Doozandeh, P. & Hedayati, S. (2022). The effect of simulation fidelity on transfer of training for troubleshooting professionals: A meta-analysis. *IISE Transactions on Occupational Ergonomics and Human Factors*, (in press). https://doi.org/10.1080/24725838.2022.2099483

The Effect of Simulation Fidelity on Transfer of Training for Troubleshooting Professionals: A Meta-Analysis

Pooyan Doozandeh, PhD[1] and Shekoofeh Hedayati, PhD[2]

[1] Corresponding Author
pooyan.doozandeh@gmail.com
https://orcid.org/0000-0001-7017-9403
College of Information Sciences and Technology
The Pennsylvania State University
E346 Westgate Building, University Park, PA 16802, USA

[2] shokoufeh.hed@gmail.com
https://orcid.org/0000-0002-5076-4004
Department of Psychology
The Pennsylvania State University
Moore Building, University Park, PA 16802, USA

Author Notes

Both authors contributed to study design and drafting the manuscript. Pooyan Doozandeh designed the experiments, searched the literature, coded data, and wrote the manuscript. Shekoofeh Hedayati helped with searching the literature, coding data, and providing extensive comments in all steps. This study was conducted in a timeframe between November 2020 and the summer of 2021. This meta-analysis was inspired by a joint project between Charles River Analytics Inc. and Penn State University, with the project title “Simulating Training Results to Understand Differing Effects of Fidelity on Learning (STRUDEL)”, but did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. No conflict of interest is reported by the authors.

Occupational Applications

People in many occupations that involve using complex tools need to know how to troubleshoot those tools in real time and with minimum cost. Training troubleshooting professionals is thus a concern for various occupational sectors—particularly the military, aviation, power plant, and in industrial processes—and large investments have been made to create and use simulators that train troubleshooting skills. In the design and evaluation of simulators for troubleshooting, this review shows that no single level of simulation realism—or *fidelity*—works best in training, and that the effect of fidelity depends on trainees' prior skill level and the type of troubleshooting system (electronic or mechanical).

Technical Abstract

Background. Due to methodological difficulties, determining the appropriate level of simulation fidelity for training has been a long-standing problem for researchers and practitioners in ergonomics and human factors as well as simulation educators.

Purpose. Our goal was to understand whether and how different levels of simulation fidelity affect transfer for training troubleshooting professionals, with a focus on practice domains such as military, chemical plants, and aviation. In analyzing the effect of fidelity, we also assessed the potential effects of two moderators: trainees' prior skill and system type (electronic and mechanical).

Method. We used quantitative (random effects) and qualitative meta-analytic techniques to address the study questions. To overcome traditional problems in measuring fidelity and transfer, instead of using quantitative measures we conducted a qualitative categorization of study variables into low, medium, and high levels. Reports from 1960 until the present (2021) that described controlled experiments were identified using online databases, which resulted in 200 reports, 25 of which satisfied our conditions and included 57 experiments with 1,481 human participants.

Results. Although the overall results favor using medium- to high-fidelity simulators, none of the low-, medium-, or high-fidelity simulations were universally superior, and the effect of fidelity depended on identified moderators. There was a positive effect of fidelity on transfer, but only for trainees with high prior skill. The same effect was also observed only for electronic systems. Of the three level of fidelity, medium-fidelity simulators produced the highest overall transfer, especially for trainees with low prior skill, and the low-fidelity simulators resulted in the lowest overall transfer.

Conclusion. In designing and evaluating simulators that train troubleshooting professionals, addressing the fidelity question is only possible by analyzing important moderators such as trainees' prior skills and system type. Researchers and practitioners should thus define such moderators and then decide on key design variables such as fidelity.

Keywords: simulation fidelity, training simulation, transfer of training, simulation-based skill acquisition, virtual environments

Manuscript word count (excluding Abstract, References, tables, and figures): ~ 7,000

1. Introduction

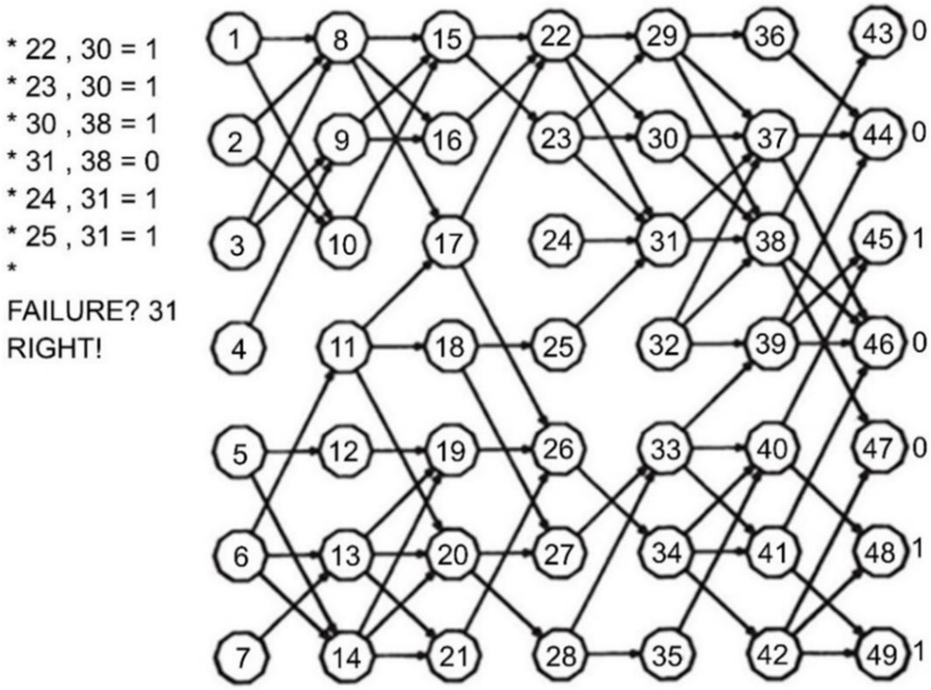
Organizations use training simulators in various occupations to train their workforce in an efficient and safe environment. For example, one study estimated $6 billion of savings in training costs for the United States Marine Corps for one year from using training simulators at a large scale (Cooley et al., 2015). Other works show the introduction and effectiveness of virtual and augmented reality tools (VR/AR) in training, for occupations such as construction (Hafsia et al., 2018), work zone inspection (Aati, et al., 2020), power line maintenance and operation (Borba et al., 2016), work-at-height engineering operations (Di Loreto et al., 2018), and in various other occupations (see also Anastassova & Burkhardt, 2009). Because of their importance, the design of training simulators has been a subject of decades of research in various disciplines, particularly in human factors and ergonomics (e.g., Allen et al., 1986; Duncan & Shepherd, 1975). One long-standing and central design problem that has baffled researchers is the level of surface realism of training materials—or *fidelity* as used in this article—that is needed in a simulator (e.g., Hays & Singer, 1989; Miller, 1954). The central question is: how realistic should the training materials be to provide successful training? Answers to the *fidelity question* can provide designers with specific guidelines in their design practice and would allow organizations to evaluate simulators.

Unlike some recent attempts to address similar questions that had a domain-general approach to the problem (e.g., Kaplan et al., 2021), we think we need to focus only on one domain of practice so that our results are readily applicable to that specific domain. A task domain that has been using training simulators for decades is troubleshooting (e.g., Hochholdinger & Schaper, 2013; Rasmussen & Rouse, 1981; Shriver et al., 1964). Various organizations—specifically the military—use sophisticated electronic and mechanical systems. To use those systems in critical situations, operators must be trained on how to diagnose faulty components and troubleshoot in a reasonable timeframe. A big part of saving costs and reducing threats in the military is due to training troubleshooting skills (Cooley et al., 2015).

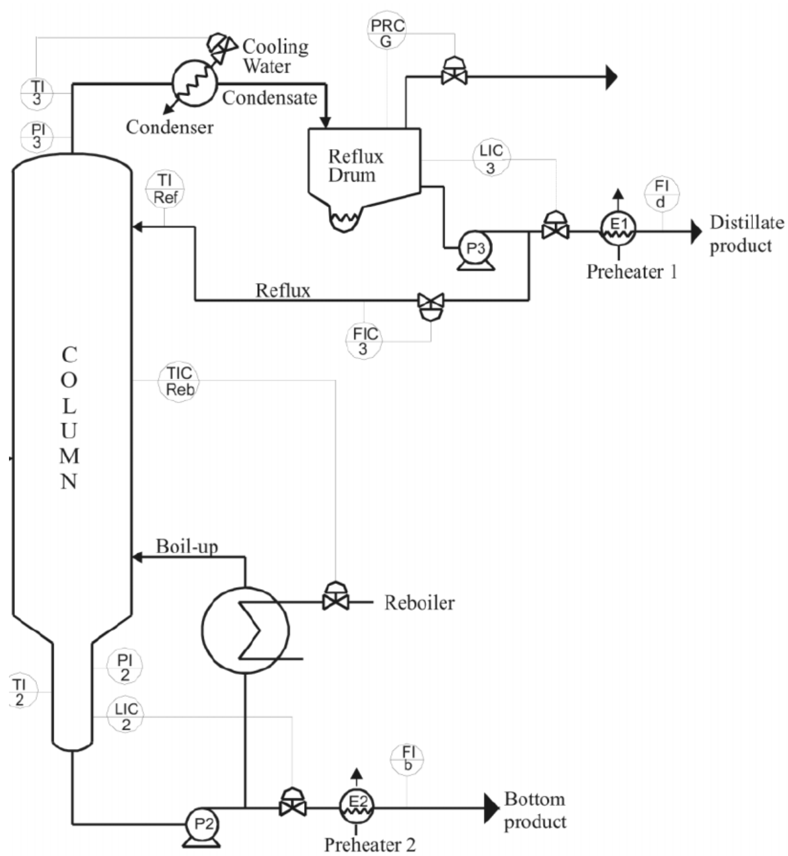
A troubleshooting professional first observes the output (or current state) of a system; if there is a fault in the output, it should be traced back through its connections to one or multiple components within the system that are responsible for the fault. Examples of troubleshooting simulators are shown in Figure 1. In electronic systems, components are resistors, diodes, capacitors, etc. that are often connected with wires in a circuit, and in mechanical systems components include thermostats, generators, fuel injectors, etc. connected by pipes, gears, or other connectors. So, finding faulty components is the main goal in troubleshooting and is our focus in this article. Often, the expertise lies in choosing a strategy—e.g., what internal components should be tested first—to find the faults resulting in an accurate, fast, and inexpensive fix for the system (see Johnson, 1988; Morris & Rouse, 1985). Mastering this skill requires practice, and for various reasons such as safety and costs, this practice is conducted through simulators (see Cooley et al., 2015). The main indicator of training success is the positive *transfer* of the acquired skills from a simulation to an operational system.

