 found that greater rates enhanced accuracy. Response rates were greatest during the Sequence-appropriate condition. This finding suggests that the greater accuracies during the colour condition were produced by the cues, not the response rate.

In summary, the present investigation replicated the procedures by Thompson (1970) and found similar results. Response rate data provided additional information and further supported the same conclusion as Thompson, showing favour for the chaining hypothesis (Skinner, 1938, p. 32). As previously mentioned [Experiment 2.1 Introduction], Thompson (1970) dimmed the keylights during the No-colour cue condition. Keylight dimming could be argued to provide an additional stimulus following each correct response. Added stimuli following correct responding has been shown to affect the accuracy of completing chain sequences (Hursh, 1977).

Hursh (1977) used a within-subject design to investigate the effects of added stimuli on monkey's chain completion accuracy using a repeated acquisition procedure. Subjects were presented with a new three-link chain each session where, following a correct response during the link, a new colour was projected onto all three keys. In addition to the colour, a correct response also produced a white spot of light superimposed in the corner of the correct key (i.e., distinctive stimulus). Stimuli were systematically removed to test the effects of the distinctive stimuli on chain completion accuracy. Hursh concluded the added distinctive stimuli improved accuracy. This outcome suggests the added keylight dimming used by Thompson (1970) may have impacted chain completion accuracy. Experiment 2.3 examined the effects of keylight dimming on chain completion accuracy.

EXPERIMENT 2.3

Introduction

Experiment 2.3 examined the effects of keylight dimming on chain completion accuracy.

Methods

Subjects

The same 6 subjects from Experiment 2.2 participated in this study.

Apparatus

The apparatus was the same as that used in Experiment 2.2.

Procedures

The present study used the same ABA reversal design within each session as was used in Experiment 2.2 with one exception. Four experimental conditions, colour cues with keylight dimming (A), colour cues without keylight dimming (B), no-colour cues with keylight dimming (C), and no-colour cues without keylight dimming (D), were arranged across four different ABA reversals (e.g., Reversals 1-4). The order of experimental conditions within Reversals 1-4 are presented in Table 2.3.1.

Results

Figure 2.3.1 shows the mean number of errors (+/- 1 *SD*) across chain type under each experimental condition. Each symbol (filled circle) represents the sum of errors from a block of ten trials and the mean was calculated across chain type. Dotted phase change lines represent the different conditions within each session and the solid lines represent a new reversal (e.g., Reversals 1-4). The number of errors was greatest during the No-colour cue conditions (C and D) of Reversals 1, 2, and 4, regardless of keylight dimming for all subjects. Number of errors decreased during the first condition of each reversal, regardless of experimental condition.

Figure 2.3.2 shows the mean response rate (+/- 1 *SD*) across chain type under each experimental condition. Each symbol (filled circle) represents the sum of responses divided by the time to complete a block of ten trials and the mean was calculated across chain type. Dotted phase change lines represent the different conditions within each session and the solid lines represent a new set of reversals (e.g., Set 1-4). Response rates were greatest during the no-colour cue conditions during reversals 1 and 2 for all subjects. Table 2.3.2 shows this finding was significant and effect size, partial eta squared, was large (Ferguson, 2009).

Table 2.3.1.

The order of experimental conditions for Reversal 1-4.

	Conditions		
Reversal 1	B	D	B
Reversal 2	A	C	A
Reversal 3	A	B	A
Reversal 4	C	D	C

Table 2.3.2

Analysis of variance results for response rates during Reversal 1 and Reversal 2 of Experiment 2.3

Comparison	MS Treatment	MS Error	df	F	p	Partial Eta Squared
			Reversal 1			
B-D-B	.10	.001	2, 10	86.92	p < .0001*	.74
			Reversal 2			
A-C-A	.123	.004	2, 10	72.39	p < .0001*	.71

