Learning a new human–computer alphabet:  
The role of similarity and practice  

Jae T. Patterson, Timothy D. Lee *

Department of Kinesiology, McMaster University, 1280 Main St. West Hamilton, Ont., Canada L8S 4K1

Received 28 February 2005; received in revised form 3 May 2005; accepted 4 May 2005

Available online 1 July 2005

Abstract

Two purposes motivated this study: (a) to quantify the difficulty in learning various symbols of the alphabet used to enter data into a personal digital assistant (PDA), and (b) to investigate the interaction of item difficulty with practice conditions that promote varying levels of cognitive effort. Levels of compatibility between members of the PDA alphabet and English were quantified through introspective ratings in Experiment 1 and objective performance measures in Experiment 2. Three levels of item compatibility were learned under conditions of proactive or retroactive augmented information in Experiment 3. Contrary to expectations, the item similarity effect did not interact with practice schedules—a retroactive augmented information condition resulted in degraded levels of acquisition performance, but superior retention levels, compared to the proactive condition. These findings are discussed in terms of the relative merits of cognitive effort in skill acquisition.

PsycINFO classification: 2340

Keywords: Learning; Human–computer interaction; Practice; Augmented information

* The research reported in this paper represents a portion of the Ph.D. thesis requirements conducted by the first author at McMaster University. The research was supported by an NSERC operating grant awarded to the second author.

* Corresponding author. Tel.: +1 905 525 9140x23579; fax: +1 905 523 6011.

E-mail addresses: jpatters@brocku.ca (J.T. Patterson), scapps@mcmaster.ca (T.D. Lee).

0001-6918/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2005.05.001
1. Introduction

Factors that impact the effectiveness of practice for novel motor tasks have been the focus of studies for almost a century. Early investigations considered factors such as the role of practice distribution (Browning, Brown, & Washburn, 1913), practice schedules (Pyle, 1919), and knowledge of results (Elwell & Grindley, 1938) in learning motor skills, and work on the effectiveness of these factors continues today. One suggestion about their effectiveness is that variables such as random practice (Shea & Morgan, 1979) and delayed augmented feedback (Swinnen, 1990), enhance the cognitive effort that is invested by learners as a result of the practice regime (Lee, Swinnen, & Serrien, 1994). Increased cognitively effortful practice tends to retard immediate performance improvements but facilitates learning (as measured in retention and transfer tests).

The role of cognitive effort in motor learning provides a satisfactory explanation for the effectiveness of some practice conditions, but numerous exceptions exist. A study by Albaret and Thon (1998) provides a good example. Their interest was the effect of contextual interference in motor learning, and the research protocol had groups of individuals practice sets of novel geometric patterns in either a blocked or random practice order. One pair of groups practiced tasks that had geometric patterns comprising of only two segments. Other pairs of random and blocked groups practiced three-segment or four-segment patterns. Albaret and Thon (1998) found that the random practice order facilitated learning the two- and three-segment patterns, but not the four-segment pattern.

One explanation for the Albaret and Thon (1998) results (and related sets of findings), suggests task complexity to be a key variable. The idea holds that more complex tasks, by their very nature, engage the learner in cognitively effortful learning processes (Guadagnoli & Lee, 2004; Wulf & Shea, 2002). Thus, practice factors that engage the learner in more cognitively effortful learning should be better suited for learning relatively simple tasks than for more complex tasks. Conversely, because of the cognitive effort required by the inherent difficulty of the task itself, more complex tasks may not benefit from these same factors, or perhaps may even benefit more from practice variables that reduce cognitive effort. Three experiments are reported here that examined these ideas. Experiments 1 and 2 were conducted in order to establish variations in the complexity of a commercially-available data-input task, using empirical means. Experiment 3 used these empirically-defined task complexity variations as the to-be-learned material, and reports a learning study using augmented information presented in two levels of cognitively effortful practice conditions. The goal of the research was to test the hypothesis that the effectiveness of cognitively effortful practice conditions is determined by the difficulty of the information to be learned (Guadagnoli & Lee, 2004; Wulf & Shea, 2002).

In the present research, task complexity was determined by the orthographic relatedness of the typographical script used to enter data into a PDA (personal digital assistant) and the English characters to which these scripts refer. Previous research has shown that the degree to which an environmental cue (e.g. PDA character) overlaps with an existing mental code in memory (e.g. English script character) defines its
conceptual compatibility (Sanders & McCormick, 1993). Further, the higher the conceptual compatibility, the faster and more accurate is the overt response (Alluisi & Warm, 1990; Eberts & Posey, 1990; Kornblum, 1992), and the easier to learn (Sanders & McCormick, 1993). Thus, the goal in the first experiment described here was to determine a range of conceptual compatibilities for PDA–English ensembles—conceptual compatibility levels that were predicted to modulate the cognitive and motor complexity of this data-input task.

Experiment 1 used introspective measures to categorize PDA script as being either high, moderate, or low in conceptual compatibility relative to its English referent. Using these introspectively-derived categories, participants in Experiment 2 provided objective measures of cognitive and motor performance when using a computer-based simulator to produce the novel PDA script. The results of Experiments 1 and 2 provided a reliable set of experimental materials that varied systematically in levels of conceptual relatedness (task complexity). These experimental materials were then used in Experiment 3, by groups whose practice conditions promoted low or high cognitive effort, to test the prediction that the effectiveness of these practice conditions would depend on the material to-be-learned.

