In *Proceedings of the International Conference on Cognitive Modeling 2020.* 229-235

# The Effect of Task Fidelity on Learning Curves

**Frank E. Ritter (frank.ritter@psu.edu)**
College of IST, Penn State
University Park, PA

**Ashley F. McDermott (amcdermott@cra.com)**
Charles River Analytics
Cambridge, MA

## Abstract

What is the effect of level of simulation fidelity on learning and then on performance in the target task? We consider an example of an electronic maintenance training system with two levels of fidelity: a high fidelity (HiFi) simulation that basically takes as much time as the real-world task and a low fidelity (LoFi) simulation with minimal delays and many actions removed or reduced in fidelity and time. The LoFi simulation initially takes about one quarter of the time, and thus starts out providing about four times as many practice trials in a given time period. The time to perform the task modifies the learning curves for each system. The LoFi curve has a lower intercept and a steeper slope. For a small number of practice trials, this makes a significant difference. For longer time periods, the differences between low and high fidelity get smaller. Learners that move from low to high appear to not be adversely affected. We note factors that could influence this transfer (i.e., subtasks included in each simulation), and how this approach could be extended.

**Keywords:** fidelity, training systems, learning curves

**Acknowledgements:** Jake Graham, Jong Kim, Jacob Oury, Fred Ryans, Martin Yeh, and two anonymous reviewers provided useful comments. This work was supported by ONR, N00014-18-C-7015 and N00014-15-1-2275.

## Introduction

What is the effect of varying the level of simulation fidelity on learning in that simulation and on the more complete learning situation? What happens to learning when a learner practices in a simpler simulation and then moves to a more realistic or higher fidelity simulation? In this paper we explore how task fidelity affects how fast a task is learned in an example task, and analyze what this means through analyses using learning curves.

We consider these questions by using an example of a maintenance training system with two levels of fidelity: (a) a simple system with minimal delays and with many actions removed or reduced in fidelity and time, and (b) a full fidelity simulation that basically takes as much time as the real world task. The higher fidelity simulations take longer to perform a more complete task including all sub-tasks. The low fidelity simulation starts out taking about one quarter of the time to complete, and thus starts out getting about four times as many practice trials in a given time period.

The task complexity in the systems influences the time to perform the task, and this in turn modifies the two learning curves, both in the intercept and in the learning rate. We will show that for a small number of practice trials, this difference in trial time makes a significant difference in the curves. For longer time periods, the differences between low and high fidelity get smaller. The amount of training the tasks being trained will receive will thus influence the choice of fidelity as well.

After briefly reviewing the effect of training system fidelity we introduce a maintenance task we have developed to study learning and retention. We will use a simple model based on ACT-R and Soar of how the task is performed and learned. Based on the learning curves we are able to draw some new conclusions about the effect of fidelity on the effectiveness of training, notably that using lower fidelity training situations help most where there is only modest time to practice, and that if there is extensive time to practice full fidelity has nearly the same outcome (but perhaps not the same costs or risks) as does starting with a simple simulation and moving to the complex simulation.

## Literature Review of Fidelity

There is a long-standing debate of the effects of fidelity on training with simulators. The early research on fidelity was based on the natural assumption that higher fidelity would necessarily lead to better learning, since the simulation would more closely resemble the actual system (e.g., Allen, Hays, & Buffardi, 1986; Miller, 1954; Noble, 2002). However, much of the research supporting this notion was conducted from the 1950s to 1980s, so it had a low ceiling for how representative high simulation fidelity could be at the time. There is also a body of research showing that higher fidelity is not always desirable to maximize learning (e.g., Dahlstrom, Dekker, van Winsen, & Nyce, 2009; Havinghurst, Fields, & Fields, 2003; Lesgold, Lajoie, Bunzon, & Eggan, 1992; Swezey, Perez, & Allen, 1991).

Delving into this literature quickly leads into the question of what fidelity actually means. The most common distinction is surface or physical fidelity versus operational or task fidelity (Allen et al., 1986; Liu, Macchiarella, & Vincenzi, 2009). Within physical fidelity there are still many dimensions, including visual clutter, visual layout, auditory fidelity, and haptic fidelity. All of these dimensions have the potential to affect both speed of learning and degree of transfer to the real task. Some of these dimensions, however, are not relevant to the task being taught. To properly learn a task, the simulation should have reasonably high fidelity on the task-relevant dimensions (Prophet & Boyd, 1970; Thorndike & Woodworth, 1901), but the irrelevant dimensions should be kept at a lower fidelity to minimize distraction from the task (Alessi, 1988).

