AN ANALYSIS OF THE WORKED-EXAMPLE EFFECT USING ANIMATED DEMONSTRATIONS

Abstract

The purpose of this study was to consider animated demonstrations in the context of cognitive load theory and to determine if demonstration users would exhibit a delayed performance decrement, described as Palmiter's animation deficit. Relative condition efficiency (RCE) was measured, but in addition a new measure, Performance efficiency (PE) was developed. Given the worked example effect, it was hypothesized that the demonstration learners would out-perform those in a practice condition. Results found the animated demonstration groups assembled the week one problem in significantly less time than the discovery-practice group, providing positive evidence for the worked example effect. Significant results were also found for relative condition efficiency and performance efficiency. Thus the performances of the demonstration groups were found to be significantly more efficient than those of the discovery-practice group. Also, group performances did not differ a week later, providing no evidence of an animation deficit.

Keywords: animation, cognitive load theory, performance efficiency

AN ANALYSIS OF THE WORKED-EXAMPLE EFFECT USING ANIMATED DEMONSTRATIONS

Abstract The purpose of this study was to consider animated demonstrations in the context of cognitive load theory and to determine if demonstration users would exhibit a delayed performance decrement, described as Palmiter's animation deficit. Relative condition efficiency (RCE) was measured, but in addition a new measure, Performance efficiency (PE) was developed. Given the worked example effect, it was hypothesized that the demonstration learners would out-perform those in a practice condition. Results found the animated demonstration groups assembled the week one problem in significantly less time than the discovery-practice group, providing positive evidence for the worked example effect. Significant results were also found for relative condition efficiency and performance efficiency. Thus the performances of the demonstration groups were found to be significantly more efficient than those of the discovery-practice group. Also, group performances did not differ a week later, providing no evidence of an animation deficit.

Keywords: animation, cognitive load theory, performance efficiency

Introduction

Sweller and Cooper (1985) compared the performance of those who had learned by studying worked examples to those who learned while solving problems and found that those who studied worked examples had a decreased solution time and a decrease in the number of mathematical errors (Cooper & Sweller, 1987; Sweller & Cooper, 1985). This was later described as "the worked example effect" (Sweller & Chandler, 1991). While the worked example effect has been replicated many times and is the most studied cognitive load effect (Sweller, 2006), there are limitations to this effect. As learners become more competent, the worked example and other cognitive load effects begin to disappear (Kalyuga, Chandler, & Sweller, 1998). This change in learner behavior was termed the "expertise reversal effect," and researchers found that those with more expertise benefited more by solving problems (Kalyuga, Ayres, Chandler, & Sweller, 2003).

Educational researchers have been recommending that novices study worked examples during early schema acquisition (Jonassen, 1997; Sweller, 1988; Sweller & Chandler, 1991) and have even argued that novices study worked examples, rather than learn through discovery problem solving (Sweller, 1988; Touvinen & Sweller, 1999). Chandler and Sweller (1992) defined worked examples as "a problem statement and the appropriate steps to solution" (p. 233). More recently, Sweller and his colleagues were much more explicit and defined worked examples as "a step-by-step demonstration of how to perform a task, or how to solve a problem" (Clark, Nguyen, Sweller, 2006, p. 190). This paper proposes this definition may apply to animated worked examples, or as they are described in the literature, animated demonstrations (Lipps, Trafton, & Gray, 1998; Palmiter, 1991; Waterson & O'Malley, 1993). Yet, the literature concerning animated demonstrations often cites Palmiter's early studies, which describe animated demonstration groups as having a decreased performance in a delayed test (Bétrancourt & Tversky, 2000; Palmiter, 1993; Tversky, Morrison, & Betancourt, 2002). This effect was also described as an "animation deficit" (Lipps et al., 1998, p.1). Therefore, if animated demonstrations act as animated worked examples, this "animation deficit" is in direct conflict with Sweller and Chandler's "worked example effect."

While the subject matter of cognitive load studies was originally confined to science and mathematics, researchers eventually considered other types of problem solving and found that the worked example effect extends to other domains [e.g. music, chess, athletics (Atkinson, Derry, Renkl, & Wortham, 2000)]. This effect was even recently extended to ill-structured domains like art and design (Rourke & Sweller, 2009). The current study analyzed the performance of novices as they learned how to use a graphics editing program. Like Touvinen and Sweller (1999) and Palmiter (1991), this study compares several instructional strategies (and combinations of these strategies) to monitor novice behavior during procedural learning.

Modality and cognitive load

In the late 1990s, Mayer and his colleagues proposed the "cognitive theory of multimedia learning" (Mayer, 1997; Mayer, 2001; Mayer & Moreno, 1998), which relies on cognitive load theory, but also draws from dual-coding theory and constructivism (Mayer & Moreno, 2002). According to the cognitive theory of multimedia learning, humans process visual and auditory information during multimedia instruction; have a limited working memory; and learn by selecting, organizing and integrating information from the environment into their long term memories (Mayer, 2001).

Mayer and his colleagues have been quite prolific over the past two decades, and are well known for their contributions to the "modality effect" or "modality principle" (Mayer, 2001; Mayer, 2005; Moreno and Mayer, 1999). These researchers have conducted many empirical studies of animated instruction. A common theme among these studies is that the instructional materials revolved around the learner's application of principles portrayed in animated instruction of causal systems. In each case, learners were shown animated representations of how a dynamic system works. These systems included hydraulic braking systems (Mayer, Deleeuw, & Ayers, 2007), a bicycle pump (Mayer & Anderson, 1991; Mayer & Sims, 1994), and lightening formation (Mayer, Bove, Byman, Mars, & Tapangco, 1996; Mayer & Moreno, 1998) [A more thorough explanation of these studies is explained in Mayer (2009)]. In each case, learners were required to answer retention and transfer questions. The required performance was to respond to verbal statements and provide written responses about causal systems as opposed to how to perform a task, or how to solve a problem. Since the development of Mayer's cognitive theory of multimedia learning, many studies have followed Mayer's lead to consider animated instruction. However, learning extends beyond an understanding of causal systems.