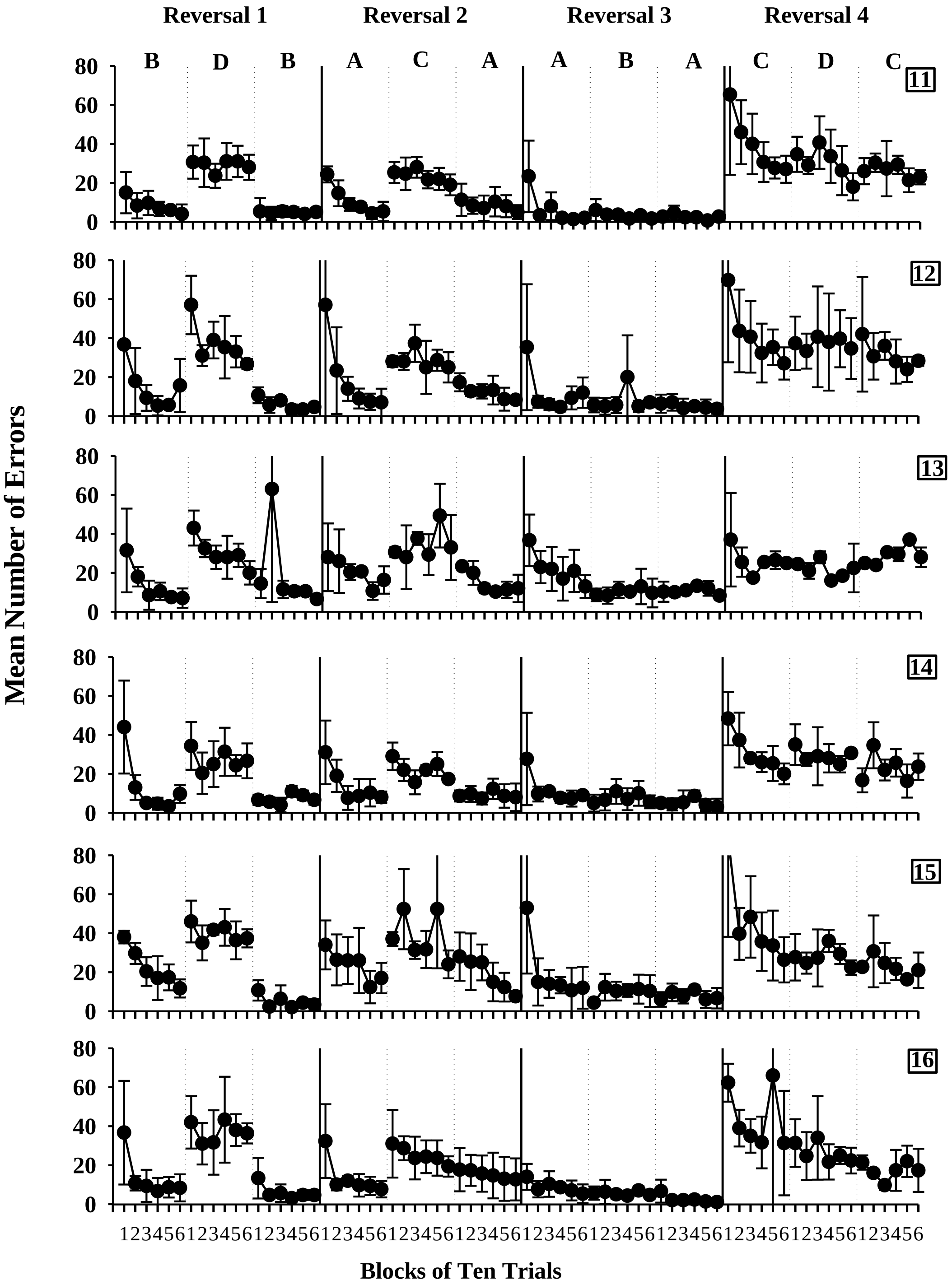


Figure 2.3.1. The mean number of errors (+/- 1 *SD*) across chain type under each experimental condition. Each symbol (filled circle) represents the sum of errors from a block of ten trials and the mean was calculated across chain type.

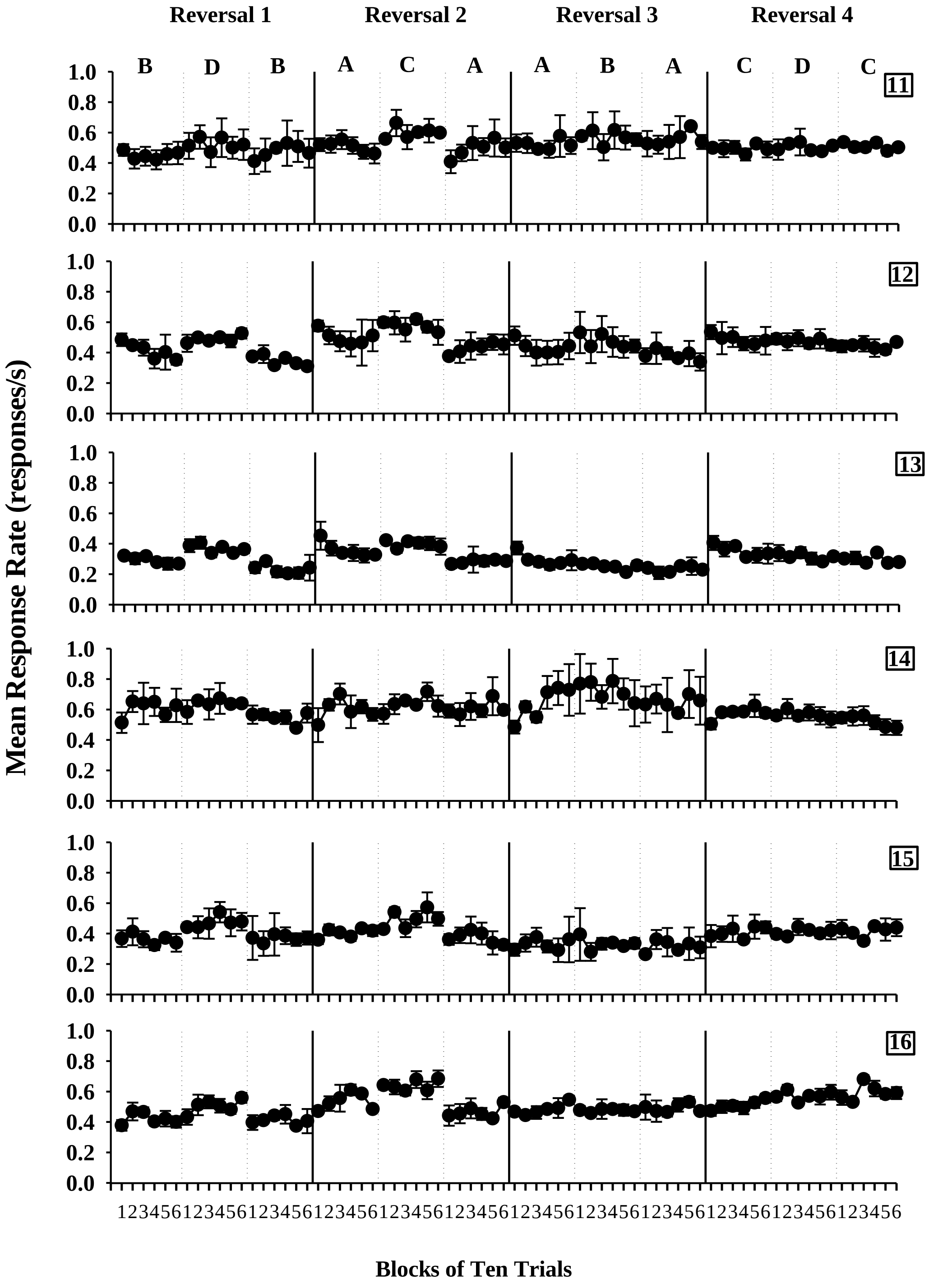


Figure 2.3.2. The mean response rate (+/- 1 *SD*) across chain type under each experimental condition. Each symbol (filled circle) represents the sum of responses divided by the time to complete a block of ten trials and the mean was calculated across chain type.

