Soldier Performance Following Distributed and Traditional Digital Skills Training

Jennifer S. Tucker, Gregory A. Goodwin
US Army Research Institute
Fort Benning, GA
Jennifer.S.Tucker@us.army.mil, GregoryGoodwin1@us.army.mil

ABSTRACT

The proliferation of digital command and control systems on the modern battlefield places a growing training requirement on Soldiers at all echelons to acquire and maintain the skills to operate those systems. With the increased demand to develop and maintain highly proficient system operators, some Army units have utilized distributed learning (dL) technologies to train digital skills. As dL instructional environments have unique training challenges and little is known about the effectiveness of training digital skills using dL, there is a critical need to know whether Soldier performance following dL instruction differs from traditional (face-to-face) classroom instruction. In this paper, we compare Force XXI Battle Command Brigade and Below (FBCB2) operator skills of Soldiers given either dL (Baseline N = 136; Retention N = 32) or face-to-face (Baseline N = 80; Retention N = 31) instruction immediately and eight weeks following the training. Although the results demonstrated some differences in proficiency on specific tasks at baseline, the declines in performance over time were very similar for both classes. For example, even though Soldiers in the traditional classes performed better on some of the tasks immediately following the course, both classes performed similarly eight weeks later. These findings provide support for dL instruction of digital skills as being comparable to that of traditional instruction. Our testing procedure also made it possible to examine performance for each step of each task (e.g., create and send a route), and these results provided a better understanding of why certain tasks were problematic (e.g., system does not cue the operator). Finally, from these findings, we suggest ways in which training developers can design courses and trainers can present materials to enhance initial performance and mitigate decrements in skill over time.

ABOUT THE AUTHORS

Dr. Jennifer Tucker is a Senior Research Psychologist / Team Leader with the US Army Research Institute for the Social and Behavioral Sciences (ARI) at Ft. Hood, TX, specializing in the areas of training research and evaluation, performance, and measurement. She holds a Ph.D. in Industrial/Organizational Psychology from Portland State University, Portland, OR.

Dr. Gregory Goodwin is a Senior Research Psychologist with the U.S. Army Research Institute for the Social and Behavioral Sciences (ARI) at Ft. Benning, GA, specializing in the areas of training and retention of digital skills. He holds a Ph.D. in Psychology from Binghamton University, Binghamton, New York.

Soldier Performance Following Distributed and Traditional Digital Skills Training

Jennifer S. Tucker, Gregory A. Goodwin
US Army Research Institute
Fort Benning, GA
Jennifer.S.Tucker@us.army.mil, GregoryGoodwin1@us.army.mil

PURPOSE

The purpose of the present research is three-fold: (1) determine the skill retention of selected digital skills following the standard Army 40-hour system operator course; (2) determine the characteristics of these system task/procedures that lead to skill decay; and (3) examine performance differences between instructional environments - distributed learning (dL) vs. face-to-face (F2F).

DIGITAL SYSTEMS

Digital systems on a wireless battlefield network allow commanders and leaders to rapidly develop a common view of the battlefield so they can make decisions faster and disseminate messages, orders, and overlays to their subordinates. Although every Army division employs the Army Battle Command System (ABCS), many units struggle to leverage the full potential of the networked capabilities (Clark, 2005). Warfighting potential is lost because of factors such as recurring hardware and software upgrades, personnel turbulence, and decay of digital skills over time. While some of these factors are beyond the control of unit commanders, providing digital training programs to sustain operator proficiency is well within a commander's purview. Knowledge of how to optimally train and sustain digital proficiency is essential to develop highly proficient ABCS operators and maintain their proficiency.

A number of mostly anecdotal reports suggest that digital skills are quickly forgotten (c.f., Goodwin, 2006). If this is true, then digital systems come with a heavy training requirement. Commanders and unit trainers must make time and other resources available for Soldiers to maintain these skills. To develop cost-effective training programs it is necessary to gain a better understanding of which digital skills Soldiers forget the most and the factors that contribute to these performance decrements. Thus, one purpose of the present research is to examine the percentage of Soldiers who remembered specific tasks 8 weeks following an intensive operator class for one of the ABCS components, Force XXI Battle Command Brigade and Below (FBCB2), so that leaders and trainers can develop more effective training for this system.