Multiple forms of learning by animation

Researchers from various fields have explored human behavior to develop a biological model of learning, based on human anatomy, physiology, and genetics (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004; Squire, 2008; Clark & Clark, 2010). By the early 1990s, the physiological literature had defined two broad classes of learning as either procedural or declarative (Squire, 1992). Declarative learning is concerned with the learning of language-based information (e.g. facts and events), while procedural learning is skills-based learning (e.g. learning how to use a computer program) (Squire & Zola, 1996). While this distinction may seem to be an oversimplification from an educator's perspective, it is based on several decades of physiological research that dealt with the learning capabilities of brain injured patients, primates, and normal humans (Squire, 2008; Scoville & Milner, 1957).

If one were to consider the educational literature for a nomenclature of learning, they might choose Clark and Mayer (2008) who describe five types of learning (facts, concepts, procedures, processes, and principles). The literature describing these five types of learning also extends back several decades and relies on many contributors, most notably Merrill (1983). The nomenclatures of Squire and Zola (1996) and Clark and Mayer (2008) are compatible, and there is quite a bit of overlap between them. The underlying message of this discussion is that educators should not think of human learning or memory as a singular entity (Merrill, 1983), but be specific about what type of learning they are trying to encourage. In doing so, it will help to develop a productive set of theories and contribute to the development of the learning sciences.

Niajar (1996) made a distinction between two different types of animated instruction. He describes animated instruction as being either “animated explanation,” or “animated demonstration.” Mayer’s instructional materials are more aptly described as animated explanation (Niajar, 1996). Mayer describes his work well when he says his experiments ask questions about scientific explanation: “By ‘explanation’ we mean a description of a causal system containing parts that interact in a coherent way, such as a description of how a pump works or how the human respiratory system works” (Mayer & Sims, 1994, p.389). Clark and Mayer (2008) even describe “two different e-learning goals,” which teach learners to “inform and perform” (p.17). We agree with this distinction and terminology to propose learning “how to do” something (perform) can be taught via animated demonstration, as opposed to learning about something (inform), which is taught via animated explanation.

Articles from outside the instructional technology or educational psychology literature, also describe these forms of learning as being quite different. John Anderson’s ACT-R framework, for instance, provides a well-accepted theoretical basis for considering explanation and demonstration as two different forms of learning (Anderson, 1993), requiring different forms of instruction. While Anderson’s original model considered a declarative only origin of procedural skill acquisition (Anderson, 1976; Anderson, 1983), his later work found that skill acquisition may also develop from example-based processing, to eventually become rule-based processing (Anderson & Finchman, 1994; Anderson, Fincham, & Douglass, 1997). Thus examples may be initially represented in working memory, and then later, after practice, generalized into production rules (Anderson et al., 1997).

This study follows on the above literature to propose a dichotomy of animated instructional materials based on the learning involved, as being either declarative or procedural. Mayer’s animated explanations (declarative) require learners to encode information about how a system works. Animated demonstrations (procedural) are similar, but learners must consider rules based upon a sequence of actions, which they eventually must perform. Thus animated demonstrations may be an important, cognitively demanding form of instruction, providing a fertile area for cognitive load research.

Animated instruction for procedural learning

For the purposes of this study, an animated demonstration is defined as a narrated presentation that shows learners how to perform a computer-based procedure. Palmiter’s dissertation project is perhaps the most cited animated demonstration study (Palmiter, 1991). In a series of experiments, Palmiter compared the learner performance of those who studied narrated animated demonstrations with those who used text-based instruction (Palmiter & Elkerton, 1991; Palmiter, 1993). Palmiter's results are quite interesting, for she found that during a training session learners who studied animated demonstrations sped skill acquisition, and performed tasks in less time and more accurately than their peers using text-based instruction (Palmiter, 1991; Palmiter & Elkerton, 1991). These results are similar to those by Sweller and Cooper (1985) who demonstrated that studying worked examples requires "considerably less time to process than conventional problems, but that subsequent problems similar to the initial ones also were solved more rapidly"(p.59).

However, Palmiter noted that, one week later, learners using animated demonstrations took longer to perform tasks and reported that even though performance may have been quicker during the initial training session, their retention was lacking a week later. This delayed performance decrement was later described as an animation deficit (Lipps et al., 1998). However there is reason to question this decrement, since several other researchers have been unable to replicate Palmiter's findings (e.g. Cornett, 1993; Harrison, 1995; Lipps et al., 1998; Reimann & Neubert, 2000; Waterson & O'Malley, 1993).

Considerations for this study

Since Sweller (1988) published his initial article outlining cognitive load theory, constructivism has garnered a strong following. Constructivists soon began to promote the idea of activity as a means of learning with technology (Brown, Collins, & Duguid, 1989; Jonassen & Rohrer-Murphy, 1999). While it has long been known that activity can reinforce learning (e.g. Skinner, 1958), Sweller and his associates questioned the timing of that activity during early schema acquisition. Tuovinen and Sweller (1999) found evidence that worked example-based instruction was more beneficial for novices than discovery learning. Even though there is considerable evidence for the worked example effect (Sweller, 2006), some question this instructional technique and promote the idea of withholding assistance or guidance from learners (Koedinger & Aleven, 2007).

Given Sweller and Cooper's early results, and Mayer's work with declarative multimedia, it was hypothesized that a narrated, animated, worked example would promote procedural learning. This research should be testable and falsifiable for there is literature describing both a positive and negative effect given animated worked examples. The overall research question of the study is: How will animated demonstrations impact student learning? In order to answer this general question a number of key variables were examined. Sweller and Cooper's early worked example studies found differences in other outcome variables like completion times and the number of errors (Sweller & Cooper, 1985). The animated demonstration literature considers similar outcome variables, but followed Palmiter's lead, to describe these variables as performance time and accuracy (Waterson & O'Malley, 1993; Lipps et al., 1998; Palmiter & Elkerton, 1991). Therefore this study used these variables to consider the worked example effect given animated demonstrations.

