Animated Demonstrations: Evidence of Improved Performance Efficiency and the Worked Example Effect

David Lewis and Ann Barron

University of South Florida,
13201 Bruce B. Downs Blvd MDC56, Tampa FL, 33612
dlewis3@health.usf.edu, Barron@tempest.coedu.usf.edu

Abstract. The purpose of this study was to assess the efficiency and effectiveness of animated demonstrations, to determine if those using animated demonstrations would exhibit the worked example effect [1], and a delayed performance decrement, described as Palmiter's animation deficit [2], [3]. The study measured relative condition efficiency (RCE) [4] and developed a construct called performance efficiency (PE). Results revealed the animated demonstration groups assembled the week one problem in significantly less time than the practice group, providing evidence for the worked example effect with animated demonstrations. In addition, subjects from the demonstration groups were significantly more efficient (given performance efficiency) than those from the practice group. Finally, group performance did not differ a week later, providing no evidence of Palmiter's animation deficit.

Keywords: Animation, cognitive load, performance efficiency.

1 Introduction

Sweller (1988) developed cognitive load theory, a theoretical framework for describing the actions of novices during problem solving. Sweller and his associates had found that those learners who studied worked examples outperformed those who learned by solving problems [5]. This was later described as the "worked example effect" [1]. This instructional strategy is recommended, as opposed to allowing learners to only learn through discovery problem solving [6]. For those who study human computer interaction, this may seem counter-intuitive, for we often try to "figure out" how to use software. In a computer environment, worked examples are commonly described as "demos" or animated demonstrations, and over the past few years, software providers (e.g. Microsoft), have begun using web-based animated demonstrations (demos) as a part of their products as documentation and support.

Lewis [7] proposed animated demonstrations act as animated worked-examples. This is because they allow learners to study solved problems, without actually using the software. This study considers the efficiency and effectiveness of animated demonstration as an instructional strategy, and like Touvinen and Sweller [6], the study compares two instructional strategies (animated demonstrations and discovery practice).

M. Kurosu (Ed.): Human Centered Design, HCII 2009, LNCS 5619, pp. 247–255, 2009.

1.1 Animated Instruction and Procedural Learning

The animated demonstration literature extends back to the early 1990s. The most notable finding from this literature is the potential for an “animation deficit” [2], [3]. Palmiter et al. [2] found learners using animated demonstrations would acquire skills in significantly less time during an initial learning phase, but a week later, these learners had difficulty and took significantly longer to reproduce the same performance. This was later described as Palmiter’s animation deficit [3]. However, other researchers were not able to replicate Palmiter’s findings [3], [8]. In addition, if animated demonstrations act as worked examples as Lewis proposed [7], then Palmiter’s animation deficit is in conflict with the worked example effect.

1.2 Relative Condition Efficiency (RCE) and Performance Efficiency (PE)

On some level, educational researchers use a medical model to “treat” ignorance, with instructional products that we hope are both efficient and effective. Those who study cognitive load theory analyze the differences between instructional conditions with a construct called “relative condition efficiency” [4]. To use this construct, researchers observe learners as they use one of several instructional conditions. Following the instruction, learners attempt a performance, and then rate their perceived mental effort during learning. Their performance scores and mental effort ratings are combined in a construct (See Equation 1) [4].

$$\text{Relative condition efficiency} = \frac{Z_{Performance} - Z_{MentalEffort}}{\sqrt{2}} \qquad (1)$$

The resulting data is then graphed in a biplot (See Fig. 2) to allow researchers to visually compare the relative efficiency of instructional conditions [4].Since its development, relative condition efficiency has become an important basis for much of cognitive load research [9] but, this measurement relies on indirect or subjective measurements [10]. Paas and van Merriënboer [4] were aware of this, and state in their original paper that this construct should be qualified with performance data.