Past investigations on the topic have provided contradictory findings regarding the use of fidelity. One group of researchers were inspired by early psychological investigations (e.g., Identical Elements Theory, Thorndike & Woodworth, 1901) and supported using high-fidelity simulators for training (e.g., Logan, 1988; Miller, 1954). Another group of researchers used different psychological theories (e.g., Problem Isomorphism of Simon & Hayes, 1976) to support low-fidelity simulators (e.g., Dale, 1958; Llaneras et al., 1989; Reiter, 1987; Swezey et al., 1991). In this respect, the answer to the fidelity question is not clear from prior work, and a comprehensive attempt is needed to provide empirical evidence to settle the fidelity question in this domain.

(A)


* 22 , 30 = 1
* 23 , 30 = 1
* 30 , 38 = 1
* 31 , 38 = 0
* 24 , 31 = 1
* 25 , 31 = 1
*
FAILURE? 31
RIGHT!

(B)


Cooling Water
Condensate
Condenser
Reflux Drum
Reflux
COLUMN
Boil-up
Reboiler
Preheater 1
Distillate product
Preheater 2
Bottom product

(C)

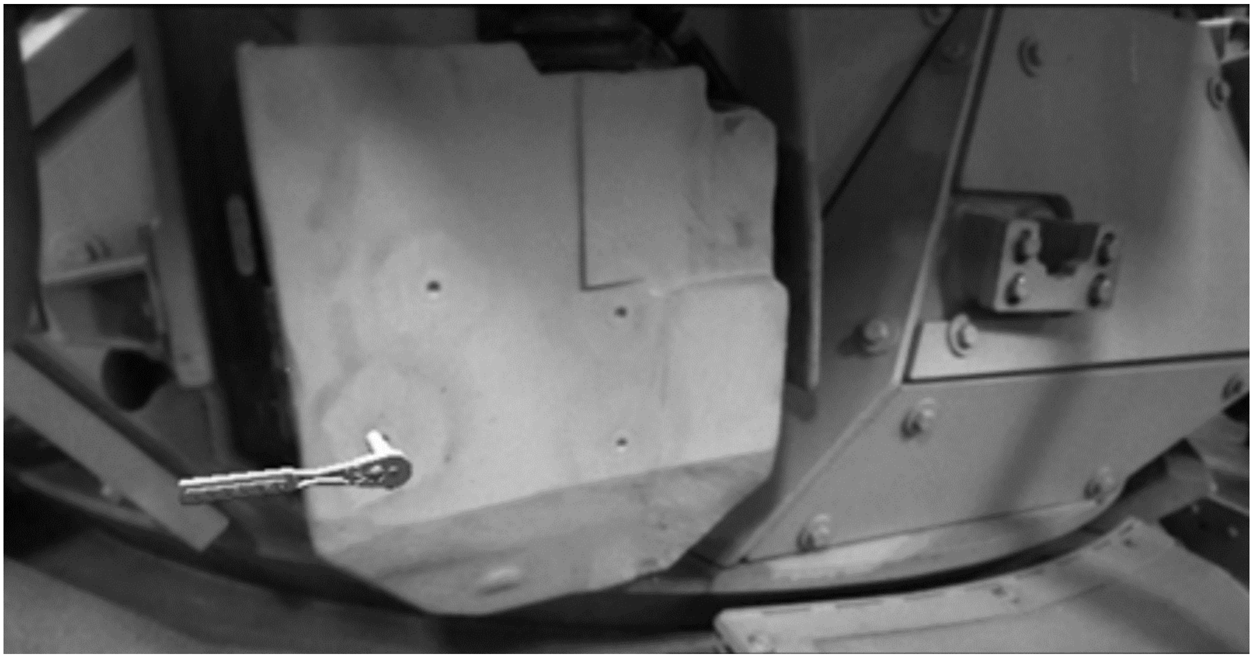


Figure 1

Examples of Low-, Medium-, and High-Fidelity Troubleshooting Interfaces

Note. Figures (A), (B), and (C) are adapted from Rouse (1981), Linou & Kontogiannis (2005), and Yarnall et al. (2015) and they represent examples of low-, medium-, and high-fidelity interfaces, respectively. (A) represents an abstract system with components as decagons and connections as arrows; from observing the output of the system on the right, trainees should diagnose the faulty component by checking connections at each step as shown on the left side of (A). (B) shows the schematic of a distillation plant. (C) shows the visuals of an augmented reality system that shows the armored vehicle in the background and an animated wrench in the fireground (horizontally located on the center left side of the image).

Because of methodological difficulties, there is currently a dearth of literature reviews and meta-analyses, particularly in the troubleshooting domain. One central difficulty is defining and measuring fidelity and transfer, and despite decades of efforts there are no unified metrics for measuring these constructs (see Adams, 1979; Hays & Singer, 1989; Roza, 2005). This situation has caused confusion in the design, evaluation, and deployment of training simulators (Beaubien & Baker, 2004; Doozandeh, 2021; Hamstra et al., 2014; Morris & Rouse, 1985; Roberts et al., 2020). Here, we introduce a categorization method that allowed us to research the question, and in response to recent calls and attempts for reviews (e.g., Lefor et al., 2020; Lintern & Boot, 2019), we take a step toward investigating the effect of fidelity on transfer for troubleshooting tasks.

In analyzing the effect of fidelity on transfer, we also investigated the moderating effect of two factors: trainees' prior skill and the type of system for troubleshooting. With respect to trainees' prior skill, researchers have argued that the level of trainees' skill can determine the transfer in a certain training tool and program (e.g., Hunt & Rouse, 1981), which is due to reasons such as expert-novice differences in using attentional resources, situation awareness, and system knowledge (see Johnson, 1988; Morris & Rouse, 1985). For example, Alessi (1988) argued that the benefit of high-fidelity simulation is not equal for "expert", "experienced", and "novice" trainees (see Figure 2). In this study, we investigated whether trainees' prior skill can moderate the effect of fidelity on transfer. We categorize trainees' prior skill to three levels of low, medium, and high, which will be discussed further below. In addition to trainees' prior skill, the type of troubleshooting system (i.e., electronic or mechanical) will be considered as the second moderator for the effect of fidelity on transfer. Researchers have shown the potential of high-fidelity materials (e.g., animation or video) in improving the training of troubleshooting for mechanical systems (e.g., Carpenter & Just, 1992; Park & Hopkins, 1992). One reason can be that the internal working of mechanical systems is often more dependent on moving parts and fluids compared to electronic systems, and such movements can be presented through animation and high-fidelity simulation. We chose these two moderating factors among all others because, based on our knowledge of the literature, they are among the most important factors that can moderate the effect of fidelity, particularly for the task of troubleshooting.

Decades of research on the topic has produced numerous empirical investigations, and we used accumulated results to conduct a meta-analysis. By using meta-analysis, we can conduct a rigorous test with a large sample size using a variety of simulators. Results are therefore not limited to one simulator or study but can be used in a general sense for the design and evaluation of simulators for troubleshooting and similar tasks.

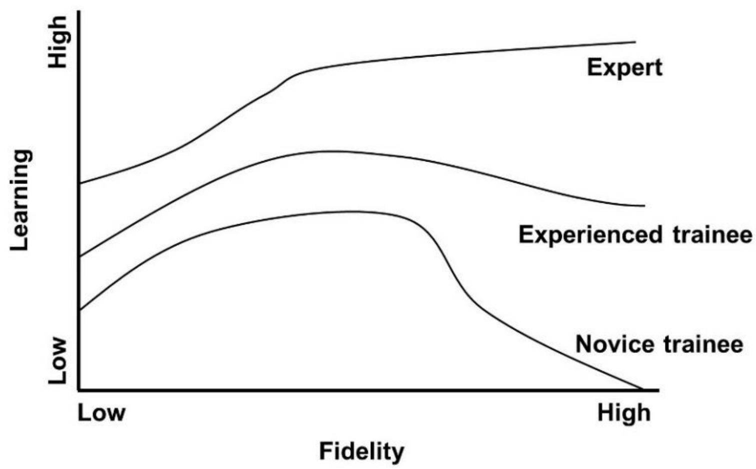


Figure 2

Relationship between level of fidelity and learning for trainees with different levels of skill as shown in Alessi (1988)

Note. The figure is adapted from Alessi (1988)

This study was intended to contribute to existing knowledge by informing researchers and practitioners about whether and how to use fidelity in the design practice, informing organizations in their evaluations and adoption of simulators, and proposing a method to researchers to investigate the fidelity question in other tasks and occupations. So, our hope is to have an impact on increasing the transfer of training for simulators that results in a skillful workforce—particularly troubleshooting professionals—with minimal costs for organizations in developing, deploying, and maintaining the simulators.

2. Methods

We used the American Psychological Association Meta-Analysis Reporting Standards (MARS; Cooper, 2010) to structure our meta-analysis, particularly in the Method, Results, and Discussion sections.

2.1. Hypotheses

We framed the study around three main hypotheses:

- *Hypothesis 1:* Fidelity can affect transfer in troubleshooting tasks.

- *Hypothesis 2:* Different levels of simulation fidelity affect the transfer of training in different ways.
- *Hypothesis 3:* The two moderating factors of trainees' prior skill and system type moderate the effect of fidelity on transfer.

The above hypotheses are addressed in the Discussions section and after presenting the results.

2.2. Search

Although the topic has been mainly studied by researchers in ergonomics and human factors, disciplines that investigated the problem include human-computer interaction, cognitive psychology, educational sciences, electrical engineering, and medical education. Therefore, we also searched for relevant articles published in outlets related to these disciplines. We used online databases that include Scopus, PsycINFO, CiteSeer^x, Google Scholar, Microsoft Academia, and Web of Science. The following combination of keywords were mainly used in our search:

(trainer OR training) AND (troubleshooting OR "problem solving" OR problem-solving OR "fault finding" OR fault-finding) AND (fidelity OR realism)

This combination of terms was used due to variation in the use of terms across countries (such as the US and UK) and between older and more recent literature (e.g., "trainer" was used more often in older literature). In addition, we used related words such as "system failure", "maintenance", and "diagnosis", among others.