Discussion

Experiment 2.3 examined the effects of keylight dimming under Colour cue and No-colour cue conditions. Errors were greatest under the No-colour cue condition, similar to the results by Thompson (1970) and Experiment 2.2. Response rates were greatest during the Sequence-appropriate condition, suggesting the effects of colour and no-colour cues, not response rate, impacted response accuracy. Contrary to our predictions, keylight dimming did not impact accuracy during any experimental condition, suggesting dimming did not govern accuracy in the experiment by Thompson. This finding is contrary to other studies showing added stimuli improve accuracy (Hursh, 1977).

In the study by Hursh (1977), the added stimuli following each correct response differed across Links 1-3. For example, a white light was projected onto the correct key following a correct response during Link 1. A second white light was projected onto the correct key during Link 2 and a third white light was projected onto the third correct key during Link 3. This arrangement added a new stimulus following each correct response while the present study dimmed all three keylights (i.e., the distinctive stimuli) in a similar manner during Links 1-3. It is possible dimming the keylights in a similar fashion during Links 1-3 in the present investigation prevented the distinctive stimuli from governing subject's chain completion accuracy. Therefore, one possible explanation as to why the keylight dimming did not impact chain completion accuracy may be because the added stimuli following each correct response was similar across chain links.

The present study showed errors decreased during the first condition of each set of reversals, regardless of experimental condition. This finding suggests that both colour cues and position of last response (i.e., the No-colour cue condition) governed response accuracy. This finding is similar to the data presented by Thompson (1970). The decrease in errors during the first set of reversals during no-colour cue conditions suggest that some of the subjects' responding was governed by the position of the previous response, supporting the unitary-response hypothesis (Hull, 1952). As discussed by Kelleher (1966), it is possible that the behaviour chains established during the repeated acquisition procedure were functioning as a single, unitary response in the absence of discriminative stimulus changes. Given the case here, and given the similarity in procedure (e.g., repeated acquisition), subjects responding in Experiment 1 was under control of both position of last response and the key colour presented during each chain link.

Experiment 2.3 found the position of last response and colour cues both governs subjects' accuracy. Both these variables have also been shown to differentially affect behaviour in anatomical studies of avian hippocampus (Watanabe, 2001). Watanabe investigated the effects of hippocampus lesions in pigeons on spatial discrimination across colour cue and no-colour cue conditions. His results showed the lesions effected accuracy only when no-colour cues were used, suggesting that the hippocampus affects pigeons spatial discrimination. These results also support the present findings; position of last response and colour cues differentially affect behaviour. Based on these and similar findings (Thompson, 1970; Snodgrass & McMillan, 1989; Watanabe, 2001), it could be argued the colour cues used during each link of the repeated acquisition procedure are a performance enhancer. The added colour cues improve accuracy, and thus, could confound the effects of the rate-building procedure.

SUMMARY

Experiment 2.1 and 2.2 replicated procedures by Snodgrass and McMillan (1989) and Thompson (1970), respectively, and found similar results, supporting the chaining hypothesis (Skinner, 1938, p. 32). The results of these studies showed that colour cues govern response accuracy during the repeated acquisition procedure, suggesting that subjects in Experiment 1 were responding to each chain link based upon colour.

Experiment 2.3 investigated the effects of keylight dimming during Colour and No-colour cue conditions, supporting both the chaining (Skinner, 1938, p. 32) and unitary-response (Hull, 1952) hypotheses. It was found the keylight dimming used by Thompson did not impact chain completion accuracy. This finding is contrary to the study by Hursh (1977), who showed added stimuli following correct responding enhanced chain completion accuracy. One reason for the different outcomes may be because keylight dimming was similar across the three chain links in the present study.

Experiment 2.3 showed that both colour cues and position of last response govern chain completion accuracy. The findings of Experiment 2.3 suggest subjects responding in Experiment 1 could have been governed by either the colour of each link or the position of last response. As mentioned previously [Experiment 2.3 Discussion], colour cues could confound the effects of the rate-building procedure.

GENERAL DISCUSSION OF EXPERIMENT 2

Porritt (2009) used a repeated acquisition procedure to overcome procedural confounds in Precision Teaching studies and concluded that rate-building improved retention accuracy. Despite its utility to provide a repeatable within-subject measure of learning, some dimensions of the repeated acquisition procedure are not well understood; two conflicting theories explain what governs responding during the repeated acquisition procedure (Hull, 1952; Skinner, 1938). Two studies have attempted to clarify the role of stimuli governing responding during the repeated acquisition procedure (Snodgrass & McMillan, 1989; Thompson, 1970).

The aim of Experiments 2.1-2.3 was to replicate the procedures of Snodgrass and McMillan (1989) and Thompson (1970) to clarify the role of stimuli in the repeated acquisition procedure. Experiments 2.1-2.3 tested the following research questions:

1. Does colour or position of last response govern chain completion accuracy?
2. Will conducting a microanalysis of response patterns during each link of the repeated acquisition procedure reveal additional effects from independent variables during the repeated acquisition procedure (Cohn & Paule, 1993; Cohn et al., 1996)?
3. Did keylight dimming affect the results by Thompson (1970)?

