LEARNING WITH SELF-CONTROLLED AND PERFORMANCE-BASED FEEDBACK

THE EFFECTS OF PERFORMANCE-BASED AND SELF-CONTROLLED FEEDBACK SCHEDULES ON MOTOR LEARNING

By

JUSTIN G BARNEY, H.B.KIN.

A thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

© Copyright by Justin Barney, August 2007

MASTER OF SCIENCE (2007) McMaster University

(Kinesiology) Hamilton, Ontario

TITLE: The Effects of Performance-Based and Self-Controlled Feedback Schedules on Motor Learning

AUTHOR: Justin G Barney, H.B.Kin. (McMaster University)

SUPERVISOR: Professor Timothy D Lee, Ph.D.

NUMBER OF PAGES: viii, 85

Abstract

This study examined the effects of self-controlled and performance-based feedback schedules on the acquisition and retention of a novel motor task. In Experiment 1 participants performed an interception task on a computer using a mouse-controlled cursor. The goal of the task was to intercept the image of a red circle as it passed through a designated area. Each trial received a score based on the speed and accuracy of the interception movement. Participants were randomly assigned to three feedback groups: Best-trial feedback, Worst-trial feedback, and Self-controlled feedback. No differences were found between groups in acquisition, however analysis of no-feedback retention and transfer tests indicated that the Worst-trial group showed the most significant improvements in performance. Experiment 2 examined the potential mechanisms contributing to the advantages of a worst-trial feedback schedule. Participants in the second experiment performed the same interception task utilized in Experiment 1 under two novel feedback conditions: Estimation feedback and Immediate feedback. These new groups were compared to the Worst-trial group from Experiment 1. Analysis of no-feedback retention and transfer tests again indicated that the Worst-trial group showed the most significant improvements in performance. These results suggest that self-controlled schedules may not be ideal when feedback is based on performance; instead, specific error information for the least successful trials appear to be most beneficial, especially when individuals have knowledge of results regarding previous attempts at the task.

Acknowledgements

Interesting that one of the first pages to appear in this thesis is the very last thing I'm writing (and maybe one of the most important). Firstly, I would like to thank my supervisor, Dr. Tim Lee for all his support and guidance throughout this process. He is the reason I began grad school and has been a fantastic mentor for me throughout these two years. I would also like to thank the members of my committee, Dig, Jim and Jae, for the time and effort they put into my work and for helping me to get the most out of this entire experience. In addition, I would like to thank my family for their support throughout this process and for dealing with me during the best and worst times. Of course, the MSc process is about both academics and the relationships you make on the way, and I have had the opportunity to work with some of the most intelligent and incredible individuals here at Mac. I would like to thank all of the members of our lab group over these two years for all their encouragement and support: Lawrence, Hansen, Passmore, D. Gonz, Cheryl, Claudia, Courtney, and Thea. However, a very special acknowledgement goes out to Katie, who has not only been an important role model for me, but also a great friend-Thank you for the laughter and the memories. Finally, I would like to thank Heidi, who has done more for me than anyone I have ever known. Your encouragement and understanding throughout this process has helped me to enjoy the good times, get through the hard times, and still come out smiling. I only hope that I can one day do as much for you as you have done for me.

Table of Contents

List of Tables

List of Figures

General Introduction

Consider a coach who watches his gymnast perform a complex series of flips and tumbles. At the end of the routine the coach provides advice to improve the movements of the athlete and (hopefully) the performance as a whole. Now consider an experienced surgeon who observes her medical student perform a mock open-heart operation. After the procedure the surgeon will also provide information regarding performance, though it will likely deal with millimeters of hand movement rather than the whole body information delivered by the gymnastics coach. In both cases, the experienced observer makes an effort to provide beneficial, post-performance information that will clarify the errors being committed and allow the learner to develop error-detection abilities in order to improve subsequent attempts at the task. In motor learning this information is typically referred to as augmented feedback.

The role of augmented feedback in the learning process has been examined extensively over the last five decades, and has evolved dramatically throughout this time. Though augmented feedback is typically defined as post-performance information pertaining to the execution of a task, two distinct types of feedback have been identified: knowledge of results (KR) and knowledge of performance (KP). KR refers to information related directly to the environmental outcome of a movement (e.g. the distance of a golf shot) (Schmidt & Lee, 2005). KP refers to information regarding the structure and kinematics of the movement pattern (e.g. speed of the golf swing or position of the hands at ball strike)

(Schmidt & Lee, 1999). Studies examining augmented feedback have manipulated a variety of different types and schedules of feedback provision in order to facilitate the learning process.

Early Studies of Augmented Feedback

Thorndike (1927) was among the first researchers to examine the role of augmented feedback in motor learning. He examined the ability of individuals to draw lines of varying lengths with and without feedback. He reported that individuals practicing the line drawing with feedback produced fewer errors in retention tests than those who practiced without feedback. This was an interesting finding given the fact that Thorndike provided only the most basic form of KR, informing individuals of whether they were right or wrong. Though logistically simple, this study provided early evidence of the advantages of augmented feedback.

Research continued to build upon the findings of Thorndike (1927). One of the areas of research receiving a great deal of attention was concerned with identifying the optimal frequency of feedback provision. In one specific study, Smode (1958) utilized a continuous tracking task to examine whether high frequency or low frequency feedback schedules resulted in superior skill acquisition. His results indicated that participants receiving feedback at an increased frequency demonstrated superior tracking ability versus those who received a low frequency feedback schedule. Additionally, through the implementation of a transfer task in which the frequency of feedback provision was switched between groups, the author was able to show that performance could be improved in the low frequency group if the feedback schedule was changed from low to high. Smode (1958) argued that increased rates of feedback provision benefited motor learning because they elevated the motivation of the learner to perform the movement well.

These results were mirrored in studies examining discrete motor tasks. Bilodeau and Bilodeau (1958) used a lever-pulling task to examine various feedback frequency schedules. In this study the number of trials with feedback was held constant for all participants, while the number of trials without feedback was varied between groups. The authors reported no differences in performance across groups, despite the fact that some groups had far more practice trials than others. They concluded that the most ideal feedback schedule should deliver information about task performance after every trial.

The results from these two studies illustrate similar advantages associated with the delivery of feedback after every practice trial. However, more recent research indicates that an individual's long-term learning is only revealed through the implementation of delayed retention and transfer tests (Salmoni, Schmidt, & Walter, 1984; Schmidt & Bjork, 1992). These delayed performance tests, which are typically administered without the inclusion of the experimental manipulation, are used to separate temporary performance enhancements that may be present in acquisition from sustained learning effects. Unfortunately neither Bilodeau and Bilodeau (1958), nor Smode (1958) employed delayed performance tests in their experimental protocols. Therefore, we can only conclude that high frequency feedback schedules are beneficial during the acquisition stage of learning. Further research utilizing delayed performance tests would be required to make definitive statements regarding the long-term learning benefits of high frequency feedback schedules.

Recent Studies of Augmented Feedback: Learning Effects of KR and KP

More recent research examining augmented feedback has incorporated these delayed performance tests to assess the long-term learning benefits of various feedback schedules. These studies have consistently shown that less feedback during acquisition actually facilitates long-term retention of motor skills (Salmoni et al., 1984; Schmidt, Young, Swinnen, & Shapiro, 1989; Viitasalo, Era, Konttinen, Mononen, Mononen, Norvapalo, 2001; Winstein & Schmidt, 1990), which is opposite to the conclusions that may have been drawn from Smode (1958) and Bilodeau and Bilodeau (1958). For example, Winstein and Schmidt (1990) compared the effects of a 100% KR schedule (feedback delivered after every trial in acquisition) with a 50%, faded KR schedule (feedback gradually decreased throughout acquisition) in the performance of a sequential timing task. The authors reported fewer timing errors for the 50% KR group in both no-KR retention tests (Experiment 2) and 100%-KR retention tests (Experiment 3) performed 24 hours after acquisition.

Similar findings have been noted in the literature examining KP. Weeks and Kordus (1998) found that participants in a 33% KP group completed a soccer throw-in skill with superior form than those receiving KP after every trial in acquisition. In addition, Young and Schmidt (1992) found that providing average KP (that is, average feedback scores for a set of trials) rather than 100% KP also resulted in superior delayed retention performance.

These learning benefits are not limited to the frequency of feedback provision, but have also been applied to the timing of feedback delivery. Research suggests that delaying the delivery of KR after trial completion facilitates motor learning when compared with providing KR immediately after each trial (Swinnen, Schmidt, Nicholson, Shapiro, 1990). In addition, providing KR in a summary format is also beneficial (Schmidt et al., 1989). In the summary KR condition participants complete a series of trials, and then receive feedback for each trial in the series at the same time (Schmidt et al., 1989). This results in a delay before feedback delivery for all but one trial (the last trial) in each series.

Though the studies discussed thus far represent a broad spectrum of the augmented feedback research, their findings can be summarized effectively using the Guidance Hypothesis outlined by Salmoni et al. (1984). This hypothesis suggests that augmented feedback, in either KR or KP form, is such a salient type of information that it has the informational capacity to guide an individual through the learning process. During the very early stages of learning, this guidance may be critical to the development of correct movement patterns, however high levels of feedback throughout acquisition may result in the development of a reliance on this information (Winstein & Schmidt, 1990). Thus, when feedback is removed in retention tests the learner no longer has this information to guide their movements and performance suffers (Salmoni et al., 1984).

The actual processes disrupted by the guiding effects of instantaneous or high frequency feedback are still under debate. Some authors have suggested that these schedules reduce the cognitive effort (defined as the mental processing conducted to evaluate performance and movement mechanics) expended by the learner, resulting in a diminished understanding of movement characteristics (Weeks & Kordus, 1998). Others have speculated that high feedback frequencies hinder the development of error-detection processes, resulting in inferior performance when feedback is removed (Swinnen et al., 1990). Yet another suggestion proposes that high frequencies of feedback disrupt the ability of the learner to accurately select movement parameters involved in the task, such as force or speed of movement, resulting in greater variability in performance from trial to trial and a negative impact on the acquisition of the motor skill as a whole (Schmidt et al., 1989). Regardless of the reasoning, this literature clearly indicates that providing feedback less frequently during acquisition and that delaying the onset of this information after task completion both facilitate the learning process.

Unfortunately, attempts to outline the processes enhanced by low frequency, delayed feedback schedules have resulted in many of the same debates, though there is more cohesion in this area. Specifically, it seems that low frequency and delayed feedback schedules encourage individuals to evaluate every trial attempt more thoroughly, enhancing their understanding of overall movement mechanics, and improving movement execution on subsequent trials (Schmidt et al., 1989, Winstein & Schmidt, 1990). Though this explanation fails to outline the exact mechanisms affected, it does provide a general overview of the benefits associated with less frequent feedback delivery.

Overall this research suggests that low frequency, delayed feedback schedules are sufficient to facilitate the acquisition of novel motor skills. However, this is a very simplistic view of the link between augmented feedback and motor learning, and leaves many questions unanswered. It is clear that providing no feedback is detrimental to the learning process (Viitasalo et al., 2001), but how frequently (or rather, infrequently) does feedback need to be delivered in order to enhance learning, while still avoiding the negative guidance effects associated with high frequency feedback schedules? Furthermore, how do individual differences impact these findings? Is it realistic to assume that an optimal level of feedback exists for all individuals, or should between-learner variations be considered? Answers to some of these questions may lie in the examination of self-controlled feedback schedules.

Self-Controlled Augmented Feedback and Motor Learning

The commonality across all of the paradigms and studies discussed thus far has been the implementation of *experimenter-imposed* feedback schedules; experimental treatments that have been pre-determined by the investigator based on the variables of research interest. Intuitively this process seems to make sense, and is mirrored in many real-world situations. For example, when a coach provides a player with feedback regarding their performance on a task, it is not customary to check if they want it first; the coach simply expects the player to incorporate their suggestions into the movement. However, this begs the question of whether the coach (or experimenter) knows when and what feedback would provide the maximum benefit to the learner.

To examine the impact of providing more freedom to the learner, researchers have introduced self-controlled feedback schedules into their experimental protocols. In a self-controlled condition participants are allowed to receive feedback at a frequency they select during acquisition; some may choose to receive feedback often while others may select a lower frequency. These conditions are then typically compared to various experimenter-imposed feedback schedules to determine which manipulations result in superior retention performance.

Janelle, Kim, and Singer (1995) conducted the first study examining a self-controlled feedback condition. These authors used a ball-tossing task to compare 5 different KP schedules: no KP, 50% KP, 5-trial summary KP, self-controlled KP, and a yoked control schedule (participants in this final group were matched to a partner in the self-controlled group and received feedback after the same trials). Results indicated that the self-controlled group performed more accurately in movement form than all other groups in an immediate retention test. However, the authors failed to employ delayed retention tests, meaning that no conclusions could be formulated to differentiate between acquisition *performance* and long-term *learning*.

Other experimenters have since expanded upon this area of research using delayed performance tests. While some have not found benefits of self-controlled feedback schedules compared to 100% KR schedules (Wrisberg & Pein, 2002), other research indicates that a self-controlled condition facilitates the learning process in both laboratory skills such as ball-tossing (Janelle, Barba, Frehlich, Tennant, & Cauraugh, 1997), and sporting skills, such as basketball free-throw shooting (Wulf, Raupach, & Pfeiffer, 2005). Self-controlled manipulations have been further extended outside of the augmented feedback literature, showing benefits in practice trial scheduling (Keetch & Lee, in press), and in the familiarization with physical aids on a ski simulator (Wulf & Toole, 1999).

In an interesting study involving dyad practice, Wulf, Clauss, Shea, and Whitacre (2001) examined whether a self-controlled condition would produce more optimal movement patterns than a yoked control condition on a ski simulator if participants in both groups were allowed to watch each other perform. In this protocol participants in the self-controlled group were allowed to use ski poles to aid their performance whenever they chose, with the yoked group matched to their selections. The authors hypothesized that visual aspects of the performance (e.g. gross movements) would be similar across groups, but that intrinsic, non-observable movements (e.g. forces exerted at the feet) would vary. The results supported these predictions, with the self-controlled group demonstrating more optimal intrinsic movement patterns in delayed retention tests.

Interestingly, self-controlled studies consistently indicate that individuals prefer to receive feedback at very low rates (7% average in Janelle et al., 1995 and 5.8% average in Wulf et al., 2001). In addition, Janelle et al. (1995, 1997) report that individuals tend to select feedback in a faded pattern, requesting performance information more frequently early in practice and less frequently as practice continues. These tendencies seem to support the suggestions forwarded by Winstein and Schmidt (1990), where feedback is more critical early in practice to develop correct movement strategies, but can result in dependency if high frequencies of feedback continue throughout practice.

There have been a number of explanations proposed to account for the reported benefits of self-controlled schedules. Janelle et al. (1997) suggested that a self-controlled schedule leads to deeper processing of task demands and increased cognitive effort during acquisition, which in turn results in an enhanced understanding of movement parameters, such as timing or force production in the limbs. Thus, when feedback is removed individuals with this enhanced knowledge perform more successfully. Wulf et al. (2005) proposed an alternative explanation. These authors suggested that individuals in a self-controlled condition would extract more useful information from the feedback when they were allowed to determine when it was delivered. Specifically, individuals would request feedback when they believe it contained certain information that would enhance their future performance, and would utilize this information to make useful adaptations to their movements.

A third explanation forwarded by Chiviacowsky and Wulf (2002) has received the most attention in the self-controlled feedback literature. These authors suggest that self-controlled schedules are beneficial to learning because they allow learners to receive feedback exactly when it is most beneficial to them. More specifically, they suggest that individuals prefer, and benefit most, from feedback delivered after good trials. This explanation is closely linked with that proposed by Wulf et al. (2005), as both hypotheses suggest that individuals extract more relevant information when they have the ability to select feedback in a self-controlled manner.