Efficiency and effectiveness may be described with dependent variables. Gagné described two general categories of dependent variables used during problem-solving studies [11]. He proposed most researchers consider (1) “the rate of attainment of some criterion performance” and (2) “the degree of correctness of this performance” [11], (p.295). In terms of the animated demonstration literature, these are described as performance time and accuracy [3]. So as Gagné suggests, another useful measure is performance time. While relative condition efficiency is a measure of efficiency, because it uses mental effort ratings, it does not include time in the equation, so it is difficult to analyze the efficiency of the performance given and instructional condition. Lewis [12] synthesized these ideas to develop yet another efficiency construct, called performance efficiency. Performance efficiency is similar to relative condition efficiency, but uses performance time rather than a mental effort rating (compare equations 2 and 3).

$$\text{Relative condition efficiency} = \frac{Z_{Performance} - Z_{MentalEffort}}{\sqrt{2}} \tag{2}$$

$$\text{Performance efficiency} = \frac{Z_{Performance} - Z_{PerformanceTime}}{\sqrt{2}} \tag{3}$$

Performance efficiency was developed to complement relative condition efficiency, and it is hoped that this metric may be used to strengthen cognitive load research. One should be aware that even though performance efficiency allows a researcher to analyze group performance, it does not include a mental effort rating, therefore it is not a measure of cognitive load. However, like relative condition efficiency, performance efficiency also may be used to analyze the relative efficiency of group performances. Performance efficiency contrasts instructional conditions in much the same manner, to combine these performance variables in a biplot with group performance times and performance scores (in this case accuracy) on the x and y axes respectively (See Fig. 1).

1.3 Hypothesis and Operational Definitions

Recall that the purpose of this study is to consider the worked example effect and Palmiter's animation deficit given animated demonstrations. This section outlines the hypotheses and operational definitions of these two effects.

The worked example effect is often defined as an improvement in learner performance given worked examples. Sweller and Cooper's early studies were the first to describe this effect [5]; [13]. They described this effect by saying a "decreased solution time was accompanied by a decrease in the number of mathematical errors" [5], (p.59). The dependent variables here are solution time and "a reduction of errors", or in terms of this study and the animated demonstration literature, performance time and accuracy [3]. In terms of the hypotheses of this study learner behavior may be described as:

H_a: During week one those using animated demonstrations will have a significant increase in performance over those only solving problems.

H_o: Learner performances will not differ during week one.

Lipps, et al. described Palmiter's animation deficit as "poorer retention despite faster learning following animation training" [2], (p. 1). In terms of this study and its dependent variables, an operational definition of Palmiter's animation deficit, was a significant increase in performance time and a simultaneous decrease in accuracy, one week after initial instruction given animated demonstrations as an instructional strategy. In terms of the hypotheses of this study learner behavior may be described as:

H_a: Those using animated demonstrations will have a decreased performance during week two.

H_o: Learner performances will not differ during week two.

2 Methods

The participants of this study were pre-service teachers. These undergraduates were enrolled in an introductory instructional technology course at a large university in the southeastern United States. 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 at a power of 0.80, with a small effect size η^2= 0.125, given an α=0.05 (α =0.05 is used throughout this study) [14].

2.1 Instructional Materials

Tuovinen and Sweller [6] provided subjects with an introductory overview to provide context. The overview in the current study was a short, narrated, non-animated web-based presentation (~ 2 minutes) developed with TechSmith Camtasia 4.0 [15]. All subjects viewed this overview which provided them with an introduction to graphic design and digital image editing. Thus all subjects were presented with screenshots of the required procedures. In addition subjects may have interacted with one of two animated demonstrations which were developed with Techsmith Camtasia Studio 4.0. The first animated demonstrations was identical to the practice problem used during week one (the mimic condition). A second “varied context” (collage-based) demonstration was also used. In each case, subjects were taught how to select, move, rotate, and hide layers within an Adobe Photoshop Elements [16] document.

2.2 Procedure

TechSmith Morae 1.1 [17] recorded all learner onscreen actions, as they interacted with the instruments and problem-solving scenarios. Upon completion of an initial demographics survey, all subjects were forwarded to an instructional overview. When the instructional overview concluded, a JavaScript randomly divided subjects into one of four instructional conditions. Two web-based animated demonstrations were used in these four instructional conditions:

- an animated demonstration group (demo), this group did solve the week one problem;
- an identical animated demonstration, and they practiced with the week one problem (demo+ practice);
- a different collage-based animated demonstration, with the week one problem as practice (demo2+practice);
- and a discovery-based practice condition (no animated demonstrations), with the week one problem serving as practice (practice).

Learners concluded their instructional condition with a post-treatment survey. This survey included a relative condition efficiency question, as documented by Paas and van Merriënboer [4]. Once learners finished the post-treatment survey, they were thanked for their participation, asked not to discuss their instruction with others, and not to use the software before the delayed test session.