2. Experiment 1

The goals in this first experiment were to collect introspective judgments of conceptual compatibility for English–PDA script pairs, and to categorize subsets of PDA characters as high, moderate, or low in conceptual compatibility. In this research, conceptual compatibility was operationally defined as the degree of spatial and motor relatedness between the tracing of the English-script and the corresponding PDA typographic symbol.

2.1. Method

2.1.1. Participants

Twenty (10 male and 10 female, mean age = 24.7 years old) right-handed volunteers from McMaster University completed a custom “Item Similarity” rating scale. All of the participants self-declared to have no experience using a PDA and no knowledge of the PDA symbols. They provided informed consent prior to their participation in the experiment and none received financial compensation.

2.1.2. The item similarity scale

The scale assessed each of the 122 PDA symbols. The scale was presented to participants on sheets of paper, with items presented on consecutive rows, divided into three columns (see Fig. 1). The far left column consisted of an English symbol (e.g., number, letter, or punctuation symbol). The middle column displayed the PDA symbol associated

with the English symbol. The right column displayed a 7-point Likert scale with the verbal anchors ‘very similar’ located below the number 1 and ‘very dissimilar’ located below the number 7. The numbers 2, 3, 4, 5 and 6 had no verbal label. Research evidence suggests that a 7-point Likert scale is the preferred number of response options, and that differences do not exist between scales with adjectives under each marking and scales with adjectives anchored at the ends (Streiner & Norman, 1995).

2.1.3. Procedure
For the initial test phase, participants were asked to trace, using a ballpoint pen, directly over the commonly used English symbol in column one, then over the PDA symbol in column 2 (a dot on the PDA symbol denotes the starting point—see examples in Fig. 1). After completing the second tracing, participants reported the perceived similarity of the tracing of the English symbol (column one) compared to the tracing of the PDA symbol (column two) on the 7-point Likert scale (in column three). Participants were informed that there was no correct or incorrect answer and were instructed to circle only one number. Approximately 20 min was required to rate all 122 items of the alphabet.

Fig. 1. Sample English–PDA pairs from the item similarity scale (Experiment 1).
One half of the participants (5 male, 5 female) were selected at random to complete the Item Similarity scale again, approximately one week later. The purpose of the second round of data collection was to gain measures of test–retest reliability of the scale. The procedure for the second test was identical to the first test.

2.1.4. Statistical analyses

The Wilcoxon Signed Ranks Test was used to assess the test–retest reliability. The Wilcoxon Signed Ranks Test, a nonparametric alternative to the repeated measures t-test, allows for the identification of items that demonstrate a statistically significant shift (in either direction) from one period in time to the next (Gravetter & Wallnau, 1996). This procedure assumes that the variables under consideration were measured on an ordinal scale and that the data are from a repeated measures design (Gravetter & Wallnau, 1996). The level of statistical significance was set at $p < 0.05$.

As a method of categorizing the compatibility of the English–PDA pairs, a comparison of the test data of the group tested once with the group participating in the retest was used to identify the English–PDA pairs sharing the same mode (i.e., the Likert scale value most frequently given for that particular PDA–English script ensemble). This was operationally defined as the common mode (ranging from 1 to 7). Ensembles that shared a common mode were organized into one of three categories: The ‘high compatibility’ English–PDA ensembles shared a common mode of either 1 or 2. The ‘moderate compatibility’ ensembles shared a common mode of 3, 4, or 5. Finally, the ‘low compatibility’ ensembles shared a common mode of 6 or 7.

2.2. Results

2.2.1. Test–retest reliability

A one-week interval resulted in no significant change in rating 98.4% (120/122) of the tested items ($p > 0.05$). Specifically, the Wilcoxon paired tests analyses indicated that of the 122 matched pairs; only two symbols demonstrated a statistically significant change from the test to the retest period. The symbols were the lower case “d” ($Z = 2.67, p = 0.0077$) and the “©” symbol ($Z = 2.023, p = 0.043$).

2.2.2. Item selection

A total of 36 English–PDA pairs were identified and categorized as high, moderate, or low compatibility ensembles, with 12 pairs per category (see Fig. 2). The Wilcoxon Matched Pairs indicated no statistically significant change ($p > .05$) from the test to the retest period for the categorized items, indicating stability of the ratings over the one-week delay.