Keywords that contained "simulation" were not used, because the combination that was used already included simulation studies, and because many studies did not use the term "simulation" to refer to the training materials. The selected date range in using online databases was 1960 – present (2021); this was used because it was difficult to categorize the fidelity of simulators prior to the 1960s, due to the fundamental differences between interfaces from that time compared to modern interfaces (e.g., introduction of televisions and radical transformations in electronics during 1960s). Additionally, there were not many articles relevant to our topic from that time, because troubleshooting professionals were often trained on the job without simulators. Our search was restricted to journal articles and conference papers, and keywords searches were applied to the title, abstract, and text of reports.

In addition to using online databases, outlets with a more particular focus on the topic were investigated more thoroughly, and we used specific search tools for those outlets to find relevant reports. These outlets included major journals in ergonomics and human factors, and conference proceedings from the Human Factors and Ergonomics Society Annual Meeting (HFES) and Interservice/Industry Training Simulation and Education Conference (I/ITSEC). We also conducted a legacy search by examining the "References" section of articles that were originally found.

It should be mentioned that the results of some relevant studies—particularly those conducted before the 1970s—did not appear in journals and conferences. Reports from those studies were mostly circulated between organizations or were published as technical reports. As a result, contemporary online search engines, as we used, might not have found those reports. We searched and included such studies to the best of our capacity, by using the names of prominent researchers and by legacy search (e.g., resulting in studies such as Miller, 1975; Shriver et al., 1964). Nonetheless, it is likely that our findings do not contain all relevant research in the area, especially from older literature.

2.3. Selection

Articles that were selected for the meta-analysis were those that investigated the relationship between fidelity and transfer for troubleshooting tasks, either directly or indirectly. Those that had a direct focus on the topic were reports whose primary concern was finding the appropriate level of fidelity for training (e.g., McDonald et al., 1982). Those with an indirect focus used simulations to train troubleshooting as part of an investigation that was not directed toward the fidelity question (e.g., Munley & Patrick, 1997), but could nonetheless provide sufficient information for our analysis.

To mitigate the problem of subjective selection of studies, each of the current authors independently searched the initial pool of studies to select those appropriate for the meta-analysis. The final studies that we included in the meta-analysis were those that were selected by both authors;

otherwise, those studies were further discussed and a decision was agreed upon by both authors regarding inclusion.

The selected studies should have clearly specified the characteristics of their simulators (i.e., via figures, videos, drawings, or clear verbal descriptions). It was only by having this information that we could categorize their fidelity levels as the independent variable in our study. As the main dependent variable, an identifiable rate of training transfer from simulation to a target task or system was needed for each simulator. Such reports on transfer rate for each simulator could be relative either to no-simulation conditions, or to simulators with different fidelity levels. We selected studies with quantifiable transfer rates (e.g., effect sizes, *p*-values); however, due to the scarcity of studies with quantifiable transfer, we also included studies with clear but qualitative interpretations of transfer (i.e., authors' interpretation of a simulator's transfer success). Information about trainees' prior skill and the type of troubleshooting system (i.e., electronic or mechanical) were the other two requirements from selected studies. In summary, studies must have had:

- at least one experiment (i.e., empirical observation) in which a training simulator was used to train troubleshooting skills,
- reports on the level of fidelity, or the surface characteristics of the simulator,
- reports on transfer rate metric for each simulator,
- random assignment of trainees and independent samples (between-subject design) for conditions with different fidelity levels and no-simulation condition,
- reports on trainees' prior skill on a specific troubleshooting task or related troubleshooting tasks,
- an identifiable system to be categorized as either an electronic or a mechanical system for troubleshooting, and
- a training phase followed by a transfer phase for each simulator.

2.4. Coding

The definition of main variables used in the meta-analysis is summarized in Table 1. We used the aggregated knowledge from past literature in creating Table 1, which aimed to represent the common knowledge on what each category of low, medium, and high means for important variables of this study. Both current authors independently validated the information in Table 1. The following few paragraphs aim to describe the rationale behind, and use of, Table 1.

We measured fidelity, transfer, and trainees' prior skill on a three-category scale: low, medium, and high. This approach was an intentional choice to address a well-known problem: difficulty in measuring constructs such as fidelity and transfer on quantitative scales (e.g., Adams, 1979; Hays & Singer, 1989; Kirkpatrick & Kirkpatrick, 2006; Roza, 2005). The problem with measuring fidelity is the difficulty in defining a dimension or direction of measurement. For example, a simulator might use simple drawings on a computer display, another might use a computer animation, and another might use a VR tool with motion simulation; how should we define a quantitative scale and compare the level of fidelity between the three simulators? Should the scale only include visual realism, or should it also include auditory, motion, and temporal aspects?

A similar difficulty exists for measuring transfer. Some studies report quantitative scales for transfer rate, based on performance criteria such as time on task, accuracy, or success in troubleshooting (e.g., Darabi et al., 2007). Other studies provide verbal interpretations or non-parametric estimates of transfer (e.g., Maddox et al., 1985). This diversity in scales and formats used in measuring and reporting transfer makes it difficult to conduct the meta-analysis, as we need a single measure to include both numerical values and verbal interpretations of transfer rates.

By acknowledging the difficulty in measuring fidelity and transfer, our solution was to define broad categories (low, medium, and high) instead of using quantitative scales. This approach was used because fidelity and transfer are inherently qualitative constructs, so our categorization provides an easy and valid characterization of studies. This solution also enabled us to implement the meta-analysis on a larger number of studies and helped alleviate the problem of validity in measurement. The same strategy was used in measuring trainees' prior skill.

Table 1

Coding Table: Operational Definitions of Main Variables

	Low	Medium	High
Fidelity level (independent variable)	Low levels of similarity with target systems, including any abstract, generalizable, and logical troubleshooting interface with simple graphical materials (drawings, logical gates, etc.; see Rouse, 1981)	Any interface that stands between low- and high-fidelity (2-D graphical simulations of components, graph diagrams, etc.; see Hegarty et al., 2003)	High levels of similarity with target systems (3-D realistic graphics, bread boards, video, virtual- and augmented-reality tools, etc.; see Hochholdinger & Schaper, 2013)
Transfer (dependent variable)	Negative to minor positive transfer to target system. Little to no performance improvement (lack of significance, with effect size from negative to less than 1; see Johnson & Rouse, 1982)	Relative success in positive transfer to the target system. Any rate between the low and high transfer (presence or absence of significance, effect size between 1 and 2; see Duncan & Shepherd, 1975)	Highly successful in positive transfer to target systems. The desired performance improvement should be achieved (significant, with effect size more than 2; see Hammell & Kingsley, 1998)
Trainees' prior skill (first moderator)	No knowledge or background in the troubleshooting task and the system in general (see Park & Gittelman, 1992)	Some knowledge with the task or the system, no practical experience (current engineering students, junior apprentices, etc.; see McDonald et al., 1982)	Practiced courses or extended training related to the system, or from technical backgrounds (see Maddox et al., 1985)
	Electronic	Mechanical	
System type (second moderator)	Any system mainly composed of electronic components and connections (radar, computer, etc.; see Hammell & Kingsley, 1998)	Any system mainly composed of mechanical components and connections (power plants, internal combustion engines, etc.; see Rouse, 1981)	

Note. "D" in "2-D" and "3-D" refers to dimensional. "Target" tasks are more complex than training tasks in all levels of fidelity. Trainees with High prior skill levels are not "experts"; they are still trainees and their expertise in troubleshooting is inferior to experts. Works that are cited here are typical examples of the categories in this table that also helped us in creating these definitions. Systems that are categorized as "Electronic" can have some mechanical components as well, but are mainly composed of electronic components; the same applies to "Mechanical" systems in which there might be some electronic components but the system as a whole and its important parts are mechanical.

The reason for using a three-scale category is that such a scale is commonly used in existing literature. For example, Figure 2 shows how Alessi (1988) categorized trainees' prior skill on the three categories of expert, experienced, and novice trainees. Similarly, training simulators are often categorized as low-, mid-, and high-fidelity, and this approach also applies to transfer (see Allen et al., 1986; Hays & Singer, 1989; Johnson & Rouse, 1982; Lintern et al., 1990; Roza, 2005).

To reduce the variability in subjective judgments when categorizing levels of fidelity, transfer, trainees' prior skill, and system type, both authors of this work independently used Table 1 to categorize each experiment. In cases of inter-coder disagreement, those experiments were reviewed and discussed in further detail to settle the differences.

2.5. Design and Data Analysis

2.5.1. Primary Analysis: Qualitative Non-Parametric Design

The overall results after using the categories in Table 1 yielded nine conditions: 3 levels of fidelity × 3 levels of transfer. In each condition, there was the number count of included experiments. To see if there was a relationship between fidelity and transfer, we used a non-parametric statistical test as described below. We further analyzed the primary results to assess moderating effects. For this, the same non-parametric test was used for each of the three levels of trainees' prior skill, and for each of the two levels of system type. Non-parametric analyses were completed using IBM SPSS Statistics (Version 24).

2.5.2. Secondary Analysis: Quantitative Effect Size Estimation

We used a quantitative meta-analysis on a fraction of selected studies that reported parametric values (i.e., means and standard deviations) for transfer. So, the categories of transfer as defined in Table 1 could not be used in this quantitative analysis. This analysis allowed us to estimate the magnitude of effect sizes between low-, medium-, and high-fidelity conditions. The number of studies that could be included in this analysis was smaller than the earlier qualitative test because only a few controlled studies in the literature reported the statistics needed for effect size calculations. Therefore, we only applied this quantitative analysis for overall data and not for the analysis of moderators. Quantitative transfer performance in most studies in the literature was measured by one of the following criteria:

- timing in finding the first or all faulty components,
- accuracy measures (e.g., total number of correct diagnoses),
- number of errors in identifying faulty components,
- total number of component tests before finding faulty components,
- post-training knowledge test of the internal working of a system,
- following proper sequence of actions in troubleshooting,
- evaluators' ratings, or
- a combination of multiple performance criteria.

Different studies use different performance criteria, and this creates a challenge in standardizing effect sizes among studies. Moreover, studies vary in what contextual factors (e.g., instructional strategies) they controlled when measuring the effect of fidelity. To overcome these difficulties, we used the random-effects meta-analysis technique, which controls for non-standardized effect sizes by adjusting the weight of the effect size from each study in calculating the summary effect (Borenstein et al., 2009). To implement the effect size analysis, we used CMA software (Borenstein et al., 2014). Figure 3 shows an overall procedure of this study. The significance threshold of $p = .05$ is used in this study.