During a second delayed session (one week later), learners returned and all groups of learners solved a different, week two problem. Learner onscreen behavior was recorded in a similar manner. Following this second set of performances, they were given a post-treatment survey which also included a relative condition efficiency question.

A researcher later reviewed the recorded video files, to score each learners attempt. Performance time and accuracy were measured using Techsmith Morae. Performance time was measured in seconds. Accuracy was scored with a rubric based upon the problem solving operators. Those who successfully completed a solution step were awarded one point, while more difficult solutions steps (flipping layers or correct layer placement) were awarded two points.

3 Results

A sample size of n=122 learners followed the instructions, completed all surveys, and attempted the required performances. A sample size of 115 participants was necessary for a four group MANOVA, given the power requirements. A sample size of this magnitude required data to be collected across two semesters. Therefore it was necessary to question if this pooled dataset would affect statistical tests.

The assumptions of a MANOVA were analyzed for the week two data set first, because all learners participated during this session. According to Glass and Hopkins [18] these data met the independence assumption. A SAS macro %MULTNORM [19] revealed non-normality (violating the normality assumption) since the Shapiro-Wilks' W= 0.76 p<0.0001 for accuracy, and for performance time the Shapiro-Wilks' W=0.95, p=0.0015. This violation was primarily due to a series of potential multivariate outliers. These learners were subsequently removed from the overall data set, using the %MULTNORM macro. Next the data were transformed. Box's M test 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 be significantly different, so there was no evidence that the homoscedasticity assumption was violated. Therefore a MANOVA was used to compare the two semester subgroups (the summer and fall subgroups). The MANOVA indicated that there was not a significant difference between the two semester subsets, since Wilks' Λ =0.95, F (2, 95) = 2.47, p = 0.09, η^2=0.05.

Given it reasonable to proceed with a pool-semester solution, the week one dataset was also analyzed with a MANOVA. First the assumptions of the MANOVA were analyzed and these data met the independence assumption [18]. Multivariate non-normality was revealed when the %MULTNORM macro revealed a Shapiro-Wilks' W= 0.62, p<0.0001 for accuracy (AC1), and for performance time (PT1) the Shapiro-Wilks' was W = 0.88, p<0.0001. Mardia skewness was found to be β_1p= 66.70, p<0.0001 and Mardia kurtosis was β_2p=3.79, p<0.0001. Outliers were retained. Box's M test was performed, and the variance-covariance matrices were not significantly different, or homogeneous, X^2(6, N=69) = 7.97, p=0.24, φ=0.34.

A MANOVA was analyzed and during 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.