To test this explanation, Chiviacowsky and Wulf (2005) examined two self-controlled groups that were asked to learn a sequential timing task. Both groups received movement time feedback when it was requested, however one group made the decision to receive feedback *before* each trial began while the other group made the decision *after* the trial was completed. Chiviacowsky and Wulf (2005) argued that if the explanation proposed by Janelle et al. (1997) was correct then both groups should perform equally well. However, if participants do in fact prefer, and benefit more from feedback for their best trials, then the group requesting feedback *after* trial completion should ask for feedback regarding their most successful trials and produce smaller timing errors overall. The authors reported findings consistent with the latter hypothesis, lending support to the explanation proposed by Chiviacowsky and Wulf (2002).

What remains unclear from the Chiviacowsky and Wulf (2002) study is whether the results were due to the effects of providing feedback after good trials (the effect of feedback when based on trial performance), or the *decision* to receive feedback after good trials (the self-control factor). While the self-controlled literature indicates that allowing individuals input into the learning process can be beneficial, common coaching practices often base the provision of feedback solely on performance, without consideration of an individual's preferences. This raises an important question: Are performance-determined feedback schedules more appropriate or advantageous than self-controlled schedules? Though these two conditions have yet to be examined together, literature on performance-determined feedback schedules does exist, and has shown some interesting results.

Performance-Determined Augmented Feedback

When one discusses the outcome of a task, it is usually done in the context of whether it was performed correctly or incorrectly. Though there may be varying levels of "correctness", generally one knows if the outcome was a desired one or if it could be considered a failure. However, the information one gains from these types of trials may be very different, and is likely to have distinct effects on learning. Research examining performance-determined feedback schedules has primarily focused on whether information regarding successful or unsuccessful trials is more advantageous to a learner.

Researchers examining verbal learning and memory have completed some interesting work in this area. In a study asking participants to learn novel word pairings, Pashler, Cepeda, Wixted and Rohrer (2005) found that individuals given the correct answer after incorrect trials recalled more word pairings than those who received no feedback and those who were only told if they were correct or incorrect. Guthrie (1971) found similar learning benefits associated with feedback following error trials in a task where participants were asked to identify missing words in various sentences. These studies both indicate that error information is critical in verbal memory tasks.

The motor learning literature also supports the benefits of providing feedback after error trials. Williams and Briggs (1962) employed a continuous, sinusoidal tracking task to examine the differences between providing feedback when an individual was on-target (performing correctly) versus when they were off-target (performing incorrectly). They found significant performance gains in no-feedback retention tests for those who received error (off-target) information during acquisition. Similarly, Lincoln (1954) found that providing participants with post-performance error information was more advantageous than delivering concurrent visual feedback when judging distances traveled using a manual rotary device.

These findings have been mirrored in non-laboratory settings. Goldstein and Rittenhouse (1954) employed a gunnery task to examine whether on-target, auditory feedback was more valuable than post-performance error information. The authors reported significant advantages in no-feedback retention tests for individuals receiving error information. Furthermore, the authors showed that error information during training prevented performance decay throughout continued trials without feedback (Experiment 4) and when transferring the skill to novel equipment (Experiment 5).

More recent studies employing bandwidth feedback schedules also lend support to this area. In these experimental designs, participants receive feedback when performance falls outside a predetermined range or bandwidth, resulting in a schedule in which specific feedback is provided after poor trials and qualitative feedback is provided after good trials (Sherwood, 1988). Sherwood (1988) compared two bandwidth feedback schedules with a 100% KR schedule in the performance of a timing task. He reported fewer timing errors in the bandwidth groups, which suggests that the error information delivered by the bandwidth schedules does enhance learning.

However, Sherwood (1988) also reported that participants in the larger bandwidth group, who consequently received less feedback, showed fewer timing errors than those in the smaller bandwidth group (though not significant). Thus, it cannot be said for certain whether error information was the critical factor or if simply receiving less feedback was key (see Winstein & Schmidt, 1990). To test these two possibilities, Lee and Carnahan (1990a) examined two bandwidth groups and two yoked control groups (who received feedback after the same trials as those in the bandwidth groups). Similarly to Sherwood (1988) these authors reported advantages in the two groups receiving less feedback. However they also reported less variable error in the bandwidth groups. This suggests that simply receiving less feedback is not the sole contributor to improved performance, but that error information associated with bandwidth feedback schedules is also critical.

These conclusions are further strengthened by research exploring reversed bandwidth feedback schedules (Cauraugh, Chen, & Radlo, 1993). As the name implies, a reversed bandwidth schedule provides detailed feedback when an individual's performance is within a predetermined range or bandwidth, and only qualitative information if performance falls outside of this range (Cauraugh et al., 1993). This is the opposite of traditional bandwidth manipulations. As a result, individuals no longer receive specific, detailed error information when performance is poor. Instead, they receive only qualitative feedback, indicating that movement patterns need to be altered to improve performance, without delivering specific information regarding where these improvements are required.

Cauraugh et al. (1993) examined traditional and reversed bandwidth schedules in a sequential timing task, and found that both manipulations resulted in superior learning compared to yoked control groups. One may argue that these results indicate that error information is not critical to learning, but that a consistent bandwidth manipulation is key. However, the authors completed a further analysis and found that the traditional bandwidth group showed the most improvement after *quantitative* information, while the reversed bandwidth group showed the most improvement after *qualitative* information. Thus, both groups showed the greatest benefits from their respective error information, even if this information was not identical.

All of the studies outlined in this section have illustrated the apparent benefits of error information in learning. However, no study has examined the reasons *why* this information seems to be so critical in the learning process. One explanation may be that individuals simply utilize this information to make changes in actions where performance is deficient. However, this would not explain the benefits associated with reversed bandwidth schedules where explicit feedback is not provided after poor trials (Cauraugh et al., 1993). Alternatively, it may be that error information encourages greater cognitive effort from the learner during acquisition, which may contribute to the reported benefits. Studies examining the reasons for these observed advantages could provide insights into the beneficial aspects associated with error information.

Experiment 1 Introduction

Augmented feedback is a critical variable influencing the learning process, whether it provides information about the outcome of a movement (KR) (Schmidt et al., 1989) or knowledge regarding specific movement parameters (KP) (Weeks & Kordus, 1998). Though early studies suggested that providing feedback after each attempt at a task would result in superior performance (Bilodeau & Bilodeau, 1958), more recent work has shown that less feedback actually facilitates sustained learning (Schmidt et al., 1989; Winstein & Schmidt, 1990). Schmidt et al. (1989) suggested that providing feedback after every trial during acquisition causes the learner to develop a dependence on this information, which then hinders performance when feedback is removed (i.e. no-feedback retention tests).

More recently, a number of studies have reported positive learning benefits associated with transferring control of feedback delivery to the learner (Chiviacowsky & Wulf, 2002; Janelle et al., 1995, 1997). In these studies the frequency of feedback delivery is not experimentally imposed, but instead participants are given power over when and how often they receive information about their performance. These self-controlled feedback schedules have resulted in superior learning compared to a variety of experimenter-imposed schedules, including 50% KP schedules (Janelle et al., 1995), summary feedback schedules (Janelle et al., 1997), and yoked controls (Chiviacowsky & Wulf, 2002).

A number of hypotheses have been forwarded to account for the benefits of self-controlled feedback schedules, however an explanation proposed by Chiviacowsky and Wulf (2002) has received the most attention. These authors suggested that self-controlled schedules are beneficial because they allow individuals to receive feedback when they think that they need it most. Specifically, they argued that the optimal time for feedback delivery is following a good performance, as this information is most useful and motivating to the learner. However, this explanation fails to specify whether providing feedback about good trials is the critical factor enhancing learning, or whether the *choice* to receive feedback is key. To date, no study has directly compared a self-controlled feedback schedule and a schedule that delivers feedback based on the outcome of a trial (i.e. a performance-determined feedback schedule) to uncover whether these schedules yield differential learning effects. Such an examination would test the explanation of self-controlled feedback benefits proposed by Chiviacowsky and Wulf (2002).

The purpose of this study was to directly examine whether individuals in a self-controlled schedule would request feedback for their best trials during acquisition, and how their performance on a novel motor task compared to groups receiving feedback on a performance-based schedule. Specifically, participants received feedback for either their best trials, their worst trials, or for trials of their choosing (the self-controlled group). If the predictions from Chiviacowsky and Wulf (2002) are accurate then individuals in the self-controlled group should request feedback for their best trials most often, and should perform equally well in acquisition and retention when compared to those receiving feedback on a best-trial schedule. Furthermore, this explanation would predict the poorest performance to belong to the worst-trial group, as this type of feedback schedule is less motivating and provides the least relevant information for task success.

Experiment 1 Methods

Participants

Thirty volunteers (24 females, 6 males: average age 21.4 ± 2.8 years) from McMaster University participated in this experimental protocol. All participants had normal or corrected-to-normal vision and were naïve regarding the purpose of the study. All but 3 participants were right handed, however all volunteers selected to perform the experimental task using their right hands. Written consent was obtained from all participants before the protocol began.

Apparatus and Task

All volunteers were tested individually using the same equipment. Participants were seated approximately 75 cm from a flat screen computer monitor measuring 40 cm in length and 33 cm in width. Located around the monitor were two computer speakers, a keyboard, and a standard mouse device. Participants were allowed to adjust this equipment for personal comfort and were instructed to perform the experimental task using their preferred hand.

Participants were asked to perform an interception task similar to that outlined by Schmidt and Young (1991). The goal of the task was to intercept the image of a red circle (displayed on the monitor) as it passed through a specified target area (denoted by a green "+"). This interception was completed using an on-screen cursor that responded to movements of the mouse. An outline of this task is illustrated in Figures 1 and 2.

Each experimental trial followed the same structure. A single trial began with the depression of the spacebar, causing the red circle to appear at the top of the monitor, 25 cm from the left edge of the screen. This display also contained the green target area, located 20 cm below the red circle, and a black line, located 7 cm from the left edge of the screen. These images remained visible and stationary for 1500 ms while three auditory tones were presented. At the conclusion of the tones, the red circle began moving in a straight, vertical line down the screen at a velocity of 20 cm/s, with zero acceleration. The completion of the tones also triggered the appearance of the mouse cursor, which was initially positioned within the green target area.

To complete the task correctly, participants were required to move the mouse cursor to the left, beyond the black line, and then back to the right so that it passed through the green target area simultaneously with the red circle. The black line remained stationary throughout all trials, acting to ensure a "backswing" with the mouse, similar to the movements outlined by Schmidt and Young (1991). If a participant failed to move the cursor beyond this line, the trial was deemed an error and was repeated.

Each trial received a total score between 0 and 100 points. This score was calculated through summation of points awarded to three performance variables: cursor speed, horizontal accuracy, and vertical accuracy. Cursor speed accounted for a maximum of 40 points, and was determined by calculating the average velocity of the mouse cursor during the 60 ms that the red circle was in contact with the target area. Horizontal accuracy also accounted for a maximum of 40 points, and was defined as the distance (along the x-axis) from the mouse cursor to the target area when the circle passed through the center of the green "+". Vertical accuracy accounted for a maximum of 20 points[1], and was defined as the distance (along the y-axis) from the mouse cursor to the target area when the circle passed through the center of the green "+".

Feedback was delivered at the conclusion of each 3-trial set. This feedback consisted of the total score (KR) for each of the three trials in the set, followed by the scores for cursor speed, horizontal accuracy and vertical accuracy (KP) for just *one* of the three trials in the set. The trial for which KP was delivered was experimentally manipulated. All participants were provided with clear instructions regarding how scoring was conducted and how to interpret the feedback they received.

Procedure

Participants were randomly assigned to 3 experimental groups (n=10): a Best-trial Feedback group, a Worst-trial Feedback group, and a Self-Controlled Feedback group. These groups differed based on the feedback delivered at the end of each 3-trial set. For the first two groups, the feedback schedules were experimentally imposed. The best-trial group received KP for their best trial in each set, while the worst-trial group received KP for their worst trial. Conversely,

---

[1] Based on pilot work, vertical accuracy produced smaller deviations than the other performance measures. As a result, it comprised a smaller proportion of the total score.

the self-controlled group was allowed to select the trial they wished to receive feedback for in each set.

The experimental protocol was conducted over two days with all groups following the same procedure. On day one participants were familiarized with the equipment and were provided with a standard set of instructions regarding the task, the calculation of their total score, and the scoring of the three performance variables. Participants were informed that their overall goal was to obtain consistently high scores. Following the instructions, participants performed 99 acquisition trials, with feedback delivered after each 3-trial set based on group assignment.

Participants returned approximately 24 hours later for retention and transfer tests. For these tests the participants' goal remained the same (to obtain consistently high scores), however feedback regarding performance was no longer provided. The retention test consisted of 18 trials of the same interception task performed in acquisition. The transfer test consisted of 24 trials in which task characteristics were altered slightly. In these trials, the presentation of the last of the preparatory auditory tones was followed by a random shift in the green target area to one of 4 different locations on the screen relative to its original position: up 3 cm, down 3 cm, left 2 cm, or right 2 cm (6 trials each). All participants received the same randomized ordering of the transfer trials. Both the temporal and spatial accuracy components of the movement were independently altered as a result of these target area perturbations.

Experiment 1 Results

Self-Controlled Trial Selection

During acquisition, each group received detailed feedback about their performance variables for 33 of the ninety-nine trials. For the self-controlled group, this resulted in the choice between receiving KP for their best trial (highest score), their worst trial (lowest score), or their intermediate trial (median score) for each 3-trial set. If this selection were completed at random, one would predict a 33% distribution for each of these trial choices. However, variations in this distribution should occur if individuals adopted an alternate strategy. For example, if participants wished to examine where their performance excelled, they would select feedback for their best trials. Alternatively, if they wished to examine aspects of performance that were weak, they would choose to receive feedback for the worst trials.

The actual distribution of trial selections for the best, worst, and intermediate trials was 66%, 22%, and 12%, respectively. This shift in the distribution indicates that participants in the self-controlled group had a strong tendency to select feedback for their best trials. In fact, only one participant elected to receive feedback for their worst trials more often than their best trials during acquisition. Figure 3 presents the distribution of feedback selections for each block in acquisition. It is clear from these data that the preference to receive KP for the highest scoring trials also remained consistent across acquisition.

Data Analysis

Three dependent measures were examined to assess performance: Total Score, Velocity Score, and a Combined Accuracy Score (sum of the vertical and horizontal accuracy scores). For acquisition, these variables were analyzed using a 3 Group (best-trial, worst-trial, self-controlled) by 11 Block (9 trials in each block) mixed analysis of variance (ANOVA), with repeated measures on the latter factor. For retention, these variables were analyzed using a 3 Group by 2 Block mixed ANOVA. For the transfer tasks, these variables were analyzed with a 3 Group by 4 Transfer Task (Up, Down, Left, Right) mixed ANOVA, with repeated measures on the latter factor. All ANOVA and post hoc tests (using Tukey's HSD) had significance levels set at $p < 0.05$.

Acquisition

Table 1 presents the means and standard deviations of the total score, combined accuracy score, and velocity score across acquisition for all three groups in Experiment 1. The analysis for total score revealed a significant main effect for Block, $F(10,270)=3.6$, $p < 0.001$. Post hoc tests indicated that Blocks 3-8, 10, and 11 were all significantly higher than Block 1. All other blocks were not statistically different. The analysis for the velocity scores also revealed a significant main effect for Block, $F(10,270)=3.29$, $p < 0.001$. Post hoc tests indicated that Block 6 was significantly higher than Blocks 1 and 2, and that Blocks 8 and 10 were also significantly higher than Block 2. All other blocks were not statistically different. The analysis for the combined accuracy scores revealed a significant main effect for Block, *F*(10,270)=2.37, $p < 0.02$. Post hoc tests indicated that Blocks 3-5, and 10 were all significantly higher than Block 1. All other blocks were not statistically different. No other effects or interactions were significant in these three analyses. Figure 4 illustrates the non-significant Group by Block interaction for the total score across acquisition.

