Designing Effective Feedback in Adaptive Training Systems

Carla R. Landsberg[1], Shannon Bailey[2], Wendi L. Van Buskirk[1], Emily Gonzalez-Holland[2], and Cheryl I. Johnson[1]

[1]Naval Air Warfare Center Training Systems Division, Orlando FL
[2]StraCon Services Group, LLC, Fort Worth, Texas

carla.landsberg@navy.mil, shannon.bailey.ctr@navy.mil, wendi.vanbuskirk@navy.mil
emily.gonzalezhollan.ctr@navy.mil, cheryl.i.johnson@navy.mil

ABSTRACT

The Department of the Navy is interested in Adaptive Training (AT) methods to improve the effectiveness and efficiency of training. AT is defined as instruction that is tailored to an individual learner's strengths and weakness either before or during training (Kelley, 1969; Park & Lee, 2003; Shute, 2008). Although AT can be implemented in a variety of ways, one method that has shown promise is tailoring the feedback that trainees receive during training (Bell, Kanar, Liu, Forman, & Singh, 2006; Landsberg et al., 2012). However, there are many parameters of feedback delivery that can be adapted (e.g., timing, content, modality), and there is little guidance in the literature to suggest which method is best within which contexts. A between-groups experiment was conducted to examine three variables related to feedback presentation within an AT system: feedback timing (participants either received immediate or delayed feedback), feedback granularity (participants received feedback based on either one event or a summary of events), and environmental feedback (participants either received environmental feedback, [intrinsic feedback inherent to task] in addition to their other feedback or they did not). One hundred forty-three participants completed scenarios in the Periscope Operator Adaptive Trainer, in which they viewed ships via a simulated periscope, estimated the ship's angle relative to their line of sight, and received feedback about their performance based on their experimental condition. Performance was measured as the difference between the angle that participants called and the true angle as well as the time latency for submitting their call. The results indicated a benefit for environmental feedback, but no interactions between feedback timing, granularity or environmental feedback. Theoretical and practical implications for the design of feedback in adaptive training systems are discussed.

ABOUT THE AUTHORS

Carla R. Landsberg is a Research Psychologist at the Naval Air Warfare Center Training Systems Division (NAWCTSD) in Orlando, Florida. She earned an M.A. and Ph.D. in Applied Experimental Human Factors Psychology from the University of Central Florida. Currently, her work includes researching innovative training methods and instructional strategies to improve training effectiveness for the warfighter and designing empirically validated adaptive training techniques.

Shannon Bailey is a doctoral candidate in Psychology at the University of Central Florida where she received an M.A. in Applied Experimental and Human Factors Psychology. She is a Research Psychologist at StraCon Services Group, LLC and currently works with NAWCTSD in Orlando, FL. Her research experience includes investigating the effectiveness of training simulations in virtual reality, gesture-based technology interaction, and spatial ability.

Wendi L. Van Buskirk is a Senior Research Psychologist, Laboratory Lead, and NAVAIR Associate Fellow at NAWCTSD in Orlando, FL. During her 16-year tenure at NAWCTSD, she has led multiple research programs to develop empirically sound instructional strategies, methods, and technologies to provide effective training for the aviation, surface, sub-surface, and Marine Corps communities. She earned a Ph.D. in Applied Experimental and Human Factors Psychology from the University of Central Florida.

Emily Gonzalez-Holland is a Research Psychologist Assistant with StraCon Services Group, LLC supporting NAWCTSD in Orlando, Florida. Emily earned her Bachelor's degree in Psychology from the University of Central

Florida. She recently was admitted to the Modeling and Simulation Master's program at UCF and will be pursuing studies pertaining to training, simulation, and team performance.

Cheryl I. Johnson is a Senior Research Psychologist at NAWCTSD in Orlando, FL, performing research on emerging training technology and adaptive training systems. She earned her M.A. and Ph.D. in Psychology from the University of California Santa Barbara and specialized in Cognition, Perception, and Cognitive Neuroscience. Dr. Johnson has over 10 years of experience in technology-based training research. Her research interests include adaptive training, instructional strategies, design principles for game-based training, and multimedia learning.