RETENTION OF DIGITAL SKILLS

Digital skills are discrete, multi-step procedures (e.g., navigation through a series of menus and submenus to set parameters and execute commands). They are predominantly cognitive in nature, ranging from simple to complex skills. Research on skill retention shows that discrete procedural tasks, like digital skills, are more perishable than continuous procedures (e.g., riding a bicycle) or declarative knowledge (Adams, 1987). From this perspective, the anecdotal evidence of the fast rate of decay of digitals skills is not surprising.

Although there is anecdotal evidence that digital skills are easily forgotten, there is little empirical evidence documenting the extent of the decay and the specific tasks that are most likely to be forgotten (Goodwin, 2006). Some empirical evidence has shown significant decreases in skill performance over approximately one month (cf., Goodwin, 2006). In particular, prior research conducted with 54 officers following a two-day FBCB2 familiarization course (Goodwin, Leibrecht, Wampler, Livingston & Dyer, 2007) indicated a 10% decline in performance across 13 tasks 28 days later. The present research builds on this prior work by examining skill retention following the standard Army 40-hour FBCB2 operator course and by using a more sensitive measure of recall.

As the second goal of the research was to identify the characteristics of the tasks that lead to skill decay, a retention measure was designed to determine recall for each step in multi-step tasks. By examining the step-level data, steps showing the most skill decay can be identified and inferences can be drawn regarding the causes for the decay. For example, if a meaningful portion of the sample consistently provided the wrong answer for a certain step in a task, then this may indicate that the system provides vague or misleading cues about how to perform that step. Additionally, by identifying points in a procedure that are particularly challenging for Soldiers to remember, trainers can be given information about where to target extra support (study guides, memory aids, supplemental training, etc.).

INSTRUCTIONAL ENVIRONMENTS

The widespread adoption and availability of digital learning technologies have made synchronous, or high-fidelity, interactions in a distributed environment possible (Bonk & Graham, 2006). Blended learning approaches combine F2F instruction with computer-mediated instruction (Bonk & Graham, 2006) and facilitate student-instructor collaborations, which at one time only occurred in F2F classrooms. Although professors in college/university settings have employed blended learning approaches to support their instruction for quite some time, few examples of this type of instruction can be found in military settings. Thus, there is a need to determine effective methods for using dL techniques to train different echelons and skills.

One notable exception is a US Army National Guard course that has blended instructional approaches to train FBCB2 operator skills. The course is designed for distributed instruction such that the Soldiers are in computerized classrooms in their home states, remote from the instructor, who is located at the Battle Command Training Center (BCTC), Camp Dodge, IA. The instructors use video-teletraining and sophisticated computer software that emulates the FBCB2 system to conduct the digital skills training. As such, the instructors employ a variety of techniques to cover the content, maintain the students' motivation, and sufficiently address students' problems/concerns.

The BCTC at Camp Dodge, IA provides a unique opportunity in which to address the final goal of the present research - to examine performance differences between dL and F2F instructional environments. Thus, the National Guard Soldiers who received FBCB2 training were the dL sample for the present research. For the present research, both the dL and F2F instructors followed the FBCB2 Program Manager's Brigade Battle Command Program of Instruction for their courses indicating that the training materials were presented similarly for both of the training environments. Although direct comparisons of the instructors who taught the classes for the present research cannot be made, past observations of both instructional environments indicated that similar training techniques were used by the instructors of these classes (see results presented in Leibrecht, Goodwin, Wampler, & Dyer, 2007 and Tucker, McGilvray, Leibrecht, Strauss, Perrault, & Gesselman, 2009).

Additionally, preliminary research investigating the different types of training techniques used by the National Guard dL instructors found that the dL instructors were able to adapt to using the technology to teach FBCB2 operations, resulting in training that is very comparable to the traditional courses (Tucker et al., 2009). In particular, positive student comments and observational results demonstrating similar training approaches and topical coverage to those of traditional classes supported these findings. The work also made several recommendations that are applicable for instructors of both types of digital classes (dL and traditional classes) in the areas of (a) leveraging student strengths, (b) emphasizing problem-centered instruction, and (c) leveraging training aids (Tucker et al., 2009).