Method

Participants

The participants of this study were undergraduate pre-service teachers (N=122) enrolled in an introductory instructional technology course at a large university in the southeastern United States. The mean age of this sample was 20.2 years (*SD*=2.89), and they were primarily female (93 females and 29 males). There were 14 freshmen, 60 sophomores, 42 juniors, and 6 seniors. Participants were randomly assigned to one of four instructional conditions, demo (n=33), demo+practice (n=29), demo2+practice (n=36), practice (n=24). An *a priori* power analysis for a four group MANOVA produced a sample size of n=115 participants. This number of participants is necessary to arrive a power of 0.80, with a small effect size η^2= 0.125, given α =0.05 (α=0.05 is used throughout this article, unless stated otherwise) (Stevens, 2002).

Materials

While the worked example effect is well documented, Koedinger and Aleven (2007) proposed this effect may be an artifact of poor control conditions. Schwonke, Renkl, Krieg, Wittwer, Aleven, & Salden (2009) considered this proposal, and found evidence that the worked example effect held, even when problem solving was well-supported by feedback and scaffolding. Therefore it is suggested that those considering the worked example effect present all learners with an identical instructional overview of the topic being considered, so that all learners are properly prepared.

Paas (1992) and Tuovinen & Sweller (1999) provided participants with an introductory overview of the subject matter. The current study continued this practice to develop an introductory overview. This overview was a narrated, non-animated, PowerPoint presentation (Microsoft, 2003) that was encoded for web presentation with TechSmith Camtasia Studio (TechSmith, 2006). It provided learners with an introduction to digital image editing with Adobe Photoshop Elements, by presenting learners with a narrated presentation of screenshots. Specifically, learners were shown how to select, move, rotate, and hide layers within an Adobe Photoshop Elements document. Thus all learners were shown the basic procedural steps required to complete the practice problem.

Two animated demonstrations were developed with TechSmith Camtasia (TechSmith, 2006). The first demonstration was identical to the week one practice problem (the mimicry condition). A second (varied context condition) used an animated demonstration that showed the construction of a photo collage document. Both were 10 minutes long and demonstrated the same underlying procedures, but within different contexts. The animated demonstrations showed learners how to select, move, rotate, and hide layers.

Two problem scenarios (the Mr. Potato head problem & the Picnic problem) were developed for this study. Each was an assembly task that required assembly within Adobe Photoshop Elements (Adobe Systems, 2002). The performance objectives of each problem required learners to select, move, rotate, and flip layers to produce a final product. During week one, three groups of learners put together the Mr. Potato head problem. One of the goals of this project was to measure what learners would retain and apply in a delayed performance. So, one week after initial instruction (week two), learners put together another Adobe Photoshop Elements document (the Picnic problem) that required learners to recall what they had learned one week prior.

Procedure

Before learners entered the test environment, TechSmith Morae Recorder (TechSmith, 2004) was turned on and was allowed to record all onscreen actions as learners interacted with the instruments and problem-solving scenarios. In addition, an image of the problem scenario was projected on a large screen in the room throughout problem construction.

Learners began by first completing an initial demographics survey. The action of submitting this survey launched the instructional overview. When the overview concluded, a JavaScript program randomly divided learners into one of four instructional conditions:

- demo - learners reviewed an animated demonstration, but did not practice with the “Mr. Potato head” problem;
- demo + practice - learners reviewed the identical animated demonstration, then practiced their newly learned skills with the “Mr. Potato head” problem;

- demo2 + practice - learners reviewed a different, collage-based animated demonstration and then practiced with the "Mr. Potato head" problem;
- practice - learners practiced with the "Mr. Potato head" problem, but did not review an animated demonstration.

Following the instructional conditions, learners were asked to complete a post-treatment survey. This survey included a relative condition efficiency question, "I invested:" with nine possible responses, from "very, very low mental effort" to "very, very high mental effort" as described by Paas and van Merriënboer (1993). Once learners had finished the post-treatment survey they were thanked for their participation, asked not to discuss the instruction and not to use the software before the delayed test session (one week later). Once learners had left the room, a researcher visited the learner's station and saved the recording for later processing.

A week later, during the delayed test session, all groups of learners solved the picnic problem. Again, an image of the completed product was projected on a large screen in the room throughout problem construction and learner on-screen actions were recorded. After attempting the picnic problem, learners took a second post-treatment survey, which also included a relative condition efficiency question.

Calculation

Paas and van Merriënboer (1993) derived the "relative condition efficiency" construct, which is composed of standardized mental effort ratings and performance scores (see Equation 1) and allows researchers to compare instructional conditions (Paas & Van Merriënboer, 1993). Learners rate their instructional conditions based upon their "perceived" mental effort during instruction. Later their performance scores and "perceived" mental effort ratings are combined in the relative condition efficiency construct (see Equation 1), and are presented graphically (see Figure 1).

Like Paas and van Merriënboer (1993), this study develops a new construct called performance efficiency (PE) (see Equation 2). Performance efficiency was developed to complement relative condition efficiency (RCE), and is used to study human behavior given instructional conditions. Performance efficiency is similar to relative condition efficiency (see Equation 2), but replaces the perceived mental effort rating with a more objective measure, performance time (compare Equations 1 and 2).

$$RCE = \frac{Z_{Performance} - Z_{MentalEffort}}{\sqrt{2}} \quad \textbf{(1)}$$

$$PE = \frac{Z_{Performance} - Z_{PerformanceTime}}{\sqrt{2}} \quad \textbf{(2)}$$

Since performance efficiency does not include a mental effort rating, it is not a measure of cognitive load; however like RCE, PE may also be used to describe group performances, relative to different instructional conditions but in a more objective manner. Performance efficiency contrasts instructional conditions in much the same manner, but combines two performance variables in a biplot, with group performance times and performance scores on the x and y axes (respectively). In keeping with the animated demonstration literature, performance in the above equations was described as accuracy.

