# Videorecording of Experts as a Method of Training-Simulator Design

Pooyan Doozandeh

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

This is a manuscript. The final version of this article appears in the journal:
*Theoretical Issues in Ergonomics Science (2021)* (issue and page numbers accessible in the journal's website)

**Abstract**

Designers of training-simulators have traditionally depended on task-analysis as the main informational resource for design. Although traditional task-analysis provides information about the necessary details of a task and the required steps in performance, it is deficient in representing a skill and its qualitative dimensions. This deficiency has led to lingering confusions for designers and resulted in using challenging constructs such as fidelity. To provide richer information about skills, this article proposes a novel design method. The method gathers information of skills directly from studying multiple experts by videorecording their performance in target environments. The recordings are then used to create a model of average expert behavior that can directly guide the design of simulators. An experimental example shows the practicality of the method in providing specific information for human factors/ergonomics practitioners in the design and improvement of training-simulators. By focusing on experts and videorecording their performance, the method provides multiple benefits to the design such as increasing the validity of the performance model, creating a qualitative model of skills, and proposing a generic solution to the design problem.

*Keywords:* training-simulation; visual methods; design; human performance modeling; learning

**Relevance to human factors/ergonomics theory**

Designing training-simulators is largely based on traditional methods of task-analysis. Although task-analysis is necessary for providing information about a task, it misses important information regarding skills in performance. This information can be captured via videorecording of experts. Because there are no standardized and verified methods of training-simulator design based on videorecording, this article presents this method and tests the method with an example. Using the videorecording method can bring significant improvements not only to training-simulator design, but has the potential to improve the practice in various other areas in human factors/ergonomics that has traditionally depended on standard task-analysis.

## Introduction

Designing training-simulators has been a subject of research and practice in human factors and ergonomics (HFE) for decades (e.g., Cream, Eggemeier, & Klein, 1978; Goode, Salmon, & Lenné, 2013; Hays & Singer, 1989; Salas, Wilson, Priest, & Guthrie, 2006; Smode, 1974). Because training-simulators are used to train the skills in certain tasks, the analysis of task informs the designer of the goal of a task, as well as the steps, tools, and strategies that are involved to accomplish the goal. This information provides the necessary details about a task, and because of this, traditional methods of task-analysis have been widely used in training-simulator design (e.g., Annett & Duncan, 1967; Stanton, 2006).

Although task-analysis has been successful in informing designers about tasks, it suffers from serious problems when applied to training-simulator design. First, in training, knowing about the skills in performance—i.e., *how* a task should be performed—is more important than the goal of the task and the steps in performance (e.g., Proctor & Vu, 2006; Schraagen, 2009; Van Merrionboer & Boot, 2009). However, traditional task-analysis methods can fall short of catching the skills, as they tend to reduce the skills to the steps that need to be taken to achieve a goal (e.g., Alliger, Beard, Bennett, & Colegrove, 2012; Card, Moran, & Newell, 1983; Militello & Hutton, 1998; Stanton, 2006). This problem is aggravated by the conventional use of text-based media in task-analysis (i.e., questionnaires, sheets, verbal protocol) because it is difficult, if not impossible, to record and model the skills through words (e.g., Nisbett & Wilson, 1977).

Additionally, task-analysis acquires information from human subjects in working with systems. However, it is not clear *who* those subjects should be. For example, in Card et al. (1983) and Diaper (2004) subjects are “users”, and in Annett and Duncan (1967) and Kirwan and

Ainsworth (1992) subjects are “operators”. Although some authors assumed that subjects should be experts in the task (e.g., John, & Kieras, 1996), this has never become a condition in recruiting subjects. In short, insufficient attention to the qualification of subjects is evident in task-analysis in general (e.g., Annett, 2004) and in training-simulator design in particular (e.g., Cream et al., 1978). In training, it is important to have criteria for selecting subjects to study as those subjects are indirectly instructing trainees in their practice, and shortcomings in the subjects’ performance can misguide the training.

These problems had consequences for the design and effectiveness of training-simulators. For the designer, in the lack of informational resources about skills, target environments and the task became the subjects of simulation, and this has engendered lingering design problems such as the “fidelity question” (e.g., Hays & Singer, 1989; Roberts et al., 2020) and the part-whole training dilemma (e.g., Wickens, Hutchins, Carolan, & Cumming, 2013). And for trainees, the result was a lack of guidance on how to perform tasks, and as a remedy, practicing with simulators were often accompanied with subject matter experts (SMEs) who provided guidance and feedback (e.g., Goode et al., 2013; Mahmood & Darzi, 2004). Although some authors argued against the sole reliance on conventional task-analysis in training (see Hays and Singer, 1989, p. 56; Schraagen, 2009), there has not been a systematic attempt to address the challenges and standardize a method specifically for training-simulator design (see also Fowlkes et al., 2009).