Although comparing dL instructional approaches to those of traditional classes is important to better understand the context in which students are learning digital skills, the comparison of student outcomes across the instructional environments is critical to determining the effectiveness of the training. It is interesting to note that the assumption that F2F training always leads to better student outcomes has *not* been supported by recent meta-analytic findings. In fact, researchers who conducted a meta-analysis with rigorous standards for study inclusion (i.e., only those studies who met certain methodological criteria were included in the analyses) reported that students who took all or part of their instruction online performed better than those in F2F instructional environments (Means, Toyama, Murphy, Bakia, & Jones, 2009). The results demonstrated that the higher performance levels attained by the online students may be due in large part to their instructors promoting more time on task throughout the courses, not the delivery medium per se.

Although the work by Means et al. (2009) shows it is possible to deliver high quality training in a dL environment, it does not indicate that all dL training is as good as all F2F instruction, nor does it demonstrate that any subject can be taught as well in either environment. As the Army uses dL technology to train an increasing number of skills, it is important to compare Soldier proficiency levels across instructional environments. This will ensure that the skills can be effectively trained in a dL environment and will identify approaches and techniques that are most effective for honing them.

The F2F sample for the present research was Active Duty Soldiers who were trained in the traditional classroom environments with about 20 – 30 students, each with a computer running FBCB2 software. Training typically involved demonstration and explanation of tasks with opportunities for the students to practice the steps on their own computers.

METHOD

Participants

The two samples for the present research consisted of a total of 216 US Army Soldiers who took the baseline test. Of those, 136 (63%) were National Guard Soldiers, who received the dL training described previously, and 80 (37%) were Active Duty Soldiers who completed traditional training. Of those who took the baseline test, 32 (24%) from the dL classes (National Guard Soldiers) and 31 (39%) from the traditional classes (Active Duty Soldiers) took the retention test. The different return rates among active duty and National Guard Soldiers probably had to do with the fact that Active Duty Soldiers were asked to return to the test site during the duty day (unless they had other priorities) whereas the National Guard Soldiers were asked to complete the retention test on their own time. Active Duty Soldiers were tested at Fort Hood and Fort Riley between August 2008 and April 2009. National Guard Soldiers were drawn from 12 different classes from March to October of 2009.

An examination of the demographic characteristics revealed some differences between the two samples (Table 1), such as rank (traditional classes had more junior enlisted personnel while the dL classes had more noncommissioned officers), branch (dL classes had more combat support personnel while traditional classes had more combat service support personnel), and prior FBCB2 use (more Soldiers in dL classes reported that they had never used the system). Soldiers in the traditional classes also were about twice as likely to have had prior FBCB2 training (38% vs. 18%). However, in both samples, enlisted ranks made up at least 90% of the sample, and approximately 75% of the Soldiers in each sample expected to use FBCB2 when deployed (either somewhat likely or very likely). Also, the number of refresher training hours during the 8 weeks between the end of the course and when the Soldiers completed the skill retention measure were the same for the dL and traditional samples (approx 2 hours). Thus, although Soldiers in the traditional classes were more junior than Soldiers in the dL classes, they had more training and experience with FBCB2. In summary, although the two samples differed on some of the demographic characteristics, these differences were not large enough to affect performance over time (see results sections).

Measures

Soldiers' performance was measured with a computer-based multiple choice test. Each question referenced an FBCB2 screenshot, and the responses matched the available choices on that screenshot. Each response was identified with a letter of the alphabet; corresponding letters were superimposed on the screenshot (Figure 1). To allow for multiple ways to complete certain procedures, several of the test questions had branch points where participants' responses determined alternate pathways to perform the task. Tasks were chosen to represent how frequently they were used in combat and criticality to unit mission (Bink, Wampler, Goodwin & Dyer, 2009) as well as relevance to the course.

Three types of performance questions were developed: multi-step full-procedure questions, multi-step partial-procedure questions, and operations screen questions. Table 2 includes a list of all the items that were included in the FBCB2 skill retention measure. For the full-procedure questions, participants had to perform all steps of seven procedures ranging from 5 to 11 steps. For the partial- procedure questions, participants performed only a subset of the steps for five procedures, ranging from 4 to 12 steps. For the operations screen questions, participants were presented with a view of the operations screen (the primary map display with the function buttons on the right) and asked to indicate the first step needed to initiate each of 14 procedures.