2.3. Discussion

The notion of compatibility commonly addresses the fundamental issues associated with stimulus–response (S–R) coding (see Kornblum, 1992; Kornblum, Hasbroucq, & Osman, 1990, for review). The essence of S–R compatibility is that some tasks are easier to perform than others based on the pairing of individual stimuli and
responses (Kornblum et al., 1990). In fact, research has shown the compatibility effects of a specific stimulus and response ensemble are attributed to the interaction, and not the independent characteristics of the stimulus or the response (Fitts & Seeger, 1953; Kornblum, 1992; Kornblum et al., 1990; Kornblum & Lee, 1995). Therefore, the compatibility of the stimulus–response ensemble can vary in degrees of conceptual compatibility to the expectations, or pre-established mental codes of the performer (Kornblum, 1992; Kornblum et al., 1990). For example, a common training strategy involved in learning a second language involves identifying orthographic and phonological similarities in the two languages (DeGroot & Keijzer, 2000). When a conceptual overlap between the native language (stimulus) and the second language (response) exists, the stimulus serves as a strong cue for the retrieval of the corresponding translation (DeGroot & Keijzer, 2000). Second language words, that share an orthographic and phonologic overlap relationship with the native language words, are learned faster and are less susceptible to forget (DeGroot & Keijzer, 2000). Learning words with minimal orthographic and phonologic similarity to the users first language results in a very difficult learning event (Schneider, Healy, & Bourne, 2002). The purpose of Experiment 2 was to examine the motor and cognitive demands of novice users producing high, moderate, or low compatibility PDA symbols in a PDA simulator.

3. Experiment 2

The second step in establishing a set of stimuli that ranged in conceptual overlap was to acquire measures of performance. As stimulus–response pairs become less compatible, the performer must invest more cognitive effort to interpret the stimulus and retrieve the correct response information from memory (Fitts & Seeger, 1953; Kornblum et al., 1990; Sanders & McCormick, 1993). As a consequence, the performer commonly demonstrates slower motor planning time (e.g., RT) and increased errors in responding (Fitts & Seeger, 1953) as compatibility declines. Thus, one objective of the present experiment was to use measures of performance (i.e., RT and error rates) to examine the validity of the compatibility data obtained in Experiment 1.

In addition, the present study collected introspective measures from the performer regarding their perceived investment of cognitive effort in a performance instance. This method of quantifying perceived cognitive exertion has received minimal investigation in motor behaviour research (cf. Rosenbaum & Gregory, 2002). Requiring participants to provide an introspective rating of their perceived investment of cognitive effort would provide a better understanding of the relation between perceived cognitive and motor demands of the English–PDA pairs and performance observations.

Based on the previous S–R research, it was predicted that as the compatibility of the English–PDA pairs decreased (from high to moderate to low), participants would show a systematic slowing of RT and increases in error rate. It was also predicted that participants would provide introspective measures that were congruent with these performance measures. The performance data and introspective measures of
effort, together with the independent measures of compatibility from Experiment 1, were expected to produce a set of categorical stimuli that could reliably be considered as differing in conceptual similarity.

3.1. Method

3.1.1. Participants

Ten participants (5 male and 5 female), from McMaster University volunteered to participate in the study. All participants were screened prior to their participation to ensure none had ever used a PDA or had any knowledge of the PDA symbols. All participants were naïve to the purposes of the experiment, provided informed consent prior to the beginning of the experiment, and were unpaid volunteers. None had participated in Experiment 1.

3.1.2. Apparatus and materials

Two computers were used to conduct the experiment. One computer monitor, located directly in front of the participant, displayed a PDA simulator—a computer program that simulated the operations of a PDA. The PDA simulator was displayed on a 38-cm colour monitor, located approximately 60 cm from the participant, on a 31-cm × 31-cm base that was elevated 17 cm above a standard desk. The PDA simulator operated on a 486-64 MB computer. The PDA simulator had a visual display of 20 cm (length) × 14.5 cm (width). Participants interacted with the PDA simulator by writing with a wireless Intuos Stroke and Inking Pen (similar to the size of a standard pen) on a 30 × 46 cm WACOM digitizing graphics tablet, serially connected to a second computer. An 11-cm × 5.5-cm writing space on the graphics tablet was surrounded by a rubber border. The size of the experimenter-defined writing space corresponded exactly to the size of a 6.5 × 4 cm PDA simulator writing space, displayed in the lower half of the PDA simulator.

Each trial started from a circular (1.5 cm × 1.5 cm) “home” button, located approximately 5 cm below the centre of the defined writing space on the WACOM digitizing graphics tablet. This home button was serially connected to the data collection computer and interacted with the E-prime data acquisition program that presented various experimental events and recorded aspects of the participant’s responses. The E-prime program was customized to present a series of digital images to participants, one above the other, for a total 9.5 cm × 7 cm display size located in the center of the second computer monitor (having 43 × 43 cm dimensions), which was elevated similar to that of the other monitor.

Twelve high, 12 moderate, and 12 low compatibility English–PDA pairs were randomly presented three times each. The stimulus presentations were grouped in triplets of three pairs, resulting in a total of 36 trials. Within the three-pair stimulus, the location of each English–PDA pair differed on each presentation. Subjects were not informed of the repeated presentation of the pairs. An English–PDA pair of different compatibility was presented on each successive trial. Participants were not informed of the categorized compatibility of the pairs.
3.1.3. Procedure

The participants’ goal was to study three, English–PDA pairs of high, moderate, and low compatibility, then to recall these by producing the PDA symbol as quickly and accurately as possible when cued by the English symbol referent. A series of 36 of these “study-test” trials were performed in a single session, lasting about 45 min.