Designing Effective Feedback in Adaptive Training Systems

Carla R. Landsberg[1], Shannon Bailey[2], Wendi L. Van Buskirk[1], Emily Gonzalez-Holland[2], and Cheryl I. Johnson[1]

[1]Naval Air Warfare Center Training Systems Division, Orlando FL
[2]StraCon Services Group, LLC, Fort Worth, Texas

carla.landsberg@navy.mil, shannon.bailey.ctr@navy.mil, wendi.vanbuskirk@navy.mil
emily.gonzalezhollan.ctr@navy.mil, cheryl.i.johnson@navy.mil

INTRODUCTION

Recently, the U.S. Navy has explored the use of advanced computer-based training techniques to minimize costs and maximize training outcomes for the fleet. Computer-based training can deliver a more personalized learning experience than traditional classroom instruction and can also be used to supplement live training exercises that can be costly to execute (Funke, 1998; Shearon, 2001). One computer-based training technique called *Adaptive Training* (AT) may be beneficial as it emulates the salient characteristics of a one-on-one tutoring relationship, which has been shown to be a highly effective form of instruction (Bloom, 1984). AT can be defined as instruction that is tailored in some way to an individual trainee, either before training begins based on trainee aptitudes and/or learning styles, or during training based on current task performance (Kelley, 1969; Landsberg et al., 2012; Park & Lee, 2003). Several studies have shown AT to be effective (Long, Hyland, & Barnieu, 2015; Romero, Ventura, Gibaja, Hervas, & Romero, 2006; VanLehn, 2011); however, it is still unclear from the literature which parameter or parameters of instruction should be adapted to create optimal training outcomes, and what individual characteristics to base those adaptations on (performance, aptitudes, learning styles, etc.).

One aspect of the instruction that can be tailored during training is the feedback learners receive. Research has shown that adjusting the amount or type of feedback received can lead to better learning outcomes when compared to non-adaptive training (Bell et al., 2006; Landsberg et al., 2012; Serge, Priest, Durlach, & Johnson, 2013). However, it is possible to tailor many aspects of feedback such as timing, modality, granularity, valence (positive or negative), etc., and the literature has not yet provided sufficient evidence to suggest which adaptations lead to the most effective training outcomes. In this study, we examined the timing of feedback presentation, the granularity of that feedback, and whether or not the trainee received environmental cues regarding their performance on a computer-based trainer that adapted to trainee performance.

Cognitive Load Theory

One model that may provide guidance on the use of feedback is the Cognitive Load Theory (CLT). According to CLT, three sources of cognitive load compete for resources in the limited capacity working memory system (Sweller, 1988). The first, *intrinsic load*, is the cognitive resources needed to understand training material and is related to the inherent complexity of the task being trained; because this type of load is related to the characteristics of the task being trained, it can be the hardest load to change. *Germane load* refers to the cognitive resources that are expended to relate new material to prior knowledge and to organize that knowledge into schemas in long-term memory (LTM). LTM has an almost limitless capacity, and moving information into this system is associated with learning. The third type of load, *extraneous load*, refers to the resources dedicated to processing information that is not related to task learning. This type of load is detrimental to learning a task and can be attributed to poor training design and other factors that distract from task learning. Thus, the goal of creating training using CLT as the framework is to reduce extraneous load and manage intrinsic load in order to increase the capacity for germane load. This framework may help inform decisions on which feedback parameters, or combinations thereof, will be the most beneficial for training.

Timing of Feedback

The literature on the timing of feedback generally refers to two types of feedback latency: immediate and delayed. Immediate feedback is given directly after a response or task performance, while delayed feedback is broadly described as any feedback that is not given immediately (the amount of delay varies between studies; Van Buskirk, 2011). Research on the timing of feedback has been conflicted; although some researchers have found support for the use of immediate feedback, others have concluded that delayed feedback is more beneficial (Kulik & Kulik, 1988; Shute, 2008).

Those who support the use of immediate feedback suggest that feedback given directly after task performance prevents errors from being encoded during the acquisition phase of learning (Shute, 2008). Several studies comparing immediate and delayed feedback for acquisition of verbal, procedural, and motor skills have shown a benefit for feedback presented immediately (Anderson, Corbett, Koedinger, & Pelletier, 1995; Corbett & Anderson, 1989; Dihoff, Brosvic, Epstein, & Cook, 2003). One meta-analysis that examined feedback in computer-based instruction found a stronger effect for immediate feedback versus no feedback (Azevedo & Bernard, 1995).

Although there is some support in the literature for immediate feedback, evidence has also shown that delayed feedback can be beneficial. Kulhavy and Anderson (1972) proposed the interference-perseveration hypothesis, which asserts that errors that are made early in training will be forgotten (and therefore not encoded) if feedback is delayed. Other supporters of delayed feedback posit that trainees become dependent on immediate feedback, causing their performance to suffer once it is removed. In a series of experiments, Schooler and Anderson (2008) found that trainees who were given delayed feedback made fewer errors on test problems and displayed more self-correcting behavior after errors were made than trainees who received immediate feedback.

Evidence from the literature suggests that immediate feedback may be more beneficial during the acquisition phase of learning, while delayed feedback may be better for transferring training skills. Two studies, Lavery (1962) and Schmidt, Young, Swinnen, and Shapiro (1989), showed that feedback given immediately after a trial produced better performance in a practice phase, but worse performance on a retention test. Because of these competing findings in the literature, we will examine both immediate and delayed feedback on both post-test and transfer tests.