Measurement of the dependent variables

This research was based in part, on a cognitive model from Human-Computer Interaction (HCI) research, the GOMS model (Card, Moran, & Newell, 1983). The GOMS model analyzes computer-based problems by reducing them to solution steps, known as problem solving operators (Dix, Finalay, Abowd, & Beale, 2003). Accuracy, one of the dependent variables, was measured by using a rubric or checklist, based upon the problem-solving operators of the problem being solved. Rubrics were developed, based on the problem scenarios, and constructed to automatically total the points scored in an Excel spreadsheet (see Table 1) (Microsoft, 2003). A researcher reviewed the Morae recordings of learner onscreen actions, and logged learner performances scores according to the rubric. Learners were either given credit for completing a problem solving operator or not. Successful completion of item movement or rotation was scored one point, and successful completion of layer placement and flipping was given two points.

In addition, performance time was measured in seconds with TechSmith Morae. This software allows a researcher to mark the timeline of a learner's onscreen recording and to designate the beginning of a performance, in this case with the first initial mouse movement. The end of the performance was also marked, and was designated as the point when the greatest number of problem solving operators had been completed.

Results

A MANOVA was used to analyze the two outcome variables, the performance time and accuracy of each performance. The assumptions of a MANOVA were analyzed for those learners who followed the instructions, completed all surveys, and attempted the required performances; this group was considered the sample (N=122). According to Glass and Hopkins (1984), these data met the independence assumption since the observations were independent of one another. A SAS macro %MULTNORM (SAS, 2007) revealed non-normality (violating the normality assumption), when the Shapiro-Wilks' W= 0.76 p<0.0001 for accuracy, and for performance time the W = 0.95, p=0.0015. Mardia skewness was found to be $\beta 1$,p= 146.5, p<0.0001 and kurtosis was $\beta 2$,p=12.01, p<0.0001 (Mardia, 1970). This violation was primarily due to a set of multivariate outliers (n=34), which were subsequently removed, by using the %MULTNORM macro. Next the data were transformed to test the "homoscedasticity" assumption. Box's M test (Box, 1954) was performed and X^2 (3, N=88) =4.50, p=0.21, φ=0.23, therefore, the variance-covariance matrices were not found to differ significantly, so there was no evidence that the homoscedasticity assumption was violated.

Recall that the demo group (n=19) did not practice during week one, so the performance of practicing learners (n=69) was analyzed. Group composition was demo+practice group (n=21), demo2+practice group (n=31), and practice group (n=17). Week one group performances (performance time and accuracy) were found to be significantly different, since Wilks' Λ = 0.68, F (2, 68) = 6.83, p <0.0001, η^2=0.32. Post hoc comparisons with Scheffé's test (p<0.025) revealed that both demonstration groups (demo+practice and demo2+practice groups) assembled the problem in significantly less time than the practice group.

The general procedure for analyzing relative condition efficiency (Paas & van Merriënboer, 1993) was used to compare groups. Group relative condition efficiency scores were compared with an ANOVA. Because the demo group did not have a week one performance, the performances of three practicing groups were compared. The assumptions of this ANOVA were analyzed and the data set met these assumptions, thus it was reasonable to consider the ANOVA. The ANOVA was conducted and revealed that there were significant differences between groups since F (2, 68) = 3.69, p=0.03.

Week one performance efficiency (see Equation 2) was also found to be significant, since F (2, 68) = 12.95, p<0.0001. Post hoc comparisons with Scheffé's test (p<0.05) revealed significant differences between groups, with the demonstration groups (demo+practice and demo2+practice) having more efficient performances than the practice group, given performance efficiency.

The week two performance was analyzed with a MANOVA to determine group differences a week after initial instruction. The results of the MANOVA found that there was not a significant difference given learner performance one week after initial instruction, since Wilks' Λ =0.96, F (3, 87) =0.64, p =0.70, $\eta 2$=0.04. During week two, relative condition efficiency was also considered, by combining the week two accuracy scores with the week two perceived mental effort ratings. Group z-scores were tested with an ANOVA and revealed that there were no significant differences between groups as F (3, 87) = 0.38, p=0.77. Week two performance efficiency was also calculated in a similar manner. An ANOVA for week two performance efficiency revealed no significant differences in group means, because F (3, 87) = 0.42, p=0.74.

Discussion

Learners who studied animated demonstrations assembled the week one problem in significantly less time than those who learned through discovery problem solving. This result is consistent with our initial hypothesis, the worked example effect and previous research (Sweller & Chandler, 1991; Tuovinen & Sweller, 1999), and it is further evidence of the worked example effect, given animated demonstrations. As expected, those learners who were provided with direct methods of instruction, in this case animated demonstrations, improved performance efficiency during early schema acquisition. Those who studied an animated demonstration, that was identical to the problem scenario (the demo+practice or mimic condition), had a lower performance time and increased accuracy over those in the practice group. In addition, those learners who studied a different animated demonstration (varied-context) significantly outperformed those who only learned through problem solving. Therefore, this study established animated demonstrations can indeed act as animated worked examples, and that they are an effective and efficient form of instruction. Instructional designers at companies like Bank of America, Amazon, and Microsoft, are all using narrated web-based animated demonstrations (demos) to train novices to use their online products and services. Therefore, it is the recommendation of this study that these developers continue to use this effective and now, evidence-based, e-learning strategy.

Group performances did not differ on the delayed test. Thus the demonstration learners retained their skills, and performed equally well one week after initial instruction. This finding does not support Palmiter's animation deficit, and is consistent with the results of other researchers (Lipps et al., 1998; Waterson & O'Malley, 1993). Tuovinen and Sweller (1999) proposed retention may be an issue given worked examples, and asked future researchers to consider the durability of learning given worked example-based instruction. This result provides some evidence for the durability of worked example-based instruction. Other researchers should also consider Tuovinen and Sweller's proposal.