With respect to the first research question, the data from Experiments 2.1 and 2.2 suggests that the colour presented during each chain link governed subjects responding. These data are consistent with the findings of Snodgrass and McMillan (1989) and Thompson (1970), supporting the chaining hypothesis (Skinner, 1938, p. 32). An extension of the Thompson procedure (Experiment 2.3) revealed that both colour and position of last response govern responding in the repeated acquisition procedure, supporting both the chaining (Skinner, 1938, p. 32) and unitary-response (Hull, 1952) hypotheses. Given the case here, and given the similarity in procedure (e.g., repeated acquisition), it is likely subjects responding in Experiment 1 was under control of position of last response and the key colour presented during each chain link. As previously discussed [Experiment 2.3 Discussion] using colour cues confounds the rate-building procedure. To avoid the confound from using colour-cues, future rate-building studies using the repeated acquisition procedure could ensure subjects respond based upon position of response only. This can be accomplished by eliminating colour cues during each chain link of the repeated acquisition procedure.

A microanalysis of subjects' response rates was conducted because results from Experiment 1 suggested that greater response rates improved accuracy. With respect to the second research question, the response rate data from Experiments 2.1-2.3 suggested response rate did not impact accuracy. A microanalysis of response latency during Links 1-3 from Experiment 2.1 revealed the greater Link 1 latencies were not a product of subjects responding based upon the previous response, but were due to post-reinforcement pausing. This finding supports the conclusion of Snodgrass and McMillan (1989). Conducting a microanalysis of latencies and response rates provided additional information and further supported the same conclusion drawn by Snodgrass and McMillan (1989) and Thompson (1970).

With respect to the third research question, the data from Experiment 2.3 suggests that the dimming used by Thompson (1970) did not impact response accuracy during the repeated acquisition procedure. These findings are contrary to studies suggesting that using additional cues facilitates learning (Hursh, 1977). As mentioned [Experiment 2.3 Discussion], procedural differences may have accounted for the contrary findings.

As previously discussed [Experiment 2.1 Introduction], Experiments 2.1 and 2.2 showed that colour cues governed responding during the repeated acquisition procedure, confirming the chaining hypothesis (Skinner, 1938, p. 32). However, an extension of the Thompson (1970) procedure revealed that response accuracy during the repeated acquisition procedure is also governed by position of last response, confirming Hull's (1952) unitary-response hypothesis. This finding suggests that both theories predict what stimuli govern responding during the repeated acquisition procedure. More research is needed to clarify the predictions made by both theories, what governs responding during the repeated acquisition procedure used in rate-building studies.

GENERAL DISCUSSION

The present investigation consisted of eleven experiments across two series of studies. The first series had two parts; part one replicated the procedures by Porritt (2007) and Porritt et al., (2009) to demonstrate the repeatability of their findings. Training component findings were similar to those of Porritt, but percentage correct under the Retention component differed. Given the contrary outcomes, the studies in part two attempted to replicate the Retention component results of Porritt by using variables that have been shown to improve retention accuracy. The results replicated Porritt only when similar behaviours were trained between the Training and Retention components. The second series of studies investigated the role of stimuli in the repeated acquisition procedure. Findings suggest that colour cues enhance accuracy, more than position of last response. However, as previously discussed [Experiment 2.3 Discussion], using colour cues within the repeated acquisition procedure confounds the rate-building procedure. While results from these studies have been discussed previously, some warrant a more general discussion

There are two different ways to define fluency. The majority of people, specifically educators, use a definition of fluency based on the topography or appearance of the behaviour. For example, the lay person may describe behaviour as fluent when it is fast, smooth, and rhythmic. This type of fluency will hereafter be termed “topographical-fluency”. Precision Teachers, however, define fluency based upon the outcomes (e.g., retention, endurance, application) of generating greater response rates. This type of fluency will hereafter be referred to as “outcome-fluency”.

There are two different ways of defining a response in Precision Teaching. Precision Teachers may define a response as an instance of behaviour, a response without a measurable duration (e.g., shorter than 1 s). For example, a student’s response when answering a math fact provides an instance of behaviour. This type of response is similar to a keypeck in the present study. A keypeck could be considered a response without a measurable duration. Hereafter, this type of response will be referred to as an “instant-response”. Precision Teachers also define a response as having duration. For example, reciting the Pledge of Allegiance could be considered a response that has duration. This type of response is similar to the time to complete Links 1-3 in the present study. Hereafter, this type of response will be referred to as an “extended-response”.

Precision Teachers investigate the outcomes of generating topographical-fluency, typically set towards a criterion, by focusing on the rate of instant-responses. For example, Hughes et al. (2007) investigated the outcomes from generating greater rates of vocabulary words, a response that has no measurable duration, by setting specific goals. Similarly, the present study investigated the outcomes of generating greater instant-responses rates (e.g., key depressions) under the No-delay condition. It can be said, therefore, that rate-building studies investigate the effects of topographical-fluency on outcome-fluency by focusing on the rate of instant-responses. Thus, when response rate is hereafter discussed, it will refer to the rate of instant-responses.