Procedure

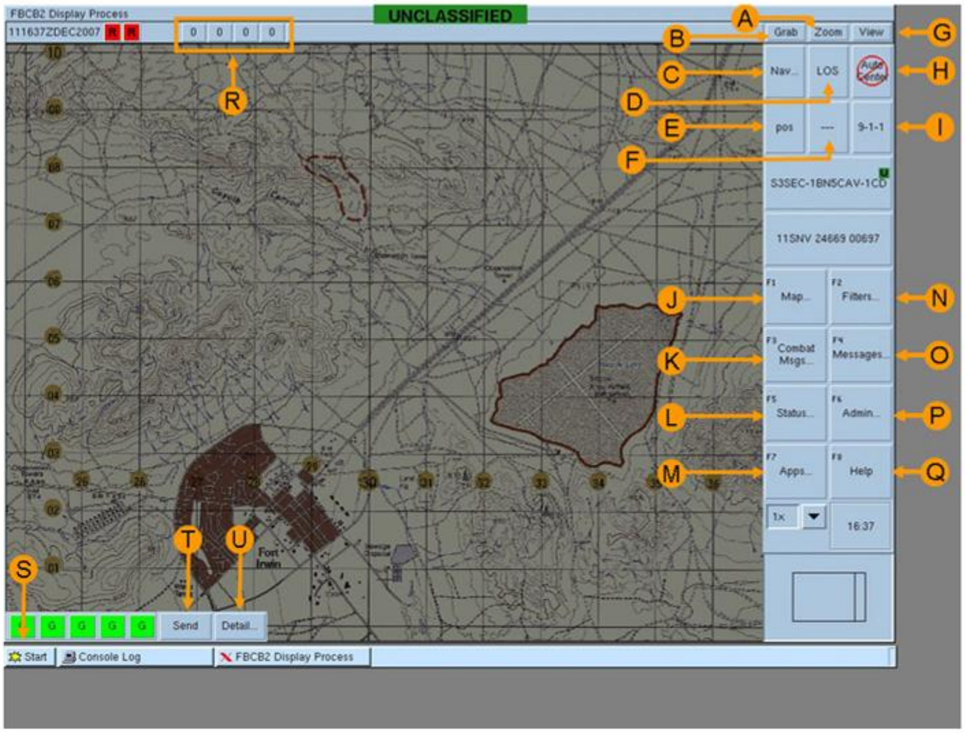
Both samples of Soldiers (F2F and dL classes) were administered the test following their respective training of the 40-hour FBCB2 operator's course and again 8 weeks later. Thus, baseline data for the present research is immediately following the course. The computer-based test could either be administered over the Internet using a standard web browser or as an executable file run from a portable storage device. The test was identical regardless of which mode of administration was used.

Table 1. Demographic Characteristics of the Traditional and dL Samples

Question		Percent of Sample	
		Traditional	dL
Current Grade/Rank:	E3 (Private)	30	5*
	E4 (Specialist/Corporal)	30	28
	E5 (Sergeant)	15	29*
	E6 (Staff Sergeant)	10	24*
	E7 (Sergeant First Class)	5	5
	E8 (First Sergeant/Master Sergeant)	0	5
	O1/O2 (Lieutenant)	9	2*
	O3 (Captain)	0	2
Branch:	Maneuver/Combat Arms	44	37
	Maneuver Support/Combat Support	24	46*
	Sustainment Support/Combat Service Support	32	17*
Likelihood of using FBCB2 if deploying immediately:	Very unlikely	17	10
	Somewhat unlikely	9	16
	Somewhat likely	37	46
	Very likely	38	28
Role/position in which FBCB2 would be used if deploying immediately:	Unknown	35	28
	Primary Operator for a leader	15	8
	Section Leader/Squad Leader	22	24
	Vehicle Commander (other than Leader/Commander)	8	10
	Platoon Leader/Platoon Sergeant	14	4*
	Company/Headquarters Support Element	4	12*
	Company/Troop Commander	0	1
	Staff Officer/ Noncommissioned Battalion or Brigade Tactical Ops Center	3	15*
Self-Ratings of Proficiency on FBCB2	Never Used	1	12*
	Basic	13	45*
	Medium	72	38*
	High	14	5*

Notes. Comparisons based on z-test for equal proportions in independent samples.

* $p < .05$.