An additional factor that affects task time and transfer of learning is the experience level of the learner (Alessi & Trollip, 1991). The experience of the learner will affect the cognitive load associated with higher fidelities and the dimensions of fidelity that could be considered task relevant (Alessi, 1988). For example, an expert who is used to using the actual interface but is doing additional training will likely experience less cognitive load with a nearly full fidelity simulation than a novice learning about the interface for the first time. Additionally, due to their experience, experts may find not having the appropriate haptic or audio cues or incorrect timings in the simulator to be a distraction to learning, while including these details would be distracting for a novice. Similarly, the age of the learner can affect what sorts of interfaces are easily usable. A low fidelity simulation could introduce interactions that are natural for younger adults but novel or slower for older adults (John & Jastrzembski, 2010).

The question of when higher fidelity is better for learning continues to be debated because it is not clear why or when lower fidelity simulations provide the most advantage. As we have discussed, experience of the learner and cognitive load are considered to be two important contributing factors, as is the type of task. In this paper we propose an additional factor, the number of repetitions of the task (or subtask) that a learner is able to complete while training.

## A Simple Task Model of Learning and Fidelity

To examine the effect of fidelity on learning we use an example simulation, the Ben-Franklin (BF) Radar (Ritter, Tehranchi, Brener, & Wang, 2019). Figure 1 shows a schematic for the BF Radar system included to show its relative complexity, not its details. The system has 35 replaceable components that can have faults, and 15 switches and a power supply that cannot have faults. The system is based on the Klingon Laser Bank task (Friedrich & Ritter, 2020; Kieras & Bovair, 1984; Ritter & Bibby, 2008) and on a functional radar system (Charvat, 2011).

The schematic shows five subsystems. The subsystems vary in their complexity and connectivity within them and across subsystems. The blue lines in Figure 1 are power connections; the red lines are information; the purple lines are both. The schematic also identifies certain components that have their status displayed on the front panel of the BF Radar.

There are several tasks that can be performed with the BF radar system. Users can turn it on; users can correctly adjust switches so that it works; users can find a single fault and replace it; and users can find and replace multiple faults. The task that we will use to examine the effect of fidelity on learning is to find a single broken component, a fault. Single broken faults create a unique light configuration and are always solvable.

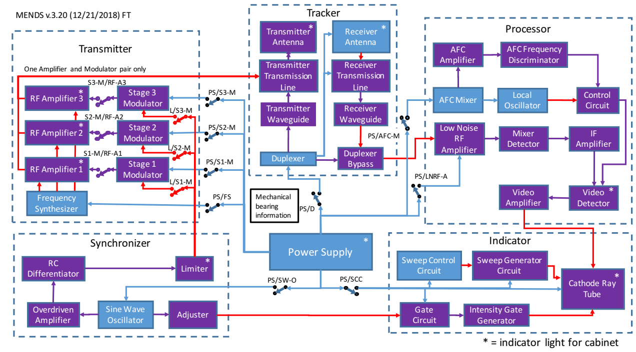


Figure 1. Schematic of the Ben-Franklin Radar simulation.

The task was created to support troubleshooting within the confines of a study, and to be more complex than the Klingon Laser Bank task, but not so complex that it would take more than an hour to learn. This system can be and has been realized in several ways with different levels of assumed fidelity.

## Task Simulations

In our analysis we examine two potential implementations of the BF Radar device. The first (Low Fidelity) is realized in software and is being used in another study. The second (High fidelity) is realized in hardware, and has been partially built.

**Low Fidelity (LoFi) Simulation** Figure 2 shows the general layout (not the details) of MENDS, a low-fidelity simulation of this system. The system is implemented in Unity. The front panel (top image) shows the subsystems and the lights in the upper right corner of each square shows which subcomponents are working. An individual tray (bottom image) shows a tray and the components that are working (yellow light and white) and the components that are not working (red light and grayed out). This system has been briefly reported before (Ritter, Tehranchi, Brener, et al., 2019).

To troubleshoot the task (the details are in Table 1) the user clicks for the next problem, examines the lights, clicks on a tray, and examines its contents. They must then choose the broken component by clicking on it and clicking done.