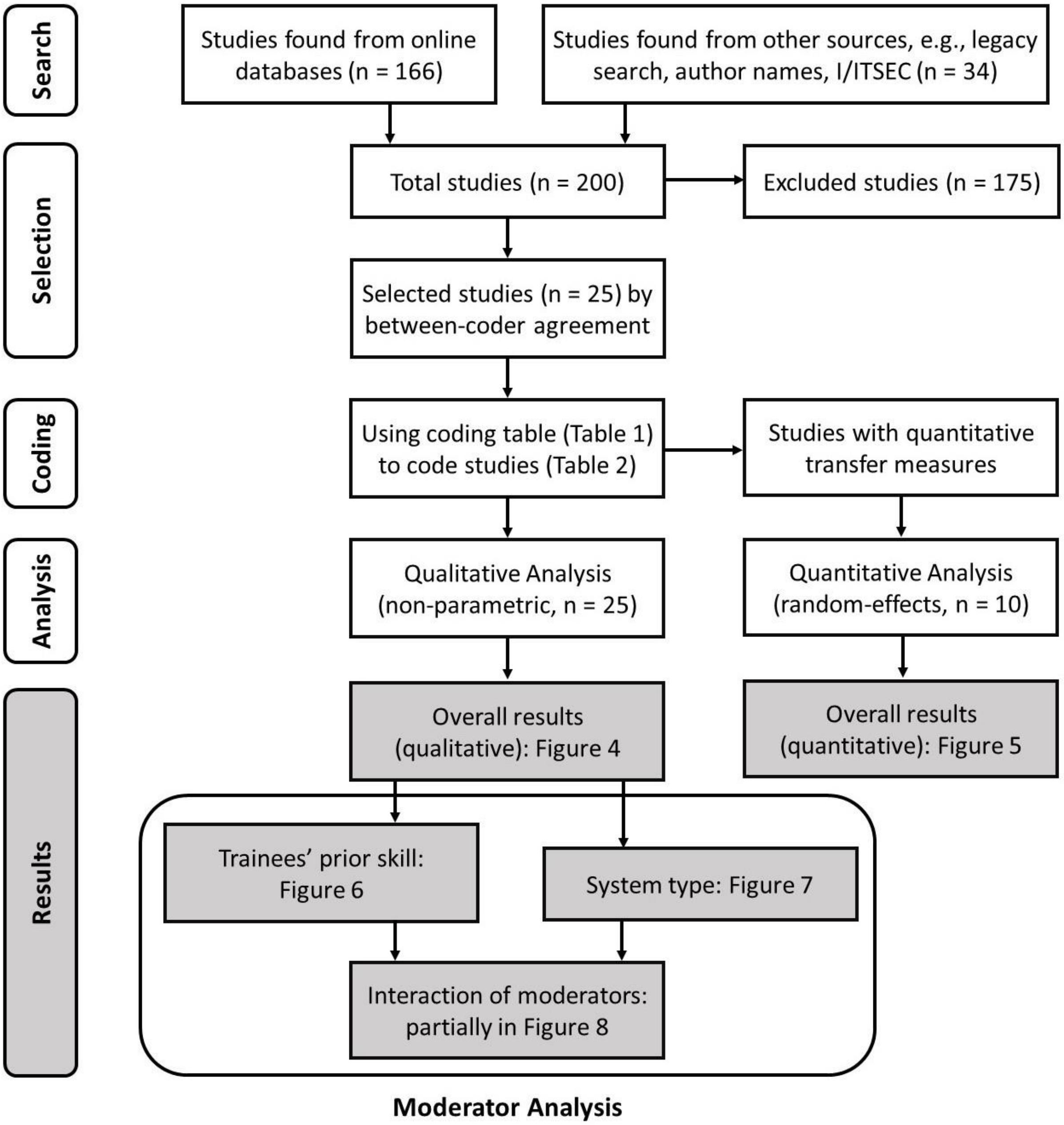


Figure 3

Flow Diagram Representing the Overall Procedure of the Meta-Analysis

Note. Results sections are highlighted in gray. The overall experimental design and this diagram is in accordance with the American Psychological Association Meta-Analysis Reporting Standards (MARS; Cooper, 2010).

3. Results

Using the search criteria resulted in 166 documents. Including articles found through the legacy search, as well as relevant reports in conferences, resulted in 200 studies in total. We reviewed these findings, and only studies relevant to the domain of electronic or mechanical troubleshooting were considered. We also excluded studies with uncontrolled experiments. Of the total number, 25 studies satisfied our conditions, which contained 57 experiments with 1,481 human participants. All 25 studies were included in the qualitative analysis, and 10 of those were included in the quantitative analysis (see the References section for listings of studies included in each of the two analyses). Table 2 summarizes all studies included in the meta-analysis; each row of the table represents one experiment.

Both current authors participated in categorizing all 25 studies and 57 experiments. IBM SPSS Statistics (Version 24) was used to calculate the intercoder agreement rate in creating Table 2; the agreement rate was considered high (Krippendorff's $\alpha > .80$), and the disagreements were entirely resolved with a repeated mutual review of those cases. Studies in Table 2 were conducted in organizations in the United States, United Kingdom, Germany, Canada, and Greece. In reporting the rate of transfer, both objective (e.g., *p* value, time) and subjective (e.g., evaluators' ratings, authors' interpretation) measures were used.

3.1. Overall Findings

3.1.1. Qualitative Analysis

For the non-parametric statistical analysis, we used the Fisher's Exact Test (also known as Fisher-Freeman-Halton Exact Test) as a more conservative alternative to the well-known χ^2 test. This decision was made because our overall sample size was relatively small, and—as will be shown in the interaction analysis—because the sample size in some conditions was either small or had zero counts (see Freeman & Halton, 1951).

Figure 4 shows an overall summary of findings in the qualitative analysis. Considering the entire set of data, there was a statistically significant relationship between fidelity and transfer (FET(N = 57) = 9.04, $p = .039$). As seen in Figure 4, the general trend favors a positive relationship between fidelity and transfer; the largest condition among the nine conditions belongs to the high-fidelity and high transfer condition (N = 17). This is followed by medium-fidelity and high-transfer (N = 13), and medium-fidelity and medium transfer (N = 12).

Table 2

Summary of Studies and Coded Results

Reference	Year	System type	Domain of practice	N	Trainees' prior skill	Fidelity	Transfer
Adams & Thomas	1987	Electronic	Aviation (F-16 flight control subsystem)	6	L	L	M
Allen et al.	1986	Electronic	Artificial (made-up circuits of relays and outputs)	30	L	L	L
				30	L	M	M
				30	L	H	H
Barnett et al.	2000	Mechanical	Aviation (aircraft fuel valve)	10	L	M	L
				10	L	H	L
				10	L	H	H
				10	L	H	H
Darabi et al.	2007	Mechanical	Chemical plant	45	M	L	L
				22	M	M	M
Duncan & Shepherd	1975	Mechanical	Chemical plant	17	L	M	H
Hammell & Kingsley	1998	Electronic	Military (AN/WSC UHF transceiver)	33	L	M	H
				13	L	H	H
Hegarty et al. (Experiments Two and Three)	2003	Mechanical	Home maintenance (toilet flushing cistern)	25	L	M	M
				25	L	H	M
				25	L	M	M
				25	L	H	M
				20	L	M	M
				20	L	H	M
Hochholdinger & Schaper (Far Transfer)	2013	Electronic	Automobile industry (production unit)	21	M	M	M
				21	M	H	M
Johnson & Fath	1983	Electronic	Military (telephone switchboard)	30	L	M	H
				35	L	H	H
Johnson & Rouse	1982	Mechanical	Aviation (aircraft power-plant troubleshooting)	12	H	L	M
				12	H	M	M
				12	H	H	H
				11	H	M	M
				11	H	H	H
Lajoie & Lesgold	1992	Electronic	Aviation	31	M	M	H
Linou & Kontogiannis	2005	Mechanical	Chemical plant (distillation column)	24	M	L	M
				12	M	M	H
Maddox et al.	1985	Mechanical	Nuclear power plant (emergency diesel generators)	4	M	M	H
				4	M	H	H
				16	H	M	M
				15	H	H	H

Reference	Year	System type	Domain of practice	N	Trainees' prior skill	Fidelity	Transfer
McDonald et al.	1982	Electronic	Military (FM radio IF amplifier)	93	M	M	H
				93	M	H	H
Miller	1975	Electronic	Military (Improved Hawk Missile)	39	L	L	M
				27	L	M	H
Munley & Patrick	1997	Mechanical	Chemical plant	8	L	L	H
Park & Gittelman	1992	Electronic	Artificial (circuits of logical gates)	45	L	M	M
				45	L	H	H
Patrick & Haines	1988	Mechanical	Chemical plant	24	L	M	H
Patrick et al.	1989	Mechanical	Chemical plant	30	L	M	L
Patrick et al.	1996	Mechanical	Hot strip steel mill	4	M	M	H
				4	H	M	M
Rouse (Experiment Eight)	1981	Mechanical	Internal combustion engine	12	M	L	H
				12	M	M	M
				12	M	H	H
Shriver et al.	1964	Electronic	Military (M33 Gunfire Control Radar)	20	M	M	H
				17	M	H	H
Swezey et al.	1991	Mechanical	Military (diesel engine)	60	L	M	H
				60	L	H	H
Thomas	1993	Electronic	Aviation (F-15 and F-16 subsystems)	183	L	M	H
Yarnall et al.	2015	Mechanical	Military (Armored vehicle maintenance and troubleshooting)	4	L	H	H
				8	L	H	H
				4	L	H	H

Note. Studies are listed in the alphabetical order of authors. There are 25 studies and 57 rows in the table; each row indicates a single experiment. L, M, and H refer to low, medium, and high, respectively, and N indicates sample size (i.e., number of participants) in each experiment.

3.1.2. Quantitative Analysis

We reviewed studies listed in Table 2 and found that 10 studies among them had controlled experiments that reported quantitative measures for transfer rates. Of those 10, four used timing and the rest used a combination of accuracy, error, quality of sequence, knowledge test, and evaluators' rating for measuring transfer performance. We standardized these criteria into a single performance measure that was used in calculating effect sizes (see the online supplementary data file). Figure 5 summarizes the results of this analysis.