Feedback Granularity

Feedback granularity (or scale) refers to feedback that is given either on a trial-by-trial basis (i.e., event-based or task-level feedback) or after a series of trials (i.e., summary-based feedback). As an example, in a training system where the trainee completes four scenarios, event-based feedback would be given after each scenario, and would contain information about that scenario only, while summary-based feedback could be given after the second, third, or fourth scenario and would contain information pertinent to either two, three, or four scenarios. The optimal granularity of information given during feedback is under debate and is one area of investigation of the current study. Some researchers have argued that feedback given frequently during practice or skill acquisition (i.e., event-based) can interfere with learning and cause detrimental effects for retention and transfer tasks (Schmidt, 1991). However, further investigation of this claim by Del Rey and Shewokis (1993) found that trainees who received event-based feedback performed better during acquisition and retention than those receiving summary feedback. This led to research on the optimal number of trials that should be presented prior to summary feedback. Guadagnoli, Dornier, and Tandy (1996) provided summary feedback to participants after each trial, five trials, or fifteen trials for either simple or complex tasks to investigate the best granularity of feedback for different task complexities. They found that the optimal granularity of feedback depended on how much practice a participant had on the task and the task complexity; frequent summaries were better early in the complex task learning, but delayed summaries were better with more task experience. Conversely, event-based summaries were always optimal for the simple task. The authors conclude that immediate guidance for each trial is beneficial when initially learning a task.

In this study, we examined both event-based and summary-based feedback and how these affect performance on a post-test and transfer task. It is important to note that if feedback is not given on a trial-by-trial basis, there is inherently a temporal separation between the event and feedback, which may have resulted in an inability to isolate the contribution of each variable to performance outcomes as reported in previous research. To limit the potential for this to confound our current study, feedback was presented either immediately or delayed for each trial or as a summary performance on multiple trials. In this way, the immediate feedback was not confounded with event feedback and delayed feedback was not confounded with summary feedback.

Environmental Feedback

Environmental feedback is a form of intrinsic outcome feedback in which the effect of a response conveys whether the response was correct or incorrect inherently as part of the task. This form of feedback models the actual relationship between the response and outcome (Johnson, Bailey, & Van Buskirk, *in press*). For example, in an electric circuit game, a learner can extrapolate whether or not s/he correctly built a circuit by seeing if the lightbulb illuminates.

A series of experiments on training a call for fire task investigated parameters of feedback (Astwood, Van Buskirk, Cornejo, & Dalton, 2008; Bolton, 2006; Van Buskirk, 2011). Their results indicated no effect of external, experimental feedback when feedback was provided by the environment/nature of the task. During the task, participants determined the highest priority target (i.e., prioritization) and then submitted a call for fire to destroy targets before they got too close to the trainee's position (i.e., targeting). No environmental feedback was available for the prioritization sub-task, however, for the targeting subtask participants could see where the munitions landed and, if a target was hit, smoke and fire appeared (if the munitions missed, the stationary targets were not damaged and/or moving targets would continue to move). Van Buskirk (2011) found that immediate, auditory, process feedback (a type of feedback that tells the trainee how to perform the task rather than just the outcome of their performance) allowed trainees to perform better than other groups on the prioritization task, but not the targeting task. Astwood et al. (2008) found similar results; participants who received process feedback outperformed participants in the other experimental conditions on prioritization performance only (with no environmental feedback) and not on targeting performance. Finally, Bolton (2006) investigated feedback timing using the call for fire task and her results also suggested an impact of environmental feedback on performance. Specifically, she found a positive effect of immediate feedback on prioritization (i.e., no environmental feedback present) performance, but not for targeting (i.e., with environmental feedback present) performance.

One possible explanation for the lack of findings on targeting performance in all three studies is that environmental feedback was provided on that sub-task. While this is a realistic feature of the task, Van Buskirk (2011) suggested that the environmental feedback may have confounded the experiment especially for delayed feedback groups. Specifically, the environmental targeting feedback provided a double stimulus exposure (Kulik & Kulik, 1988); participants in the delayed feedback groups received feedback on the accuracy of their targeting calls twice - once immediately during the scenario from the environmental feedback and a second time at the end of the scenario. On the other hand, participants had no way of knowing whether or not they correctly prioritized until they received the explicit feedback message. Bolton (2006) posited that for a task that has no intrinsic (i.e., environmental) feedback, providing immediate, extrinsic feedback is necessary to train effective strategies. Based on the results of these studies, it appears that there is anecdotal evidence for an effect of environmental feedback on performance. However, this variable has not been investigated systematically in the literature. This study is a first attempt to determine environmental feedback's relationship with other feedback variables in order to provide guidance and empirical support for effective feedback delivery strategies within adaptive training systems.