However it should be clearly stated that the animated demonstrations used in this study were not like those of Palmiter's study, for they were narrated. Narration is heavily suggested given the modality and split-attention effects (Chandler, & Sweller, 1992; Moreno & Mayer, 1999; Penney, 1989; Mousavi et al., 1995). The addition of narration to an animated sequence is beneficial to learners and is described as the "modality effect" (Mayer, 2001). In an asynchronous e-learning environment, an instructor's role may be compromised because of an inability to communicate with an e-learner, but the use of a narrated animated demonstration allows that instructor to overcome that barrier to provide e-learners with guidance, support and "just-in-time" training. Palmiter may have been correct, asserting that only providing learners with animation produces mimicry, what Ausubel may have described as rote learning (Ausubel, 1963). The importance of adding narration to an animated demonstration should not be underestimated, for it promotes what Mayer (2001) described as "multimedia learning," now even with procedure-based instruction. Narrated animated demonstrations provide learners with a verbal narrative, which directs a learner's attention during the presentation, and more importantly, provides the learner with an expert level explanation.

Palmiter chose to study students who learned individual discrete tasks. Each of Palmiter's articles describe tasks (e.g. copy button or copy field) in isolation, and not in the context of an overall problem; whereas, the current study investigated learners *in situ*. That is, learner performance was studied given an authentic context, as they used their skills as a part of a larger project. This was necessary to gather data concerning a learner's cognitive load during actual problem solving, but also to measure learning in an ecologically valid manner.

Animated demonstrations present learners with a problem and a solution for solving that problem (the underlying schema). Those who use animated demonstrations rarely find a problem solution that exactly matches the problems they encounter, so they must extract a schema from an animated demonstration to later apply that schema in a new context. The results from the varied context condition (demo2+practice) showed learners do indeed extract that problem schema and improved their subsequent performance, to the point where their group performance was significantly different from those who discovered a problem solution. Given these results, Palmiter's critique of all animated demonstrations is unfounded.

It should be reiterated that the worked example effect has its limitations. Recall that Kalyuga et al. (2003) found evidence of an expertise reversal effect. In other words, those learners who had more experience with the subject matter benefited less from worked example-based instruction. Learners with more expertise benefited more from problem solving. This limitation has implications for animated demonstrations. Because of the expertise reversal effect, it would be expected that animated demonstrations would prove to be redundant and less useful for those with more expertise; perhaps this question can be answered by future researchers.

Relative condition efficiency and other similar metrics have become a basis for much of cognitive load measurement in instructional technology (Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Tuovinen & Paas, 2004), but this measurement relies on indirect or subjective measures, therefore it is unclear how this construct relates to actual cognitive load (Brünken, Plass, & Leutner, 2003). This study developed a new metric for measuring the efficiency of learner performance given an instructional condition. Performance efficiency is similar to relative condition efficiency, but does not include perceived mental effort in the equation. While perceived mental effort is related to performance, performance efficiency focuses on the efficiency of a performance. There probably is a relationship between the mental effort endured during an instructional condition and performance related outcome variables, but performance efficiency does not consider that relationship. It is only a way to compare and relate the performances of groups given instructional conditions. Performance efficiency is a useful metric, because it is generalizable. It is conceivable that any performance measure could be contrasted with its performance time, graphed and analyzed. Therefore, this metric may be used outside of the cognitive load literature to be considered by other groups of researchers. However, it has its limitations and should only be applied to procedure-based instructional conditions. It is hoped researchers will use both of these metrics to refine instructional materials and improve learner performance.

The finding of no significant differences a week after initial instruction may seem to be a reason to discount the need for animated worked examples. However the finding of improved performance efficiency during the initial training period is reason to consider this presentation strategy. An efficient performance is important for it showed this group of learners is functioning at a higher level, early in schema acquisition. This improved performance is evidence that this group of learners is better prepared. Therefore it could be argued that they are better able to learn additional procedures whereas, the practice group may very well be overloaded by additional training.

Another impact of this study is that the recording methodology may be generalized to any e-learning environment. Researchers may record learner on-screen actions to document how novices behave and react when tasked with an unfamiliar learning environment. In this study, recordings documented learner errors and the problem solving operators that they employed. Thus educational researchers may use recording technologies to consider learner performance, given various instructional strategies and environments. They may do so from a cognitive load perspective, to contrast the relative condition efficiency of instructional alternatives. Now, they may do so from a performance perspective, to contrast the "performance efficiency" of instructional strategies. It is hoped these tools and methodologies will be used together to promote the learning sciences.

In conclusion, it has been proposed that learners will be more productive if they teach themselves how to perform procedures (e.g. Bruner, 1961), but like Tuovinen and Sweller (1999), this study cannot support that position. This position diminishes the role of an instructor or instructional designer. These educators have purpose. As Kirschner, Sweller & Clark (2006) proposed, they provide guidance and support. Given the results of this study, it is more ethical for an instructor to take action and guide learners during early schema acquisition, rather than withhold that assistance to leave them to teach themselves.

References

Adobe Photoshop Elements (Version 2.0) [Computer program] Mountain View, CA: Adobe Systems

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y . (2004). An integrated theory of the mind. *Psychological Review 111* (4). 1036-1060. doi:10.1037/0033-295X.111.4.1036

Anderson, J.R., & Fincham J.M. (1994). Acquisition of procedural skills from examples. *Journal of experimental psychology: Learning, memory, and cognition, 20* (6) 1322-1340. doi:10.1037/0278-7393.20.6.1322

Anderson, J.R., Fincham, J.M. & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of experimental psychology: Learning, memory, and cognition*, 23(4) 932-945. doi:10.1037/0278-7393.23.4.932

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: a proposed system and its control processes. In K.W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol. 2* (pp. 89–195). New York: Academic Press.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. W. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*, 181–214. doi:10.2307/1170661

Ausubel, D.P. (1963). *The psychology of meaningful verbal learning; an introduction to school learning*. New York: Grune and Stratton.