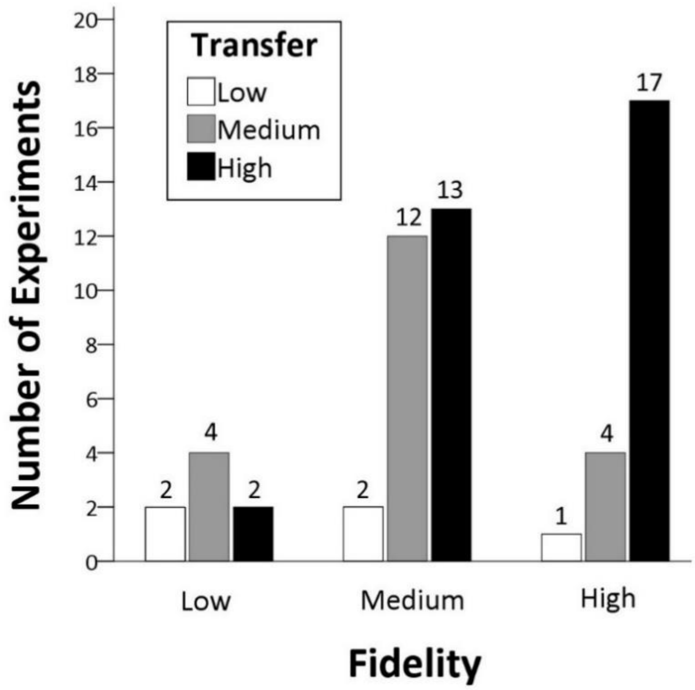
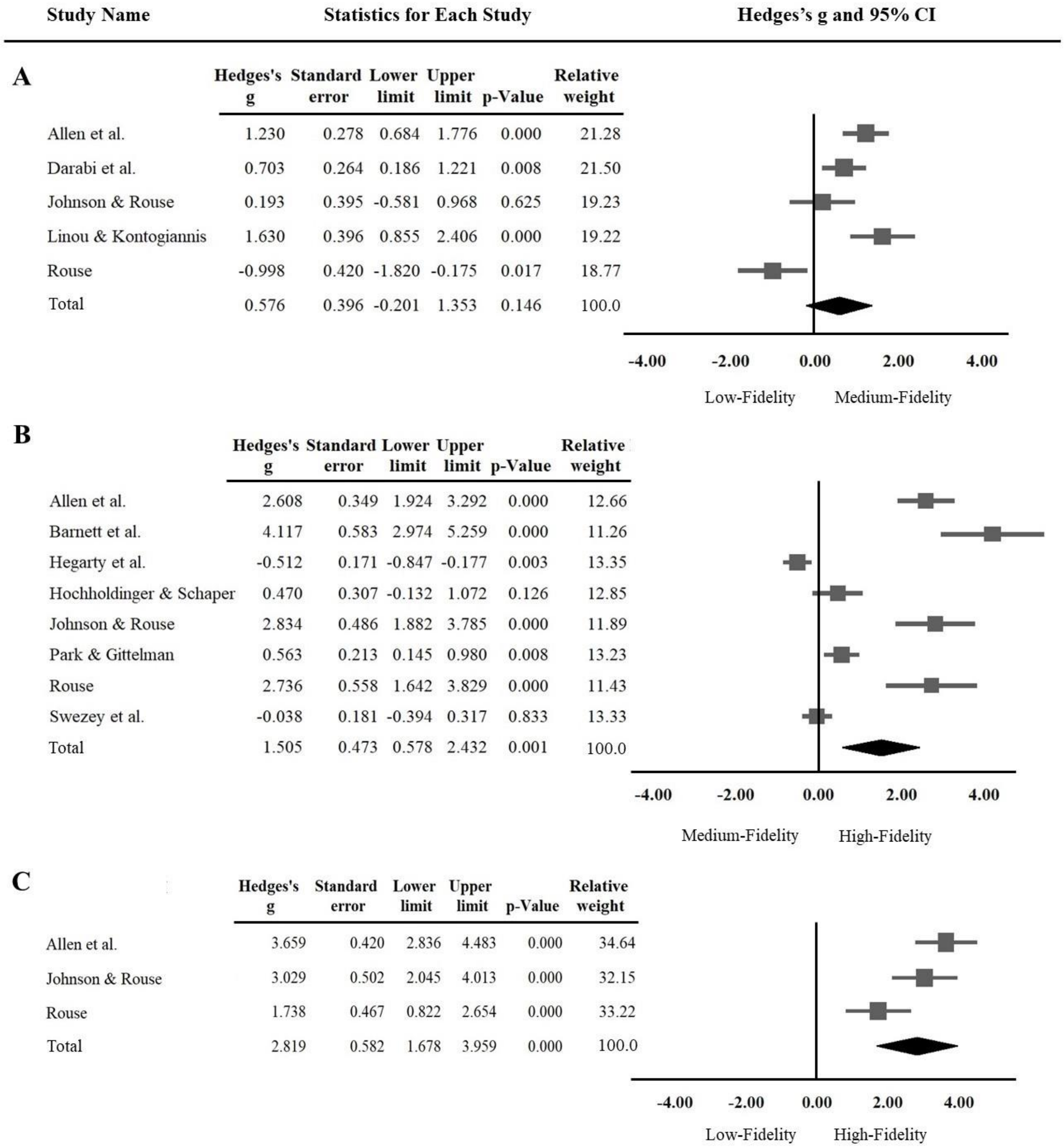


Figure 4

Summary of Overall Results in Qualitative Analysis

A weak improvement in transfer can be seen from low- to medium-fidelity conditions (g = 0.58, 95% CI = -0.20 - 1.35; Figure 5A), and from medium- to high-fidelity conditions (g = 1.51, 95% CI = 0.58 - 2.43; Figure 5B). A strong effect was observed in the comparison between low- and high-fidelity conditions (g = 2.82; 95% CI = 1.68 - 3.96; Figure 5C). To test the likelihood of publication bias and missing studies in this analysis, we ran Egger's test. Results of this test suggested no evidence for statistical asymmetry for low- versus medium-fidelity comparisons (intercept = -5.36, 95% CI = -24.68 - 14.00, p = .220) or for low- versus high-fidelity comparisons (intercept = -11.66, 95% CI = -281.10 - 257.79, p = .340). However, this asymmetry was significant when comparing between medium- and high-fidelity conditions (intercept = 9.99, 95% CI = 5.91 - 14.07, $p < .001$). To address this issue, we used the Trim and Fill method (Duval & Tweedie, 2000), and the result slightly decreased the magnitude of the summary effect between medium- and high-fidelity conditions (g = 0.39, 95% CI = -0.57 - 1.34).

For the low- versus medium-fidelity comparison, heterogeneity was evident ($Tau^2 = 0.66$; Q = 27, df = 4; $p < .001$; $I^2 = 85$), and the test for overall effect was not significant (Z = 1.45; p = .146). For the medium- versus high-fidelity, heterogeneity was evident (Tau^2 = 1.65; Q = 158, df = 7; *p* = .000; I^2 = 96, and the test for overall effect was significant (Z = 3.18; *p* = .001). The low- versus high-fidelity comparison resulted in significant heterogeneity (Tau^2 = 0.80; Q = 9, df = 2; *p* = .009; I^2 = 79) and a significant overall effect (Z = 4.85; *p* < .001).



A

Study Name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value	Relative weight
Allen et al.	1.230	0.278	0.684	1.776	0.000	21.28
Darabi et al.	0.703	0.264	0.186	1.221	0.008	21.50
Johnson & Rouse	0.193	0.395	-0.581	0.968	0.625	19.23
Linou & Kontogiannis	1.630	0.396	0.855	2.406	0.000	19.22
Rouse	-0.998	0.420	-1.820	-0.175	0.017	18.77
Total	0.576	0.396	-0.201	1.353	0.146	100.0

B

Study Name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value	Relative weight
Allen et al.	2.608	0.349	1.924	3.292	0.000	12.66
Barnett et al.	4.117	0.583	2.974	5.259	0.000	11.26
Hegarty et al.	-0.512	0.171	-0.847	-0.177	0.003	13.35
Hochholdinger & Schaper	0.470	0.307	-0.132	1.072	0.126	12.85
Johnson & Rouse	2.834	0.486	1.882	3.785	0.000	11.89
Park & Gittelman	0.563	0.213	0.145	0.980	0.008	13.23
Rouse	2.736	0.558	1.642	3.829	0.000	11.43
Swezey et al.	-0.038	0.181	-0.394	0.317	0.833	13.33
Total	1.505	0.473	0.578	2.432	0.001	100.0

C

Study Name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value	Relative weight
Allen et al.	3.659	0.420	2.836	4.483	0.000	34.64
Johnson & Rouse	3.029	0.502	2.045	4.013	0.000	32.15
Rouse	1.738	0.467	0.822	2.654	0.000	33.22
Total	2.819	0.582	1.678	3.959	0.000	100.0

Figure 5

Forest Plot of Overall Results in Quantitative Analysis

Note. The analysis is the comparison between the simulators with (A) low and medium, (B) medium and high, and (C) low and high fidelity.

Overall, this analysis of effect sizes indicates weak improvements of transfer from low- to medium-fidelity and from medium- to high-fidelity simulators, and this outcome mirros findings in the non-parametric test (Figure 4). The improvement in transfer was stronger between low- and high-fidelity simulators. It should be mentioned that the result of comparing between low- and high-fidelity effect sizes should be taken with precaution, as the small number of studies could decrease the accuracy of results (Borenstein et al., 2009, p. 84). Nevertheless, because random-effects models are conservative, and because of the pattern of overall results, it is plausible to see a stronger summary effect between low- and high-fidelity conditions.

Although these general findings indicate a positive effect of fidelity on transfer, they have little explanatory power for practical purposes. The following paragraphs describe how considering moderators can provide more definitive results of practical value.

3.2. Trainees' Prior Skill as Moderator

We tested if trainee's prior skill moderates the effect of fidelity on transfer. For each of the three levels of trainees' prior skill, (low, medium, and high) we used a separate fidelity-transfer test. Figure 6 shows the results of this analysis. Although a positive effect of fidelity on transfer is visible for trainees with low prior skill, this effect was not statistically significant ($FET(N = 32) = 3.72$, $p = .445$). The same lack of significance was found for trainees with medium prior skill ($FET(N = 17) = 4.26$, $p = .524$), and the trend in Figure 6 supports this finding. However, when we considered trainees with high prior skill, the fidelity-transfer relationship became significant ($FET(N = 8) = 6.87$, $p = .018$). For high prior skill trainees, there was only one experiment with low-fidelity and medium transfer; the remainder of the medium transfer results (N = 4) were the result of medium-fidelity simulators, and all high-fidelity simulators resulted in high transfer (N = 3). So, the fidelity-transfer relationship for trainees with high prior skill was almost perfectly linear.