Retention

Table 2 presents the means and standard deviations of the total score, combined accuracy score, and velocity score in the retention and transfer tests for all three groups in Experiment 1. The analysis for total score revealed a significant Group x Block interaction, *F*(2,27)=4.77, $p < 0.02$, which is illustrated in Figure 5. Post hoc tests indicated that all three groups had equivalent scores in Block 1, however in Block 2 the worst-trial group produced significantly higher scores than the best-trial group, with the self-controlled group intermediate and not significantly different from either group. The analysis for the combined accuracy scores also revealed a significant Group x Block interaction, *F*(2,27)=5.56, $p < 0.01$, which is illustrated in Figure 6. Post hoc tests indicated that the three groups had equivalent scores in Block 1, however in Block 2 all groups were significantly different from each other, with the worst-trial group producing the highest scores, followed by the self-controlled group and the best-trial group, respectively. No other effects or interactions were significant in these two analyses. No effects or interactions were significant for the analysis of the velocity scores.

Transfer

The analysis for total score revealed a significant main effect for transfer type, $F(3,81)=10.25$, $p<0.001$. Post hoc tests indicated that the Down, Left, and Right shifts all produced significantly higher scores than the Up shift, but were not different from each other. The Group effect also approached conventional levels of significance ($p<0.065$), with the worst-trial group recording the highest scores, followed by the self-controlled and best-trial groups, respectively. The analysis for the velocity scores also revealed a significant main effect for transfer type, $F(3,81)=13.68$, $p<0.001$. Post hoc tests indicated that the Down, Left, and Right shifts were all significantly lower than the Up shift. In addition, the Down shift was also significantly lower than the Left shift. The analysis for the combined accuracy scores revealed a significant main effect for transfer type, $F(3,81)=17.32$, $p<0.001$. Post hoc tests indicated that the Down, Left, and Right shifts all scored significantly higher than the Up shift, but were not different from each other. No other effects or interactions were significant.

As a result of the Group trend noted in the total score analysis, four additional analyses were conducted on the transfer data. Each transfer task was analyzed separately using a 3 Group, between-subject ANOVA. The results of these analyses indicated group differences for the Up ($F(2,27)=3.48$, $p<0.05$) and Down ($F(2,27)=3.48$, $p<0.05$) target area shifts. In both cases post hoc tests revealed that the worst-trial and self-controlled groups scored significantly higher than the best-trial group, but were not different from each other. The analyses for the Left and Right transfer tasks were not significant.

Experiment 1 Discussion

The results of Experiment 1 provide an interesting addition to the self-controlled literature, as it is the first study to compare a self-controlled feedback schedule with performance-based schedules. The purpose was to discover whether individuals in a self-controlled condition prefer feedback regarding their best trials, and whether this information results in learning advantages compared to best-trial and worst-trial schedules. In accordance with predictions made by Chiviacowsky and Wulf (2002), we formulated three hypotheses. The first hypothesis predicted that individuals in a self-controlled schedule would select feedback for their best trials more often than worst or intermediate trials. The second hypothesis predicted that self-controlled and best-trial feedback schedules would result in equivalent performance in retention and transfer tests, as these conditions provide similar information to the learner. The third hypothesis predicted that both self-controlled and best-trial feedback schedules would show superior learning advantages compared to a worst-trial schedule, as the latter schedule is less motivating and provides less relevant information to the learner. Examination of feedback selection revealed that the self-controlled group requested feedback for their best performance most often overall (Best=66%, Worst=22%, Intermediate=12%), which is in agreement with the findings reported by Chiviacowsky and Wulf (2002) and the first experimental hypothesis.

However, these preferences did not translate into the predicted performance outcomes. Contrary to the second hypothesis, the self-controlled group showed superior learning advantages compared to the best-trial group, producing higher accuracy scores in retention and higher total scores in two of the four transfer tests. In addition, the self-controlled and best-trial groups failed to produce higher scores than the worst-trial group in any of the dependent measures analyzed, which directly contradicts the third hypothesis. Thus, it appears as though the explanation forwarded by Chiviacowsky and Wulf (2002) to explain the benefits of self-controlled feedback schedules was only partially supported.

There are two factors that may explain why the findings in Experiment 1 contradict those reported by Chiviacowsky and Wulf (2002). The first factor is the provision of KR. In their experimental protocol, Chiviacowsky and Wulf (2002) did not provide any feedback to participants after trial completion unless it was requested. As a result, individuals in the self-controlled group were requesting feedback when they *thought* they were performing well, and may have been conducting more detailed self-evaluations of their movements after each trial in order to make this determination. Conversely, the current study delivered KR for every trial regardless of which trial participants requested detailed feedback for. As a result, participants in the self-controlled group always knew precisely which trial was their best. It may be that the process of evaluating performance after trial completion contributed to the benefits reported by Chiviacowsky and Wulf (2002) in the self-controlled group.

The second factor that could be contributing to these contradictory findings is the ability (or rather, the inability) of individuals to make judgments of their learning. A number of studies have reported that individuals fail to make accurate estimations of their personal levels of learning when performing a novel motor task (Baddeley & Longman, 1978; Simon & Bjork, 2001, 2002). Specifically, these studies have shown that performance in acquisition can influence an individual's overall evaluation of their abilities; good results in acquisition leads to overestimation of learning while poor results leads to underestimation (Simon & Bjork, 2001, 2002). It may be that choosing to receive feedback for the best performance most often caused participants in the self-controlled group to overestimate their abilities, resulting in inferior no-feedback retention performance. This may also explain why the self-controlled group failed to show superior learning advantages compared to the worst-trial group, as individuals receiving feedback for their most unsuccessful trials were likely to underestimate their level of learning in acquisition, and avoid similar performance decrements as the self-controlled group in retention.

The absence of performance benefits in the self-controlled group compared to the worst-trial group is, in itself, an interesting finding. Research examining self-controlled feedback has consistently shown superior performance in retention and transfer tasks under this condition (see Chiviacowsky & Wulf, 2002 and Janelle et al., 1995, 1997). However, these studies have primarily focused on *when* and *how often* to deliver feedback by providing learners with control over the timing and frequency of feedback provision, respectively. The current study differed fundamentally by examining the *type* of feedback that plays the most crucial role in learning, focusing on how the outcome of a trial affects feedback utilization. It may be that certain performance-based feedback schedules are superior to self-controlled schedules when timing and frequency of feedback provision are equated. Specifically, the results of this study indicate that the worst-trial group had the most significant learning advantages in retention, producing higher accuracy scores than all other groups, and higher total scores than the best-trial group.

The primary difference between the worst-trial group and the other two groups was the error information associated with trial feedback; the worst-trial group always received feedback for the most poorly executed trial in each set. This may have provided the worst-trial group with a distinct learning advantage as the delivery of error information has been shown to be beneficial in both verbal memory research (Pashler et al., 2005) and motor learning (Lincoln, 1954; Williams & Briggs, 1962). Though it has been suggested that information regarding poor performances can be less motivating to a learner (Chiviacowsky & Wulf, 2002), it appears that the error information contained in this feedback is somehow beneficial to the learning process.

The results of Experiment 1 add to the early motor learning literature that describes the benefits of error information in the acquisition and retention of novel motor skills (Lincoln, 1954; Williams & Briggs, 1962). It also agrees with more recent research employing bandwidth feedback schedules (Cauraugh et al., 1993; Lee & Carnahan, 1990a). These studies have consistently shown performance benefits associated with detailed feedback delivered following error trials, even when controlling for the frequency and timing of feedback delivery using yoked control groups (Lee & Carnahan, 1990a). However, no study has examined *why* error information is beneficial to motor learning. Experiment 2 explores the reasons for why feedback describing poor performance is so valuable to the learning process.

Experiment 2 Introduction

It is clear from previous motor learning research and the results from Experiment 1 that feedback about the errors committed when performing a novel movement can have significant, positive impacts on the learning process (Cauraugh et al., 1993; Lee & Carnahan, 1990a; Williams & Briggs, 1962). This is good news for coaches and instructors who often utilize error information to guide changes in the performance of their pupils. However, the characteristics inherent to error information that encourage enhancements in learning are not completely clear. An understanding of the mechanisms driving these observed benefits could aid researchers to further enhance the learning process. The purpose of Experiment 2 is to examine the factors that may contribute to the beneficial learning effects of error information.

Based on current literature and the results from Experiment 1 there may be several explanations for the learning advantages associated with the provision of error information. The first possibility is that providing feedback regarding the worst aspects of performance highlights the specific area(s) where the most improvement is necessary (or where it could be made at all). Theoretically, error information would draw the learner's attention to the specific aspects of the movement that were executed poorly and encourage positive changes in the movement mechanics of those areas. Research has shown that novices performing a new skill benefit from focusing on movement mechanics during task execution (Beilock, Carr, MacMahon, & Starkes, 2002; Perkins-Ceccato, Passmore, & Lee, 2003). Conversely, feedback for the best trials may not reveal areas of weakness as this schedule only provides information on successful aspects of performance.

For this explanation to be accurate, the feedback delivered must clearly indicate the characteristics of the movement that are deficient. In Experiment 1, the detailed KP delivered for the single trial at the end of each 3-trial set clearly indicated the relative success for each of the parameters of performance. However, Lee and Carnahan (1990b) found that providing error information for a single aspect of a movement does not benefit motor learning. Specifically, these researchers examined two groups performing a 3-segment timing task: a worst-segment group that received feedback for the most poorly performed section of the movement, and a yoked control group that received feedback for the same section, regardless of how it was performed. They reported no benefits of the specific error information, as the worst-segment group showed no significant movement time improvements compared to the yoked control group. Thus, it appears that further examination of this explanation is required.

A second possibility is that individuals in the worst-trial feedback group were estimating their performance before receiving feedback. Research has shown that estimating trial outcome prior to receiving feedback is beneficial to learning, and can actually counteract the negative effects associated with instantaneous feedback provision (Liu & Wrisberg, 1997; Swinnen et al., 1990). Liu and Wrisberg (1997) suggest that performance estimations are valuable because they allow the learner to intrinsically evaluate the outcome of their movements and assimilate task requirements more efficiently. In terms of Experiment 1, it may be that the participants evaluated each trial in a set, ranking them mentally based on perceived performance, and then utilized feedback to confirm or reject their predictions. However, if this hypothesis is correct, it is unclear why the self-controlled and best-trial groups would fail to conduct these performance estimations as well. It may be that the process of receiving feedback for the best trials suppressed these spontaneous estimations, however it is unclear how or why this may be.

A final possibility is that individuals in the worst-trial group used the KR provided for each trial to compare and contrast performance in order to improve subsequent attempts at the task. This explanation is supported by research examining bandwidth feedback schedules where qualitative feedback is delivered after good trials (i.e. participants do not receive feedback, therefore performance was acceptable and should be repeated) and quantitative feedback is delivered after poor trials (i.e. participants receive detailed feedback indicating the aspects of performance that are lacking) (Sherwood, 1988). In these studies, participants compare performance on good and bad trials to determine the strategies that produce ideal results. The worst-trial group in Experiment 1 may have employed a similar comparison. Conversely, individuals in the best-trial group may not make these comparisons as their attention is primarily directed towards the successful aspects of their performance, not the areas where improvements are required.

Based on the current literature, it is unclear which of these explanations are most likely contributing to the benefits of delivering error information. To examine each of these three possibilities, Experiment 2 compared the worst-trial group from Experiment 1 with two new groups performing the same task. One group received feedback immediately following the completion of each 3-trial set while the other group was asked to make estimations about their performance before receiving feedback. In addition, these two new groups only received KR and KP for *one* trial in each set; KR was not provided for the other two trials. Through the comparison of these three groups each of the explanations outlined above were individually examined. As such, three hypotheses were formulated to coincide with each of the explanations.

The Attention to Error Hypothesis: The only common factor linking these three groups is the delivery of feedback regarding the worst trials in each set. As a result, if all the groups showed similar retention benefits then we could conclude that feedback regarding the worst aspects of performance focuses the learner's attention on areas where improvement is necessary, allowing them to adjust their tactics accordingly.

The Error Estimation Hypothesis: As mentioned, individuals in the estimation group were asked to predict aspects of their performance before receiving detailed feedback. In Experiment 1, the worst-trial group also had time to estimate their performance before receiving detailed feedback, though these instructions were not explicitly given. Conversely, those receiving feedback immediately were not given this opportunity. As a result, if the worst-trial and estimation groups exhibited retention scores that were superior to the immediate feedback group, we could conclude that performance estimation plays a critical role in the benefits observed from error information.

The Performance Comparison Hypothesis: Neither the estimation nor immediate feedback groups received any knowledge regarding their performance on the intermediate and best trials in each 3-trial set; only the worst-trial group from Experiment 1 was exposed to this information. As a result, if the worst-trial group outperformed the other two groups, we could conclude that the additional KR was used in conjunction with the detailed KP to compare and develop optimal movement strategies, and improve overall performance.

Experiment 2 Methods

Participants

Twenty volunteers (16 females, 4 males: average age 21.7 ± 3.5 years) from McMaster University participated in this experimental protocol. Participants were recruited from the same population as Experiment 1, using similar methods. All participants had normal or corrected-to-normal vision and were naïve regarding the purpose of the study. All but 3 volunteers were right handed, however only one participant selected to perform the experimental task using their left hand. Written consent was obtained from all participants before the protocol began.

Apparatus and Task

Participants were tested under the same environmental conditions, using the same equipment as Experiment 1. Individuals were allowed to adjust all equipment for personal comfort and were again instructed to use their preferred hand for the experimental task. The same interception task was performed, with the structure and scoring system for each phase (acquisition, retention, and transfer) remaining consistent across both studies. Feedback was still delivered at the conclusion of each 3-trial set, however the quantity and timing of this feedback was experimentally manipulated differently then was the case in Experiment 1.

Procedure

Participants were randomly assigned to 2 experimental groups (n=10): a Worst-trial Immediately group, and a Worst-trial Estimation group. These groups differed from the worst-trial group outlined in Experiment 1 based on the *quantity* of feedback received at the end of each 3-trial set. Specifically, the immediate and estimation groups received less KR; instead of receiving total scores for all three trials in the previous set, individuals in the estimation and immediate groups only received KR for a single trial (their worst trial) in each set. The quantity of KP was not altered, with both groups continuing to receive feedback on cursor speed, horizontal accuracy, and vertical accuracy for their worst trials.

The immediate and estimation groups differed from each other based on the *timing* of feedback. For the immediate group, KR and KP were provided for the worst trial in each set immediately following completion of the third trial. For the estimation group, there was a delay between the completion of the last trial in a set and the provision of feedback. During the delay participants in this group were required to make three estimations about their performance on the previous trials. Specifically, they were asked to predict their worst trial in the set, to estimate their total score for that trial, and to identify which aspect of their performance was the poorest (cursor speed, horizontal accuracy, or vertical accuracy). Once these estimations were complete, feedback consisting of KR and KP for the actual worst trial was provided.

Experiment 2 Results

Data Analysis

The estimation and immediate groups were compared to the worst-trial group from Experiment 1. Four dependent measures were analyzed to assess performance: Total Score, Velocity Score, Horizontal Accuracy Score, and Vertical Accuracy Score. For acquisition, these variables were analyzed using a 3 Group (Worst-trial, Estimation, Immediate) by 11 Block (9 trials in each) mixed ANOVA, with repeated measures on the latter factor. For retention, these variables were analyzed using a 3 Group by 2 Block mixed ANOVA. For the transfer tasks these variables were analyzed with a 3 Group by 4 Transfer Task (Up, Down, Left, Right) mixed ANOVA, with repeated measures on the latter factor. All ANOVA and post hoc tests (using Tukey's HSD) had significance levels set at $p < 0.05$.