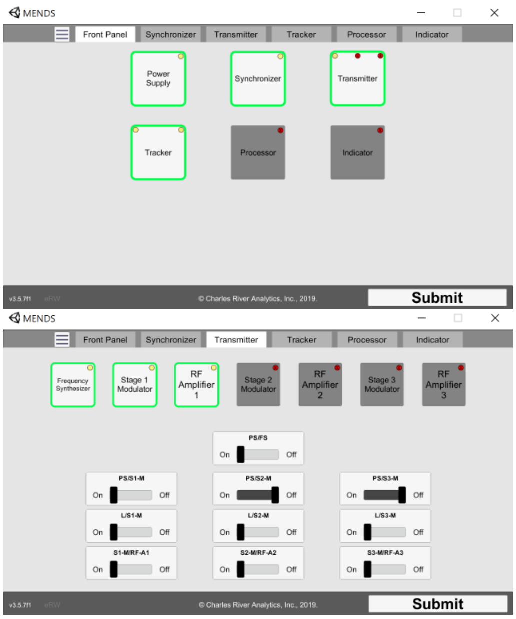


Figure 2. The initial interface of the MENDS low-fidelity task (top) and a tray (bottom). In these pictures, the fault is in the Processor subsystem.

**High Fidelity (HiFi) Simulation** Figure 3 depicts the higher fidelity version of this system. This system is realized in an approximate 2 ft physical metal and component cube using Raspberry Pi's and in Unity 3D. It has a cabinet holding trays for each subsystem. The top tray provides a summary of the system, including the indicator lights. The other trays each hold one subsystem.

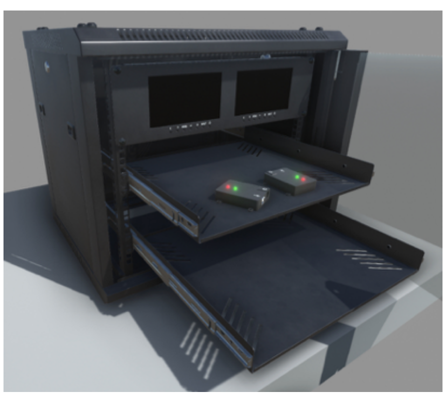

Figure 3. The high-fidelity realization of the BF Radar, showing the cabinet (with its door removed), two racks that will hold one subsystem each, and two components that will be inserted into component holders on the tray when the tray is built-out.

To troubleshoot the task (as an overview, the details are in Table 1) the user must first put on a grounding strap, and then examine the lights, open the cabinet door, pull out a tray, and examine its contents. They must then choose the broken component, find the replacement part, and replace the broken component. To set up a trial, the experimenter must have the user look away, replace a working component with a broken component, and then close the tray and the door.

## The Task Assumptions

We assume that the user has been taught the BF Radar schematic and has it available, either in their head or on a sheet of paper. Table 1 shows the subtask times and the total time. The times are broken up into Learnable and Fixed tasks. The learnable tasks improve with practice; the fixed do not. These steps and their times (similar to and taken from the Keystroke-Level Model, Card, Moran, & Newell, 1983) are shown in Table 1. The user will start with the front display panel, and will have to examine the lights to know what tray and component to examine. Each step takes time, and we assume is error free. We use the Overdriven Amplifier, an early component in the system, as the example fault for this analysis.

## The Learning Theory

The time to perform a task is broken down into two types of time: skills to be learned and skills that are already learned (or, essentially learned). Skills that are to be learned get improved with practice. On this task, learnable skills include: recognizing the lights and their implications. Skills that are essentially already learned are moving the mouse and clicking, and system response times include replacing faults or inserting faults by the system.

These times are used to compute the time to do the task using Eq. 1. This equation is consistent with Soar's (Newell, 1990) and ACT-R's (Anderson, 2007; Ritter, Tehranchi, & Oury, 2019) learning theories, and the learning curve in general (Ritter & Schooler, 2001). The times are computed in 10-minute blocks. Thus, the first 10 min. block of trials are done at 2.7 (LoFi) and 1.0 (HiFi) trials per minute, then in the second block the pace is updated to reflect what is learned after 10 min. This is repeated for nine more blocks.

(1) $\text{Time} = \text{Fixed tasks} + \text{learned tasks}\ (\text{Trial})^{-\alpha}$

The choice of α (alpha) was arbitrarily chosen as 0.2. This value of α is consistent with values from Newell and Rosenbloom (1981, 0.06 - 0.81, a variety of tasks); and similar to values from Delaney, Reder, Staszewski, and Ritter (1998, 0.265 - 0.510, mental arithmetic); and Kim and Ritter (2016, 0.4 – 1.2, spreadsheet tasks).