As opposed to focusing on response rate, decreasing the duration of a response may also lead toward topographical-fluency (Howell and Lorson-Howell, 1990). Although the duration of extended-responses can be measured in Precision Teaching studies, they typically aren’t. Based on the suggestion by Howell and Lorson-Howell, perhaps a more relevant measure of topographical-fluency is measuring the duration of chain completion. Thus, extended-responses will hereafter be referred to as response duration. The present discussion suggests that both response rate and duration may generate topographic-fluency. Understanding the role of rate and duration in creating topographical-fluency may shed light on how to best generate outcome-fluency, the goal for Precision Teachers.

It is assumed that response duration and rate are correlated, an example can be found within traditional Precision Teaching methods. Traditionally, Precision Teachers use one-minute timings. Participants are encouraged to respond as fast and accurate as possible, typically set to a criterion. As topographical-fluency is developed, response duration decreases. In this case response duration and rate are correlated. These two measures can also be independent from one another. In an applied example, a participant can practice a song on the piano twice per day. While the rate of practice remains constant across months, the duration of practice (e.g., the response) will decrease as performance becomes more topographically-fluent. Thus, response duration and rate are correlated, but not in all cases. This raises the question whether response duration or rate better predict topographical-fluency.

Investigating the outcomes of conditions in which delays are imposed within or between responses is one way to test whether response duration or response rate better predicts topographical-fluency. In the present study, response duration represents the duration from the illumination of Link 1 keylights to the emission of the third correct keypeck (i.e., the response during Link 3). The present study imposed a delay following each correct keypeck (e.g., Within-chains delay) and after a response (e.g., Between-chains delay). Results showed that, while response rates were similar under these two conditions, response durations differed; the condition generating shorter response durations (e.g., Between-chains delay) produced the greatest topographical-fluency. Analysis of this result from Experiment 1.3 shows significant differences between experimental conditions, F (2, 6) = 29.4, p < .05, η_p^2 = .91. This result suggests that response duration has a significant effect on topographical-fluency. This analysis is important because it suggest that response duration, rather than response rate, generated topographical-fluency. Thus, an alternative interpretation of the data presented in this thesis suggests that response duration, not response rate, led towards topographical-fluency. This finding suggests that reducing duration (Howell & Lorson-Howell, 1990) or topographical fluency, rather than increasing rate, may be the critical variable for developing outcome-fluency.

The correlation between response duration and rate has led Precision Teachers to investigate the outcomes of generating topographical-fluency, typically set towards a criterion, by focusing on response rate. An interpretation of the data in the present study suggests focusing on duration-reduction also leads towards topographical-fluency. Reaction time is another measure, favoured by cognitive researchers, used to show topographical-fluency (Deary, Liewald, & Nissan, 2011).

Reaction time, or the time from the onset of a stimulus to a response (Sternberg, 1969), has been the measure of automaticity since the 1890's (Deary et al., 2011). Like topographical-fluency, automaticity refers to the ability to respond quickly and effortlessly (Dougherty & Johnston, 1996; Logan, 1978). However, unlike behavioural researchers, cognitivists state that automaticity is produced by inner changes of associative and perceptual processes (LaBerge & Samuels, 1974). These inner changes are revealed in reaction time data; shorter reaction times suggest shortened perceptual processing, and the more “automatic” the response becomes (LaBerge, 1973). Thus, it can be argued that response rate, duration, and reaction time are different measures from two scientific approaches (e.g., behavioural and cognitive), all aimed at producing topographical-fluency. While there is a lot of research focused on response rate and reaction time, little research exists focusing on duration-reduction in generating topographical-fluency. It may be that focusing on duration-reduction will aid Precision Teachers in producing outcome-fluency.

Generating topographical-fluency by focusing on duration-reduction changes the landscape of Precision Teaching’s quest to produce outcome-fluency. The first change is the dependent measure. As previously discussed [Experiment 1.1 Introduction], Precision Teachers focus on response rate to generate topographical-fluency. An interpretation of the data in the present study suggests, rather than response rate, duration-reduction should also be a considered as a dependent measure in developing topographic-fluency. Future research could use a performance standard that would be based upon response duration. For example, the performance standard could be completing ten consecutive chains, with each chain being completed under 2 s. The retention accuracy generated by different performance standards could then be assessed.

Interestingly, treating duration as a dependent measure impacts the initial question of this thesis. Viewing the effects of response rate while controlling for extra practices has been the focus of Precision Teaching discussions (Doughty et al., 2004, Binder, 2003; Kubina, 2005) and research (Campbell, 2012; Cohen, 2008; Holding, 2011; Fox & Ghezzi, 2003; Porritt, 2007; Porritt et al., 2009; Wheetley, 2005) for the past several years. For example, the present study used procedural controls to ensure response rate, not extra practices, accounted for changes in retention accuracy. The argument is that response rate and practices are correlated because extra practices are required for greater response rates (Doughty et al., 2004). This argument does not hold for duration because, as previously mentioned [General Discussion], response rate and response duration are not necessarily correlated. Thus, the initial argument of whether response rate or extra practices account for outcome-fluency is weakened if response duration is used as a dependent measure.

Overall, eleven experiments divided between two series of studies were aimed at, first, replicating the findings of Porritt (2007) and Porritt et al. (2009) and, then, gaining a better understanding of using the repeated acquisition procedure in animal analogue learning studies. Findings from Series 1 of the first part of replications suggest that greater response rates, when number of practices and reinforcement rate are controlled, enhance training accuracy. However, the greater response rates did not improve retention accuracy, a failure to replicate. Findings from the second part of the first series replicated Retention Component findings from Porritt (2007) and Porritt et al. (2009) when similar behaviours were trained and tested. Due to the differing procedures used between studies, it is still unclear why Porritt (2007) and Porritt et al. (2009) obtained different Retention Component results than the present study. An interpretation of the Series 1 data suggests that response duration may contribute towards topographical-fluency. Thus, it may be that a focus on duration-reduction leads towards greater retention accuracy. Findings from the second series of experiments suggest attention should be paid to the use of cues when the repeated acquisition procedure is used in learning experiments. Microanalysis from Series 2 data showed colour cues act as performance enhancers; if variables are to be studied that effect acquisition, the present findings suggest using no-colour cues in the repeated acquisition procedure. The present investigation began by asking how Porritt obtained his results. Overall, the present study found that focusing on duration-reduction, in an animal analogue study using a repeated acquisition procedure with no-colour cues, may reveal the prime contributor to greater retention in Precision Teaching.