The pre-experiment protocol began with a series of instruction screens for the participant to read, followed by a series of practice trials using PDA symbols not later used in the experimental session. Participants practiced these symbols using the graphics pen. The final component of the pre-experiment practice required subjects to perform three practice trials of the experimental protocol and were invited to ask any question regarding the experiment.

An experimental trial began with subjects cued to depress the home button with the stylus. After a three second delay, subjects were presented a stimulus display consisting of either three high-, three moderate-, or three low-compatibility English–PDA pairs. Participants studied the three-pair display for a self-determined amount of time (hereafter called the ST, or “study time”) then indicated verbally to the experimenter when they believed that they had successfully memorized the three pairs. Following the ST, the experimenter removed the stimulus display and, after a 250 ms delay, presented one English symbol from the three-pair display as the “recall cue”. Upon seeing the recall cue, the task was to produce the corresponding PDA symbol by lifting the stylus from the home position (this period hereafter referred to as the RT, or “reaction time”) and writing the symbols in the defined writing space on the digitizing tablet. Once completed, the subjects again depressed the home button to complete the trial. The duration from button release to button press defined the MT, or “movement time”. The entire process of responding to the recall cue was to be performed as quickly and as accurately as possible.

Recent research has found that providing augmented feedback about an outcome, before an introspective measure of perceived effort is made, can bias the rating (Rosenbaum & Gregory, 2002, Experiment 1). Therefore, participants in the present experiment provided introspective effort ratings prior to the receipt of qualitative feedback regarding the success of their response. Upon completion of every response, participants were presented two Likert scales that requested an estimate of perceived cognitive effort. They first estimated the perceived amount of cognitive effort to plan their motor response, operationally defined as “the mental energy required to mentally organize the movements into a movement plan that included the mental effort employed in the study of the three pairs of symbols, and the recall of the PDA symbol from memory”. Second, participants were asked to rate their perceived level of cognitive effort to execute the motor response, operationally defined as “the mental energy required to physically execute the motor plan”. Using a numeric keypad to the left of the tablet, participants entered a number from one to seven, corresponding on the far left (numerical rating of one) to minimal cognitive effort and on the far right (numerical rating of seven) to maximal cognitive effort.

Upon completion of the introspective scales, subjects were provided verbal and visual KR regarding their motor performance. Verbally, subjects were informed that their motor response was ‘correct’ or ‘incorrect’. Visually, participants were able to
view their produced PDA character and the resulting English-script on the PDA simulator. If the produced symbol was performed correctly, the required English symbol was displayed on the PDA simulator. If the motor action was performed incorrectly but similar to that for a different English symbol, then that English symbol was displayed, indicating an incorrect response. Otherwise, no English character was displayed on the PDA simulator screen. The PDA simulator automatically provided the visual KR (knowledge of results).

To capture the type of error committed by the subjects, a third screen asked participants to characterize the success of their motor response in one of the three categories. If subjects correctly performed the motor response, subjects selected a ‘0’ on the numeric keypad. If the subject had recalled what PDA symbol to produce, but performed it incorrectly, the subject selected a ‘1’. Finally, if the subjects were unable to recall the required PDA symbol, they were required to select ‘2’. Completion of all three scales took approximately 1 min. Upon completing the scales, subjects were cued to depress the home button and prepare for the next three-pair presentation.

3.1.4. Dependent variables and data analyses

Measures of ST, RT, and MT were recorded for all 36 trials for each participant. Subjective measures of cognitive effort to plan and execute the motor response were recorded for all 36 trials. To summarize: ST was the self-selected time from the presentation of the three English–PDA pairs to the presentation of the single English stimulus character; RT was duration from the presentation of the single stimulus character to movement initiation; and MT was the interval from removal of the stylus from the home button to the completion of the motor response and depression of the home button to end the trial. Response success was recorded as either a correct response (0), a motor production error (1), or a cognitive retrieval error (2).

All measures were analyzed separately in one-way, repeated measures ANOVA (analyses of variance) for PDA symbol compatibility (high, moderate, low). A significance level of $p<0.05$ was used for all statistical tests. Post-hoc comparisons were conducted using a Tukey HSD.

3.2. Results

3.2.1. Objective performance measures

Participants demonstrated longer STs for items of low compatibility ($M = 15,801$ ms) and moderate compatibility ($M = 14,209$ ms) compared to high compatibility PDA symbols ($M = 11,689$ ms). However, these differences failed to reach statistical significance, $F(2, 18) = 2.55, p > 0.05$.

For RT, the ANOVA indicated a significant effect for compatibility, $F(2, 18) = 8.37, p < 0.05$. A post-hoc analysis indicated motor planning times were significantly slower for the low compatible PDA symbols ($M = 1867$ ms) compared to the high compatible PDA symbols ($M = 1230$ ms). The RT to moderate symbols ($M = 1521$ ms) was not different compared to the other two types.
The MT ANOVA also indicated a significant effect for compatibility, $F(2,18) = 6.95$, $p < 0.05$. The high compatible PDA symbols ($M = 3802$ ms) were performed faster than the low compatible PDA symbols ($M = 4336$ ms). Comparisons to the moderate compatibility PDA symbols ($M = 4268$ ms) were not statistically significant.