Hypotheses

The literature on feedback parameters is mixed, and it may be the case that the combination of feedback parameters that is more effective depends on how those factors are combined. From a CLT perspective, the optimal combination of these parameters will be the one that reduces intrinsic load as much as is possible for a given task, and minimizes extraneous load to allow trainees to engage in germane processing. As such, we hypothesized a 3-way interaction that is depicted in Figure 1. When environmental feedback was not present and feedback was presented immediately, we predicted that 1a) trainees who received event-based feedback would outperform those who receive summary-based feedback on a post-test and transfer task, and 1b) when the presentation of feedback was delayed, we predicted the inverse (i.e., those who received summary-based feedback would outperform those that received event-based feedback). According to CLT, event-based feedback presented immediately should be processed more easily and cause less extraneous load than summary-based feedback presented immediately, leading to better performance. Similarly, summary-based feedback that is given after a delay should cause less extraneous cognitive load than event-based feedback presented after a delay. In theory, this is because there is less information to process simultaneously in event-based immediate feedback and summary-based delayed feedback than either event-based delayed feedback or summary-based immediate feedback.

The second set of hypotheses refers to trainees who receive environmental feedback. In these conditions, we predicted that 2a) trainees would perform better if they were given event-based feedback rather than summary-based feedback. This is because participants may be better able to make sense of the environmental feedback given after each event if they are given feedback that corresponds to each event rather than a summary, reducing extraneous load by providing relevant feedback in the form of environmental cues. Additionally, we predicted that 2b) in environmental feedback conditions, participants who received immediate feedback would outperform those who received delayed feedback, because they would be better able to make sense of the environmental feedback for each event when informational and environmental feedback reinforce each other immediately after the performance, reducing the extraneous load of having to hold on to additional information in working memory.

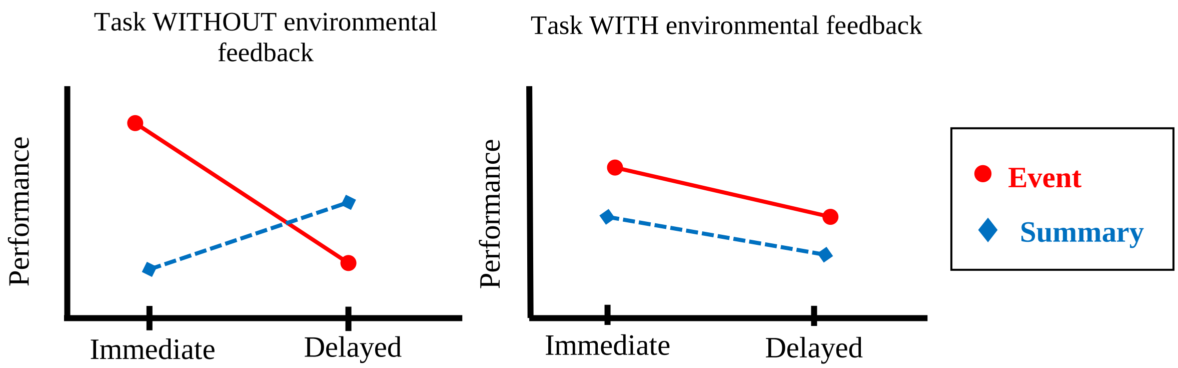

Figure 1. Illustration of hypothesized results for trainee performance in the various feedback conditions.

METHOD

Testbed

We used PC-based periscope simulation software called the *Periscope Operator Adaptive Trainer* (POAT) as the testbed for this experiment. In the simulation, participants viewed ships (hereafter referred to as *contacts*) through a simulated periscope and identified a series of Angle on the Bow (AOB) calls, which is an estimation of a contact's orientation in degrees relative to the trainee's point of view (See Figure 2). Each scenario in the trainer contained a single contact presented at an angle between 0° (aimed straight at ownship) and 180° (aimed directly away from ownship) on either the port (left) or starboard (right) side of the ship. Determining a contact's AOB is an important periscope task that helps to ascertain whether a contact presents a collision threat to the submarine. Therefore, both accuracy and timeliness are important aspects of the call.

Figure 2. POAT interface

For each AOB call, the contact was pre-centered in the simulated periscope view, and the participants were able to move the periscope crosshairs and change the camera type (e.g., infrared, color) and magnification. Participants had one minute to view the contact and enter the call before the simulation displayed the next scenario.