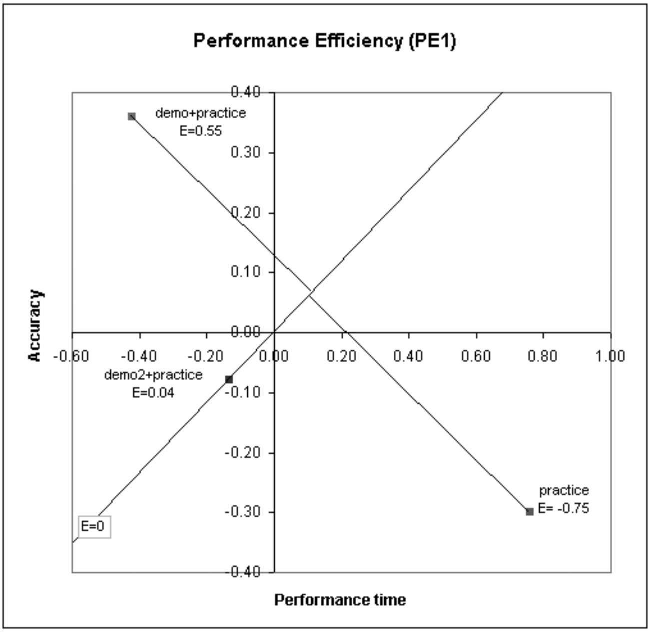

Fig. 1. Performance Efficiency

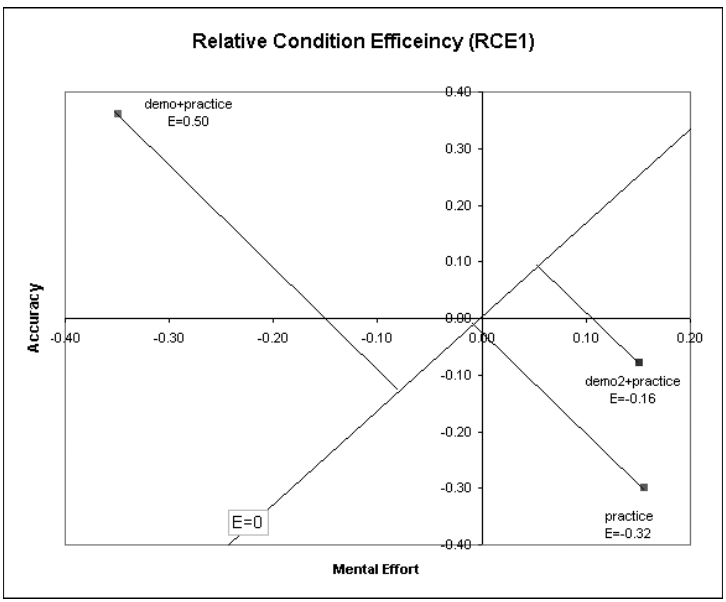

Fig. 2. Relative Condition Efficiency

Post hoc comparisons with Scheffé's test ($p<0.025$) revealed the demonstration groups (demo+practice and demo2+practice groups) assembled the problem, in significantly less time than the practice group. Also week one relative condition efficiency was considered and significant differences between conditions were revealed, since F (2, 68) = 3.69, $p=0.03$ (See Fig. 2). However, post hoc comparisons with Scheffé's test ($p<0.05$) found no differences between groups (given relative condition efficiency).

Finally, week one performance efficiency (See Equation 3) was also found to be significantly different, since F (2, 68) = 12.95, $p<0.0001$ (See Fig.1). However during week two, group differences were not found (See Table 2).

Table 1. Week one dependent variables

	demo	demo+practice	demo2+practice	practice
n	19	21	31	17
Perf. time				
M	NA	19.66	22.40	28.62
SD	NA	6.35	6.28	9.01
Accuracy				
M	NA	0.56	0.99	1.44
SD	NA	0.79	1.99	1.13

Table 2. Week two dependent variables

	demo	demo+practice	demo2+practice	practice
n	19	21	31	17
Perf time				
M	34.10	31.92	33.29	32.09
SD	3.78	4.93	4.57	3.44
Accuracy				
M	6.55	6.55	6.54	6.50
SD	0.26	0.25	0.22	0.21

4 Discussion and Conclusion

One view, the expository approach, is that learners should be guided during early instruction (instructor led). The alternative perspective is that learners should be allowed to discover problem solutions on their own (discovery learning). In short, Bruner [20] says "Practice in discovering for oneself teaches one to acquire information in a way that makes that information more readily viable in problem solving. So goes the hypothesis" [20], (p. 26). This hypothesis was tested in this study, and like many other worked example studies (e.g. [6]), it found that those learners who studied worked examples performed significantly better than their peers who learned through discovery problem solving.

Lewis [7] claimed animated demonstrations act as animated worked examples. This study justified that claim for it found that those who studied animated demonstrations assembled the week one problem in significantly less time, than those who learned through discovery problem solving; a result that is consistent with the worked example effect [1]. As expected, those who studied an identical animated demonstration (the demo or "mimic" condition) significantly out-performed their peers. However, this study also found those who studied a varied-context demonstration, significantly outperformed their problem-solving peers. Therefore this study has established animated demonstrations act as animated worked examples, and are an effective, efficient method of instruction.

Companies like Bank of America, Amazon.com, and Microsoft are all using web-based narrated animated demonstrations (demos) to train potential clients (novices) to use their products and services. The results of this study verified that this practice improves learner performance during the initial stages of learning. Therefore, it is the recommendation of this study that developers continue to use this effective and now evidence-based, e-learning strategy.

The study also found learner performance did not differ a week later, a finding that does not support Palmiter's animation deficit, further confirmation of the results by other researchers [2], [8]. However Palmiter may have been correct, that only providing learners with animation, produces mimicry; which Ausubel described as rote learning [21]. Non-narrated animation provides little guidance since learners only have self explanations in what may already be an overwhelming learning environment. The alternative is an instructor-led, narrated multimedia environment, which provides learners with a verbal narrative that simultaneously directs attention and provides an expert-level explanation. The importance of adding narration to an animated demonstration should not be underestimated, for it promotes what Mayer describes as "multimedia learning" [22].