We also looked at the time if users were to move back to the HiFi trainer at the end of each 10-min. block. That is, if a user were to train on the LoFi simulator and then move to the HiFi simulator at the end of each block. This curve is thus not a learning curve, but shows how well the learner would perform in the HiFi simulator after that much practice in the LoFi simulator. Equation 2 shows how that time is computed. We include the power law effect for the new task, but on the subtasks in the HiFi simulation, they have not been learned, and thus they are trial 1. This is just the subtask time itself, no learning has occurred.

(2) Time = Fixed tasks(Hi) + Learnable tasks only in Hi $(1)^{-\alpha}$ + (learnable tasks in both) $(\text{Trial})^{-\alpha}$

Table 1. Task analysis. The fault modeled is the Overdriven Amplifier fault. (times are in s.)

| | | MENDS Low Fidelity | | 3D High Fidelity | |
|---|---|---|---|---|---|
| | | Skills | | Skills | |
| | | Learnable | Fixed | Learnable | Fixed |
| | Sub total time (s) | 15.75 | 7.55 | 32.75 | 26.00 |
| | Total | 23.30 | | 58.75 | |
| | Trials per min. | 2.6 | | 1.0 | |
| Step | Setup | | | | |
| 0 | Insert fault | | 0.05 | 12 | 0 |
| | | click "Next" | | insert fault | |
| 1 | Approach System | | 1.5 | | 5 |
| | | | | | |
| 2 | Open cabinet | 1 | 1.5 | 1 | 3 |
| | | click to open | | lift lid | |
| 3 | Ground yourself | 0 | 0 | 4 | 6 |
| | | not required | | | |
| | Find subsystem | | | | |
| 4 | Check front panel | 5.25 | | 5.25 | |
| | scan first panel, 0.25 per light plus Mop | | | | |
| 5 | Choose tray | 1.5 | 0 | 1.5 | 0 |
| | | Mop | | | |
| 6 | Open tray | 1 | 1.5 | 2 | 4 |
| | | move+click | | reach etc. | |
| | Tray loop | | | | |
| 7 | Check light | 0.25 | | 0.25 | |
| | | scan first tray, 0.25 per light | | | |
| 8 | If off, have answer | 1.5 | | 1.5 | |
| | If on, go to next light | | | | |
| 9 | Check light | 0.25 | | 0.25 | |
| | Overdriven amp fault | | | | |
| 10 | Recognize fault | 2 | | 2 | |
| | Replacement | | | | |
| 11 | Indicate fault | 1.5 | 1.5 | 1.5 | 6.5 |
| | | Mop+move+click | | replace physically | |
| | Confirm done | 1.5 | 1.5 | 1.5 | 1.5 |
| | | Mop+move+click | | say "done" | |

## Simulation Results

Table 2 shows the number of repetitions of the tasks that arise across the ten 10-minute blocks. The LoFi simulator has a much larger number of reps, and this difference is maintained across the total training time, although the ratio between HiFi and LoFi decreases with practice.

Table 2. Number of total task repetitions over ten 10-min. blocks.

| | Total Repetitions | | |
|---|---|---|---|
| **Block** | **LoFi** | **HiFi** | **Ratio** |
| 1 | 26 | 10 | 0.26 |
| 2 | 64 | 23 | 0.24 |
| 3 | 105 | 37 | 0.23 |
| 4 | 149 | 51 | 0.22 |
| 5 | 194 | 66 | 0.22 |
| 6 | 240 | 81 | 0.22 |
| 7 | 287 | 96 | 0.22 |
| 8 | 334 | 111 | 0.22 |
| 9 | 382 | 127 | 0.22 |
| 10 | 431 | 142 | 0.22 |

Figure 4 shows the learning curves for the HiFi and LoFi simulations, in linear and log-log coordinates. There is an additional line showing the response time for a user that practiced with the LoFi simulator and then moved to the HiFi simulator.

The plots show that the low fidelity users would get extremely fast on the material that is taught (green triangle, dashed line) compared to the high fidelity (blue square, solid line). The intercepts are different; the low-fidelity group starts out faster. And the slopes are different (best shown in the log-log plot), the low-fidelity group learns at a faster rate because they get an increasingly large number of repetitions because they are using a faster interface. With increasing practice, the low fidelity group remains faster, but the difference decreases as the power law effect is applied; that is, it takes increasingly larger amounts of practice for decreasing gains.