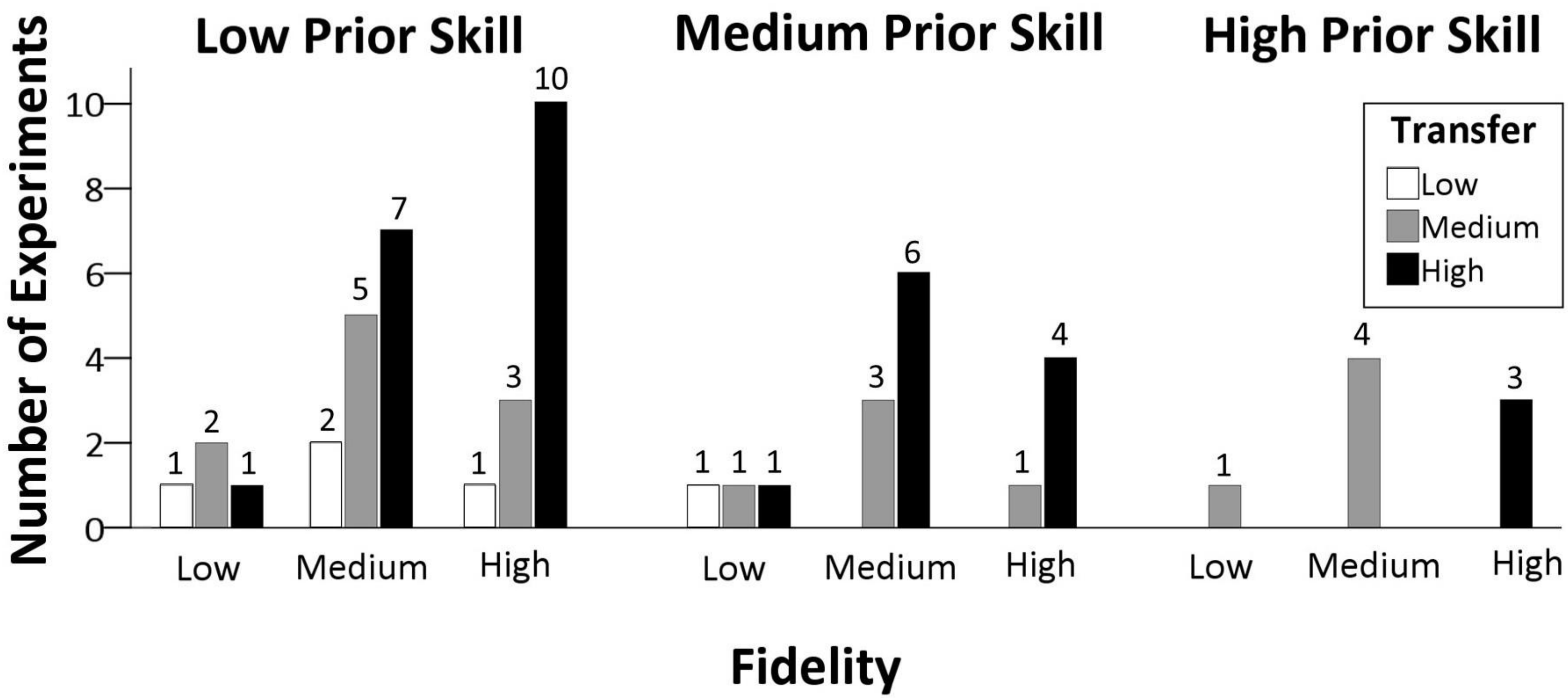


Figure 6

Summary of Results for Trainees with Low, Medium, and High Prior Skill Levels

3.3. System Type as Moderator

We similarly tested the fidelity-transfer relationship for each of the two system types: electronic and mechanical. Figure 7 summarizes the results of this analysis.

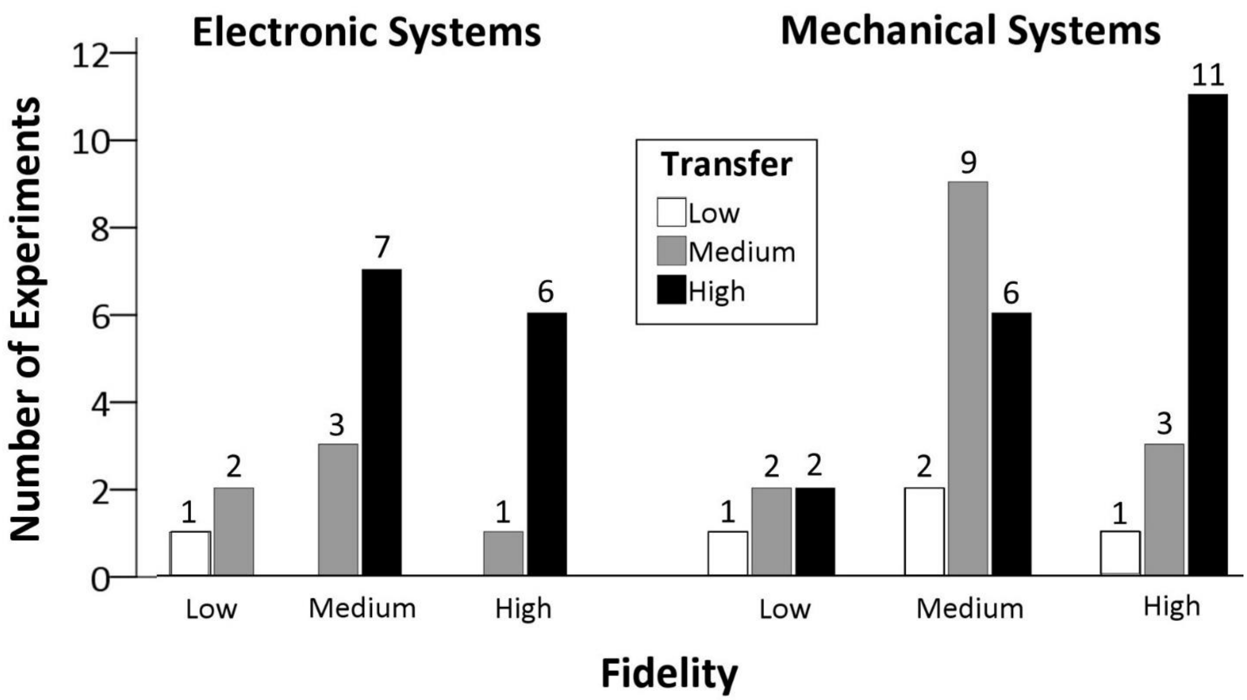


Figure 7

Summary of Results for Electronic and Mechanical Systems

For electronic systems, there was a significant relationship between fidelity and transfer (*FET*(N = 20) = 7.95, p = .047). For low-fidelity simulators, there was one experiment with low transfer and there was no high transfer. Medium-fidelity simulators had the highest number of medium (N = 3) and high transfer (N = 7) reports, and this was followed by high-fidelity simulators (medium transfer: N = 1, high transfer: N = 6). A visible trend for mechanical systems can be seen in Figure 7 that favors a positive relationship between fidelity and transfer. Mechanical systems, however, did not result in a significant relationship (*FET*(N = 37) = 5.66, p = .215).

As a side note, it should be mentioned that we conducted the non-parametric tests because we believed that low-, medium-, and high-fidelity systems belong to three categories of systems that have qualitative differences, and that this is also the case for transfer and trainees' level of skill. However, although the data were categorical, categories have a direction: low is lower than medium, which is lower than high. So, one might suggest the use of correlation (i.e., Pearson r) to confirm the results reported above. To address this concern, we verified all the results above using the Pearson r, and by assigning the values of 1, 2, and 3 instead of the categories of low, medium, and high, respectively. For each of the tests above, the significance of the Pearson r was in complete agreement with those from the FET, which confirms the accuracy of results.

3.4. Interaction of Trainees' Prior Skill and System Type

A question that arises at this point is: do the two moderators interact in explaining the relationship between fidelity and transfer? Table 3 summarizes the results of testing the interaction between the two moderators. The only condition in which the fidelity-transfer relationship was significant was for trainees with high prior skill who were trained on the mechanical type of systems.

Table 3

Summary of Fidelity-Transfer Relationship for Interaction Between Trainees' Prior Skill and System Type

	Electronic systems	Mechanical systems
Low prior skill	***p* = .052** (N = 13)	*p* = .773 (N = 19)
Medium prior skill	*p* = 1.00 (N = 7)	*p* = .750 (N = 10)
High prior skill	No data	***p* = .018** (N = 8)

Note. Significance values refer to the significance of the fidelity-transfer relationship in each condition, and are calculated using Fisher's Exact Test. Boldened numbers are what we considered as practically-significant results.

The *p* value between low prior skill and electronic systems is larger than the .05 threshold (p = .052). Nevertheless, it approached significance and we think this result is practically meaningful for two reasons. First, because Fisher's Exact Test is generally a conservative test with sufficient sample size for this condition (N = 13). Second, as Figure 8 shows, there is a clear pattern in the data for this interaction favoring medium- and high-fidelity simulations.

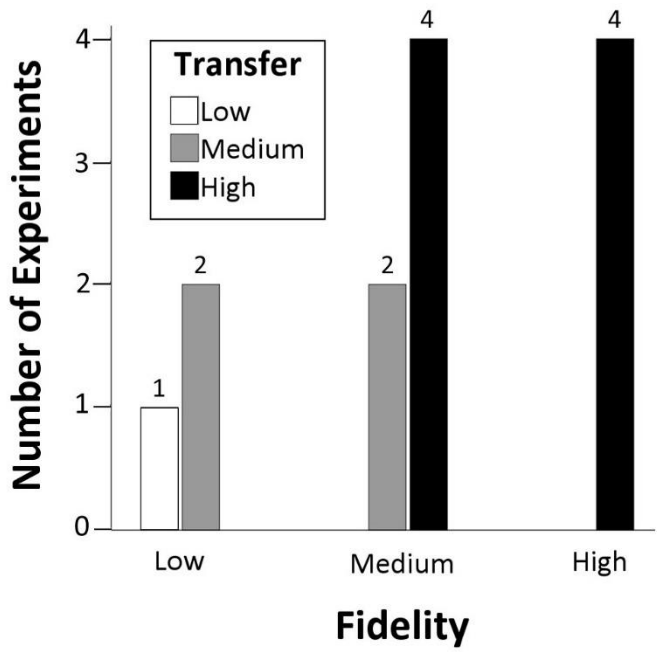


Figure 8

Summary of Results for Intersection of Low Prior Skill and Electronic Systems

4. Discussion

4.1. Practical Implications

Before discussing the practical implications of the study in detail, we address the three general hypotheses that were posed in the Methods section:

- *Result for hypothesis 1:* Fidelity affected transfer in training troubleshooting tasks.
- *Result for hypothesis 2:* Medium- to high-fidelity simulators resulted in higher rates of transfer compared to low-fidelity simulators.
- *Result for hypothesis 3:* Trainees' prior skill level and system type moderated the effect of fidelity on transfer.