Finally, some have proposed experience is the best teacher [20], but this position diminishes the role of an instructor. Instructors have purpose in any environment for they provide guidance and support [23]. In an e-learning environment, this role may be hindered because of an inability to communicate with an "e-learner," but the use of animated demonstrations allows an instructor to overcome the obstacles of time and space, to provide "e-learners" with guidance during "just in time" training.

References

1. Sweller, J., Chandler, P.: Evidence for Cognitive Load Theory. Cognit Instruct 8(4), 351–362 (1991)
2. Animation as documentation: A Replication with Reinterpretation, `http://www.stc.org/proceedings/ConfProceed/1998/PDFs/00006.PDF`
3. Palmiter, S.L., Elkerton, J., Baggett, P.: Animated demonstrations vs. written instructions for learning procedural tasks: a preliminary investigation. Int. J. Man Mach. Stud. 34, 687–701 (1991)
4. Paas, F.G.W.C., van Merrienboer, J.J.G.: The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. Hum Factors 35(4), 737–743 (1993)
5. Sweller, J., Cooper, G.A.: The use of worked examples as a substitute for problem solving in learning algebra. Cognition and Instruction 2(1), 59–89 (1985)

6. Tuovinen, J.E., Sweller, J.: A comparison of cognitive load associated with discovery learning and worked examples. J. Educ. Psychol. 91(2), 334–341 (1999)
7. Lewis, R.D.: Demobank: a method of presenting just-in-time online learning. In: The Proceedings of the Association for Educational Communications and Technology (AECT) Annual International Convention, Orlando, FL, October 2005, vol. 2, pp. 371–375 (2005)
8. Waterson, P.E., O'Malley, C.E.: Using animated demonstrations in multimedia applications: Some suggestions based upon experimental evidence. In: Salvendy, G., Smith, M.J. (eds.) Human-Computer Interaction: Software and Hardware Interfaces, Proceedings of the Fifth International Conference on Human-Computer Interaction (HCI International 1993), pp. 543–548 (1993)
9. Paas, F., Tuovinen, J.E., Tabbers, H.K., Van Gerven, P.W.M.: Cognitive load measurement as a means to advance cognitive load theory. Educ. Psychol. 38(1), 63–71 (2003)
10. Brünken, R., Plass, J.L., Leutner, D.: Direct measurement of cognitive load in multimedia learning. Educ. Psychol. 38(1), 53–61 (2003)
11. Gagné, R.M.: Problem solving. In: Melton, A.W. (ed.) Categories of human learning. Academic Press, New York (1964)
12. Lewis, R.D.: The acquisition of procedural skills: An analysis of the worked-example effect using animated demonstrations. Unpublished doctoral dissertation, University of South Florida (2008)
13. Cooper, G., Sweller, J.: Effects of schema acquisition and rule automation on mathematical problem-solving transfer. J. Educ. Psychol. 79(4), 347–362 (1987)
14. Stevens, J.: Applied multivariate statistics for the social sciences. Erlbaum, Mahwah (2002)
15. Techsmith, TechSmith Camtasia Studio 4.0 [Computer program]. Okemos, MI (2006)
16. Adobe Systems: Adobe Photoshop Elements 2.0 [Computer program]. Mountain View, CA (1990-2002)
17. Techsmith: TechSmith Morae 1.0.1 [Computer program]. Okemos, MI (2004)
18. Glass, G., Hopkins, K.: Statistical methods in education and psychology, 2nd edn. Allyn and Bacon, Boston (1984)
19. Macro to test multivariate normality, `http://support.sas.com/kb/24/983.html`
20. Bruner, J.S.: The act of discovery. Harv. Educ. Rev. 31(1), 21–32 (1961)
21. Ausubel, D.P.: The psychology of meaningful verbal learning; an introduction to school learning. Grune and Stratton, New York (1963)
22. Mayer, R.: Multimedia Learning. Cambridge University Press, Cambridge (2001)
23. Kirschner, P.A., Sweller, J., Clark, R.E.: Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educ. Psychol. 41(2), 75–86 (2006)