But, where would the new learners be on the whole task (HiFi) if they move to the HiFi after working with the LoFi simulator? When the low-fidelity group moves back to the high fidelity interface (black circle, dotted line) the effect of practice with the LoFi simulator is most pronounced early on. The black line shows not practicing all the tasks in the HiFi simulator can lead to faster times, but that this effect decreases with practice. And, if there was one or several learnable tasks in the HiFi task that were not in the LoFi task, the LoFi transition line could conceivably come in higher than the HiFi task at some point.

## Human Participant Data that We Have So Far

We have three sets of data related to this task. On the original Laser Bank task, Ritter and Bibby (2008) saw reaction times ranging from 20 s initially to around 7 s when practiced. Friedrich and Ritter (2020) reported similar times.

In the MENDS task (LoFi interface), Ritter et al. (2019) saw a subject with 10 minutes of practice that went from 60 s to 22 s. The initial trials were thus slower, but the final time after 10 min. is approximately accurate.

We are currently running a study that will gather more data on the low fidelity version. We have run 8 out of 115 human participants so far.

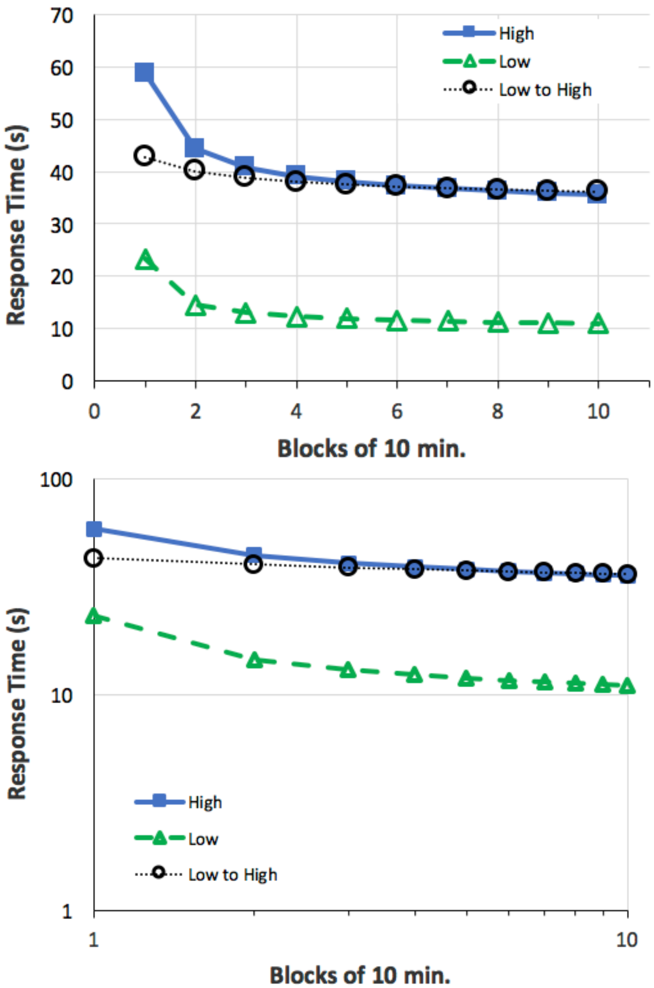

Figure 4. Response time for the 10 blocks for the high, low, and transitioned training schedules (linear, top; log-log, bottom).

## Discussion and Conclusion

This analysis can help explain why there are still discussions about whether to choose low or high fidelity simulations. The analysis is sensitive to different assumptions about time costs of the two training systems. The analysis shows that the expected factors influence the amounts of learning: previous training on a task, setup costs for a training task, what can be transferred, what can be trained, and how much training is required. Each will influence the learning curves and the differences between the two levels of fidelity. This analysis points out that it is probably worthwhile to note and document what tasks are being trained in each system, and how many repetitions they are getting.

What is also clear is that time to train is an important measure. When there is a lot of training time (i.e., a large number of training trials is available), a low fidelity trainer does not offer as much benefit as when training time on the full system is limited. If a low fidelity trainer is available, it might not so much save time but save money (or lives or equipment if the situation being trained is dangerous). Lower fidelity training systems, if they cost less, can also lead to large learning gains even when transitioned to more complex tasks, and this has been seen before (Alluisi, 1991; Caro, Isley, & Jolley, 1973). It would be interesting to put those situations into these analyses.