REFERENCES

Alberto, P. A., & Troutman, A. C. (2003). *Applied behavior analysis for teachers* (6th ed.). Upper Saddle, NJ: Merrill.

Baddeley, A. D. & Longman, D. J. (1978). The influence of length and frequency of training sessions on the rate of learning to type. *Ergonomics, 21*, 627-635.

Baldwin, R. L., Chelonis, J. J., Prunty, P. K., & Paule, M. G. (2012). The use of an incremental repeated acquisition task to assess learning in children. *Behavioural Processes, 91*, 103-114.

Baron, A., Perone, M., & Galizio, M. (1991). The Experimental Analysis of Behavior: Indispensable, ancillary, or irrelevant? *The Behavior Analyst, 14,* 145-155.

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

Becker, W. C. (1977). Teaching reading and language to the disadvantaged – What we have learned from field research. *Harvard Educational Review, 1977, 47*, 518-543.

Belmont Report (1979). Ethical principles and guidelines for research involving human subjects. Retrieved from http://videocast.nih.gov/pdf/ohrp_belmont_report.pdf

Berens, K. N., Boyce, T. E., Berens, N. M., Doney, J. K., & Kenzer, A. L. (2003). A technology for evaluating relations between response frequency and academic performance outcomes. *Journal of Precision Teaching and Celeration, 19,* 20-34.

Bickel, W. K., Higgins, S. T., & Hughes, J. R. (1990). Development of repeated acquisition methodologies: Implications for the detection of drug-induced disruption in human learning. *NIDA Research Monographs, 100,* 99-111.

Bickel, W. K., Higgins, S. T., & Hughes, J. R. (1991). The effects of diazepam and triazolam on repeated acquisition and performance of response sequences with an observing response. *Journal of the Experimental Analysis of Behavior, 56*, 217-237.

Binder, C. V. (1988). Precision Teaching: measuring and attaining exemplary academic achievement. *Youth Policy, 10,* 12-15.

Binder, C. V. (1990). Precision Teaching and curriculum based measurement. *Journal of Precision Teaching, 7,* 33-35.

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

Binder, C. V. (2003). Doesn't everybody need fluency? *Performance Improvement, 42,* 14-20.

Binder, C. V., & Bloom, C. (1989, February). Fluent product knowledge: Application in the financial services industry. *Performance and Instruction,* 17-21.

Binder, C. V., & Watkins, C.L. (1990). Precision Teaching and Direct Instruction: Measurably superior instructional technology in schools. *Performance Improvement Quarterly, 3,* 74-96.

Binder, C. V., Haughton, E., & Van Eyk, D. (1990). Increasing endurance by building fluency: Precision teaching attention span. *Teaching Exceptional Children, 22*, 24-27.

Bloom, C. & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research, 74,* 245-248.

Branch, M. (1991). On the difficulty of studying "basic" behavioral processes in humans. *The Behavior Analyst*, *14*, 107-110.

Brown, P. L. & Jenkins, H. M. (1968). Auto-shaping of the pigeon's key-peck. *Journal of the Experimental Analysis of Behavior, 11*, 1-8.

Bryan, W.L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review, 6*, 345-375.

Bucklin, B. R., Dickinson, A. M., & Brethower, D. M. (2000). A comparison of the effects of fluency training and accuracy training on application and retention. *Performance Improvement Quarterly, 13*, 141–163.

Burke, L. M. (2012). States must reject national education standards while there is still time. *Backgrounder, 2680*, 1-9.

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

Campbell, T. (2012). Effect of response rates on non-distracted and distracted conditional discrimination performance. Unpublished master's thesis, University of Waikato, Hamilton.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354-380.

Chapman, S. S., Ewing, C. B., & Mozzoni, M. P. (2005). Precision Teaching and fluency training across cognitive, physical, and academic tasks in children with traumatic brain injury: A multiple baseline study. *Behavioral Interventions, 20,* 37-49.

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

Cohen, J. (2008). Retention, endurance, stability, and application of learned performances as a function of training condition. Unpublished master's thesis, University of North Texas, Denton.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Cohn, J. & Paule, M. G. (1993). Repeated acquisition of response sequences: The analysis of behavior in transition. *Neuroscience and Biobehavioral Reviews, 19*, 397-406.

Cohn, J., MacPhail, R. C., & Paule, M. G. (1996). Repeated acquisition and the assessment of centrally acting compounds. *Cognitive Brain Research, 3,* 183-191.

Crawford, E., & Olson, J. (1990). Precision teaching with the physically impaired: They can chart too! *Journal of Precision Teaching, 7,* 4-8.

Deary, I. J., Liewald, D., & Nissan, J. (2011). A free, easy-to-use, computer-based simple and four-choice reaction time programme: The Deary-Liewald reaction time task. *Behavior Research and Therapy,43,* 258-268.

Dougherty, K. M. & Johnston, J. M. (1996). Overlearning, fluency, and automaticity. *The Behavior Analyst, 19,* 289-292.

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

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77,* 615-622.

Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology*. New York: Columbia University.