The response success ANOVA revealed that although participants were generally better in retrieving from memory and producing high compatible PDA symbols ($M = 0.76$) compared to moderate ($M = 0.71$) and low ($M = 0.63$) PDA symbols, the differences were not statistically significant, $F(2,18) = 2.77$, $p > 0.05$.

### 3.2.2. Subjective measures

Participants estimated that the cognitive effort to plan a motor response for high-compatible symbols ($M = 2.4$) was rated as less demanding than for moderate ($M = 2.9$) and low compatible PDA symbols ($M = 3.1$). The one-way ANOVA confirmed these differences as significant, $F(2,18) = 13.52$, $p < 0.05$. However, the post-hoc results indicated the cognitive effort ratings between the moderate and low compatibility PDA symbols were not statistically different.

Participants estimated that PDA symbols of high compatibility required less cognitive effort to execute the motor plan ($M = 2.1$) than moderate ($M = 2.5$) and low compatibility PDA symbols ($M = 2.7$). A one-way ANOVA indicated a main effect for compatibility, $F(2,18) = 7.34$, $p < 0.05$. The high compatibility PDA symbols were perceived to require significantly less cognitive effort to execute the motor plan compared to moderate and low compatibility PDA symbols (which did not differ in the post-hoc test).

### 3.3. Discussion

The superior performance (e.g., faster and more accurate responses) in planning and producing high-compatible PDA symbols is consistent with the notion of a direct transfer of information from the stimulus (e.g. English script symbol) to the response (e.g. PDA symbol) (Fitts & Seeger, 1953; Kornblum et al., 1990; Kornblum & Lee, 1995; Sanders & McCormick, 1993). However, the increased processing demands (e.g., RT) and response errors when producing moderate and low compatible ensembles supports the notion of an increased investment of cognitive effort required for encoding and executing a motor response (Alluisi & Warm, 1990; Eberts & Posey, 1990; Kantowitz, Triggs, & Barnes, 1990). In fact, according to Kornblum et al. (1990), presentation of a stimulus from a low compatible ensemble may result in the retrieval of an alternate response that is an incorrect conceptual match to the stimulus already existing as a memory code. Therefore, the performer must consciously inhibit this initially retrieved response, then retrieve a new response (Kornblum & Lee, 1995). As a consequence of this process, the performer sacrifices both speed and accuracy at the expense of encoding the information (Alluisi & Warm, 1990).

Learning research has inferred the degree of invested cognitive effort in a performance instance as a function of objective performance scores. However, this methodology has afforded an understanding of cognitive effort that is circular: slow or inaccurate levels of performance indicate a relatively difficult task, which requires greater cognitive effort for performance, resulting in slower and more inaccurate
performance levels. We attempted to circumvent this circularity by asking participants in Experiment 1 to provide introspective measures of the perceived congruency between the English–PDA pairs. Performance measures from a separate group of participants were obtained in Experiment 2 using these congruency categories. On the basis of these two experiments, we were now able to select an independently established set of stimulus items that varied in task complexity (English–PDA congruency), to be used in a learning experiment to examine the interaction of task complexity and the role of cognitive effort. This represented the goal of Experiment 3.

4. Experiment 3

Previous research suggests that: (a) practice factors differ in the amount of cognitive effort that is promoted in the learner, and (b) tasks of varying complexity differ inherently in the amount of cognitive effort required to perform them (e.g., Lee et al., 1994; Sherwood & Lee, 2003; Weeks, Hall, & Anderson, 1996). Further, previous research and theoretical arguments have suggested that, (c) practice conditions that promote more (compared to less) cognitive effort are likely to benefit the learning of inherently simple tasks, but that, (d) for complex tasks, practice conditions that promote greater cognitive effort are likely to be no more effective, or perhaps less effective, than conditions that promote less cognitive effort (e.g., Albaret & Thon, 1998; Guadagnoli & Lee, 2004; Wulf & Shea, 2002).

Experiments 1 and 2 in the present research established a range of task materials to empirically assess predictions (c) and (d) above. The practice conditions used to establish lower and greater cognitive effort demands are similar to that of a previous study that varied the temporal location of augmented information (Richardson & Lee, 1999; Weeks et al., 1996). Participants in the Richardson and Lee study, for example, attempted to learn the gestures of the American manual alphabet. In one practice condition, defined as the proactive timing of augmented information, participants observed a model perform manual gestures prior to the required motor response. Presenting the augmented information prior to the participants’ motor response required a minimal investment of cognitive effort because motor planning processes were obviated by the modeled solution currently present in short-term memory. In a second practice condition, augmented information was withheld until after the completion of the motor response, termed retroactive. In this condition, participants were required to formulate and execute a motor response prior to the receipt of the augmented, modeled task information. Thus, the retroactive practice condition required an increased investment of cognitive effort by the participant because information regarding the solution to the problem needed to be generated, rather than read out from the working memory. The results of this experiment revealed that the heightened cognitive effort observed under retroactive, compared to proactive conditions, was the determining factor in learning the manual gestures.