This approach thus offers a calculus, a way for choosing how and why to use different levels of simulations. It can provide support for how much more training can be obtained from each type of simulation.

It could also be used to avoid the awkward situation where spending effort to make the simulation/training more faithful to the external environment by including behavior that is not greatly influenced by learning would none-theless lead to learning less. Tasks that do not get faster and do not get learned are cases where fidelity could be dropped.

In this task, there does not appear to be a cost to starting low and going to the high fidelity training situation. This approach can save substantial time and resources. This approach shows that for this task there appears to be no cost to starting low and going high, unless there are essential skills that are learned and that are not in the LoFi simulator. Putting on a grounding strap, for example, if it was learnable and not taught in the LoFi simulator, could have an important role in this story.

We have run this analysis of the grounding strap as an example task only in the HiFi task. The curve indicates that in the first few training blocks, the LoFi interface still leads to faster performance. As amount of training increases, there is a cross-over point where the low fidelity performance is dominated by not having practice on the unpracticed task, and when the user transfers to the HiFi task, they are slower than the full task for the same training time. This effect should be explored further.

### More Repetitions Are Important Early

The analysis shows that if you have only a short period to train, it is better to have learners on a low fidelity system. Figure 4 shows that the low fidelity when transitioned back to high was faster than only high fidelity because the learner had more practice on what could be learned. Performing more repetitions in the same period of time has a greater effect on learning when there is not a lot of trials. On the other hand, at larger amounts of practice, learners on the low fidelity do not gain as much relative to the high fidelity as they do at low practice time. The low fidelity is still faster, but the effect is smaller. In some situations this will still greatly matter (where differences in response time are important, such as adversarial tasks), and in some situations this will not (perhaps in safety tasks where doing the task correctly and slow is good enough).

### Limitations

There are several limitations to this analysis. We have revised the task in Table 1 numerous times. Thus, there are likely to still be some inconsistencies. The model's general predictions appear to be robust against these changes, however. As we updated Table 1 while writing this paper, the curves in Figure 4 did not substantially change.

These analyses do not account for other differences in training systems such as cost, risk to the learner, environment, and equipment, time to get to the system, and so on. These are important considerations, and will have an important impact on training system choice.

## Future Work

As a next step we are moving this analysis to R and doing a more detailed analysis. We will examine different tasks (faults) as well. We continue to run the study of the LoFi condition. The physical apparatus will provide more detailed empirical results to support this approach.

There are several analyses that we would like to do in the future. We would like to explore what happens when there are more tasks that are not trained in the LoFi simulation. This may lead to a situation where coming back to the HiFi task from the LoFi simulator is slower than staying on the HiFi curve. There currently is the use of the grounding strap as an example subtask. There could be numerous tasks like this in other situations. Other considerations such as cost may also be important.

We would like to generate a set of plots showing the effect of changes in learning rate (e.g., 0.1 to 1.0 in 0.1 steps). Exploratory analyses show that higher learning rates can alter the curves and the relative value to each level of simulator fidelity. It would also be interesting to see the net cost of the LoFi curves, either in training time or training costs.

It would be useful to make this analysis even easier to use. It could then be used to analyze more realistic, complex tasks, for example, as IMPRINT does (Booher & Minninger, 2003). This analysis could include the costs of building the additional LoFi interface. This tool could even go so far as to predict the cost of each component in the LoFi interface (e.g., building it out more could cost a little more but lead to greater learning savings, system saving, or system effectiveness). This approach can also be informed by tools to model users in interfaces automatically (John & Jastrzembski, 2010; Wallach, Fackert, & Albach, 2019), and could be potentially included in them.

This work can lead to a better method to determine optimal simulator training time based on examining performance improvement through using learning curves.

## References


Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction, 15*(2), 40–47.

Alessi, S. M., & Trollip, S. R. (1991). *Computer-based instruction: Methods and development*. Englewood Cliffs, NJ: Prentice-Hall.

Allen, J. A., Hays, R. T., & Buffardi, L. C. (1986). Maintenance training simulator fidelity and individual differences in transfer of training. *Human Factors, 28*(5), 497-509.

Alluisi, E. A. (1991). The development of technology for collective training: SIMNET, a case history. *Human Factors, 33*(3), 343-362.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Booher, H. R., & Minninger, J. (2003). Human systems integration in Army systems acquisition. In H. R. Booher (Ed.), *Handbook of human systems integration* (pp. 663-698). Hoboken, NJ: John Wiley.