Bétrancourt, M., & Tversky, B. (2000). Effect of computer animation on users' performance: a review. *Le travail Humain, 63*(4) 311-330.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. *Annals of Mathematical Statistics*, *25,* 290-302. doi:10.1214/aoms/1177728717

Brown, J.S., Collins,A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18* (1) 32-42. doi:10.2307/1176008

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*(1) 21–32.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning, *Educational Psychologist, 38*(1), 53–61. doi:10.1207/S15326985EP3801_7

Card, S., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction.* Lawrence Erlbaum Associates, Hillsdale, NJ

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293-332. doi:10.1207/s1532690xci0804_2

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233-246.

Clark, R.E., & Clark, V. P. (2010). From Neo-Behaviorism to neuroscience: Perspectives on the origins and future contributions of cognitive load research. In J. Plass, R. Moreno, & R. Bruken (Eds.), *Cognitive load: Theory and application*. New York: Cambridge University Press.

Clark, R.C., & Mayer, R. E. (2008). *E-learning and the science of instruction* (2nd ed). San Francisco: Jossey-Bass.

Clark, R.C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: evidence-based guidelines to manage cognitive load*. San Francisco: Pfeiffer.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79(4), 347-362. doi:10.1037/0022-0663.79.4.347

Cornett, L. L. (1993). *Animated demonstrations versus text: a comparison of training methods* (Unpublished Master's Thesis). Rice University, Houston, TX.

Dix, A., Finlay, J., Abowd, G. & Beale, R. (2003). *Human-Computer Interaction*. (3rd ed). Essex: Pearson Education Limited

Harrison, S. M. (1995). A comparison of still, animated, or nonillustrated on-line help with written or spoken instructions in a graphical user interface. In I. R. Katz, R. Mack, L. Marks, M. B. Rosson, & J. Nielsen (Eds.), *Proceedings of the ACM conference on human factors in computing systems* (pp. 82-89). Denver, CO: Association for Computing Machinery.

Jonassen, D.H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development, 45*(1), 65–94. doi: 10.1007/BF02299613

Jonassen, D.H & Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments. *Educational Technology Research and Development, 47* (1) 61-79 doi: 10.1007/BF02299477

Kalyuga, S. Ayres, P., Chandler, P. & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1) 23–31. doi:10.1207/S15326985EP3801_4

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40* (1) 1-17. doi:10.1518/001872098779480587

Kirschner, P. A., Sweller, J., and Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist. 41*(2) 75-86. doi: 10.1207/s15326985ep4102_1

Koedinger, K. R., & Aleven, V. (2007). Exploring the Assistance Dilemma in Experiments with Cognitive Tutors. *Educational Psychology Review 19*(3) 239–264 doi: 10.1007/s10648-007-9049-0

Lipps, A., Trafton, J., & Gray, W. (1998). Animation as documentation: A Replication with Reinterpretation. In the *Proceedings for the 45th annual conference of the Society for Technical Communication Arlington, VA*. Retrieved from http://www.stc.org/proceedings/ConfProceed/1998/PDFs/00006.PDF

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3) 519-530. doi:10.1093/biomet/57.3.519

Mayer, R.E. (1997). Multimedia Learning: Are We Asking the Right Questions? *Educational Psychologist, 32*(1) 1-19. doi:10.1207/s15326985ep3201_1

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge: Cambridge University Press

Mayer, R.E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist. 59*(1) 14-19. doi: 10.1037/0003-066X.59.1.14

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed). New York: Cambridge University Press.

Mayer, R. E., & Anderson, R.B. (1991). Animations need narrations: an experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*(4) 484-490. doi:10.1037/0022-0663.83.4.484

Mayer, R.E., Bove, W., Bryman, A., Mars, R. and Tapangco, L. (1996). When less is more: meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88*(1) 64-73. doi:10.1037/0022-0663.88.1.64

Mayer, R.E., & Chandler, P. (2001). When learning is just a click away: does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93* (2) 390-397. doi:10.1037/0022-0663.93.2.390

Mayer, R.E., Deleeuw, K.E., & Ayers, P. (2007). Creating retroactive and proactive interference in multimedia learning. *Applied Cognitive Psychology, 21* 795–809. doi: 10.1002/acp.1350

Mayer, R. E., & Moreno, R. (1998, April). A Cognitive Theory of Multimedia Learning: Implications for Design Principles . Paper presented at the annual meeting of the ACM SIGCHI Conference on Human Factors in Computing Systems, Los Angeles, CA. Retrieved from http://www.unm.edu/~moreno/PDFS/chi.pdf

Mayer, R.E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1) 107–119. doi:10.1016/S0959-4752(01)00018-4

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 84, 389–401. doi:10.1037/0022-0663.86.3.389

Merrill, M.D. (1983). Component Display Theory. In C. Reigeluth (Ed.), *Instructional Design Theories and Models*. Hillsdale, NJ: Erlbaum Associates.

Microsoft Excel 2003. (Version 11)[Computer program]. Redmond, WA: Microsoft

Moreno, R., & Mayer, R. E. (1999). Multimedia-supported metaphors for meaning making in mathematics. *Cognition and Instruction, 17*(3) 215–248. doi:10.1207/S1532690XCI1703_1

Mousavi, S.Y., Low, R. & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87(2) 319-334. doi:10.1037/0022-0663.87.2.319

Najjar, L. J. (1996). *The effects of multimedia and elaborative encoding on learning* (GIT-GVU-96-05). Retreived from Georgia Institute of Technology, Graphics, Visualization - Usability Center. http://hdl.handle.net/1853/3491

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal Educational Psychology, 84* (4), 429-434. doi:10.1037/0022-0663.84.4.429

Paas, F. G. W. C., & van Merrienboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*(4), 737-743.