Acquisition

Table 3 presents the means of the total score, vertical accuracy score, horizontal accuracy score, and velocity score across acquisition for all three groups in Experiment 2. The analysis for total score revealed a significant main effect for Block, $F(10,270)=6.49$, $p<0.001$. Post hoc tests indicated that Blocks 2-11 were all significantly higher than Block 1. All other blocks were not statistically different. The analysis for horizontal accuracy scores revealed a significant main effect for Block, $F(10,270)=2.27$, $p<0.02$. Post hoc tests indicated that Block 4 was significantly higher than Block 1. All other blocks were not statistically different. The analysis for vertical accuracy score revealed a significant main effect for Block, $F(10,270)=5.16$, $p<0.001$. Post hoc tests indicated that Blocks 3-11 were all significantly higher than Block 1. All other blocks were not statistically different. The analysis for velocity scores revealed a significant main effect for Block, $F(10,270)=4.32$, $p<0.001$. Post hoc tests indicated that Blocks 5-11 were all significantly higher than Block 1. In addition, Blocks 10 and 11 were also significantly higher than Block 2. All other blocks were not statistically different. No other effects or interactions were significant in these four analyses. Figure 7 illustrates the non-significant Group by Block interaction for the total score across acquisition.

Retention

Table 4 presents the means of the total score, vertical accuracy score, horizontal accuracy score, and velocity score in the retention and transfer tests for all three groups in Experiment 2. The analysis for total score revealed a significant main effect for Block, $F(1,27)=4.59$, $p<0.05$. Post hoc tests indicated that Block 2 was significantly higher than Block 1. The interaction between Group and Block also approached conventional significance levels, $F(2,27)=3.00$, $p=0.067$. The analysis for vertical accuracy scores revealed a significant Group by Block interaction, $F(2,27)=4.64$, $p<0.02$, which is illustrated in Figure 8. Post hoc tests indicated that the three groups had equivalent scores in Block 1, however in Block 2 the worst-trial group produced significantly higher scores than the estimation and immediate groups, which did not differ from each other.

No other effects or interactions were significant. The analyses for horizontal accuracy and velocity scores failed to reveal any significant effects or interactions.

The retention data for total score seemed to indicate a similar trend as the vertical accuracy analysis. As a result, a further analysis on the total score was conducted. A 3 Group (Worst-trial, Immediate, Estimation) by 2 Day (Acquisition, Retention) by 2 Block (9 trials in each) mixed ANOVA, with repeated measures on the last two factors, was performed. This analysis revealed a significant Day by Block interaction, $F(1,27)=4.8, p<0.04$, and a significant Group by Day by Block interaction, $F(2,27)=3.96, p<0.04$. The 3-way interaction is illustrated in Figure 9. Post hoc tests for this interaction indicated that the three groups had equivalent scores over both blocks in acquisition. However, in retention Block 1 the worst-trial and immediate groups were significantly higher than the estimation group, but similar to each other. In retention Block 2 the worst-trial group produced significantly higher scores than both the estimation and immediate groups, which did not differ from each other. No other effects or interactions were significant.

Transfer

The analysis for total score revealed a significant main effect for transfer type, $F(3,81)=5.65, p<0.002$, and a significant Group by Transfer Type interaction, $F(6,81)=2.27, p<0.05$. This interaction is illustrated in Figure 10. Post hoc tests for the main effect indicated that the Left task produced significantly higher scores than the Up task, with the Down and Right shifts intermediate and statistically similar to all transfer types. Post hoc tests for the interaction indicated that the score for the worst-trial group was higher than the estimation group in the Down transfer task and higher than both the estimation and immediate groups for the Right transfer task. The analysis for horizontal accuracy revealed a significant main effect for transfer type, $F(3,81)=5.6, p < 0.002$. Post hoc tests indicated that the Left task was significantly higher than the Up task, with the other two shifts intermediate and statistically similar. The analysis for vertical accuracy revealed a significant main effect for transfer type, $F(3,81)=9.81, p < 0.001$. Post hoc tests indicated that the Down, Left, and Right tasks were all significantly higher than the Up task, but not different from each other. The analysis for velocity scores revealed a significant main effect for transfer type, $F(3,81)=14.1, p < 0.001$. Post hoc tests indicated that the Down, Left, and Right tasks were all significantly lower than the Up task, and that the Down task was also lower than the Left task. No other effects or interactions were significant.

Experiment 2 Discussion

The purpose of Experiment 2 was to determine why the error information associated with a worst-trial feedback schedule was found to benefit motor learning more than best-trial and self-controlled schedules in Experiment 1. In accordance with current literature and the findings from Experiment 1, we suggested three explanations for these observed benefits. 1) Providing feedback regarding the worst aspects of performance draws a learner's attention to areas that are flawed, allowing them to make corrections to these areas on subsequent trials. 2) Participants may estimate trial outcome before receiving detailed feedback, resulting in an increased cognitive effort during task acquisition, an enhanced understanding of ideal movement parameters, and superior performance overall (Liu & Wrisberg, 1997). 3) Participants in Experiment 1 may have utilized the detailed, worst-trial feedback provided in conjunction with the KR delivered for each completed trial make comparisons between trials and improve performance by contrasting movement strategies until an optimal pattern was identified.

The Attention to Error Hypothesis

Three hypotheses were formulated to coincide with these explanations. The first hypothesis was referred to as the direction of attention hypothesis, and dealt with the suggestion that feedback for poor trials directs a learner's attention to areas where movement mechanics are weak. This hypothesis predicted that the worst-trial, estimation, and immediate feedback groups would show similar levels of performance, as each schedule delivered error information. However, contrary to these predictions, the three groups did not show equivalent performances in retention tests. Instead, the performance of the worst-trial group was better than the estimation group in both retention blocks, and better than the immediate group in retention block 2. In addition, significant differences were noted between the groups in two of four transfer tests, with the worst-trial again performing with the highest total scores. Based on this evidence, the direction of attention hypothesis was not supported.

One could argue that the failure to support this hypothesis may have occurred because the feedback delivered was not specific enough to provide participants with the information required to accurately adjust their performance. Specifically, the detailed KP delivered may have been too general to accurately alter movement parameters. For example, this information may have indicated where corrections were required, but not the *extent* of these corrections (e.g. participants were not informed if they needed to move twice as fast to receive a perfect velocity score). However, based on the findings from Lee and Carnahan (1990b), this explanation seems unlikely. These authors examined a sequential, multisegment timing task to determine if the delivery of detailed feedback for the worst segment of the movement would enhance the learning process. They found that delivering detailed movement time feedback for the worst segment did not improve learning compared to yoked control groups. These results suggest that even specific feedback regarding the worst aspects of performance may not have been adequate to produce meaningful learning advantages.

Thus, it appears that the direction of attention hypothesis does not provide an accurate explanation for the learning benefits of error information. Interestingly, this finding agrees with some suggestions forwarded in the focus of attention literature. While some authors have reported advantages associated with directing attention to the mechanics of a movement during acquisition (Gray, 2004; Perkins-Ceccato et al., 2003), others have indicated that this focus may actually impede the learning process (Wulf, Shea, & Park, 2001; Wulf, Weigelt, Poulter, McNevin, 2003). Zachry, Wulf, Mercer, and Bezodis (2005) suggested that focusing specifically on the mechanics of a movement results in maladaptive, short-term changes to task execution, and leads to the development of inefficient movement patterns. They further suggest that these inefficient patterns cause noise in the muscular system during task performance, producing negative effects on motor learning. These studies indicate that directing attention to the aspects of performance that are completed poorly (or even correctly) may hinder the learning process.

The Error Estimation Hypothesis

The second hypothesis, termed the estimation benefit hypothesis, addressed the suggestion that individuals were estimating their performance outcomes prior to receiving detailed KP, resulting in an enhanced understanding of movement requirements and improved performance. This hypothesis predicted two outcomes: First, the worst-trial and estimation groups would perform equally well, as both groups had the opportunity to estimate levels of performance prior to receiving feedback. Second, these two groups would show superior learning compared to the immediate feedback group, as the latter group did not have time to make performance predictions before feedback delivery.

These predictions were partially supported by the findings of Experiment 2. The retention tests revealed that the worst-trial group did outperform the immediate group in block 1. In addition, the estimation and worst-trial groups showed similar levels of performance on the Up and Left transfer tests. However, these two groups failed to show equal performances across all retention and transfer measures. Instead, the worst-trial group outperformed the estimation group in retention blocks 1 and 2, vertical accuracy scores, and the Down and Right transfer tests. In addition, the estimation group also produced lower total scores in retention block 1 compared to the immediate feedback group.

The lack of complete support for the estimation benefit hypothesis may have resulted from the number of estimations that participants were required to perform. In total, participants in the estimation group were asked to make three predictions about various aspects of their performance, rather than the single prediction pertaining to trial outcome that is generally used in estimation studies (Swinnen et al., 1990). Conversely, the worst-trial group from Experiment 1 was not forced to make as many estimations. However, this explanation seems unlikely, given the findings of Liu and Wrisberg (1997). Their results suggest that the additional predictions should have benefited the estimation group by further enhancing their understanding of the task, which in turn would result in superior learning advantages.

An alternative, but related explanation is that the number of predictions completed by the estimation group may have delayed the onset of feedback to an extent where the error information was no longer useful. If participants could not accurately recall their performance, then these estimations may have been ineffective at producing learning benefits. However this explanation also seems unlikely as previous research has shown that delaying the delivery of feedback after trial completion is actually beneficial, not detrimental, to motor learning (Schmidt et al., 1989; Swinnen et al., 1990). In addition, research examining performance estimations has shown similar learning benefits when comparing schedules in which feedback is delivered immediately following predictions of trial outcome, and schedules where these predictions are followed by a delay before feedback is delivered (Liu & Wrisberg, 1997).

Thus, it appears as though the estimation benefit hypothesis does not provide a full explanation of the learning advantages reported for worst trial feedback schedules. Though outcome estimations resulted in some benefits early in retention, it failed to maintain these benefits, or to transfer these advantages to similar tasks. Given the findings from Experiment 1, these results are not surprising. There was no indication that the worst-trial group was spontaneously performing these estimations after each trial. In addition, there was no reason to believe that the best-trial and self-controlled groups would not have spontaneously conducted such estimations. It is clear then that additional processes must be involved in the utilization of error information.

The Performance Comparison Hypothesis

The third hypothesis addressed the suggestion that individuals in the worst-trial group were utilizing the KR for each trial to conduct comparisons between movement strategies, and then using detailed KP to develop tactics that would produce ideal movement patterns. This hypothesis was referred to as the performance comparison hypothesis, and predicted that the worst-trial group would show superior performance advantages compared to the estimation and immediate feedback groups because of the provision of KR for each trial. The results from Experiment 2 supported this hypothesis. The worst-trial group consistently showed equal or superior performance in both retention and transfer tests. Specifically, the worst-trial group performed significantly better than both groups in retention block 2 and the Right transfer task, and better than the estimation group in retention block 1 and the Down transfer task.

These results agree well with studies examining bandwidth feedback schedules, which is not surprising given the similarities between the two conditions. In both cases, explicit, quantitative information is delivered when performance is poor. In addition, both schedules allow for the active comparison between successful and unsuccessful trials, which is clearly an advantage based on the results from Experiment 2. The most important difference between these two schedules is the provision of KR; participants in the worst-trial group received explicit KR after successful trials while participants in bandwidth schedules receive only qualitative information after good trials (i.e. no feedback indicates that the previous performance was acceptable) (Lee & Carnahan, 1990a). However, this variation in KR provision may not be a significant difference. Though KR was delivered to the worst-trial group, without the provision of KP there was no way for participants in this group to know for certain the areas of the movement that were performed accurately and those where improvements were necessary. Thus, these individuals were limited to using the KP delivered for the worst performances to make comparisons between trials. These restrictions are very similar to those imposed by bandwidth feedback schedules (Lee & Carnahan, 1990a).

Overall, the results from Experiment 2 indicate that the performance comparison hypothesis provides the most accurate explanation of the learning benefits associated with error information. It appears as though knowledge pertaining to the outcome of a movement plays a critical role in the utilization of error information, as it allows learners to make comparisons between trials and establish strategies that may lead to successful task execution. In addition, learning appears to suffer without this information, as illustrated in the retention and transfer scores for the estimation and immediate feedback groups.

General Discussion

The Limitations of a Performance-Based Self-Controlled Schedule

Since the work of Thorndike (1927), research examining the effects of augmented feedback provision on motor learning has continued to produce interesting and complex results. Frequency, timing, type, and scheduling of augmented feedback delivery have all been identified as critical variables that influence the acquisition and retention of novel motor skills. More recently, studies examining self-controlled schedules have added a further level of complexity to the augmented feedback literature by demonstrating the benefits of relinquishing control of feedback delivery to the learner. These studies have consistently reported learning benefits associated with allowing participants to self-select the frequency and timing of augmented feedback (Janelle et al., 1995, 1997; Wulf et al., 2005). In an attempt to explain these findings Chiviacowsky and Wulf (2002) suggested that feedback utilization is optimized when it follows and outlines a successful performance of a task. However, these authors did not specify whether the timing of feedback delivery (i.e. directly following a good performance), or the information regarding successful trials was the more critical factor contributing to self-controlled benefits.

To examine these mechanisms, Experiment 1 employed a performance-based self-controlled schedule in which participants were required to make feedback selections based strictly on trial performance, and not on the frequency or timing of feedback delivery. This design allowed us to directly examine how the type of feedback selected during acquisition (i.e. feedback outlining the most successful or least successful trials) affected learning, while limiting any influences of the timing of feedback delivery. This design also allowed for the direct comparison of a self-controlled condition with two performance-based schedules (worst-trial and best-trial feedback schedules). Based on the explanation proposed by Chiviacowsky and Wulf (2002), we hypothesized that individuals in the self-controlled group would select feedback for their best trials most frequently, as the information contained in these trials should deliver the most beneficial information about overall performance. In addition, we predicted that the self-controlled and best-trial groups would show equivalent performance scores in retention, as each group would receive similar information, and superior scores compared to the worst-trial group, as this group would not have the benefits associated with feedback describing successful performances.

The findings from Experiment 1 indicated that a self-controlled schedule was not ideal when feedback selections were performance-based. This is a very interesting finding given the fact that self-controlled schedules have been shown to be beneficial to motor learning in a variety of augmented feedback studies (Janelle et al., 1995, 1997; Wulf et al., 2005) and in practice scheduling (Keetch & Lee, in press). We can therefore conclude that the type of feedback selected by an individual during acquisition is not the sole contributor to self-controlled learning benefits. However, based on these results, it appears that the type of feedback delivered does have *some* impact on self-controlled schedules (Janelle et al., 1995, 1997; Wulf et al., 2005). In the current study, the self-controlled group did show some learning advantages in the Up and Down transfer tasks compared to the best-trial feedback group, and indicated a clear preference to receive feedback after their most successful trials. However, the self-controlled feedback schedule failed to result in significant performance advantages in retention.

Instead, the results from Experiment 1 indicate that the type of feedback *alone* is not the driving force behind the learning advantages associated with a self-controlled schedule. Specifically, when control over feedback timing was removed, the benefits associated with a self-controlled schedule diminished.

Thus, it may be that the self-selection of feedback based on trial outcome is only beneficial when the timing of feedback delivery is also under the learner's control.