2. Your GPS is not operating at the optimum level and you need to troubleshoot the problem. The hardware has been checked and there were no problems. You need to check the software. **What is the first step?**

○ A - Zoom	○ D - LOS	○ G - View	○ J - F1 Map	○ M - F7 Apps...	○ P - F6 Admin...	○ S - Start
○ B - Grab	○ E - pos	○ H - Auto Center	○ K - F3 Combat Msgs...	○ N - F2 Filters...	○ Q - F8 Help	○ T - Send
○ C - Nav...	○ F - ---	○ I - 9-1-1	○ L - F5 Status...	○ O - F4 Messages...	○ R - 0 0 0 0	○ U - Detail...

Previous Question   Next Question

Exit & resume testing later

Figure 1. Example of a Question on the Computer-Based Test

Table 2. Items Included in the FBCB2 Skill Retention Measure

Item Name	Number of Steps Tested
Multi-step Whole Procedure Items (7 Items)	
Create a message folder	5
Set default addressing for a SPOT (size activity location time) Report	8
Clear logs and queues	11
Create, save and send a SPOT report	7/4*
Create and send a route	11
Set screen to display all enemy units and only current, friendly units	6/7*
Show a specified vehicle on the display	6
Multi-step Partial Procedure Items (5 Items)	
Create and save an overlay object group	12/11*
Create and save a Combat Services Support overlay	11
Attach an overlay to an OPORD (operations order)	6
Display a satellite image on SA display	3
Display MGRS (military grid reference system) gridlines on map	4
Operations Screen Items (14 Items)	
Check GPS status	1
Manually place your vehicle icon on the map	1
Enter the MEDEVAC (medical evacuation) call sign and voice net frequency	1
Create a periodic reminder	1
Create an address group	1
Assign message to quick-send button	1
Create and save a position report	1
Activate driver's display for a route	1
Use the circular line of sight tool	1
Display an overlay message	1
Create and send an NBC1 report	1
You must create and send a Mayday report	1
Transmit current platform status/SITREP (situation report)	1
Edit a location folder	1

Note. *Items had two possible correct solutions with different numbers of steps.

BASELINE COMPARISON RESULTS

To examine overall differences in baseline performance for all of the *multi-step procedures* (i.e., whole and partial procedure questions), the percent steps correct for all 12 tasks was averaged for all participants. This score was analyzed in a two factor, instructional environment (dL vs. F2F) by number of measures taken (baseline only vs. baseline and retention), analysis of variance (ANOVA). There were no significant effects of either instructional environment (dL vs. F2F) or number of measures taken (baseline only vs. baseline and recall). However, there was a significant interaction of these two factors, $F(1, 210) = 13.67$, $p < .01$. The Soldiers in the dL classes who took the retention test performed better at baseline than those Soldiers in the dL classes who didn't, $F(1,210) = 10.01$, $p < .01$, whereas Soldiers in the traditional classes who took the retention test performed worse than those Soldiers in the traditional classes who didn't, $F(1,210) = 4.24$, $p < .05$ (see Table 3).

Table 3. Average Performance on the Multi-step Procedures

Instructional Environment	Number of Measures: Baseline Only	Number of Measures: Baseline and Retention
dL	67%	76%*
Traditional	72%	66%*

Note. * $p < .05$ as compared to baseline only group.

The results from analyzing the 12 multi-step procedures individually (full results for all of the analyses presented in this paper are available from Goodwin, Tucker, Wampler, Gesselman, & Johnson, in preparation) revealed few differences between the instructional environments at the procedure-level. As there were no systematic differences between the instructional environments (i.e., the dL classes scored higher on two tasks whereas the traditional classes scored higher on three), a post hoc explanation may be that slight differences in how the classes were executed (i.e., greater emphasis on specific tasks) may have yielded higher performance on certain tasks at baseline.

For the *operations screen questions*, a two factor, instructional environment by number of measures, ANOVA revealed only a significant effect of instructional environment, $F(1, 210) = 18.94$, $p < .01$, with Soldiers in the traditional classes performing better (74% correct) than the Soldiers in the dL classes (60% correct).

The results from analyzing each of the 14 operations screen questions individually indicated that Soldiers in the traditional classes scored higher on seven questions than the Soldiers in the dL classes. A post hoc explanation for these findings may stem from prior research indicating that the instructors for the traditional classes had more operational experience with FBCB2 (cf., Leibrecht et al., 2007; Tucker et al., 2009). Based on this experience, the instructors of the traditional classes may have provided more operational examples throughout the course that allowed Soldiers to better remember the first steps of these tasks.