Paas, F., Tuovinen, J. E., Tabbers, H. K., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38* (1), 63–71. doi:10.1207/S15326985EP3801_8

Palmiter, S. (1991). *Animated Demonstrations for learning procedural, computer-based tasks*. (Doctoral Dissertation). Available from ProQuest Dissertation and Theses database. (UMI No. 9124076)

Palmiter, S. (1993). The effectiveness of animated demonstrations for computer-based tasks: a summary, model and future research. *Journal of Visual Languages and Computing, 4* 71-89. doi:10.1006/jvlc.1993.1005

Palmiter, S., & Elkerton, J. (1991). An evaluation of animated demonstrations of learning computer-based tasks. *In the Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*. New Orleans, Louisiana, USA. p. 257 – 263. doi:10.1145/108844.108906

Penney, C.G. (1989). Modality effects and the structure of short-term memory. *Memory and Cognition 17*(4) 398-442. Retrieved from http://www.psychonomic.org/backissues/7537/mc/Issue%204/mc-17-398.pdf

Reimann, P. & Neubert, C. (2000). The role of self-explanation in learning to use a spreadsheet through examples. *Journal of Computer Assisted Learning, 16*(4), 316-325. doi:10.1046/j.1365-2729.2000.00144.x

Rourke, A., & Sweller, J. (2009). The worked example effect using ill-defined problems: learning to recognise designers' styles. *Learning and Instruction, 19*(2) 185–199. doi:10.1016/j.learninstruc.2008.03.006

SAS (2007). Macro to test multivariate normality. [Computer program]. Retrieved from http://support.sas.com/kb/24/983.html

Schwonke, R., Renkl, A., Krieg C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior, 25*(2) 258–266. doi:10.1016/j.chb.2008.12.011

Scoville, W.B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20, 11-21. doi:10.1136/jnnp.20.1.11

Skinner, B. F. (1958). Teaching machines: From the experimental study of learning come devices which arrange optimal conditions for self-instruction. *Science, 128* (3330) 969-977. doi:10.1126/science.128.3330.969

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences, 93*, 13515-13522.

Squire, L.R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99* (2) 195-231. doi:10.1073/pnas.93.24.13515

Squire, L.R. (2008). Introduction to Memory Section. In Pomerantz, J.R. (Ed.) Topics In *Integrative neuroscience: from cells to cognition*. (pp. 239-242) Cambridge: Cambridge University Press

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257-285. doi:10.1207/s15516709cog1202_4

Sweller, J. (2006). The worked-example effect and human cognition. *Learning and Instruction, 16*(2) 165-169. doi:10.1016/j.learninstruc.2006.02.005

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction, 8*(4) 351-362. doi:10.1207/s1532690xci0804_5

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59-89. doi:10.1207/s1532690xci0201_3

TechSmith Morae (Version 1.0.1) [Computer program] Okemos, MI: TechSmith

TechSmith Camtasia Studio (Version 4.0) [Computer program] Okemos, MI: TechSmith

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*(2), 334-341. doi:10.1037/0022-0663.91.2.334

Tuovinen, J. E., & Paas, F. G. W. C. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science, 32*(1-2) 133–152. doi:10.1023/B:TRUC.0000021813.24669.62

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology*, 91(2), 334-341. doi:10.1037/0022-0663.91.2.334

Tversky, B., Morrison, J. B., & Betancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human–Computer Studies, 57*, 247–262. doi:10.1006/ijhc.2002.1017

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction, 7*(1), 1-39. doi:10.1207/s1532690xci0701_1

Waterson, P. & O'Malley, C. (1993). Using animated demonstrations in multimedia applications: Some suggestions based upon experimental evidence. In G. Salvendy & M. Smith (Eds.), *Human-Computer Interaction: Software and Hardware Interfaces, Proceedings of the 5th International Conference on Human-Computer Interaction* (August 8-13, Orlando, Florida) Vol. 2 (p. 543-548). Amsterdam: Elsevier.