Materials

Several trials of simulated contacts presented under varying environmental conditions (e.g., rain, cloud cover) and angles were created in order to cover a wide range of AOB estimation difficulty. In addition, contact distance, sea state, time of day, and number of distinguishing visual features on the contacts were varied to create three different levels of difficulty: easy, medium, and hard. Input from Subject Matter Experts (SMEs) and data from previous studies were utilized to empirically validate the difficulty of each scenario. The scenarios were presented to participants adaptively; all participants started the training with scenarios at medium difficulty, and were presented with either easier, the same level of difficulty, or harder scenarios based on their performance on the task. For example, trainees who performed well on the medium scenarios were given harder scenarios in their next set, trainees who performed fairly well stayed at the medium level, while trainees who did not perform well were given easier scenarios in their next set.

The participants also completed a Mental Effort Rating scale to assess cognitive load, in which they rated their mental effort on the task on a Likert scale ranging from 1 (Very, very low) to 9 (Very, very high). This questionnaire was given three times during the study. A demographics questionnaire was used to assess participants' level of education and experience with computers and video games.

Participants

One hundred and forty-three students (62 males, 81 females, M_{age}= 21.76, SD=4.41 years) from the University of Central Florida participated in the study. Students were paid $10 an hour for up to three hours of participation. Participants were randomly assigned to conditions.

Procedure

Eight conditions were created to test our hypotheses (See Table 1). Participants received either event or summary-based feedback, either immediate or delayed feedback, and either present or absent environmental feedback. Event-based feedback consisted of information on how far off the participant's called AOB was from the correct AOB for each trial, and summary feedback consisted of the percentage of calls within a certain range of correctness (i.e., calls < 10° off, calls between 11° - 29° off, and calls that were > 30° off). Immediate feedback, presented as either event or summary form, was given as soon as the participant responded to a trial, whereas delayed feedback, again presented either event or summary information, was given after every 15 trials. Environmental feedback consisted of an animation following each call projecting what would happen to ownship as a result of the trainees' answer. When a trainees' call was correct, the submarine submerged successfully to avoid the collision threat, and when their answer was incorrect, the trainee saw an animation of their submarine colliding with the contact. In the beginning of the experimental session, participants were randomly assigned to one of the eight experimental conditions. They were then provided with a PowerPoint tutorial that explained the simulation interface and the AOB task, which included visual cues they could use to determine a contact's AOB. Next, they were given five practice scenarios within the simulation.

Next, all participants completed a pre-test consisting of 20 scenarios. The AOB difficulty varied randomly in this section, and participants did not receive feedback on these trials. Following the pre-test, participants performed the experimental portion in which they completed 90 training scenarios and received feedback according to their condition. For all participants, the difficulty of the scenarios was adapted based on their performance on prior scenarios. Following the experimental session, the participants filled out the first Mental Effort Rating Scale. Participants then completed a post-test containing the same 20 scenarios as the pre-test presented in a random order, with no feedback. They then completed the second Mental Effort Rating Scale and the Demographics questionnaire. Finally, participants completed a near transfer task consisting of 20 scenarios that included contacts never before seen by participants, without feedback and then completed the third Mental Effort Rating Scale.

Table 1. Experimental Conditions

	Type of Feedback	*Description*
1	Event Immediate With Environmental	Feedback is provided immediately after each scenario. Participants also receive environmental feedback after each scenario.
2	Event Delayed With Environmental	Feedback on each scenario is provided after each testlet of 15 scenarios. Participants also receive environmental feedback immediately after each scenario.
3	Summary Immediate With Environmental	Feedback is provided as a summary of performance immediately after each scenario. Participants also receive environmental feedback after each scenario.
4	Summary Delayed With Environmental	Feedback is provided as a summary of performance after each testlet of 15 scenarios. Participants also receive environmental feedback immediately after each scenario.
5	Event Immediate With No Environmental	Feedback is provided immediately after each scenario. Participants do not receive environmental feedback after each scenario.
6	Event Delayed With No Environmental	Feedback on each scenario is provided after each testlet of 15 scenarios. Participants do not receive environmental feedback after each scenario.
7	Summary Immediate With No Environmental	Feedback is provided as a summary of performance immediately after each scenario. Participants do not receive environmental feedback after each scenario.
8	Summary Delayed With No Environmental	Feedback is provided as a summary of performance after each testlet of 15 scenarios. Participants do not receive environmental feedback after each scenario.

RESULTS

Participants' scores for pre-test, post-test, and transfer task were calculated as the average AOB estimate error (difference between the actual AOB and the estimate they entered). As such, when referring to these dependent variables, lower numbers indicate better performance.