Evans, S. S., Merger, C. D., & Evans, W. H. (1983). The relationship of frequency to subsequent skill acquisition. *Journal of Precision Teaching, 4,* 28-34.

Evans, S. S., & Evans, W. H. (1985). Frequencies that ensure skill competency. *Journal of Precision Teaching, 6,* 25-30.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*, 532-538.

Fields, A. (2005). *Discovering statistics using SPSS* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Fox, E. J., & Ghezzi, P. M. (2003). Effects of computer-based fluency training on concept formation. *Journal of Behavioral Education, 12,* 1-21.

Galizio, M., McKinney, P., Cerutti, D. T., & Pitts, R. C. (2009). Effects of MDMA, methamphetamine and methylphenidate on repeated acquisition and performance in rats. *Pharmacology, Biochemistry, and Behavior, 94,* 305-311.

Gallagher, E. (2006). Improving a mathematical key skill using precision teaching. *Irish Educational Studies, 25,* 303-319.

Gibbon, J., Baldock, M. D., Locurto, C. M., Gold, L., & Terrace, H. S. (1977). Trial and intertrial durations in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes, 3,* 264-284.

Gillespie, M. (2002). The effect of overlearning on the acquisition and retention of the overhand throw for children with and without mental retardation. *Clinical Kinesiology. 56,* 25-31.

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

Haring, N. G., & Liberty, K. (1978, October). *What do I do when? Stages of learning and facilitating instructional events*. Presented at a meeting of the American Association for the Education of the Severely and Profoundly Handicapped, Baltimore, MD.

Haughton, E. (1972). Aims: Growing and sharing. In J. B. Jordan & L. S. Robbins

(Eds), *Let's try doing something else kind of thing* (pp. 20-39). Arlington, Virginia: Council on Exceptional Children.

Higgins, S. T, Woodward, B. M, Henningfield, J. E. (1989). Effects of atropine on the repeated acquisition and performance of response sequences in humans. *Journal of the Experimental Analysis of Behavior, 51*, 5-15.

Holding, E., Bray, M. A., Kehle, T. J. (2011). Does speed matter? A comparison of the effectiveness of fluency and discrete trial training for teaching noun labels to children with autism. *Psychology in the Schools. Volume 48*, 87–213.

Howell, K. W., & Lorson-Howell, K. A. (1990). What's the hurry?: Fluency in the classroom. *Teaching Exceptional Children, 22*, 20-23.

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

Hull, C. L. (1952). *A Behavior System*. New Haven: Yale Univ. Press.

Hursh, S. R. (1977). The conditioned reinforcement of repeated acquisition. *Journal of the Experimental Analysis of Behavior, 27*, 315-326.

Ivarie, J. J. (1986). Effects of proficiency rates on later performance of a recall and writing behavior. *RASE: Remedial and Special Education, 7*, 25–30.

Johnson, K., & Layng, T. V. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist, 47*, 1475-1490.

Johnson, K. & Street, E. M. (2014). Precision teaching: The legacy of Ogden Lindsley. In F. K. McSweeney & E. S. Murphy (Eds.), *The wiley-blackwell handbook of operant and classical conditioning* (pp. 581-609). Oxford, England: Wiley Blackwell.

Johnston, J. M. & Pennypacker, H. S. (1993). *Strategies and Tactics of Behavioral Research*. New York: Routledge.

Kelleher, R. T. (1966). Chaining and conditioned reinforcement. In W. K. Honig (Ed.), *Operant behaviour: Areas of research and application* (pp. 160-212). New York: Appleton-Century-Crofts.

Keller, F. S. & Schoenfeld, W. N. (1950). *Principles of Psychology*. Acton, Massachusetts: Copley.

Kessissoglou, S., & Farrell, P. (1992). Whatever happened to Precision Teaching.

British Journal of Special Education, 22, 60-63.

Kirk, R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurements, 56*, 746-759.

Kubina, R. M. Aho, D., Mozzoni, M. P., & Malanga, P. (1998). A case-study in re-teaching a traumatically brain injured child handwriting skills. *Journal of Precision Teaching and Celeration, 15*, 32-40.

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

Kubina, R., Ward, M. C., & Mozzoni, M. P. (2000). Helping one person at a time: Precision Teaching and traumatic brain injury rehabilitation. *Behavioral Interventions, 15*, 189-203.

Kubina, R., & Wolfe, P. (2005). Potential applications of behavioral fluency for students with autism. *Exceptionality, 13*, 35-44.

Kubina, R., Young, A., & Kilwein, M. (2004). Examining an effect of fluency: Application of letter sound writing and oral word segmentation to spelling words. *Learning Disabilities, 13*, 17-23.

LaBerge, D. & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.

Liberty, K. A., & Paeth, M. A. (1990). Self-recording for students with severe and multiple handicaps. Teaching *Exceptional Children, 22*, 73-75.

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

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

Lindsley, O. R. (1991). Precision Teaching's unique legacy from B. F. Skinner. *Journal of Behavioral Education, 1*, 253-266.

Locke, E.A., & Latham, G.P. (1990). *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice-Hall.

Logan, G. D. (1978). Attention in character-classification tasks: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General, 107,* 32-63.

Malabello, G. (1998). IMAGINE: The first six months. *Journal of Precision Teaching and Celeration, 15,* 41-49.

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

McDowell, C., & Keenan, M. (2002). Comparison of two teaching structures examining the effects of component fluency on the performance of related skills. *Journal of Precision Teaching and Celeration, 18,* 16-29

McNeir, G (1993). Outcome-based education: Tools for restructuring. *Oregon School Study Council Bulletin 36,* 1-29.