In summary, when comparing the two instructional environments, the Soldiers in the traditional classes performed better than the Soldiers in the dL classes on the operations screen questions but not on the multi-step procedures. This may reflect either slight differences in the experience of the instructors or differences in the prior experience / training of the two samples or both.

RETENTION RESULTS (ONLY THOSE SOLDIERS AT 8 WEEKS)

Comparisons were made across the two instructional environments for those participants who took the retention measure at 8 weeks. For the *multi-step procedures*, the percent of steps correct was averaged across all 12 items (whole & partial procedures). This average score was analyzed in a two factor, time by instructional environment, ANOVA. This analysis revealed a significant effect for time [71% to 65%; Wilks' $\Lambda = .80$, $F(1, 59) = 14.35$, $p < .01$] but not for instructional environment, Wilks' $\Lambda = .97$, $F(1, 59) = 1.88$, $p = .18$, *ns* (Figure 2). The lack of a significant interaction indicates that the Soldiers in the dL and traditional classes forgot at the same rate overall.

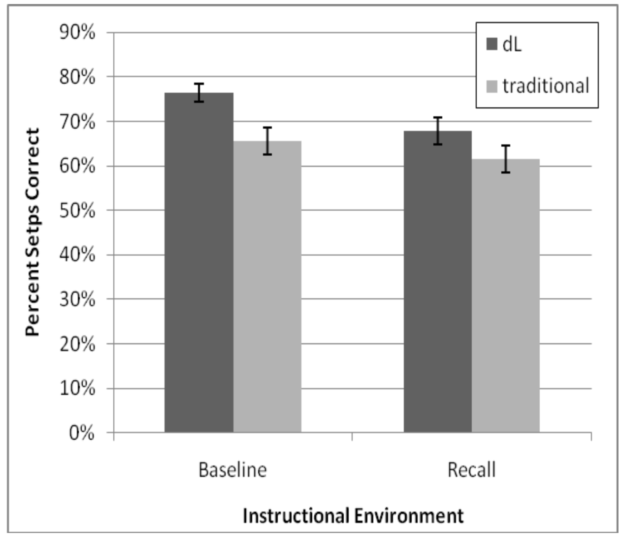


Figure 2. Soldier Performance Over Time on the Multi-step Procedures

An examination of each of the multi-step procedures that significantly declined over time for both groups (based on univariate ANOVAs) revealed that significant decay may be associated with the objective of the task. For example, as junior enlisted personnel were the primary recipients of the training, it is reasonable to expect greater performance declines for the tasks that were primarily leader tasks such as *Create and save a Combat Services Support overlay*. On the other hand, the results revealed very little decay for tasks that were more operator-oriented such as *Display gridlines on the map* and *Create and send a route*. Other factors contributing to decay are discussed below.

For the *operations screen items*, an average percent correct score for all 14 items was analyzed using a 2 (instructional environment) x 2 (time) repeated measures ANOVA. The results indicated a significant effect for time (Wilks' $\Lambda = .44$, $F(1, 61) = 78.82$, $p < .01$) but not instructional environment (Wilks' $\Lambda = .96$, $F(1, 61) = 2.82$, $p = .10$, *ns*), indicating a similar pattern of skill decay for both instructional environments (Figure 3). Factors contributing to the skill decay of individual operations screen questions are described in the next section.

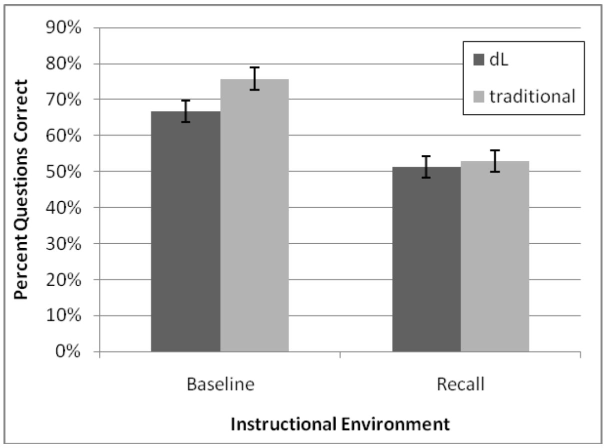


Figure 3. Soldier Performance Over Time on the Operations Screen Questions

Although three of the questions had somewhat different decay rates across the instructional environments, the differences were not systematic across the instructional environments.