Cognitive Load and Amount of Time on Feedback

The mental effort measure that assessed the cognitive load imposed during a set of trials did not correlate significantly with the outcome measures or the feedback conditions (all $p>.05$). The amount of time participants viewed the feedback varied between the Immediate and Delayed groups and the Summary and Event groups. Participants in the delayed feedback group, on average, viewed the feedback for longer than those in the immediate group ($M_{delayed}$= 17.11 sec, $SD_{delayed}$=11.60; $M_{immediate}$= 3.00 sec, $SD_{immediate}$= 1.22), $t(1,138)=10.13$, $p<.001$. Likewise, participants in the event condition looked at their feedback longer than participants in the summary condition (M_{event}= 13.78 sec, SD_{event}=13.44 $M_{summary}$=6.11 sec, $SD_{summary}$=4.69), $t(1,141)=4.45$, $p<.001$. There were no significant correlations between the environmental feedback conditions and feedback reading time.

Feedback Conditions

To test the hypotheses that type of feedback affects learning outcomes, a 2 (Timing: Immediate or Delayed) X 2 (Granularity: Event or Summary) X 2 (Environmental: Present or Absent) between-subjects Analysis of Covariance (ANCOVA) was conducted with the dependent variable of post-test score and a covariate of pre-test score. Means and standard deviations are presented in Table 2. The covariate of pre-test score was significant, $F(1,134)=72.76$, $p<.001$, $\eta^2= .04$, indicating that those who performed better on the pre-test also performed better on the post-test. There was also a significant main effect for environmental feedback, $F(1,134)=3.78$, $p=.03$ (one-tailed), $\eta^2= .002$, such that participants who received environmental feedback (M=22.52, SD=7.55) performed better than those who did not (M=25.31, SD=9.40). There were no other significant main effects or interactions (all $p>.05$). Another 2X2X2 ANCOVA was conducted on the same feedback conditions with the transfer test as the dependent variable. The covariate of pre-test score was significant, $F(1,132)= 40.32$, $p<.001$, $\eta^2=.02$ but there were no significant main effects or interactions (all $p>.05$).

To examine how feedback type affects time to make an AOB call, a 2x2x2 ANCOVA was conducted on the eight feedback conditions with the post-test call latency as the dependent variable. Pre-test latency was used a covariate, and it was found to be significant, $F(1,134)=25.43$, $p<.001$, $\eta^2= .002$. There was a significant main effect for feedback timing, $F(1,134)=2.70$, $p=.05$(one-tailed), $\eta^2= .002$, such that participants who received delayed feedback made their post-test calls faster (M=15.85 sec, SD=5.56 sec) than those who received immediate feedback (M=19.05 sec, SD=8.40 sec). There was also a main effect for feedback granularity, $F(1,34)=3.34$, $p=.03$ one-tailed, $\eta^2= .003$, such that participants who received summary feedback made their post-test calls faster (M=16.47 sec, SD=6.34 sec) than participants who received event feedback (M=18.38 sec, SD=8.00 sec). A similar 2x2x2 ANCOVA was conducted using transfer task call latency as the dependent variable. Pre-test time was found to be a significant covariate, $F(1,132)= 37.73$, $p<.001$, $\eta^2= .03$; however, no significant main effects or interactions were found for transfer task call times (all $p>.05$; See Table 3).

Table 2. Descriptive statistics for post-test accuracy score (AOB error) and call latency (sec) in each condition (lower numbers indicate better performance)

	Condition	*n*	*Accuracy M*	*Accuracy SD*	*Time M*	*Time SD*
1	Event Immediate w/ Environmental	18	21.78	8.90	17.15	6.63
2	Event Delayed w/ Environmental	18	22.78	8.06	17.41	6.37
3	Summary Immediate w/ Environmental	18	20.29	6.85	19.82	5.50
4	Summary Delayed w/ Environmental	18	25.17	6.01	14.39	6.75
5	Event Immediate No Environmental	20	24.58	12.02	22.68	10.77
6	Event Delayed No Environmental	19	24.26	8.85	15.66	5.13
7	Summary Immediate No Environmental	16	26.30	8.90	15.55	8.05
8	Summary Delayed No Environmental	16	26.41	7.14	15.96	3.03
	Total	143	23.92	8.61	17.45	7.28

Table 3. Descriptive statistics for transfer task accuracy score (AOB error) and call latency (sec) in each condition (lower numbers indicate better performance)

	Condition	*n*	*Accuracy M*	*Accuracy SD*	*Time M*	*Time SD*
1	Event Immediate w/ Environmental	18	40.28	16.34	18.80	8.25
2	Event Delayed w/ Environmental	18	38.36	11.12	17.63	5.38
3	Summary Immediate w/ Environmental	17	35.50	9.49	20.71	6.44
4	Summary Delayed w/ Environmental	18	43.28	13.83	17.01	10.49
5	Event Immediate No Environmental	20	39.08	15.20	20.70	10.09
6	Event Delayed No Environmental	18	40.00	13.85	18.63	5.69
7	Summary Immediate No Environmental	16	42.72	13.17	19.22	6.96
8	Summary Delayed No Environmental	16	40.22	12.63	17.87	4.08
	Total	141	39.90	13.23	18.84	7.50