Both qualitative and quantitative analyses of the overall data showed a small effect of fidelity on transfer; nonetheless, our analysis of moderators and their interaction showed that *none of the low-, medium-, or high-fidelity simulations provided a universal solution in training troubleshooting*. This conclusion is based on the observation that if any of the three levels of fidelity are considered, there was at least one condition in which that level of fidelity was not better than other levels in terms of transfer. This outcome, however, does not downgrade the effect of fidelity on transfer; *the effect of fidelity should be investigated with an analysis of moderators*.

Of the three levels of trainees' prior skill, *higher levels of fidelity resulted in higher transfer only for trainees with high prior skill* (see Figure 6). A positive and nearly linear fidelity-transfer relationship for trainees with high prior skill confirms established knowledge that experienced trainees often need high-fidelity simulators for improving performance. This effect may be because, unlike novices, increased environmental details of high-fidelity simulators do not distract experienced trainees, and the simplicity of the task in low-fidelity trainers seem unattractive and boring to experienced trainees (see Alessi, 1988).

Regarding system type as the second moderator, *the relationship between fidelity and transfer was not statistically supported for mechanical systems*. On the other hand, *a relationship was observed for electronic systems, in which medium-fidelity simulators resulted in the highest transfer and were followed by high-fidelity simulators*. This outcome might be because, for electronic systems, medium-fidelity has advantages over low-fidelity simulations in aspects such as showing the direction of the electric current in a circuit and the perceptual correspondence of components in a simulator with those in reality. Similarly, high-fidelity simulation might not add useful information but only increase the visual details of electronic components with less relevant functions for troubleshooting performance. For example, in Allen et al. (1986) the output of the high-fidelity simulator (e.g., "fan, water pump, solenoid valve assembly, three lights, TV monitor, and sound generator and speaker") could easily distract trainees in their troubleshooting, while in simulators of lower fidelity the output was often represented as 0/1, or active/inactive. Although this finding contradicts our initial predictions, it is not unprecedented, because the benefit of animation over static training materials for electronic and mechanical systems has been a controversial issue (e.g., Abich et al., 2021; Bailey et al., 2017; Johnson & Rouse, 1982; Spangenberg, 1973; Swezey et al., 1991).

Our interaction analysis showed that breaking the data of mechanical systems into the three levels of trainees' prior skill yielded useful results. Specifically, *for training the troubleshooting of mechanical systems, unlike participants with low and medium prior skill levels, those with high prior skill benefited from increased fidelity* (see Table 3). The "High Prior Skill" section of Figure 6 shows this trend; in our data, all of the eight experiments that enrolled participants with high prior skill used mechanical systems (see Table 2). Finally, for low prior skill and electronic systems, we can argue that for trainees with low prior skill, medium- and high-fidelity simulations resulted in higher transfer than low-fidelity simulations (although this effect was not statistically significant). Considering costs and practical implementations, we conclude that *medium-fidelity simulations can be the best choice for training the troubleshooting of electronic systems to trainees with low prior skill*.

Practical takeaways can also be gained with respect to the domains of practice that had multiple studies in our data (see Table 2). In the (non-aviatory) Military domain (7 studies), except for one experiment that used a low-fidelity simulator (Miller, 1975, first experiment), all other experiments (N = 14) used medium-fidelity (N = 6) and high-fidelity (N = 8) simulators that resulted in high transfer rates. This conforms with the results that were discussed earlier, as most studies in this domain (5 of the total 7 studies) used electronic systems. With considering the costs of high-fidelity simulations, this evidence suggests that *using medium-fidelity simulators could be the best solution for training troubleshooting in non-aviatory military tasks* (e.g., armored vehicle circuits, diesel engines). Also, the vast majority of experiments in the Military domain (N = 14 of the N = 15 total) resulted in high transfer rates. Though our sample is limited, we think there are two main reasons for the success of training in the military domain: being one of the oldest domains that has been using training simulators for decades, and having little to no financial constraints in using sophisticated simulators, experienced designers, and experienced training scientists. Our results are not as meaningful for the domain of Chemical plant (6 studies), but for Aviation (5 studies) all experiments that resulted in high-transfer rates (N = 6) were the result of medium-fidelity (N = 2) and high-fidelity (N = 4) simulators. So, we conclude that *training troubleshooting in aviation benefits from increased fidelity*.

4.2. Limitations

We selected two moderators (i.e., trainees' prior skill and system type) among all other potential moderators that could influence the effect of fidelity on training. We provided the reason for our choice earlier. However, in addition to these two, there are many more factors that could affect transfer and moderate the effect of fidelity. For example, participant age might moderate the effect of fidelity for various reasons. Younger trainees who are more likely to have the experience of being exposed to simulated environments in video games might find it easier to practice with similar simulators compared to older people with less exposure to video games (see Mead & Fisk, 1998). Only a few of the studies we found reported on participant age, and so more studies are needed to determine the moderating effect of age in this respect.

Likewise, fatigue as the result of training could directly affect transfer. Only two studies among this in our data discussed the effect of fatigue (i.e., Patrick et al., 1996; Rouse, 1981). Fatigue acts against transfer, and it often occurs for longer durations of practice and in medium- to high-fidelity simulators (Rouse, 1981). Because our results favor using medium- to high-fidelity simulators, including fatigue might not dramatically change the direction of our results.

Likewise, aspects of training protocols can directly affect transfer, such as instructional features, the presence and type of feedback, practice duration, frequency of training, specific modality of training, whole- or part-task training, and other contextual factors, and this could occur regardless of the level of fidelity (see Salas et al., 1998). We controlled those factors to the best of our capacity; for example, we avoided including studies in which training contexts were radically different from each other, and using the random-effects quantitative analysis helped us in mitigating this concern. However, our focus on only two factors among all others makes our analysis susceptible to confounding effects. Therefore, our results should be used with precautions.

Our definition of high-fidelity simulation (see Table 1) includes VR/AR tools, and most studies that used computer-based simulations could also have presented the same training materials through VR/AR media that can result in similar or enhanced transfer (e.g., Barnett et al., 2000). This approach could be particularly useful for practice domains that can benefit from the mobility VR/AR tools can provide (e.g., chemical plants, nuclear reactors). Nonetheless, in our analysis only Yarnall et al. (2015) used an AR tool, in training the troubleshooting of armored vehicles to participants with low prior skill, which resulted in a successful transfer. Further experiments are needed to study the potentials that are particular to VR/AR tools, especially for troubleshooting (see Borsci et al., 2015).

One of our assumptions in selecting studies was that they should measure transfer on a target operational system. However, this condition was not entirely satisfied in all included studies. For example, Allen et al. (1986) and Darabi et al. (2007) used the same simulator both in one condition of training and in measuring transfer. In fact, those studies assumed high levels of transfer between the simulator and the target system, and this is the basis that transfer is measured on the same simulator that was used in training. Although there is evidence that this assumption might not be valid (see Logan, 1988), we think the results from those studies can still provide useful evidence regarding the effect of fidelity on transfer as used in our meta-analysis.

We used limited combinations of search terms to find relevant articles in our original search. There might exist prior experimental work applicable to our analysis that used different terminology, and as a result we could not find those works. This outcome is, to some extent, due to the nature of research in this area, and the fact that the field lacks a standardized set of key terms and concepts to use consistently in research. Likewise, although we maintained the validity of our results by defining our own measures of fidelity and transfer (see Table 1), many of the works included in our meta-analysis are dated more than two decades ago, when training technology was much simpler than now. This was mainly because experimental works in this domain that satisfied our inclusion criteria were hard to find, and because in contrast to the 1980s and 1990s there is currently a dearth of experimental work and meta-analysis on this topic in recent years and with modern training technology. The limitations above can be addressed by conducting future experimental studies with modern technology, which that would tackle the specific question of fidelity by controlling contextual factors and further refining the moderators.

We should also stress the scope of the task domains in our investigation: troubleshooting electronic and mechanical systems. Although the tasks and skills in troubleshooting and similar procedural tasks might have similarities with other task domains, caution should be exercised when using the results of this study for other task domains. Even the slightest differences in the nature of tasks can make investigations in one domain inapplicable for another domain. Likewise, of the 25 studies included in our meta-analysis, 18 were from three domains of practice: military (non-aviatory, 7 studies), chemical plant (6 studies), and aviation (5 studies). So, the specific results of this meta-analysis are best applicable to troubleshooting in specific domains of practice (see Table 2), particularly in military, chemical plants, and aviation. More research is needed in other task and practice domains to investigate the effect of fidelity on transfer.

5. Conclusions

With the introduction of novel technology tools to training (e.g., VR/AR devices) in various occupations, there is a growing need to investigate how simulation fidelity can affect transfer in training simulators. Our results showed that, for training troubleshooting professionals, fidelity can play an important role in transfer. Medium-fidelity simulators performed better than our expectations, and we found that low-fidelity simulators performed poorly in nearly all conditions. More importantly, designers, researchers, and educators need to appreciate the importance of moderators. Here, the factors that moderated the effect of fidelity were trainees' prior skill and the type of troubleshooting system. These moderators determined how fidelity should be used, and dispelled myths about relying on high- or low-fidelity simulators as universal solutions for training. Researchers can also use the methods described here to conduct meta-analyses that help alleviate the problem of errors in measuring fidelity and transfer. Finally, from an organizational standpoint, the long-standing goal of saving costs in creating and using simulators for troubleshooting tasks can be achieved by choosing the appropriate level of fidelity as shown in our results. In choosing between simulators, organizations can now evaluate simulators with respect to their fidelity level. We hope to see future meta-analyses and experimental investigations that address the fidelity question in other domains of task and practice (e.g., medical training), and that further refine the moderators we used or consider new moderators to investigate the effect of fidelity.

References

References marked with * are studies included in the qualitative meta-analysis, and references marked with ¶ are studies included in the quantitative meta-analysis (random-effects model).

Aati, K., Chang, D., Edara, P., & Sun, C. (2020). Immersive work zone inspection training using virtual reality. *Transportation Research Record, 2674*(12), 224–232.

Abich, J., Parker, J., Murphy, J. S., & Eudy, M. (2021). A review of the evidence for training effectiveness with virtual reality technology. *Virtual Reality, 25*(4), 919–933.

* Adams, L., & Thomas, C. (1987). F-16 flight control system training game. In *Proceedings of the Interservice/Industry Training Systems Conference (I/ITSC)* (pp. 172–174). Orlando, Florida: National Training and Simulation Association.

Adams, J. A. (1979). On the evaluation of training devices. *Human Factors, 21*(6), 711–720.

*¶ 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.