In addition, the frequency of feedback provision may have an important impact in a self-controlled schedule. Though the explanation forwarded by Chiviacowsky and Wulf (2002) failed to address how often feedback should be delivered, most studies examining self-controlled schedules allow participants to determine both the timing and frequency of feedback provision. Conversely, Experiment 1 restricted the frequency of feedback provision across all groups, which may have contributed to the results observed in the self-controlled group. Thus, it may be that a combination of control over all three variables (timing, frequency, and performance) is necessary to produce the sustained learning advantages reported in previous self-controlled feedback studies (Chiviacowsky & Wulf, 2002; Wulf et al., 2005).

Does Perception of Performance Influence Self-Controlled Feedback Schedules?

Alternatively, the results from Experiment 1, in combination with the findings reported by Chiviacowsky and Wulf (2002), may suggest another explanation for the self-controlled benefits reported in previous literature. It may be that selecting to receive feedback based on the *perception* of performance is a critical factor influencing learning. This hypothesis can be explained more clearly by examining the differences in the delivery of KR across these two studies. Chiviacowsky and Wulf (2002), like most authors examining self-controlled schedules, did not provide any feedback to participants at the conclusion of a trial unless it was specifically requested. As a result, participants had no explicit knowledge about how they were performing, but had to rely on their perception of performance to determine whether they would select feedback for the current trial, or to go directly to the next trial.

Making these decisions based on the perception of performance may have encouraged participants to engage in active evaluations of each trial in order to determine whether their performance was successful or not. Swinnen et al. (1990) suggested that performing such active evaluations could have beneficial affects on motor learning. Specifically, these authors hypothesized that conducting self-evaluations of performance prior to feedback delivery can improve error-detection processes leading to more efficient movement patterns. Furthermore, they suggest that these evaluations can decrease the guiding effects associated with high frequencies of feedback delivery, resulting in less dependence on this information, and improved performance when feedback is removed.

In Experiment 1, participants assigned to the self-controlled group were provided with KR at the conclusion of each 3-trial set, regardless of the trial for which they selected to receive detailed KP. As a result, these individuals always knew the precise outcome of each trial, in terms of their total scores (KR). It could be argued that providing KR for each trial served to suppress the spontaneous self-evaluations and negated any of the benefits associated with selecting feedback based on the perception of trial outcome. This hypothesis could explain why the self-controlled group in Experiment 1 failed to show learning advantages compared to the performance-based feedback groups.

Worst-Trial Feedback Schedules: Why is KR for Every Trial Advantageous?

The results from Experiment 1 indicated that a worst-trial feedback schedule was the most advantageous performance-based schedule when frequency and timing of feedback delivery were controlled. The findings from Experiment 2 extended these results by suggesting that worst-trial schedules are most beneficial when learners receive KR for each performance of a task prior to receiving detailed error information for the least successful trials. However, this finding begs the question as to why the extra information is advantageous to motor learning.

There are a number of reasons that could explain why KR for each performance benefits worst-trial feedback schedules. The first explanation is that participants in the worst-trial group simply took advantage of the increased quantity of feedback they received to improve their performance. Specifically, the worst-trial group received KR for each trial in a set while the estimation and immediate feedback groups received KR for only a single trial. However, this explanation seems unlikely, as higher quantities of feedback have been shown to be detrimental to the learning process (Weeks & Kordus, 1998; Winstein & Schmidt, 1990). In fact, based on this literature, we would have predicted the lowest performance proficiency for the worst-trial group. Thus, it appears that simply receiving more feedback does not account for these learning advantages.

Another explanation for the findings of Experiment 2 involves the manner that KR was presented to participants in the worst-trial group. Specifically, this information was delivered in a summary format (all three trials appearing at the same time after completion of a set). Schmidt et al. (1989) suggested that providing feedback in a summary form decreases the guiding effects associated with this information, allowing the learner to utilize the feedback without becoming dependent upon it. However, this explanation also seems doubtful given the results of Experiment 1 where all three groups (best-trial, worst-trial and self-controlled) received KR in a summary form, but only the worst-trial group showed significant learning advantages in retention. Thus, it appears that the manner by which KR was presented did not contribute to the learning advantages in the worst-trial group.

The results of Experiment 2 may be best explained with the Performance Comparison hypothesis. According to this hypothesis, providing KR for every trial in a set before delivering detailed KP enhances performance in two ways: First, it allows participants to compare performance across all trials in order to determine the general movement parameters that contribute to successful task execution. Second, it allows participants to examine and address specific areas where movement mechanics are weak by providing detailed error information regarding the least successful trials. This explanation seems the most reasonable given the superior performance of the worst-trial group in Experiments 1 and 2. In addition, this explanation agrees well with studies examining bandwidth feedback schedules where qualitative information is delivered following successful performances of a task, and detailed quantitative information is delivered following unsuccessful attempts (Sherwood, 1988).

Practical Applications of Worst-Trial Feedback Schedules

The strength of the findings from Experiment1 lies in its application to practical coaching situations. In many coaching scenarios, the delivery of feedback is influenced to a large extent by the performance of an athlete, and not their individual preferences (as would be the case with a self-controlled schedule). It is often commonplace for coaches to provide feedback when a skill is performed poorly rather than when performance is satisfactory. The results from the current study indicate that this is a sound practice to employ when coaching individuals on their performance of a new skill.

However, the more interesting findings are those associated with Experiment 2. In most athletic performances it is realistic to assume that individuals will have the ability to observe the outcome of their attempts at a task. Be it a golf shot, a free throw, or a soccer penalty kick, the athlete will be able to view the consequences of their actions and obtain of general idea of whether the attempt was successful or not. The results from Experiment 2 indicate that these performance observations play a critical role in the benefits associated with worst-trial feedback schedules. While the literature examining performance estimations may suggest that the removal of this information is important to enhance the learning process, Experiment 2 suggests that these performance observations actually provide positive motor learning benefits.

Future Research Considerations

The results from Experiment 1 suggest that a performance-based self-controlled feedback schedule may be insufficient on its own to enhance motor learning. Instead, it appears that the benefits of self-controlled schedules are also contingent on the timing and frequency of feedback provision. However, it is unclear whether the self-determination of one or both of these factors is necessary to enhance motor learning. Future studies examining the timing and frequency of feedback delivery individually under a self-controlled condition may help researchers understand the exact mechanisms contributing to self-controlled learning advantages.

In addition, the validity of the Performance Comparison hypothesis also deserves further examination. Specifically, it remains to be determined whether this hypothesis is applicable to realistic sport situations, where athletes often know the outcome of their performance almost immediately after movement execution. If the Performance Comparison hypothesis is accurate, then detailed information regarding poorly performed trials should continue to provide the most significant learning benefits, even when the only KR available is an athlete's observation of their performance.

Conclusion

The results from Experiment 1 indicate that performance-based self-controlled feedback schedules may not be advantageous to motor learning when the frequency and timing of feedback provision are kept constant. Rather, it appears that a worst-trial feedback schedule yields superior learning benefits under these conditions. This finding agrees with common coaching practices that often rely heavily on the delivery of error information in order to improve an athlete's overall performance.

Experiment 2 examined the mechanisms contributing to the observed advantages of a worst-trial feedback schedule. The results from this study suggested that worst-trial schedules are most effective when detailed KP for the least successful attempts at a task are combined with KR for every trial completed. We suggest that this feedback structure allows learners to compare their performances on every trial in order to gain a general understanding of correct movement mechanics, and enables them to use the detailed error information to make specific adaptations to areas where movement mechanics are weak.

References

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

Beilock, S.L., Carr, T.H., MacMahon, C., & Starkes, J.L. (2002). When paying attention becomes counterproductive: impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied, 8,* 6-16.

Bilodeau, E.A., & Bilodeau, I.M. (1958). Variable frequency knowledge of results and the learning of a simple motor skill. *Journal of Experimental Psychology, 55,* 379-383.

Cauraugh, H., Chen, D., & Radlo, S.J. (1993). Effects of traditional and reversed bandwidth knowledge of results on motor learning. *Research Quarterly for Exercise and Sport, 64,* 413-417.

Chiviacowsky, S., & Wulf, G. (2002). Self-controlled feedback: Does it enhance learning because performers get feedback when they need it? *Research Quarterly for Exercise and Sport, 73,* 408-415.

Chiviacowsky, S., & Wulf, G. (2005). Self-controlled feedback is effective if it is based on the learner's performance. *Research Quarterly for Exercise and Sport, 76,* 42-48.

Goldstein, M., & Rittenhouse, C.H. (1954). Knowledge of results in the acquisition and transfer of a gunnery skill. *Journal of Experimental Psychology, 48,* 187-196.

Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied, 10,* 42-54.

Guthrie, J.T. (1971). Feedback and sentence learning. *Journal of Verbal Learning and Verbal Behaviour, 10,* 23-28.

Janelle, C.M., Barba, D.A., Frehlich, S.G., Tennant, L.K., & Cauraugh, J.H. (1997). Maximizing performance feedback effectiveness through videotape replay in a self-controlled learning environment. *Research Quarterly for Exercise and Sport, 68,* 269-279.

Janelle, C.M., Kim, J., & Singer, R.N. (1995). Subject-controlled performance feedback and learning of a closed motor skill. *Perceptual and Motor Skills, 81,* 627-634.

Keetch, K.M., & Lee, T.D. (in press). Effects of self-regulated and experimenter-imposed practice schedules on motor learning for tasks of varying difficulty. *Research Quarterly for Exercise and Sport.*

Lee, T.D., & Carnahan, H. (1990a). Bandwidth knowledge of results and motor learning: more than just relative frequency effect. *Quarterly Journal of Experimental Psychology, 42A,* 777-789.

Lee, T.D., & Carnahan, H. (1990b). When to provide knowledge of results during motor learning: Scheduling effects. *Human Performance, 3,* 87-105.

Lincoln, R.S. (1954). Learning a rate of movement. *Journal of Experimental Psychology, 47,* 465-470.

Liu, J., & Wrisberg, C.A. (1997). The effect of knowledge of results delay and the subjective estimation of movement form on the acquisition and retention of a motor skill. *Research Quarterly for Exercise and Sport, 68,* 145-151.

Pashler, H., Cepeda, N.J., Wixted, J.T., & Rohrer, H. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 3-8.

Perkins-Ceccato, N., Passmore, S.R., & Lee, T.D. (2003). Effects of focus of attention depend on golfers' skill. *Journal of Sport Science, 21,* 593-600.

Salmoni, A.W., Schmidt, R.A., & Walter, C.B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin, 95,* 355-386.

Schmidt, R.A., & Bjork, R.A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207-217.

Schmidt, R.A., & Lee, T.D. (2005). *Motor control and learning: A behavioral emphasis.* Champaign, IL: Human Kinetics.

Schmidt, R.A., & Young, D.E. (1991). Methodology for motor learning: A paradigm for kinematic feedback. *Journal of Motor Behavior, 23*, 13-24.

Schmidt, R.A., Young, D.E., Swinnen, S., & Shapiro, D.C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 352-359.

Sherwood, D.E. (1988). Effect of bandwidth knowledge of results on movement consistency. *Perceptual and Motor Skills, 66,* 335-342.

Simon, D.A., & Bjork, R.A. (2001). Metacognition in motor learning. *Experimental Journal of Exercise Psychology: Learning, Memory, and Cognition, 27,* 907-912.

Simon, D.A., & Bjork, R.A. (2002). Models of performance in learning multisegment movement tasks: Consequences for acquisition, retention, and judgments of learning. *Journal of Experimental Psychology: Applied, 8,* 222-232.

Smode, A.F. (1958). Learning and performance in a tracking task under two levels of achievement information feedback. *Journal of Experimental Psychology, 56,* 297-303.

Swinnen, S.P., Schmidt, R.A., Nicholson, D.E., & Shapiro, D.C. (1990). Information feedback for skill acquisition: Instantaneous knowledge of results degrades learning. *Experimental Psychology: Learning, Memory, and Cognition, 16,* 706-716.

Thorndike, E.L. (1927). The law of effect. *American Journal of Psychology, 39,* 212-222.

Weeks, D.L., & Kordus, R.N. (1998). Relative frequency of knowledge of performance and motor skill learning. *Research Quarterly for Exercise and Sport, 69*, 224-230.

Williams, A.C., & Briggs, G.E. (1962). On-target versus off-target information and the acquisition of tracking skill. *Journal of experimental Psychology, 64,* 519-525.

Winstein, C.J., & Schmidt, R.A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 677-691.

Wrisberg, C.A., & Pein, R.L. (2002). Note on learners' control of the frequency of model presentation during skill acquisition. *Perceptual and Motor Skills, 94,* 792-794.

Wulf, G., Shea, C., & Park, J. (2001). Attention and motor performance: Preferences for and advantages of an external focus. *Research Quarterly for Exercise and Sport, 72*, 335-344.

Wulf, G., Clauss, A., Shea, C.H., & Whitacre, C.A. (2001). Benefits of self-control in dyad practice. *Research Quarterly for Exercise and Sport, 72*, 299-303.

Wulf, G., Raupach, M., & Pfeiffer, F. (2005). Self-controlled observational practice enhances learning. *Research Quarterly for Exercise and Sport, 76*, 107-111.

Wulf, G., & Toole, T. (1999). Physical assistance devices in complex motor skill learning: Benefits of a self-controlled practice schedule. *Research Quarterly for Exercise and Sport, 70*, 265-272.

Wulf, G., Weigelt, M., Poulter, D., McNevin, N. (2003). Attentional focus on suprapostural tasks affects balance learning. *The Quarterly Journal of Experimental Psychology, 56A,* 1191-1211.

Viitasalo, J.T., Era, P., Konttinen, N., Mononen, H., Mononen, K., Norvapalo, K. (2001). Effects of 12-week shooting training and mode of feedback on shooting scores among novice shooters. *Scandinavian Journal of Medicine and Science in Sports, 11,* 362-368.

Young, D.E., & Schmidt, R.A. (1992). Augmented kinematic feedback for motor learning. *Journal of Motor Behavior, 24*, 261-273.

Zachry, T., Wulf, G., Mercer, J., & Bezodis, N. (2005). Increased movement accuracy and reduced EMG activity as a result of adopting an external focus of attention. *Brain Research Bulletin, 67,* 304-309.

Table 1: Means and standard deviations of the total score, combined accuracy score, and velocity score across acquisition for the worst-trial, best-trial, and self-controlled groups in Experiment 1.

		Acquisition Block										
		1	2	3	4	5	6	7	8	9	10	11
Total Score	Worst-Trial (Mean)	33.6	38.8	41.8	43.4	41.8	42.2	41.5	42.4	40.7	44.0	41.3
	(STDev)	*9.3*	*11.1*	*8.4*	*6.4*	*8.4*	*7.7*	*7.3*	*5.1*	*9.8*	*10.4*	*8.2*
	Best-Trial (Mean)	32.5	39.6	37.3	40.7	39.5	36.9	40.0	35.1	36.2	38.5	37.4
	(STDev)	*9.0*	*9.1*	*11.0*	*9.3*	*7.1*	*5.6*	*9.1*	*11.2*	*9.7*	*8.2*	*7.7*
	Self-Controlled (Mean)	32.6	36.4	39.8	40.0	38.9	40.2	38.4	41.1	37.4	41.7	40.9
	(STDev)	*10.3*	*8.2*	*10.1*	*7.3*	*8.2*	*6.5*	*5.9*	*4.9*	*5.6*	*6.5*	*7.7*
Combined Accuracy	Worst-Trial (Mean)	29.9	35.4	38.2	40.1	38.7	38.4	38.1	37.0	36.2	38.4	36.9
	(STDev)	*10.4*	*13.4*	*11.0*	*8.8*	*10.4*	*10.1*	*9.3*	*9.1*	*11.6*	*12.9*	*11.5*
	Best-Trial (Mean)	29.6	37.1	33.6	36.2	33.9	29.9	34.6	30.7	30.2	33.0	31.2
	(STDev)	*9.6*	*10.5*	*11.5*	*11.0*	*10.0*	*8.3*	*12.5*	*12.9*	*11.7*	*11.4*	*10.6*
	Self-Controlled (Mean)	28.2	31.6	33.7	34.0	33.1	33.3	32.3	34.2	31.2	36.0	34.7
	(STDev)	*10.3*	*9.0*	*10.7*	*9.7*	*10.0*	*10.4*	*9.5*	*7.2*	*7.8*	*8.9*	*9.0*
Velocity Score	Worst-Trial (Mean)	3.7	3.4	3.6	3.3	3.1	3.8	3.4	5.5	4.5	5.6	4.4
	(STDev)	*2.6*	*3.4*	*3.2*	*3.1*	*2.3*	*2.9*	*2.5*	*5.8*	*2.9*	*4.0*	*4.2*
	Best-Trial (Mean)	2.8	2.5	3.8	4.4	5.6	7.0	5.4	4.4	6.0	5.6	6.1
	(STDev)	*1.2*	*1.6*	*2.9*	*2.5*	*4.2*	*5.6*	*4.1*	*4.1*	*6.0*	*3.6*	*4.6*
	Self-Controlled (Mean)	4.4	4.8	6.1	6.0	5.9	6.9	6.1	6.9	6.2	5.8	6.2
	(STDev)	*1.8*	*2.5*	*5.4*	*4.2*	*2.6*	*4.9*	*4.4*	*4.3*	*2.8*	*2.9*	*3.6*

Table 2: Means and standard deviations of the total score, combined accuracy score, and velocity score in retention and transfer for the worst-trial, best-trial, and self-controlled groups in Experiment 1.