DISCUSSION

The goal of this experiment was to systematically examine the effects of feedback granularity, feedback timing, and environmental feedback on performance in an adaptive simulation-based training task. We hypothesized that when trainees were not given environmental feedback, the effectiveness of granularity would depend on when the feedback was presented. Specifically, we predicted an interaction such that when feedback was presented immediately, trainees who received event-based feedback would outperform those who received summary-based feedback on a post-test and transfer task (1a). However, when the presentation of feedback was delayed, we predicted that those who received summary-based feedback would outperform those that received event-based feedback (1b). These hypotheses were not supported. We also predicted that trainees who received environmental feedback would perform better if they were given event-based feedback rather than summary-based feedback (2a). Additionally, we predicted that in environmental feedback conditions, participants who received immediate feedback would outperform those who received delayed feedback (2b). These hypotheses were also not supported. However, there was a significant main effect for environmental feedback on post-test accuracy score, such that those who received environmental feedback performed better on the post-test than those who did not.

An issue inherent in the use of environmental feedback is that participants who were in the environmental delayed feedback conditions essentially received both immediate and delayed feedback, providing a double-stimulus exposure (Kulik & Kulik, 1988). Although we predicted that immediate feedback would be better for training than delayed feedback, the double feedback exposure in the delayed environmental condition may have attenuated the differences between the delayed and immediate groups, which is consistent with findings reported by Bolton (2006), Astwood et al. (2008), and Van Buskirk (2011). Moreover, the post-test results indicate that participants who received immediate environmental feedback did perform better than those who received delayed environmental feedback; however, this result was not statistically significant. For those who did not receive environmental feedback it is possible that the lack of significant findings was due to the content of the feedback they received. For instance, the participants may have found that for this task, which was visuospatial in nature, the text-based feedback was either not useful or distracting, causing higher extraneous load than intended regardless of feedback timing. It is possible that in the environmental feedback groups, the more task-congruent visual environmental feedback was sufficient to help them improve their performance. This may have been particularly true for the immediate feedback groups if the environmental feedback helped provide a more salient context for their feedback. Additionally, a different type of feedback may have helped participants (especially the no environmental groups) perform better. Several studies have shown a benefit of process feedback over outcome feedback (the type used in this study). Process feedback gives the trainee information on how to perform the task better rather than solely informing them of the outcome of their performance (Kluger & DeNisi, 1996). In some cases, process feedback has been demonstrated to be more beneficial for training, particularly for complex tasks than outcome feedback (Astwood et al., 2008; Buff & Campbell, 2002). For instance, process feedback could help with schema development, thus increasing task-relevant germane load that is necessary for meaningful learning. Finally, another issue that may have contributed to the lack of statistically significant findings is that statistic power for this study was too low to reveal the hypothesized 3-way interaction.

Although we also did not find a significant 3-way interaction on post-test call latencies, we did find that participants who received delayed feedback made their post-test calls faster than those who received immediate feedback. The finding that the delayed groups viewed their feedback for longer may suggest that they may have engaged in more active processing than the immediate groups, which helped them perform faster on the post-test.

Based on these results, it is apparent that continued research is needed to investigate various combinations of feedback parameters, especially when environmental feedback is provided naturally within the environment, to determine which are the most effective for improving performance. Future research could explore these feedback parameters using task-congruent feedback (i.e., environmental feedback instead of text-based feedback). For instance, a future study may provide immediate and delayed feedback both in the form of environmental and text-based feedback in order to examine if task-feedback congruence is more effective. It may also be useful to examine these feedback parameters in a different domain, for instance, in a task that is more procedural in nature. This would allow us to develop guidance and empirical support for effective feedback strategies to be used within adaptive training systems. As Bolton (2006) stated, "If consistent findings are discovered for certain combinations of variables, it may be possible to build these findings up to higher order, generalizable principles and theories – something that has been lacking for the topic of feedback despite the established importance of it (p. 138)."

ACKNOWLEDGEMENTS

We gratefully acknowledge Dr. James Sheehy who sponsored this work through the Section 219 Naval Innovative Science and Engineering (NISE) Basic and Applied Research (BAR) program. We would also like to thank the Submarine Research Applications Team, and especially Derek Tolley, who helped us create our testbed and research stimuli. Finally, we would like to thank Dr. Randy Astwood and Ms. Alyssa Mercado who assisted on this project. The views expressed herein are those of the authors and do not reflect the official position of the organization with which they are affiliated.

REFERENCES

Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*, 167-207.

Astwood, R. S., Van Buskirk, W. L., Cornejo, J. M., & Dalton, J. (2008). The Impact of different feedback types on decision-making in simulation based training environments. *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society* [CD-ROM].