The purpose of Experiment 3 was to examine the effects of differing levels of task complexity in the acquisition and retention of the PDA symbols varying in degrees of conceptual compatibility. Similar to Richardson and Lee (1999), task difficulty was
defined by the temporal location of the augmented information (e.g., the complete English–PDA pair display), either before (e.g., proactive) or after (e.g., retroactive) the participant’s response. Based on previous research (e.g., Albaret & Thon, 1998) and theoretical arguments (Guadagnoli & Lee, 2004; Wulf & Shea, 2002), we predicted that the retroactive practice condition would be most effective for the learning of easy items (i.e., highly or moderately compatible English–PDA pairs), but that the proactive practice condition would benefit the learning of the most difficult items (i.e., the low compatibility pairs).

4.1. Method

4.1.1. Participants

Twelve participants ($M = 22.3$ years old), all self-declared right-handed, from McMaster University volunteered to participate in the study. All participants
verbally reported no experience using a PDA or any knowledge of the PDA symbols. Participants were naïve to the purposes of the experiment, and provided informed consent prior to their participation. They did not receive financial compensation for their participation and none had participated in either of the previous experiments.

4.1.2. Apparatus

Identical to Experiment 1.

4.1.3. Characteristics of the stimulus display

The objective measures of ST, RT and MT as well as the subjective measures of perceived mental and physical effort from Experiment 2 were used to create the stimulus sets. Six pairs, demonstrating the fastest objective and lowest subjective scores were defined as the high compatibility items; six pairs demonstrating median objective and subjective scores were defined as the moderate compatibility items; and six symbols demonstrating the slowest objective and highest subjective measures were selected as the low compatibility items for this experiment.

4.1.4. Procedure

The participants’ goal was to produce PDA symbols of high, moderate, and low compatibility as accurately as possible using the PDA simulator. Two practice conditions were created, which differed only in terms of when the complete English–PDA symbol pair was presented—either prior to (proactive) or after (retroactive) an attempted recall. A timeline of events for a trial in each group is presented in Fig. 2. For participants in the proactive group, a trial began with participants studying a complete symbol pair, then producing the PDA symbol on the digitizing tablet. Subjects in the retroactive conditions were required to produce a PDA symbol from memory before viewing the complete pair.

An experimental trial began with subjects cued to depress the home button with the stylus. After a 3 s temporal delay, subjects were presented with a stimulus display consisting of either a high, moderate, or low compatibility pair. A typical trial in the proactive condition required subjects to study one complete pair for a self-determined amount of study time. Once the participant verbally indicated ‘ready’, the monitor screen was blanked and followed 200 ms later by only the English referent as a cue for recall of the corresponding PDA symbol.

Participants initiated their response by lifting the stylus from the home position, then producing the PDA symbol on the digitizing tablet. The response was completed by returning to depress the home button. Participants were provided feedback regarding the success of their performance immediately following the depression of the home button. The subjects then completed the remaining 17 symbol pairs in immediate succession.

A typical trial for the retroactive group required participants to first view a single English character without the corresponding PDA symbol. After viewing this display, participants would leave the home button and attempt to recall the correct
The trial was completed when the home button was pressed again. Immediately after the home button was depressed, the complete English–symbol pair was presented and participants were allowed to study it for a self-determined duration (measured as the ST duration), after which feedback about the success of the trial was provided. The next trial began after the study period was complete.

Participants were tested individually and completed 6 trials of each of the 18 English–PDA pairs for a total of 108 trials in one testing session. The practice session lasted approximately 45 min. The instruction screens and experiment familiarization (e.g. practice trials) for participants were identical to Experiment 2. Retention tests were administered approximately 10 min (“immediate retention”) and 48 h (“delayed retention”) after the last acquisition trial. In each of the retention tests the participants were presented an English symbol and asked to recall the correct PDA by writing it on the digitizing tablet. Subjects were required to recall and produce one-half of the learned symbols (3 high, 3 moderate, and 3 low compatibility symbols) in each of the retention tests, counterbalanced across subjects. Qualitative augmented feedback was not presented to the participants in the cued recall tests.

4.1.5. Dependent variables

The dependent variables for the acquisition trials were ST, RT, and recall success. Only RT and recall success were measured in retention.

4.1.6. Data analyses

Analyses of variance were conducted on the acquisition data using a 2 (group: proactive/retroactive) × 3 (compatibility: high, moderate, low) × 6 (trials) model with repeated measures on the last two factors. A similar ANOVA model was used to analyze the retention data, together with the last trial in acquisition. A significance level of $p < 0.05$ was used for all statistical tests. Post-hoc tests (Tukey HSD test) were conducted on all main effects and interactions to determine differences between means.

4.2. Results

4.2.1. Acquisition

4.2.1.1. ST (study time). The time taken by the experimental groups to study the English–PDA pairs is presented in Fig. 3. Main effects were found for group, $F(1,10) = 7.84, p < 0.05$, symbol compatibility, $F(2,20) = 19.39, p < 0.001$ and trial, $F(5,50) = 23.84, p < 0.001$. These main effects were superseded by interactions between group and symbol compatibility, $F(2,20) = 4.56, p < 0.05$, and between group and trial, $F(5,50) = 6.19, p < 0.001$.