Caro, P. W., Isley, R. N., & Jolley, P. B. (1973). *Research on synthetic training: Device evaluation and training program development* (Technical Report 73-20). Fort Rucker, AL: Human Resources Research Organization (HumRRO), Division No. 6 (Aviation).

Charvat, G. L. (2011). MIT IAP 2011 laptop based radar: Block diagram, schematics, bill of material, and fabrication instructions. from https://ocw.mit.edu/resources/res-ll-003-build-a-small-radar-system-capable-of-sensing-range-doppler-and-synthetic-aperture-radar-imaging-january-iap-2011/

Dahlstrom, N., Dekker, S., van Winsen, R., & Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomics Science, 10*(4), 305–314.

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy specific nature of improvement: The power law applies by strategy within task. *Psychological Science, 9*(1), 1-8.

Friedrich, M. B., & Ritter, F. E. (2020). Understanding strategy differences in a diagrammatic reasoning task. *Cognitive Systems Research, 59*, 133-150.

Havinghurst, L. C., Fields, L. E., & Fields, C. L. (2003). High versus low fidelity simulations: Does the type of format affect candidates' performance or perceptions? In *27th Annual IPMAAC Conference on Personnel Assessment: Exploring New Horizons in Assessment*.

John, B. E., & Jastrzembski, T. S. (2010). Exploration of costs and benefits of predictive human performance modeling for design*.* In *Proceedings of the 10th International Conference on Cognitive Modeling*, 115-120. Philadelphia, PA: Drexel University.

Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning how to operator a device. *Cognitive Science, 8*, 255-273.

Kim, J. W., & Ritter, F. E. (2016). Microgenetic analysis of learning a task: Its implications to cognitive modeling*.* In *Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016)*, 21-26. University Park, PA: ACS Lab, Penn State.

Lesgold, A. M., Lajoie, S., Bunzon, M., & Eggan, E. (1992). SHERLOCK: A coached practice environment for an electronics troubleshooting job. In J. Larkin, R. Chabay & C. Scheftic (Eds.), *Computer assisted instruction and intelligent tutoring systems: Establishing communication and collaboration*. Hillsdale, NJ: Erlbaum.

Liu, D., Macchiarella, N. D., & Vincenzi, D. A. (2009). Simulation fidelity. In *Human factors in simulation and training* (pp. 61-73): Taylor & Francis.

Miller, R. B. (1954). *Psychological considerations in the design of training equipment*. (WADC Report No. 54-563, AD 71202). Springfield, OH: Carpenter Litho & Prtg. Co.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-51). Hillsdale, NJ: Erlbaum.

Noble, C. (2002). The relationship between fidelity and learning in aviation training and assessment. *Journal of Air Transportation, 7*, 34–54.

Prophet, W. W., & Boyd, H. A. (1970). *Device-task fidelity and transfer of training: Aircraft cockpit procedures training* (Technical Report 70-10). Fort Rucker, AL: Human Resources Research Organization (HumRRO), Division No. 6 (Aviation).

Ritter, F. E., & Bibby, P. A. (2008). Modeling how, when, and what is learned in a simple fault-finding task. *Cognitive Science, 32*, 862-892.

Ritter, F. E., & Schooler, L. J. (2001). The learning curve. In W. Kintch, N. Smelser & P. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 13, pp. 8602-8605). Amsterdam: Pergamon.

Ritter, F. E., Tehranchi, F., Brener, M., & Wang, S. (2019). Testing a complex training task. In *Proceedings of the 17th International Conference on Cognitive Modeling (ICCM 2019)*, 184-185.

Ritter, F. E., Tehranchi, F., & Oury, J. D. (2019). ACT-R: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(3), Paper e1488.

Swezey, R. W., Perez, R. S., & Allen, J. A. (1991). Effects of instructional strategy and motion presentation conditions on the acquisition and transfer of electromechanical troubleshooting skill. *Human Factors, 33*(3), 309-323.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions Part I. *Psychological Review, 8*(247–261).

Wallach, D. P., Fackert, S., & Albach, V. (2019). Predictive prototyping for real-world applications: A model-based evaluation approach based on the ACT-R cognitive architecture. In *DIS '19: Proceedings of the 2019 on Designing Interactive Systems Conference*, 1495–1502.