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

Anastassova, M., & Burkhardt, J.-M. (2009). Automotive technicians' training as a community-of-practice: Implications for the design of an augmented reality teaching aid. *Applied Ergonomics, 40*(4), 713–721.

*¶ Barnett, B., Helbing, K., Hancock, G., Heininger, R., & Perrin, B. (2000). An evaluation of the training effectiveness of virtual environments. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Orlando, Florida: National Training and Simulation Association.

Beaubien, J. M., & Baker, D. P. (2004). The use of simulation for training teamwork skills in health care: How low can you go? *Quality and Safety in Health Care, 13*(1), 51–56.

Bailey, S. K, Johnson, C. I., Schroeder, B. L., & Marraffino, M. D. (2017). Using virtual reality for training maintenance procedures. In *Proceedings of the Interservice/Industry Training Systems Conference (I/ITSC)* (No. 17108), Orlando, Fl: National Training and Simulation Association.

Borba, E. Z., Cabral, M., Montes, A., Belloc, O., & Zuffo, M. (2016). Immersive and interactive procedure training simulator for high risk power line maintenance. In *ACM SIGGRAPH '16 VR Village*, 1–1.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). *Comprehensive meta-analysis* (Version 3.3.070) [Computer software]. Biostat. https://www.meta-analysis.com

Borsci, S., Lawson, G., & Broome, S. (2015). Empirical evidence, evaluation criteria and challenges for the effectiveness of virtual and mixed reality tools for training operators of car service maintenance. *Computers in Industry*, *67*, 17–26.

Carpenter, P. A., & Just, M. A. (1992). *Understanding mechanical systems through computer animation and kinematic imagery*. (Report No. ONR92-1) Arlington, VA: Office of Naval Research.

Cooley, T., G. Seavers, S. Gordon, J. Roth, & Rodriguez, J. (2015). Calculating simulation-based training value: Cost avoidance and proficiency. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Orlando, Florida: National Training and Simulation Association.

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed., Applied Social Research Methods Series, Vol. 2). Thousand Oaks, CA: Sage.

Dale, H. C. A. (1958). Fault-finding in electronic equipment. *Ergonomics*, *1*(4), 356–385.

*¶ Darabi, A. A., Nelson, D. W., & Palanki, S. (2007). Acquisition of troubleshooting skills in a computer simulation: Worked example vs. conventional problem solving instructional strategies. *Computers in Human Behavior*, *23*(4), 1809–1819.

Di Loreto, C., Chardonnet, J. R., Ryard, J., & Rousseau, A. (2018). Woah: A virtual reality work-at-height simulator. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 281–288.

Doozandeh, P. (2021). From surface realism to training considerations: A proposal for changing the focus in the design of training systems. *Theoretical Issues in Ergonomics Science, 22*(6), 689–728.

* Duncan, K. D., & Shepherd, A. (1975). A simulator and training technique for diagnosing plant failures from control panels. *Ergonomics, 18*(6), 627–241.

Duval, S. J., & Tweedie, R. L. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463.

Freeman, G. H., & Halton, J. H. (1951). Note on exact treatment of contingency, goodness-of-fit and other problems of significance. *Biometrika, 38*, 141–149.

Hafsia, M., Monacelli, E., & Martin, H. (2018). Virtual reality simulator for construction workers. In *Proceedings of the Virtual Reality International Conference*, 1–7. New York, NY: Association for Computing Machinery.

* Hammell, T. J., & Kingsley, L. C. (1998). Nonresident computer-based training: Effectiveness evaluation. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Orlando, Florida: National Training and Simulation Association.

Hamstra, S. J., R. Brydges, R. Hatala, B. Zendejas, & Cook., D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine, 89*(3), 387–392.

Hays, R. T., & Singer, M. J. (1989). *Simulation fidelity in training system design: Bridging the gap between reality and training*. New York, NY: Springer Science & Business Media.

*¶ Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animation and external animation in understanding mechanical systems. *Cognition and Instruction, 21*(4), 325–360.

*¶ Hochholdinger, S., & Schaper, N. (2013). Training troubleshooting skills with an anchored instruction module in an authentic computer based simulation environment. *Journal of Technical Education (JOTED), 1*(1), 7–22.

Hunt, R. M., & Rouse, W. B. (1981). Problem-solving skills of maintenance trainees in diagnosing faults in simulated powerplants. *Human Factors*, *23*(3), 317–328.

Johnson, S. D. (1988). Cognitive analysis of expert and novice troubleshooting performance. *Performance Improvement Quarterly, 1*(3), 38–54.

* Johnson, W. B., & Fath, J. L. (1983). Design and initial evaluation of a mixed-fidelity courseware for maintenance training. In *Proceedings of the 27th Annual Meeting of the Human Factors Society* (pp. 1017–1021). Norfolk, VA: Human Factors Society.

*¶ Johnson, W. B., & Rouse, W. B. (1982). Training maintenance technicians for troubleshooting: Two experiments with computer simulations. *Human Factors, 24*(3), 271–276.

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors*, *63*(4), 706–726.

Kirkpatrick, D. L., & Kirkpatrick J. D. (2006). *Evaluating training programs*. San Francisco, CA: Berrett-Koehler.

* Lajoie, S. P., & Lesgold, A. M. (1992). Apprenticeship training in the workplace: Computer-coached practice environment as a new form of apprenticeship. In M. J. Farr & J. Psotka (Eds.), *Intelligent instruction in computer: Theory and practice* (pp. 15–36). Washington, DC: Taylor & Francis.

Lefor, A. K., Harada, K., Kawahira, H., & Mitsuishi, M. (2020). The effect of simulator fidelity on procedure skill training: A literature review. *International Journal of Medical Education, 11*, 97–106.

*¶ Linou, N., & Kontogiannis, T. (2005). A training approach to the acquisition and retention of fault-finding skills: Making instructions "visible" on the interface. *International Journal of Human-Computer Interaction*, *18*(1), 59–84.

Lintern, G., & Boot, W. R. (2019). Cognitive training: Transfer beyond the laboratory? *Human Factors, 63*(3), 531–547.

Lintern, G., Roscoe, S. N., & Sivier, J. E. (1990). Display principles, control dynamics, and environmental factors in pilot training and transfer. *Human Factors*, *32*(3), 299–317.

Llaneras, R. E., Swezey, R. W., and Allen, J. A. (1989). Motion as an instructional feature in maintenance training. In *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 1305–1309). Santa Monica, CA: Human Factors Society.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*(4), 492–527.

* Maddox, M. E., Johnson, W. B., & Frey, P. R. (1985). A transfer of training study in the nuclear power industry: From simulation to live performance. In *Proceedings of the 29th Annual Meeting of the Human Factors Society* (pp. 1022–1026). Baltimore, MD: Human Factors Society.

* McDonald, L. B., Waldrop, G. P., & Reynolds, R. (1982). Fidelity of simulation vs. transfer of training on a maintenance trainer. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 26, No. 8, pp. 741–745). Los Angeles, CA: SAGE Publications.

Mead, S., & Fisk, A. D. (1998). Measuring skill acquisition and retention with an ATM simulator: The need for age-specific training. *Human Factors, 40*(3), 516–523.

* Miller, E. E. (1975, July). *Instructional strategies using low-cost simulation for electronic maintenance* (Tech. Report HumRRO-FR-WD(TX)-75-20). Alexandria, VA: Human Resources Research Organization.

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.

Morris, N. M., & Rouse, W. B. (1985). Review and evaluation of empirical research in troubleshooting. *Human Factors, 27*(5), 503–530.

* Munley, G. A., & Patrick, J. (1997). Training and transfer of a structural fault-finding strategy. *Ergonomics, 40*(1), 92–109.

*¶ Park, O. -C., & Gittelman, S. S. (1992). Selective use of animation and feedback in computer-based instruction. *Educational Technology Research and Development, 40*(4), 27–38.

Park, O. -C., & Hopkins, R. (1992). Instructional conditions for using dynamic visual displays: A review. *Instructional Science, 21*(6), 427–449.

* Patrick, J., & Haines, B. (1988). Training and transfer of fault-finding skill. *Ergonomics, 31*(2), 193–210.

* Patrick, J., Haines, B., Munley, G., & Wallace, A. (1989). Transfer of fault-finding between simulated chemical plants. *Human Factors, 31*(5), 503–518.

* Patrick, J., James, N., & Friend, C. (1996). A field study of training fault-finding skill. *Le Travail Humain, 59*, 23–44.

Rasmussen, J. & Rouse, W. (Eds). (1981). *Human detection and diagnosis of system failures*. New York, NY: Plenum.

Reiter, R. (1987). A theory of diagnosis from first principles. *Artificial Intelligence 32*(1), 57–95.

Roberts, A. P. J., Stanton, N. A., Plant, K. L., Fay, D. T., & Pope. K. A. (2020). You say it is physical, I say it is functional; let us call the whole thing off! Simulation: An application divided by lack of common language. *Theoretical Issues in Ergonomics Science, 21*(5), 507–536.

*¶ Rouse, W. B. (1981). Experimental studies and mathematical models of human problem solving performance in fault diagnosis tasks. In J. Rasmussen & W. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 199–216). New York, NY: Plenum.

Roza, Z. C. (2005). *Simulation fidelity theory and practice: A unified approach to defining, specifying and measuring realism of simulations* (Doctoral thesis, Delft University of Technology, Delft, The Netherlands). Retrieved from https://repository.tudelft.nl/islandora/object/uuid:a5afd816-4b04-459d-b54c-da93e3b8a0d7/datastream/OBJ

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology, 8*(3), 197–208.

* Shriver, E. L., Fink, G. D., & Trexler, R. C. (1964). *Forecast systems analysis and training methods for electronics maintenance training* (No. HUMRRO-RR-13). Alexandria, VA: Human Resources Research Office.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology, 8*(2), 165–190.

Spangenberg, R. W. (1973). The motion variable in procedural learning. *AV Communication Review, 21*(4), 419–436.

*¶ 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.

* Thomas, T. K. (1993). Using cognitive simulations in multimedia for maintenance troubleshooting training: Practical, cost-effective simulations. In *Proceedings of the Interservice/Industry Training Systems Conference (I/ITSC)* (pp. 723–736). Orlando, Florida: National Training and Simulation Association.

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

* Yarnall, L., Vasquez, S., Werner, A., Kumar, R., Samarasekera, S., Acharya, G., Murray, G., Wolverton, M., Zhu, Z., Branzoi, V., Vitovitch, N., & Carpenter, J. (2015). Human performance in content design for interactive augmented reality systems. In *Proceeding of Annual Meeting of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*. Orlando, Florida: National Training and Simulation Association.