Factors Affecting Recall

To determine possible factors affecting recall, we examined the responses on each *step* (individual items) within the multi-step tasks for good and poor recall. Steps with scores of 90% and higher were considered high recall questions (23 steps) whereas steps with scores 50% and lower were considered low recall questions (27 steps). These two cut points were chosen because they are reasonable from a grading standpoint and included about 20% of the steps.

Questions with high rates of recall were divided into three general categories (Table 4). The largest category was comprised of questions that were logically linked to the procedure or step. An example of this category is the step in which Soldiers chose "clear logs and queues" from the start menu for the procedure *clear logs and queues.* Another example is selecting "SPOT report" from the list of report types when performing *create, save, and send a SPOT report*. Another example was completing a step of a procedure such as saving a message after creating it by selecting the "save" button or entering the details of a report in the appropriate data fields. Soldiers also did well when the system prompted them to verify an action such as going offline and on steps that completed items such as "close" or "apply."

There also were three categories of questions that were poorly recalled. The largest category was comprised of questions for which cues in the system did not indicate what needed to be done. Several questions in this category resulted from confusion between the "F5 Status," "F6 Admin," "F7 Apps," and "Start" buttons. These four buttons access a range of administrative, troubleshooting, and miscellaneous functions. For example, for *create a periodic reminder*, the correct choice was "F7 Apps," but a common error was "F6 Admin." When asked to *troubleshoot a malfunctioning GPS*, the correct choice was "F5 Status" but frequent errors were "F6 Admin" and "Start."

Table 4. Characteristics of Well- and Poorly-Recalled Questions

Categories of Recall Factors	Percent of questions
Factors contributing to good recall	
Action in step is logically linked to the overall procedure or question	52%
Verification steps: e.g., "are you sure" or "ok"	39%
Item Completion: e.g., "close" or "apply"	13%
Factors contributing to poor recall	
Vague cues that don't clearly indicate what needs to be done	44%
Misleading cues point to the wrong choice	33%
Forgot to perform some part of the procedure or repeated a step	22%

In other cases, the availability of options was dependent on an action that was not clearly cued. For example, when *attaching an overlay to an operations order*, the system does not indicate that the selection of the "Order Thread" tab activates the "attachments" button. Moreover, some system cues were misleading. For example, when *manually placing their vehicle icon on the map*, many Soldiers chose "F1 Map." The correct choice was "F6 Admin." When asked to *use the circular line of sight* button, many Soldiers chose the "LOS" (line of sight) button when the correct choice was "F7 Apps." When asked to *enter the MEDEVAC callsign*, many Soldiers chose "F3 Combat Messages" where they see the callsign displayed on the MEDEVAC message. To change the call-sign, they had to choose "F6 Admin." In summary, these types of operator mistakes when using the system are likely caused by poor system design features and have serious implications for training programs (discussed in more detail below).

The final category of errors was forgetting a step or repeating a step that had previously been completed. For example when asked to *create, save, and send a message*, some Soldiers created and sent the message and then attempted to close before saving it. In another example, Soldiers attempted to add an icon to an overlay a second time.

DISCUSSION

One goal of this research was to compare the training of a digital system for two different instructional environments: dL and F2F. Despite some demographic differences, student performance over time was comparable across the instructional environments. The strongest differences were found at baseline for the operations screen items. Overall, the findings indicate that the instructional environment only affected baseline performance on the operations screen questions with the Soldiers in the traditional classes performing better than the Soldiers in the dL classes. The performance advantage of Soldiers in the traditional classes disappeared over time. Although there were some significant differences in the multi-step procedures at baseline, these effects also disappeared over time.

Although there were no differences between the instructional environments on the overall baseline score for the multi-step procedures (average of tasks), performance differences were found for Soldiers who took the retention test. The Soldiers in the dL classes who took the retention test had higher overall performance on the multi-step procedures (average of tasks). These findings may be the result of a self-selection bias such that the higher performers at baseline were the ones who completed the second measure. Some evidence of this was found such that the National Guard Soldiers who completed the second measure (n = 31) scored 9% higher on the average multi-step procedure score at baseline compared to the entire National Guard baseline sample (n = 134).