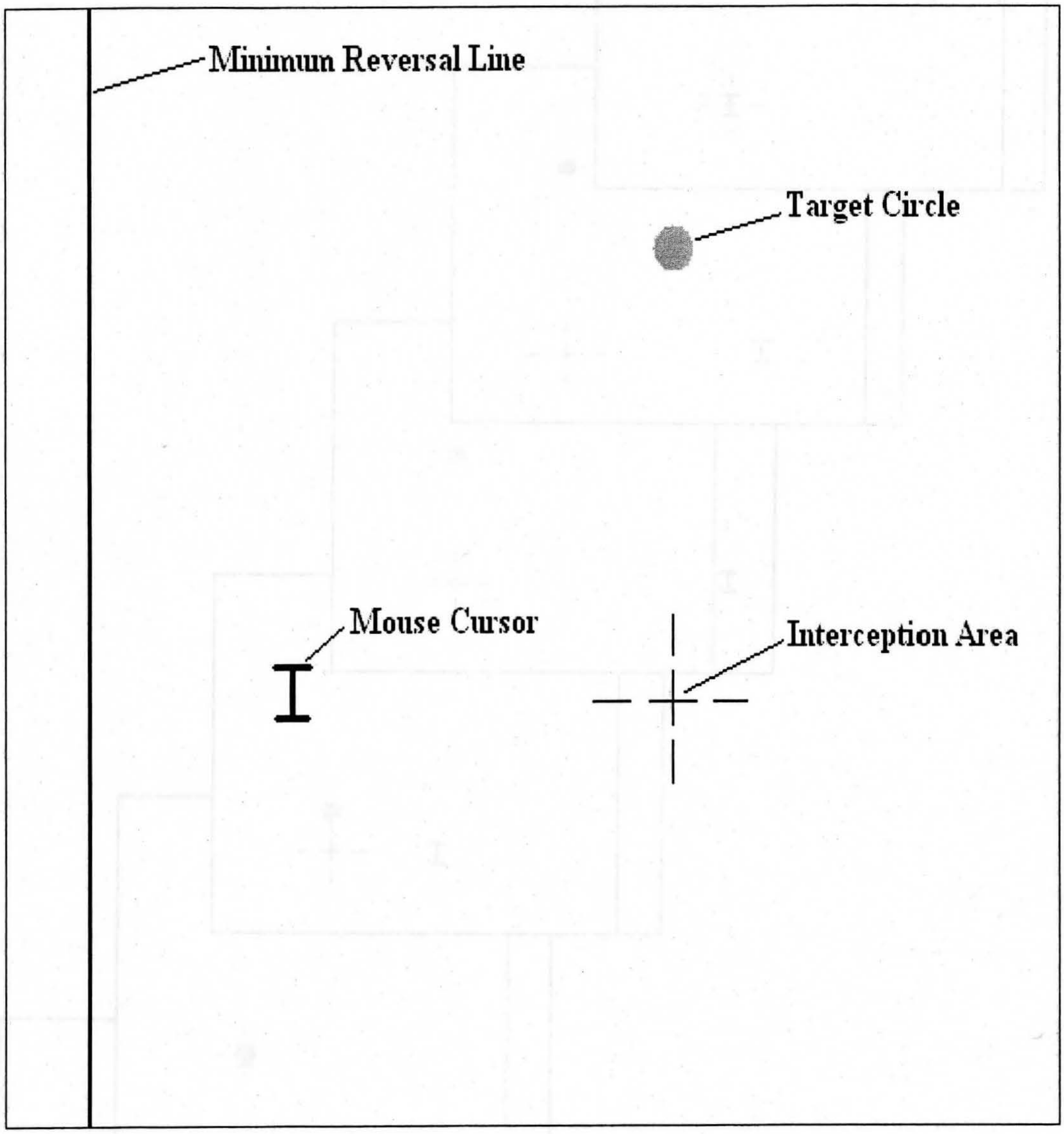
		Retention Block		**Transfer Type**			
		1	2	Up	Down	Left	Right
Total Score	Worst-Trial (Mean)	41.7	46.0	33.6	41.7	42.8	42.8
	(STDev)	*7.9*	*5.4*	*8.1*	*12.1*	*9.9*	*6.9*
	Best-Trial (Mean)	41.9	36.4	25.5	30.1	38.5	35.6
	(STDev)	*7.1*	*11.8*	*9.1*	*13.4*	*9.2*	*12.5*
	Self-Controlled (Mean)	41.0	41.3	34.6	40.7	40.2	40.7
	(STDev)	*6.8*	*5.4*	*8.0*	*6.2*	*9.4*	*10.5*
Combined Accuracy	Worst-Trial (Mean)	37.0	40.5	27.4	38.4	39.0	38.7
	(STDev)	*12.2*	*10.1*	*11.7*	*14.2*	*13.1*	*10.1*
	Best-Trial (Mean)	35.8	30.0	17.9	26.7	31.4	29.1
	(STDev)	*10.1*	*14.9*	*9.2*	*14.3*	*13.4*	*12.0*
	Self-Controlled (Mean)	35.2	35.0	27.3	36.8	34.9	36.9
	(STDev)	*8.3*	*7.0*	*9.5*	*7.9*	*10.9*	*11.1*
Velocity Score	Worst-Trial (Mean)	4.8	5.6	6.1	3.2	3.8	4.1
	(STDev)	*4.9*	*6.5*	*4.2*	*2.9*	*3.6*	*3.6*
	Best-Trial (Mean)	6.0	6.4	7.6	3.4	7.1	6.5
	(STDev)	*4.7*	*4.3*	*3.0*	*3.0*	*5.2*	*6.1*
	Self-Controlled (Mean)	5.8	6.3	7.3	3.9	5.4	3.9
	(STDev)	*2.8*	*2.9*	*2.6*	*3.2*	*3.4*	*2.4*

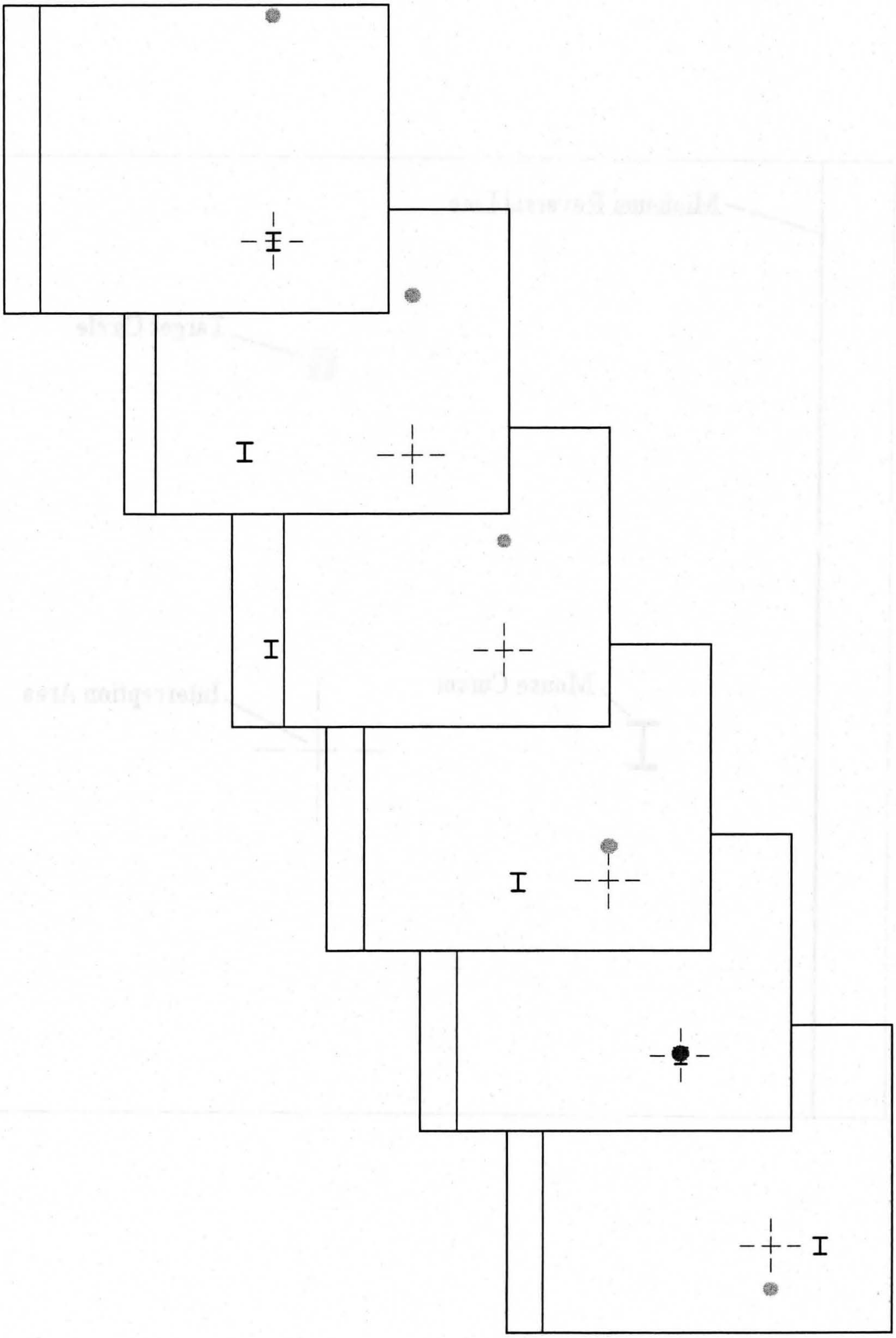
Table 3: Means and standard deviations of the total score, vertical accuracy score, horizontal accuracy score, and velocity score across acquisition for the worst-trial, estimation, and immediate groups in Experiment 2.

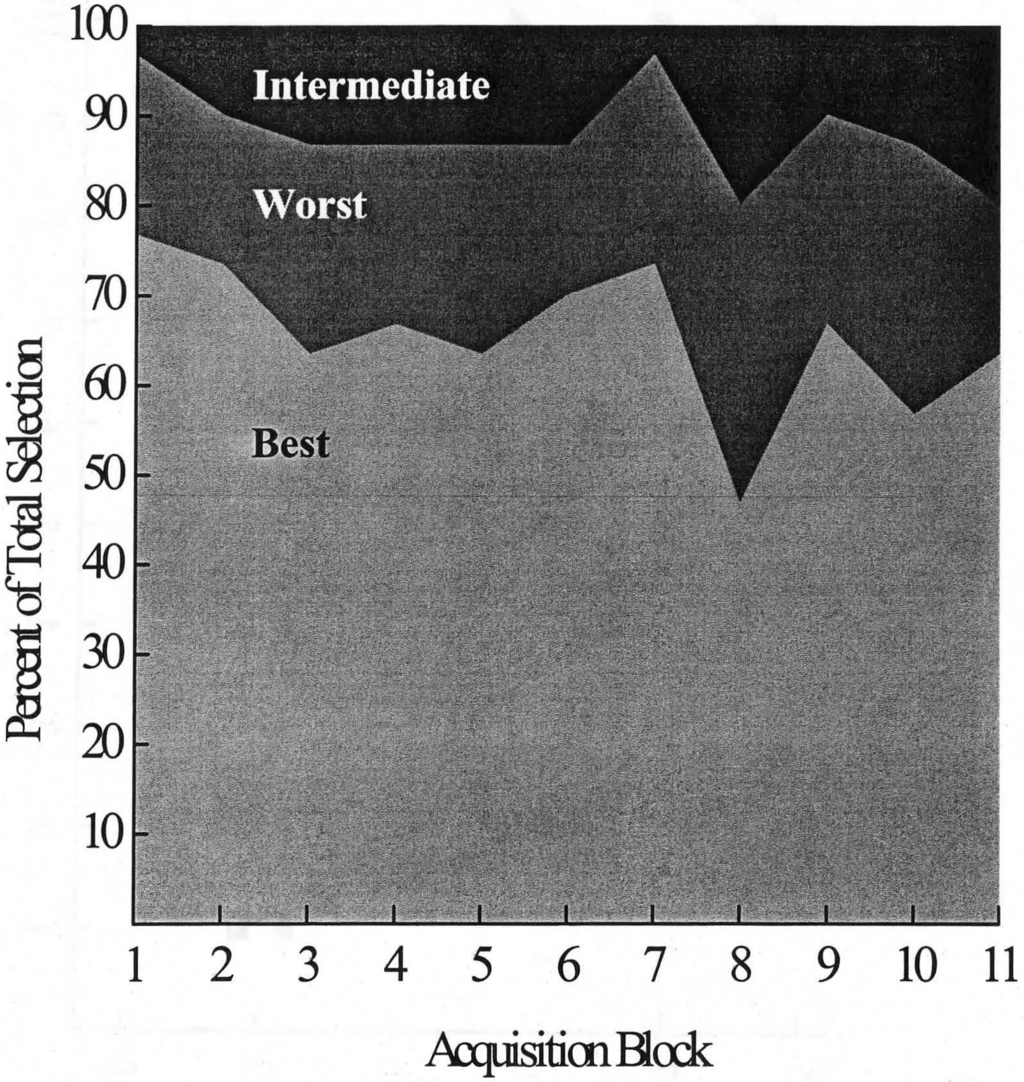
		Acquisition Block										
		1	2	3	4	5	6	7	8	9	10	11
Total Score	Worst-Trial (Mean)	33.6	38.8	41.8	43.4	41.8	42.2	41.5	42.4	40.7	44.0	41.3
	(STDev)	*9.3*	*11.1*	*8.4*	*6.4*	*8.4*	*7.7*	*7.3*	*5.1*	*9.8*	*10.4*	*8.2*
	Estimation (Mean)	28.6	29.1	32.6	38.6	35.2	34.0	36.9	36.5	37.1	39.9	41.4
	(STDev)	*10.6*	*11.4*	*11.6*	*8.7*	*12.7*	*8.5*	*10.3*	*7.6*	*8.1*	*7.1*	*10.2*
	Immediate (Mean)	29.4	41.9	41.0	40.8	39.8	35.8	43.1	38.2	40.7	40.8	41.4
	(STDev)	*7.3*	*4.7*	*8.1*	*9.2*	*8.7*	*10.8*	*8.1*	*11.9*	*6.9*	*13.0*	*9.4*
Vertical Accuracy	Worst-Trial (Mean)	8.9	10.7	12.3	11.5	12.2	12.0	11.9	11.6	11.1	12.5	11.4
	(STDev)	*3.9*	*4.1*	*2.2*	*2.8*	*3.5*	*3.3*	*2.0*	*3.7*	*3.8*	*4.2*	*3.5*
	Estimation (Mean)	7.8	7.5	9.7	10.1	10.0	9.1	11.2	11.1	10.7	11.2	11.5
	(STDev)	*3.1*	*3.6*	*3.3*	*3.4*	*4.0*	*3.5*	*3.9*	*3.2*	*3.7*	*2.5*	*3.4*
	Immediate (Mean)	7.7	11.4	12.0	11.4	11.7	10.2	11.4	11.6	11.6	12.1	12.7
	(STDev)	*2.2*	*1.9*	*3.1*	*4.2*	*2.3*	*3.1*	*3.5*	*4.0*	*2.2*	*3.2*	*3.7*
Horizontal Accuracy	Worst-Trial (Mean)	21.1	24.7	25.9	28.6	26.5	26.4	26.2	25.3	25.1	25.9	25.5
	(STDev)	*8.2*	*11.3*	*9.5*	*7.7*	*8.0*	*8.2*	*8.2*	*7.3*	*9.1*	*9.8*	*9.6*
	Estimation (Mean)	17.6	16.6	18.1	22.7	17.5	18.7	19.0	19.2	20.1	21.7	21.9
	(STDev)	*7.5*	*8.7*	*9.7*	*7.6*	*9.7*	*7.3*	*8.9*	*6.6*	*7.5*	*7.4*	*8.4*
	Immediate (Mean)	19.6	27.3	24.8	24.1	23.4	20.7	26.3	22.3	23.5	23.9	23.6
	(STDev)	*5.3*	*5.2*	*8.2*	*7.9*	*9.7*	*11.9*	*9.3*	*11.6*	*10.0*	*12.5*	*9.8*
Velocity Score	Worst-Trial (Mean)	3.7	3.4	3.6	3.3	3.1	3.8	3.4	5.5	4.5	5.6	4.4
	(STDev)	*2.6*	*3.4*	*3.2*	*3.1*	*2.3*	*2.9*	*2.5*	*5.8*	*2.9*	*4.0*	*4.2*
	Estimation (Mean)	3.3	5.0	4.8	5.8	7.7	6.2	6.7	6.2	6.3	6.9	7.9
	(STDev)	*2.4*	*3.6*	*1.5*	*2.6*	*5.1*	*4.1*	*2.7*	*4.0*	*4.0*	*5.4*	*5.0*
	Immediate (Mean)	2.0	3.2	4.2	5.3	4.7	4.9	5.4	4.3	5.6	4.9	5.1
	(STDev)	*2.0*	*2.5*	*3.9*	*5.9*	*3.2*	*4.1*	*5.9*	*3.4*	*5.6*	*4.7*	*3.1*