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*(2), 111-127.

Bell, B. S., Kanar, A., Liu, X., Forman, J., & Singh, M. (2006). *Adaptive guidance: Effects on self-regulated learning in technology-based training*. Unpublished manuscript, Center for Advanced Human Resource Studies, Cornell University, Ithaca, NY. Retrieved from http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1455&context=cahrswp.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16.

Bolton, A. E. (2006). *Immediate versus delayed feedback in simulation based training: Matching feedback delivery timing to the cognitive demands of the training exercise*. Unpublished doctoral dissertation, University of Central Florida, Orlando.

Buff, W. L., & Campbell, G. E. (2002). What to do or what not to do?: Identifying the content of effective feedback. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society* (pp.2074-2078), Santa Monica, CA: HFES

Corbett, A. T., & Anderson, J. R. (1989). Feedback timing and student control in the LISP intelligent tutoring system. *In Proceedings of the Fourth International Conference on AI and Education* (pp. 64-72).

Del Rey P., & Shewokis P.A. (1993). Appropriate KR summary lengths for learning timing tasks under conditions of high and low contextual interference. *Acta Psychologica*, 83, 1-12.

Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2003). The role of feedback during academic testing: The delay retention test revisited. *The Psychological Record, 53*, 533-548.
Funke, J. (1998). Computer-based testing and training with scenarios from complex problem solving research: Advantages and disadvantages. *International Journal of Selection and Assessment, 6*(2), 90-96.
Guadagnoli, M. A., Dornier, L. A., & Tandy, R. D. (1996). Optimal length for summary knowledge of results: The influence of task-related experience and complexity. *Research Quarterly for Exercise and Sport*, *67*(2), 239-248.
Johnson, C. I., Bailey, S. K. T., & Van Buskirk, W. L. (*in press*). Designing effective feedback messages in serious games and simulations: A research review. In P. Wouters & H. van Oostendorp (Eds.), *Techniques to Improve the Effectiveness of Serious Games.* Switzerland: Springer International Publishing.
Kelley, C. R. (1969). What is adaptive training? *Human Factors*, *11*(6), 547-556.
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.
Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79-97.
Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology, 63*, 505-512.
Landsberg, C. R., Astwood, R. A., Van Buskirk, W. L., Townsend, L. N., Steinhauser, N. B., & Mercado, A. D. (2012). Review of adaptive training system techniques. *Military Psychology, 24*, 96-113.
Lavery, J. J. (1962). Retention of simple motor skills as a function of type of knowledge of results. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *16*(4), 300.
Long, R., Hyland, J., & Barnieu, J. (2015). Development and evaluation of mobile adaptive training technologies. *Interservice/Industry Training Simulation, and Education Conference (I/ITSEC) 2015.*
Park, O., & Lee, J. (2003). *Adaptive instructional systems*. Handbook of Research for Educational Communications and Technology, 651-684. Retrieved from http://www.aect.org/edtech/ed1/25.pdf .
Romero, C., Ventura, S., Gibaja, E. L., Hervas, C., & Romero, F. (2006). Web-based adaptive training simulator system for cardiac life support. *Artificial Intelligence in Medicine, 38*(1), 67-78.
Schmidt, R.A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In J. Requin & G.E. Stelmach (Eds.), *Tutorials in Motor Neuroscience* (pp. 59-75). Springer, Netherlands.
Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology Learning, Memory, and Cognition, 15*(2), 352-359.
Schooler, L. J., & Anderson, J. R. (2008). The disruptive potential of immediate feedback. *Carnegie Mellon University Research Showcase,* 702-708. Retrieved from http://repository.cmu.edu/cgi/viewcontent.cgi?article=1079&context=psychology.
Serge, S. R., Priest, H. A., Durlach, P. J., & Johnson, C. I. (2013). The effects of static and adaptive performance feedback in game-based training. *Computers in Human Behavior,29*(3), 1150-1158. Retrieved from http://dx.doi.org/10.1016/j.chb.2012.10.007.
Shearon, B. T. (2001). *The cost effectiveness of west coast distributed simulation training for the Pacific Fleet* (Master's Thesis). Retrieved from from dtic.mil. (ADA406612).
Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. Retrieved from http://myweb.fsu.edu/vshute/pdf/shute%202008_b.pdf.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285. Retrieved from http://onlinelibrary.wiley.com/doi/10.1207/s15516709cog1202_4/pdf.
VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems*.* *Educational Psychologist*, *46*(4), 197-221.
Van Buskirk, W. L. (2011). *Investigating the optimal presentation of feedback in simulation-based training: An application of the cognitive theory of multimedia learning*. Unpublished doctoral dissertation, University of Central Florida, Orlando.