National Center for Education Statistics, U.S. Department of Education. (2012). The condition of education. Retrieved March 23, 2011, from the National Center for Education Statistics Web Site: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2012045

Nevin, J. A., & Grace, R. C. (2005). Resistance to extinction in the steady state and in transition. *Journal of Experimental Psychology: Animal Behavior Processes, 31*, 199-212.

Odum, A. L., Shahan, T. A., & Nevin, J. A. (2005). Resistant to change of forgetting functions and response rates. *Journal of the Experimental Analysis of Behavior, 84*, 65-75.

Palmer, D.C. & Donahoe, J.W. (1991). Shared premises, different conclusions. *The Behavior Analyst*, *14*, 123-127.

Picker, M., & Poling, A. (1984). Effects of anticonvulsants on learning: Performance of pigeons under a repeated acquisition procedure when exposed to phenobarbital, clonazepam, valproic acid, ethosuximide, and phenytoin. *Journal of Drugs and Experimental Therapeutics, 230,* 307-316.

Porritt, M. (2007). Fluency in pigeons: The effects of response rate on learning (Doctoral dissertation). Retrieved from PubMed, ISBN # 054903367X, 9780549033677.

Porritt, M., Wagner, K., & Poling, A. (2009). Effects of response spacing on acquistion and retention of conditional discriminations. *Journal of Applied Behaviour Analysis, 42*, 295-307.

Postman, L. (1962). Retention as a function of degree of overlearning. *Science, 135*, 666-667.

Roberts, W. A. (1972). Short-term memory in the pigeon. *Journal of Experimental Psychology, 94*, 74-83.

Roberts, W. A., & Grant, D.S. (1976). Studies in short-term memory in the pigeon using the delayed matching-to-sample procedure. In D. L. Medin, W.A. Roberts & R. T. Davis (Eds.), *Processes of animal memory* (pp. 79-112). Hillsdale, NJ: Erlbaum.

Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19*, 361-374.

Seijts, G.H. & Latham, G.P. (2001). The effect of distal learning, outcome, and proximal goals on a moderately complex task. *Journal of Organizational Behavior, 22*, 291-307.

Shimamune, S., & Jitsumori, M. (1999). Effects of grammar instruction and fluency training on the learning of *the* and *a* by native speakers of japanese. *The Analysis of Verbal Behavior, 16*, 3-16.

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.

Sidman, M. & Rosenberger, P. B. (1967). Several methods for teaching serial position sequences to monkeys. *Journal of the Experimental Analysis of Behavior, 10*, 467-478.

Skinner, B. F. (1938). *The Behavior of Organisms*. New York: Appleton-Century-Crofts.

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

Snodgrass, S. H. & McMillan, D. E. (1989). Repeated acquisition of behavioural chaines: Response sequences or conditional discriminations? *Journal of the Experimental Analysis of Behavior, 51*, 233-241.

Stebbins, L.B., St. Pierre, R.G., Proper, E.C., Anderson, R.B., & Cerva, T.R.

(1977). *Education as Experimentation: A Planned Variation Model* (Vol. IV-A). Cambridge, MA: Abt Associates.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57*, 421-457.

Sterne, J. A C & Smith, G. D. (2001). Shifting the evidence – what's wrong with significance tests? *British Medical Journal, 322*, 226-231.

The National Commission on Excellence in Education (1983). *A Nation at Risk: The Imperative for Educational Reform*. U. S. Government Printing Office, Washington, DC.

Thompson, D. M. (1970). Repeated acquisition as a behavioural baseline. *Psychonomic Science, 21*, 156-157.

Turkkan, J. S., & Hienz, R. D. (1992). Behavioral performance effects of verapamil in normotensive and renovascular hypertensive baboons. *Integrative Physiological and Behavioral Science, 27,* 142-150.

Vaughan, M. E. (1985). Repeated acquisition in the analysis of rule-governed behavior. *Journal of the Experimental Analysis of Behavior, 44,* 175-184.

Walker, M K. (1981). Effects of dextroamphetamine sulphate on repeated acquisition behavior and mood in humans: A preliminary report. *The Psychological Record, 31*, 29-41.

Watanabe, S. (2001). Effects of hippocampal lesions on repeated acquisition of spatial discrimination in pigeons. *Behavioural Brain Research, 120,* 59-66.

Weiss, K. (1978). A comparison of forward and backward procedures for the acquisition of response chains in humans. *Journal of the Experimental Analysis of Behavior, 29,* 255-259.

Wheetley, B. (2005). The effects of rate of response on retention, endurance, stability, and application performance on a match-to-sample task. Unpublished master's thesis, University of North Texas, Denton.

White, G. W. (1985). Characteristics of forgetting functions in delayed matching to sample. *Journal of the Experimental Analysis of Behavior, 44*, 15-34.

White, G. W. (2001). Forgetting functions. *Animal Learning and Behavior, 29,* 193-207.

White, G. W., & Wixted, J. T. (1999). Psychophysics of remembering. *Journal of the Experimental Analysis of Behavior, 71*, 91-113.

White, O. R. (1986). Precision Teaching – Precision Learning. *Exceptional Children, 52*, 522-534.

White, O. R., & Haring, N. G. (1976). *Exceptional teaching*. Columbus, Ohio: Charles E. Merrill. Wood, S., Burke, L., Kunzelmann, H., & Koenig, C. (1978). Functional criteria in basic math skill proficiency. *Journal of Special Education Technology, 2*, 29-36.

Wilkinson, L. & Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594-604.

Wright, L. K. M. & Paule, M. G. (2007). Response sequence difficulty in an incremental repeated acquisition (learning) procedure. *Behavioural Processes, 75*, 81-84.

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*, 73-86.