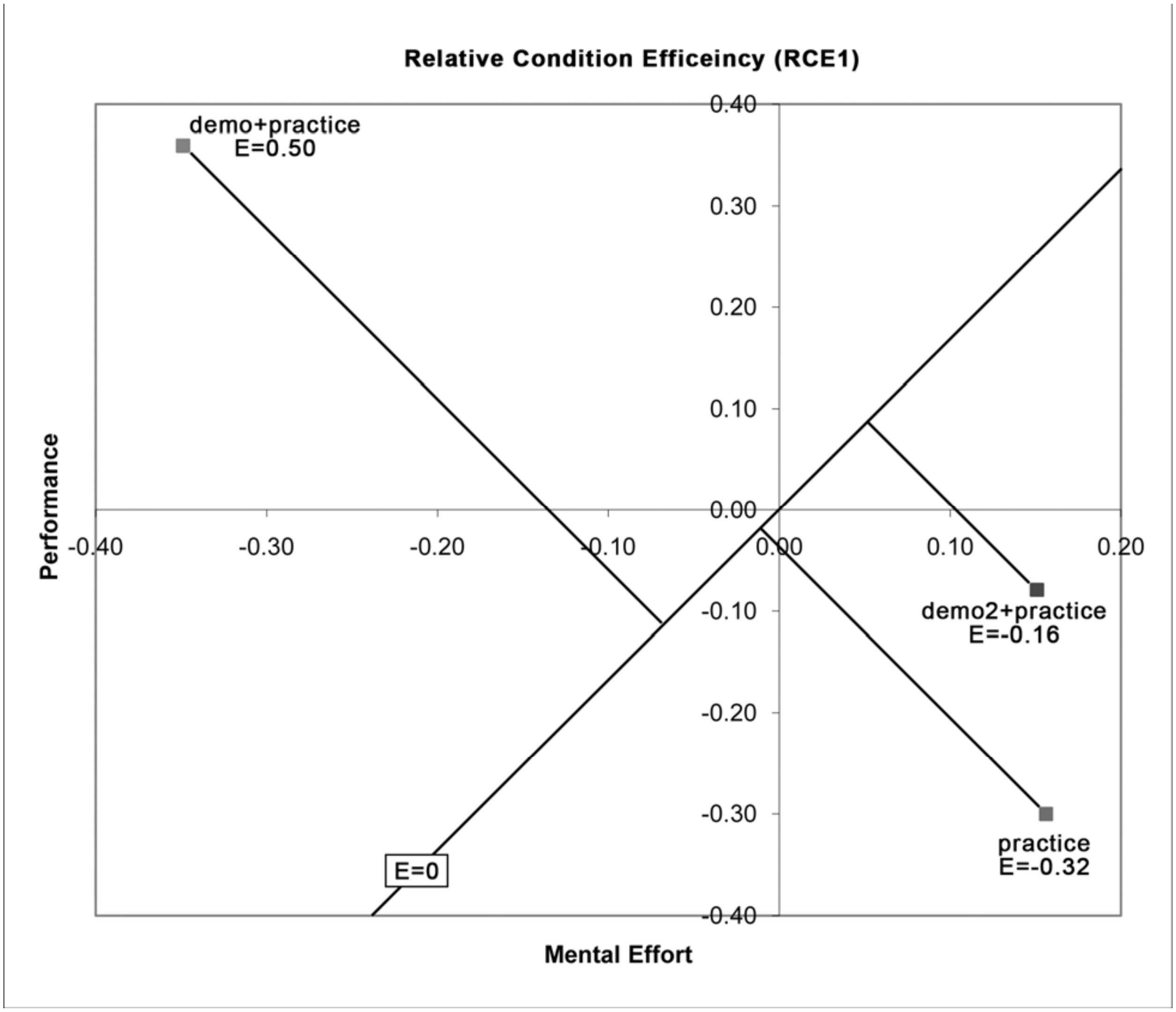

Figure 1. Week 1 Relative Condition Efficiency (RCE1)

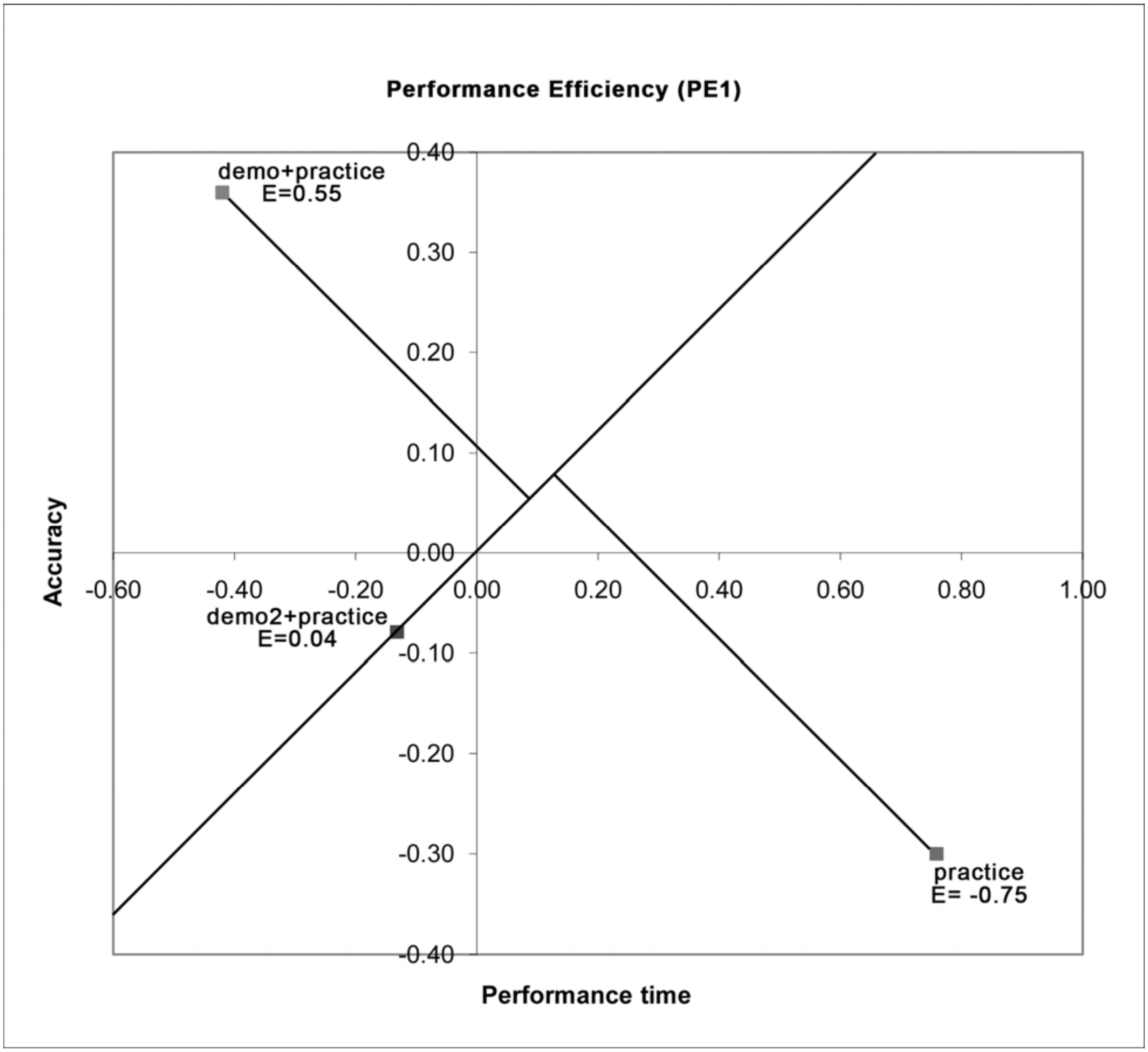

Figure 2. Week 1 Performance Efficiency (PE1)

Table 1

Picnic problem accuracy rubric

flip	**layer**	**rotate**	**move**	**item**
	***			umbrella
	***			tshirt
***	***			head
***	***	***		right leg
***	***			head 2
***	***			purple shirt
***				hat
***	***	***		s left leg
***	***	***		bent right leg
***	***	***		left leg
***				green shorts
***	***	***		arm 2
***				pink shorts
***	***	***		left arm
***	***			body
***	***			picnic basket
***	***	***		arm
***	***			right arm
***	***			torso
***	***			table
***	***			bird3
***	***	***		bird2
***	***			bird1
0	0	0	0	0