For the proactive group, participants voluntarily studied the high (3224 ms), moderate (3546 ms) and low (3746 ms) compatibility pairs for equivalent amounts of time. The retroactive group however, used less study time for the high compatibility pairs (4977 ms) than for either the moderate (6078 ms) or low (6397) compatibility pairs.

2 Note that the very first trial in the retroactive condition required the participant to guess the corresponding PDA symbol.
The group x trial interaction resulted from significant study time differences between the two groups until trial 5 in acquisition.

4.2.1.2. RT (reaction time). The RT taken to initiate a response to the presentation of the English referent is presented in Fig. 4. The ANOVA revealed main effects for group, $F(1,10) = 36.24$, $p < 0.001$, symbol compatibility, $F(2,20) = 11.93$, $p < 0.001$, and trial, $F(5,50) = 4.66$, $p < 0.01$. More importantly, the ANOVA also found significant interactions between group and symbol compatibility, $F(2, 20) = 12.41$, $p < 0.001$, and between group and trial, $F(5, 50) = 4.35$, $p < 0.01$.

The nature of the two interactions can be seen clearly in Fig. 4. No acquisition differences in RT to symbols of differing compatibility were observed for the proactive group. However, for the retroactive group, participants responded with longer RTs to the moderate (6060 ms) and low (5515 ms) symbols compared to the high compatibility symbols (3691 ms).

Similarly, no differences across trials were observed for the proactive group. However, an interesting pattern of results emerged for the retroactive group. Overall, RT on trial 2 was significantly higher than either trials 1 or 3. This finding was very likely the result of guessing on trial 1 (see footnote 1), which was undertaken with a small RT, followed by a more informed choice in trial 2 (involving a longer RT), and subsequent reductions in RT thereafter.

4.2.1.3. Success rate. The acquisition success rates are displayed in Fig. 5. The ANOVA indicated main effects for group, $F(1,10) = 22.27$, $p < 0.001$; symbol compatibility $F(2,20) = 29.48$, $p < 0.001$, and trial, $F(5,50) = 41.02$, $p < 0.001$. The group x trial interaction was also significant, $F(5,50) = 21.84$, $p < 0.001$. This interaction resulted
from significantly better recall performance for the proactive compared to the retro-active group on trials 1–5, but not on the last acquisition trial.
4.2.2. Retention

4.2.2.1. Reaction time (RT). The retention results are illustrated in the right side of each panel in Fig. 4. The ANOVA revealed two main effects and two interactions. Significant main effects were found for symbol compatibility, $F(2, 20) = 11.40, p < 0.001$, and trial, $F(2, 20) = 12.56, p < 0.001$. The interactions involved group and trial, $F(2, 20) = 9.83, p < 0.001$, as well as a three-way interaction between group, symbol compatibility and trial, $F(4, 40) = 3.50, p < 0.05$. The three-way interaction resulted from the significantly slower RTs for the retroactive group than the proactive group on trial 6 of acquisition, but significantly faster RTs in both retention tests. This pattern of results was true for the moderate and low compatibility symbols, but not for the high compatibility symbols, for which the proactive and retroactive groups showed similar RTs in retention.

4.2.2.2. Recall success. The results of the immediate and delayed retention tests are illustrated in the right side of each panel in Fig. 5. The ANOVA resulted only in two main effects: for group, $F(1, 10) = 17.08, p < 0.01$, and symbol compatibility, $F(2, 20) = 15.28, p < 0.001$. Of most importance, the findings revealed that the recall success for the retroactive group was significantly better than the proactive group, regardless of the symbol compatibility of the stimuli to be recalled.

4.3. Discussion

Experiment 3 examined the prediction that learning PDA symbols that varied in degrees of compatibility to the commonly used English symbol would differ as a function of the cognitive effort that was promoted by the conditions of practice. Specifically, we had predicted that high compatibility symbols would benefit most from conditions of highest cognitive effort (e.g., the retroactive group) and that low compatibility symbols would benefit most from conditions of low cognitive effort (e.g., the proactive group). These predictions were not supported by the results of the present experiment. For recall success, the high cognitive effort group (retroactive condition) retained the information better than the low effort group (proactive condition), regardless of the difficulty of the information to be learned. This general finding was true for RT as well, with the only exception being the highly compatible pairs, for which no RT difference between groups was found. However, this single exception was in a direction that was opposite to what had been predicted—that the benefit for conditions promoting high cognitive effort would be the largest for the most highly compatible symbols. Thus, the present findings are inconsistent with the arguments by Wulf and Shea (2002) and Guadagnoli and Lee (2004) regarding the efficacy of practice conditions relative to the information to be learned.

Throughout the acquisition period, presenting augmented information in a retroactive schedule resulted in longer RTs and diminished recall for all PDA symbols, regardless of compatibility. However, congruent with the performance—learning paradox, participants required to retrieve a motor plan from memory resulted in superior retention of the PDA symbols, contrary to the characteristics of the acquisition
data. Therefore, these results suggest that the cognitively effortful processes required in the retrieval of a motor plan for the retroactive condition facilitated a more permanent memory for the PDA symbols. Mental code permanence for the retroactive condition is suggested by the stability of the motor planning time and recall effectiveness over both retention intervals. In contrast, presenting augmented information proactively resulted in a weak memory representation, as revealed by the diminished recall success data, relative to the level that had been achieved in acquisition. These results are congruent with the results of Richardson and Lee (1999), who suggested that a proactive schedule inhibits the cognitively effortful processes required for the development of a strong memory representation.