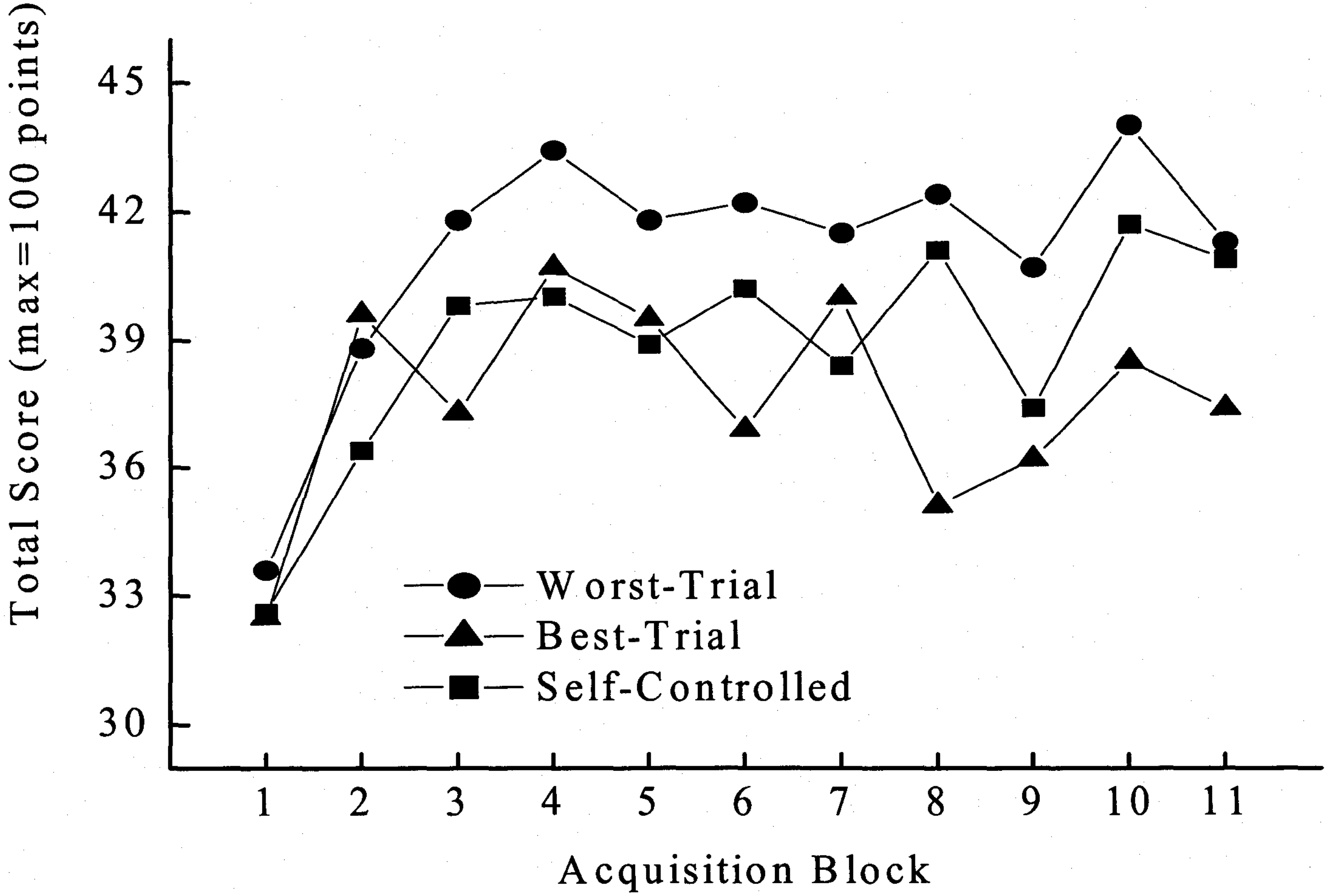
Table 4: Means and standard deviations of the total score, vertical accuracy score, horizontal accuracy score, and velocity score in retention and transfer for the worst-trial, estimation, and immediate groups in Experiment 2.

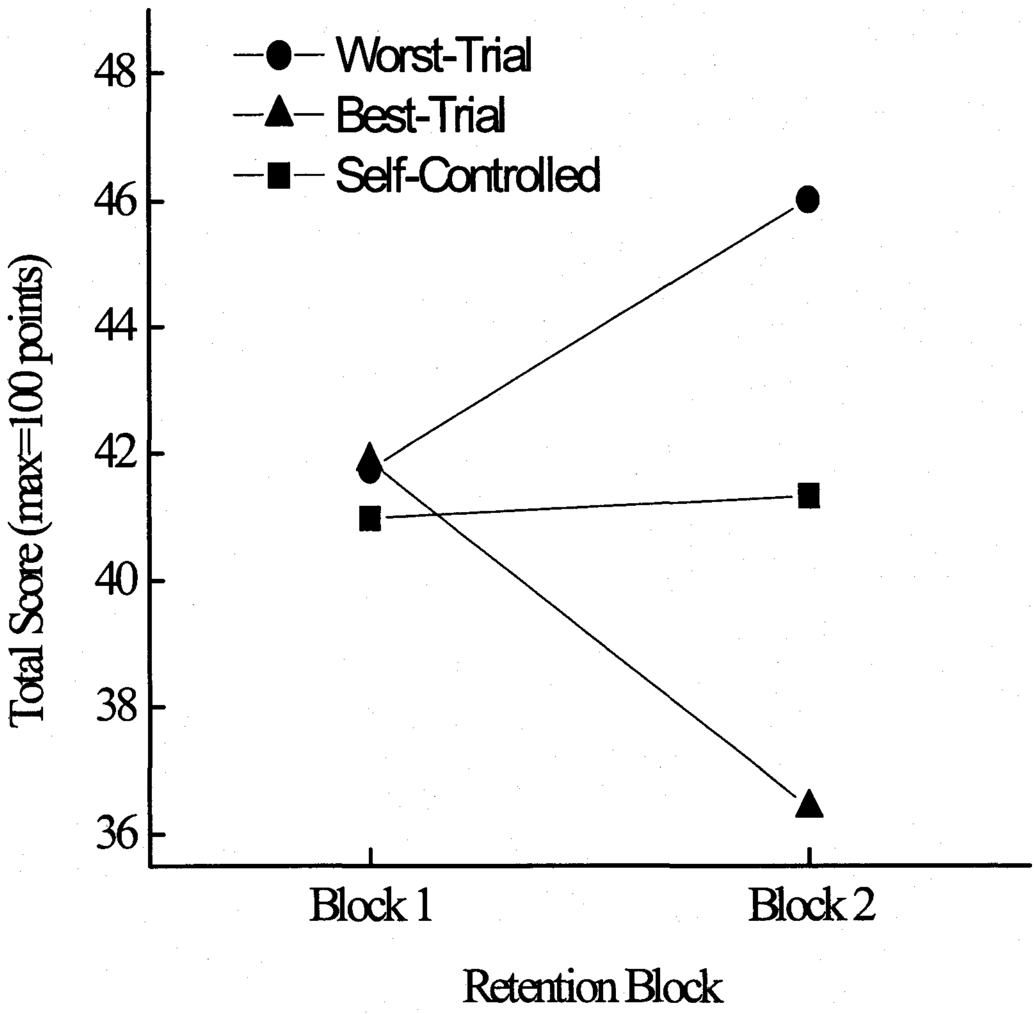
		Retention Block		**Transfer Type**			
		1	2	Up	Down	Left	Right
Total Score	Worst-Trial (Mean)	41.7	46.0	33.6	41.7	42.8	42.8
	(STDev)	*7.9*	*5.4*	*8.1*	*12.1*	*9.9*	*6.9*
	Estimation (Mean)	33.5	39.2	32.4	32.1	41.3	31.8
	(STDev)	*11.7*	*9.9*	*5.4*	*11.1*	*7.4*	*10.4*
	Immediate (Mean)	40.1	38.5	35.7	36.5	38.0	35.2
	(STDev)	*7.8*	*10.4*	*9.9*	*11.7*	*12.7*	*12.8*
Vertical Accuracy	Worst-Trial (Mean)	11.3	12.7	7.8	11.5	10.8	11.6
	(STDev)	*2.2*	*2.1*	*2.9*	*3.4*	*2.9*	*3.0*
	Estimation (Mean)	10.5	10.4	8.1	10.1	11.3	9.1
	(STDev)	*3.1*	*3.9*	*2.8*	*3.7*	*2.7*	*3.2*
	Immediate (Mean)	11.4	10.3	7.0	9.8	9.7	9.3
	(STDev)	*2.7*	*2.9*	*2.8*	*3.9*	*3.4*	*4.0*
Horizontal Accuracy	Worst-Trial (Mean)	25.9	28.0	19.9	26.6	28.7	27.2
	(STDev)	*10.6*	*9.0*	*9.8*	*11.8*	*10.3*	*10.1*
	Estimation (Mean)	17.0	19.9	18.4	18.6	23.4	18.3
	(STDev)	*8.0*	*8.5*	*5.4*	*8.8*	*7.5*	*8.6*
	Immediate (Mean)	23.7	23.1	21.5	24.3	24.4	22.7
	(STDev)	*8.6*	*11.2*	*8.9*	*10.7*	*11.6*	*10.8*
Velocity Score	Worst-Trial (Mean)	4.8	5.6	6.1	3.2	3.8	4.1
	(STDev)	*4.9*	*6.5*	*4.2*	*2.9*	*3.6*	*3.6*
	Estimation (Mean)	6.0	8.9	5.9	3.5	6.6	4.4
	(STDev)	*3.5*	*5.6*	*2.3*	*2.2*	*3.9*	*2.5*
	Immediate (Mean)	5.1	5.0	7.3	2.4	3.9	3.2
	(STDev)	*3.4*	*3.9*	*4.7*	*1.7*	*2.6*	*2.7*