In summary, this research demonstrated that it is possible to train a digital system equally well using either dL or F2F instructional techniques. Additional research is needed to determine whether these findings generalize to other digital systems and dL contexts. That is, the dL technologies available at the BCTC, Camp Dodge, IA enabled dL instructors to maintain a high level of interaction with the students. There may be greater differences in student performance on digital systems for dL courses with more limited instructor-student interactions. Further, as the present research did not assess the Soldiers' performance prior to the instruction, future research should conduct these baseline assessments to more rigorously test for group differences.

Factors Contributing to Skill Decay

An important message for unit leaders is that Soldiers are in need of refresher training on FBCB2 as early as 8 weeks following classroom instruction. Instructors need to provide Soldiers with better ways to remember some of the procedures. Suggestions derived from the research findings for instructors of both dL and traditional courses include developing training to call attention to places where system cues are especially vague, inconsistent, or misleading. For example, confusion about the functions accessed through certain FBCB2 buttons appeared to be a cause of poor performance on a number of steps. Calling attention to the distinctions between functions under these buttons (e.g., Start, F5 Status, F6 Admin, and F7 Apps) or developing exercises and job aids to help Soldiers remember the functions under these buttons would be helpful. Further, there are many different ways to add objects and icons to the map and to overlays. In the overlay toolbox, on the "group setup" tab, the "add icon" button adds a selected object to the overlay object group. On the "object" tab, the "add" button allows the user to place a selected object on the map, and in other places, icons are placed on the map by use of a "map" button. It is easy to see how confusion arises over time regarding the functions of these buttons.

In summary, as system cueing was found to play a significant role in skill decay, the identification and characterization of these cueing problems should help both instructors to develop better training for the current system and system developers to improve the design of the most widely used digital system in the Army. As the results demonstrated that poor interface design affects skill decay, these findings call for better collaboration between engineers and end users when systems are being designed and developed. Although instructors can develop training aids and supplemental material to better support Soldiers as they operate the system, a more effective approach may be to ensure the system features are easy to use.

These findings may be especially useful for trainers involved in new equipment training (NET), as poor system cuing likely will affect the acquisition of knowledge regarding new system use. Additionally, the findings are helpful in understanding that performance on end-of-course exams may not be the best measure of the degree to which operators learned the tasks and procedures. Resource investments in finding out how Soldiers are performing on the job and which procedures prove to be most difficult to remember may be worthwhile to modify training programs to attain greater transfer and higher performance over time.

ACKNOWLEDGEMENTS

The authors are indebted to the Project Officer and FBCB2 instructors at Camp Dodge, IA and Forts Hood and Benning for their assistance with the project.

REFERENCES

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention , and transfer of human motor skills. *Psychological Bulletin*, 101(1), 41-74.

Bink, M. L., Wampler, R. L., Goodwin, G. A., & Dyer, J. L. (2009). *Combat veterans' use of Force XXI Battle Command Brigade and Below (FBCB2)* (ARI Report 1888). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Bonk, C. J., & Graham, C. R. (2006). *The handbook of blended learning: Global perspectives, local designs.* San Francisco, CA: Pfeiffer.

Clark, J. E. (2005). *Solving a command and control system education and training dilemma for the modular force (A white paper).* Washington, DC: Army Joint Support Team.

Goodwin, G. A. (2006). *The training, retention, and assessment of digital skills: A review and integration of the literature* (ARI Report 1864). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Goodwin, G. A., Leibrecht, B. C., Wampler, R. L., Livingston, S. C. & Dyer, J. L. (2007). *Retention of selected FBCB2 operating skills among Infantry Captains Career Course (ICCC) students.* (ARI Research Report 1872). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Goodwin, G. A., Tucker, J. S., Wampler, R. L., Gesselman, A. N., & Johnson, V. (ARI Research Report in preparation). *Soldier performance following distributed and traditional digital skills training*.

Leibrecht, B. C., Goodwin, G. A., Wampler, R. L., & Dyer, J. L. (2007). Techniques and practices in the training of digital operator skills. (ARI Report 1878). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development: Policy and Program Studies Service.

Tucker, J. S., McGilvray, D. H., Leibrecht, C. B., Strauss, C., Perrault, A., & Gesselman, A. N. (2009). *Training digital skills in distributed classroom environments: A blended learning approach* (ARI Report 1893). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.