5. General discussion

The purpose of the present research was twofold. First, the purpose of Experiments 1 and 2 was to quantify and categorize the perceptual–motor experiences associated with the symbols commonly used in human–PDA interactions. Second, research examining the role of practice conditions have recently suggested that task complexity is an important, modifying variable in learning motor skills (Albaret & Thon, 1998; Guadagnoli & Lee, 2004; Wulf & Shea, 2002). Therefore, the purpose of Experiment 3 was to explore this predicted interaction using the PDA symbols investigated in Experiments 1 and 2.

Traditional motor learning research has commonly defined task complexity based on the overt objective behaviour of the learner characterized by the performer’s movement errors or the temporal duration involved in planning (e.g. RT) or executing (e.g. MT) the motor response. The results of Experiment 1 contribute to the importance of using introspective measures as a method of facilitating a complete understanding of the cognitive and motor demands of a motor task being experienced by the performer. Overall, the results of Experiment 1 augment recent motor behavior research suggesting that performers have the ability to accurately introspect on the degree of invested effort in a performance instance (Rosenbaum & Gregory, 2002).

Contrary to the predictions of recent reviews and theoretical discussions, the retention data of Experiment 3 revealed no interactive effects of the practice orders with symbol compatibility. This finding may be related to findings in previous cognitive learning research. Specifically, the process of retrieval is complicated when a retrieval cue is linked to more than one item in memory, such that both items in memory compete with the single retrieval cue for access to conscious awareness (Anderson, Bjork, & Bjork, 1994). When a retrieval cue does not facilitate the activation of the appropriate mental code (e.g., in low compatible English–PDA pairs), the performer must consciously inhibit the response, and subsequently retrieve the correct response (Kornblum et al., 1990). This type of inhibition has been termed retrieval inhibition (Bjork, 1989). The behavioural consequence of retrieval inhibition, is increased time errors in responding (Fitts & Seeger, 1953; Kornblum, 1992; Kornblum et al., 1990; Kornblum & Lee, 1995).
However, the processes associated with cognitively effortful item retrieval is believed to be a potent learning process, such that repeated retrieval attempts facilitate the ease of future retrieval attempts (Anderson et al., 1994; Bjork, 1988, 1994). The absence of an interactive effect of symbol compatibility in Experiment 2 for the proactive conditions in the acquisition period suggests that having the information available in working memory at the time of “retrieval” circumvented the cognitive costs associated with the experimenter-defined conceptual compatibility of the PDA symbols. In contrast, for the retroactive conditions, the multiple attempted retrievals of the PDA symbols, regardless of compatibility and success rate, facilitated the eventual strength and retrievability of the memory code of the symbols.

When examining the structure of the repetition schedule in the retroactive conditions, participants experienced a temporal delay of intervening trials before having the opportunity to retrieve an updated mental code of the information. Perhaps this temporal delay before the subsequent repetition to produce the PDA symbols was not an optimal repetition schedule to facilitate learning. That is, it could well be the case that a retroactive repetition schedule would be facilitated by the opportunity to use the just-modeled information in an immediate retrieval attempt. According to the Challenge Point Framework, motor tasks that are inherently difficult place increased demands on the cognitive processes of the learner, independent of the repetition schedule. In fact, repetition schedules that further challenge the processing demands of the performer have the potential to overwhelm, rather than facilitate, the cognitive processes required for learning complex tasks (e.g., moderate and low compatibility symbols). Therefore, based on the tenets of the Challenge Point Framework, it could be predicted that participants receiving augmented information retroactively who are also afforded the opportunity to retrieve the updated mental code of the PDA symbol on an immediate repetition (e.g. paired repetition) will experience enhanced learning of the complex task content. Future research is required to understand the unique rules of retrieval practice for learning complex motor tasks.

Battig (1979) characterized memory research as the ‘psychology of memory for relatively easy items’ (p. 36). Battig (1979) also speculated that the beneficial use of a complex task is to make the task as similar as possible to that of memory tasks required in everyday living. More recently, it has been postulated that to better understand motor learning, theoretical constructs developed through simple motor tasks must be extended and investigated in complex, ecological motor tasks (Wulf & Shea, 2002). Further, when examining the learning of tasks commonly found outside the laboratory, the nominal overlap between the training and real-world environments is less important than the functional overlap regarding the processing required of the performer outside the laboratory (Bjork, 1994). In fact, practice factors facilitating the processes of memory retrieval are believed to be the functional overlap between the processing demands in the laboratory and the processing demands in tasks outside the laboratory setting (Bjork, 1994). The results outlined in the present document revealed that learning a motor task commonly used in daily living was facilitated by practice factors that simulated the processing demands experienced by the performer in a real-world setting. Further research examining the cognitive and motor demands of motor tasks commonly used in daily living is needed.
References