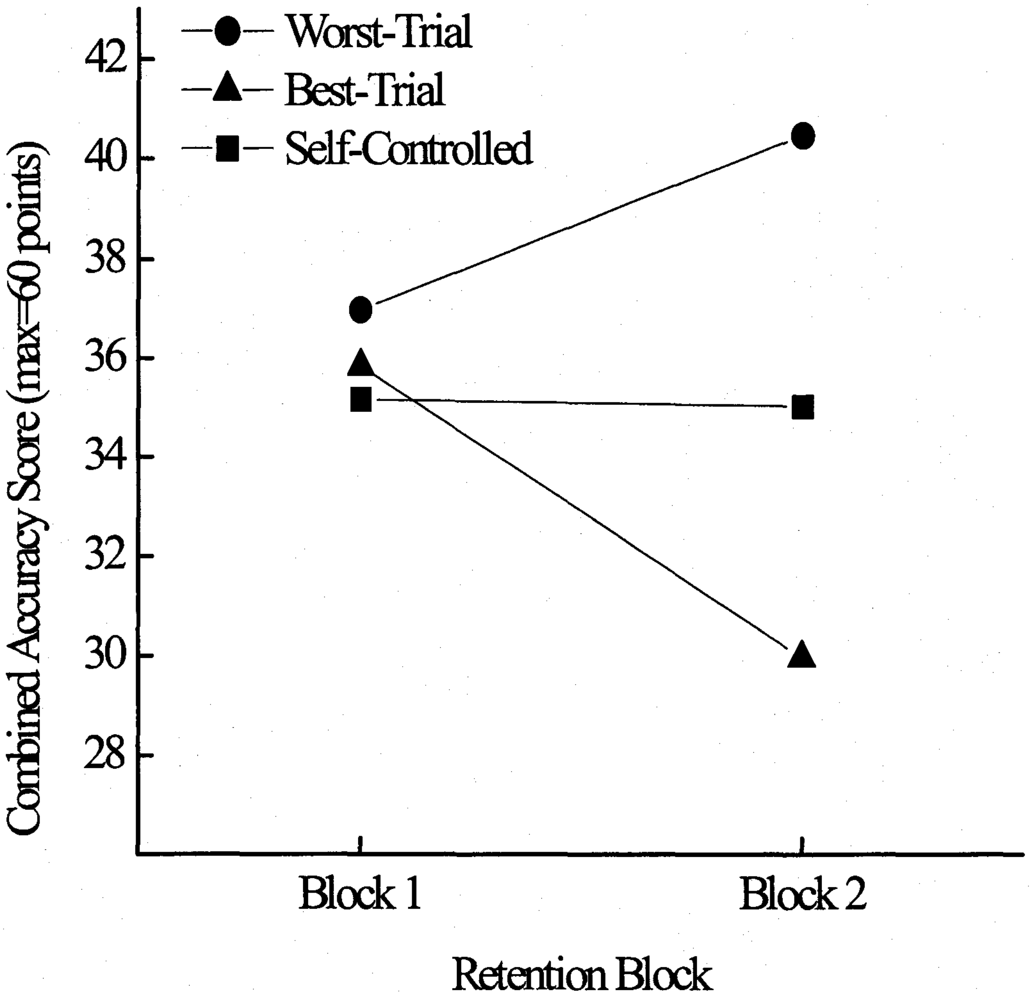

Minimum Reversal Line
Target Circle
Mouse Cursor
Interception Area

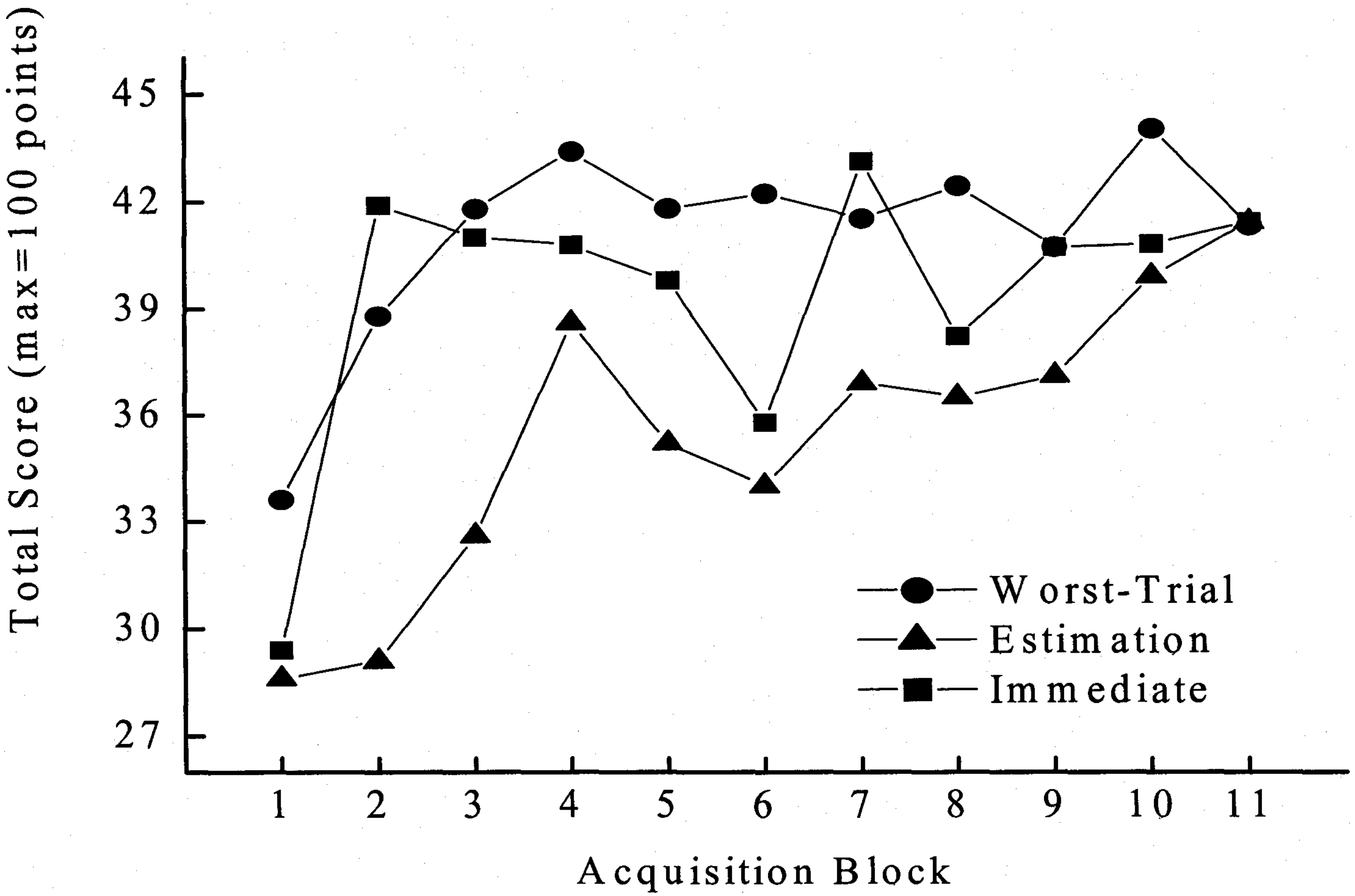


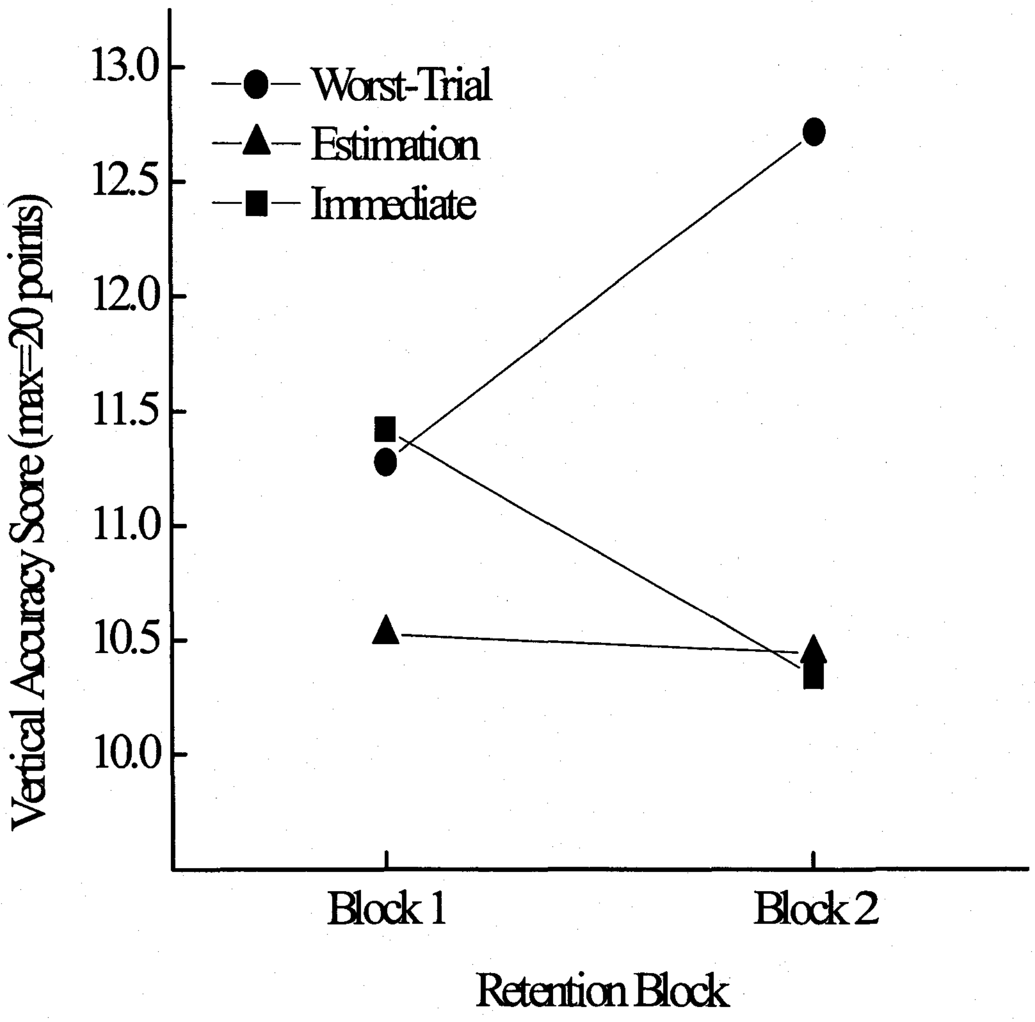

Intermediate
Worst
Best
Percent of Total Selection
100
90
80
70
60
50
40
30
20
10
1
2
3
4
5
6
7
8
9
10
11
Acquisition Block

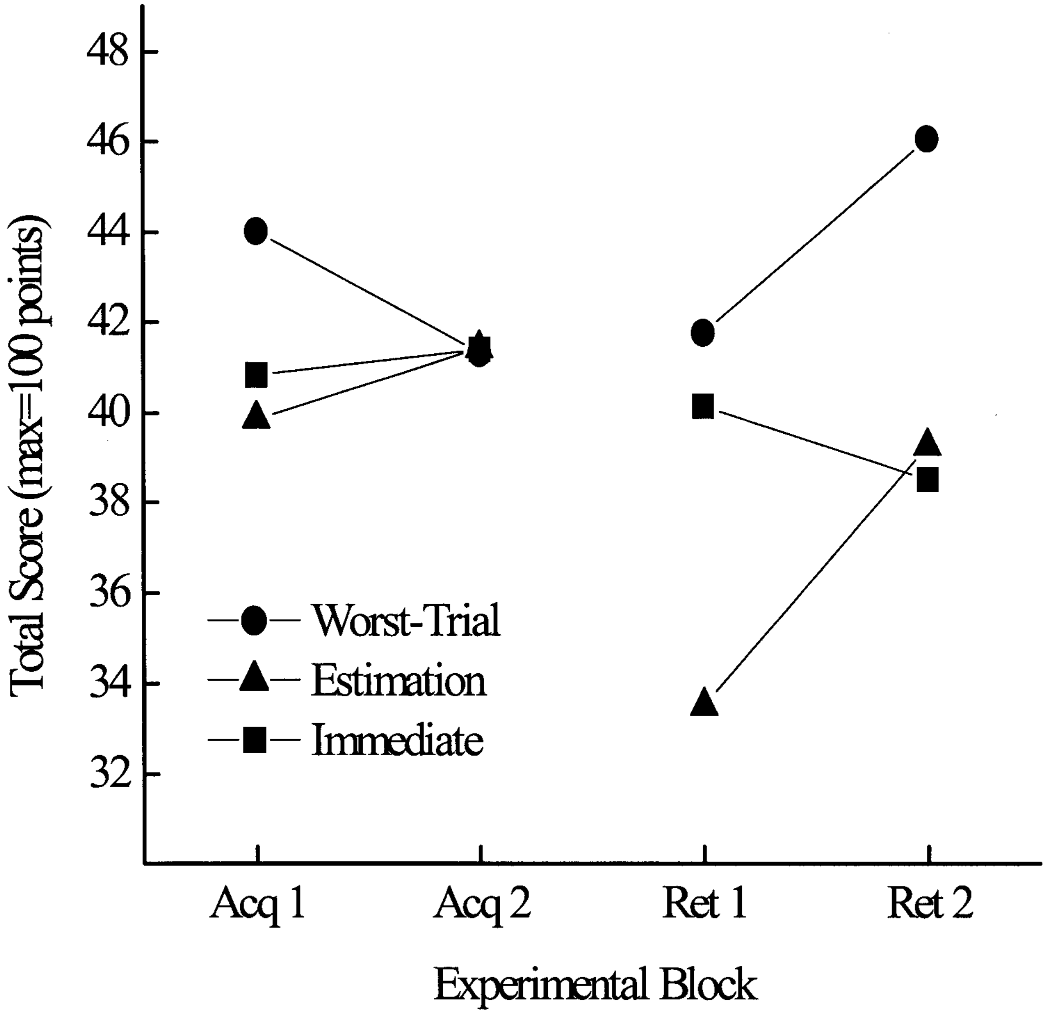

Total Score (max=100 points)
45
42
39
36
33
30
1
2
3
4
5
6
7
8
9
10
11
Acquisition Block
Worst-Trial
Best-Trial
Self-Controlled

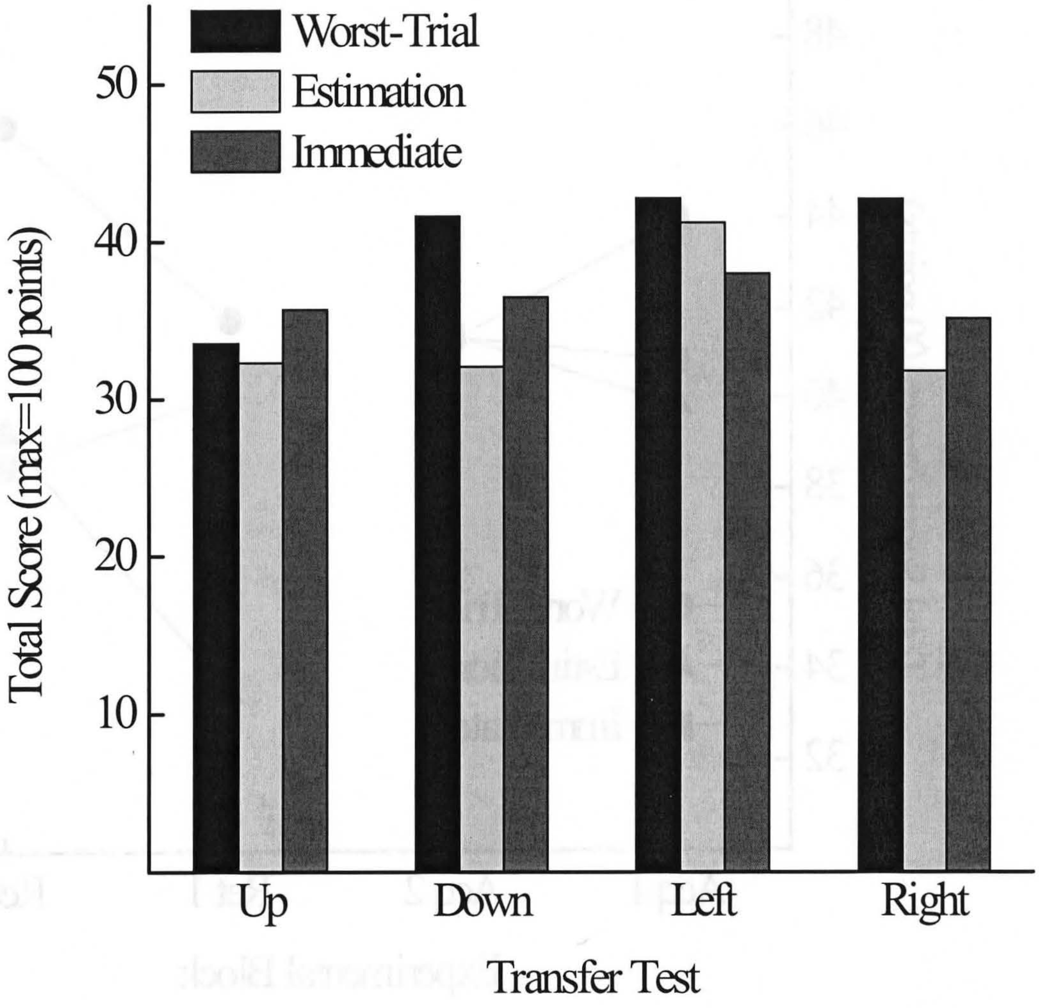

Worst-Trial
Best-Trial
Self-Controlled
Total Score (max=100 points)
48
46
44
42
40
38
36
Block 1
Block 2
Retention Block


Worst-Trial
Best-Trial
Self-Controlled
Combined Accuracy Score (max=60 points)
42
40
38
36
34
32
30
28
Block 1
Block 2
Retention Block


Total Score (max=100 points)
45
42
39
36
33
30
27
1
2
3
4
5
6
7
8
9
10
11
Acquisition Block
Worst-Trial
Estimation
Immediate


Worst-Trial
Estimation
Immediate
13.0
12.5
12.0
11.5
11.0
10.5
10.0
Vertical Accuracy Score (max=20 points)
Block 1
Block 2
Retention Block


48
46
44
42
40
38
36
34
32
Total Score (max=100 points)
—●— Worst-Trial
—▲— Estimation
—■— Immediate
Acq 1
Acq 2
Ret 1
Ret 2
Experimental Block


Worst-Trial
Estimation
Immediate
Total Score (max=100 points)
50
40
30
20
10
Up
Down
Left
Right
Transfer Test

2464 17