To resolve these shortcomings, we first need a medium that would inform us of skills and presents us with a direct and accurate picture of performance without modifications. And we need criteria to select subjects to study their performance. To resolve the first problem, this article uses the videorecording of performance in target task environments. This is because video can provide a direct picture of how a task is performed. This method—also known as “video-ethnography”—has been occasionally used by researchers in some areas of research (e.g., Blomberg, Giacomi, Mosher, & Swenton-Wall, 1993; Clancey, 2006; Engström & Medbo, 1997; Goldman, Pea, Barron, & Derry, 2014; Gilbreth, 1911; Hall, 2000; Heath, Hindmarsh, & Luff, 2010; Jordan & Henderson, 1995; Mathiassen, Liv, & Wahlström, 2013; Pink, 2013). Although observation and videorecording are considered as one of the traditional methods of data collection in HFE (e.g., Flanagan, 1954; Stanton et al., 2005/2018), it has not been standardized and presented as a data-collection method for training-simulator design (see Kirwan, & Ainsworth, 1992; Petersen, Nyce, & Lützhöft, 2011).

And to resolve the second problem, this article aims to study the performance of skillful individuals—or experts. Because the goal of most training programs is to make trainees perform like experts, modeling experts provides the criteria of performance and a direction that can guide trainees in their practice (e.g., Davids, Button, & Bennett, 2008; Jansson, Erlandsson, & Axelsson, 2015; Klein & Borders, 2016). The importance of studying experts and their characteristics has a long history in psychology (e.g., de Groot, 1965; Ericsson & Charness, 1994), and this article will use expert performance as a resource in design. This proposal would also reduce the reliance on the presence and explicit judgments of experts and SMEs regarding how to perform the task (e.g., Hoffman, Crandall, & Shadbolt, 1998) as those judgments can be inaccurate and so can misguide the training (e.g., Collins, 2006; Polanyi, 1966; Robinson, 1974).

These solutions are presented in the form of a method that shows what specifically should be done to use expert performance in designing training-simulators. The *videorecording method* is based on two key factors: videorecording and experts. It records the performance of experts, models their performance, and uses the model to guide the design of simulators. Other important elements of training such as trainees, the task, and tools are also considered, and their interplay with experts would constitute the basis of the proposed method of design. Although both studying experts and videorecording was practiced by previous research, a unified method of design for training-simulators is lacking. This is the goal of this study: to provide a method that uses the videorecording of experts to guide the design of training-simulators. This would help HFE researchers to advance the theory by providing comparative analyses of future methods, and would also help designers to have a practical reference for the design.

The next section introduces the method and discusses its five phases to design. Following the presentation of the method, an experimental example is provided to test the applicability and feasibility of the method. The strengths and limitations of the method are discussed toward the end of the article.

# The Videorecording Method

The details of the videorecording method are divided in five phases that are discussed toward the rest of this section: Identification of Elements, Recording Performance, Analyzing the Recordings, Validation of Behaviors, and Implementation.

## Identification of Elements

The following paragraphs describe the four essential elements that are needed in the design method.

### *Task*

Often, the designer is provided with the task to be trained. What is important on the designer's side is to clearly define the goal of the task as it tells us what should be trained by the simulator and what aspects of performance are important. This is not task-analysis, but a clear specification of the task and its goal. Because this would inform and guide our recordings and analysis in later phases, it is important to clearly identify the task and its goals.

### *Experts*

Experts and their performance are the focus of our recordings. Experts are individuals who have the experience of practicing tasks over an extended period of time in various conditions, and are recognized as having skills in performance. Despite attempts at defining general qualifying criteria (e.g., 10-year-rule of practice: Ericsson & Charness, 1994), the qualifications of an individual as expert for our recording does not follow fixed criteria. Rather, the experts' level of skill depends on the goal of the training-simulator and should be relative to the trainees' level of skill (Vygotsky, 1978). Because we will model their performance, the identified experts will determine and define the criteria of performance and practice in simulators.

### *Tools*

Devices, interfaces, and systems with which experts operate during performance are tools. For instance, for a pilot, tools can be the interface of a cockpit (e.g., hardware and software systems in the control panel) and the airplane as a whole. What distinguishes experts from novices is largely in their relationship with tools, and how tools become a part of experts' performance (e.g., Polanyi, 1966). Thus, if the goal is to derive behaviors that represent the expertise, we should identify and record tools and other environmental details that are important in performance.

### *Trainees*

Trainees are the users of the simulator and similar to the task, they are often determined by factors outside of the designer's control. The central characteristic that we need to know about trainees is their level of skills when they start using the simulator. This might directly affect our choices in the method such as what to record and how to record them. For example, recording a certain micro-scale behavior might only be important to design a simulator for highly-skilled trainees. Depending on the task and the training context, we might also consider videorecording trainees in working with target devices or when practicing with existing simulators. Analyzing such recordings can increase our familiarity with trainees regarding the characteristics of their performance such as their common errors or distracting elements during training sessions, and this familiarity can directly influence our technical specifications of the simulator.

## Recording Performance

After identifying and learning about the elements, we need to start the recordings. The goal of this phase is to record the performance of experts via video. Ideally, for a certain driving scenario, all experts should be given the same task in similar environmental conditions. For example, all expert drivers should be asked to drive in the same itinerary, traffic, weather, and visibility conditions. We can later expand this by adding more scenarios with different sets of conditions (e.g., night driving, adverse weather condition). To record their performance, for each expert, we need one or multiple video pieces—each taken from a different camera—from their performance (as shown in Figure 1).

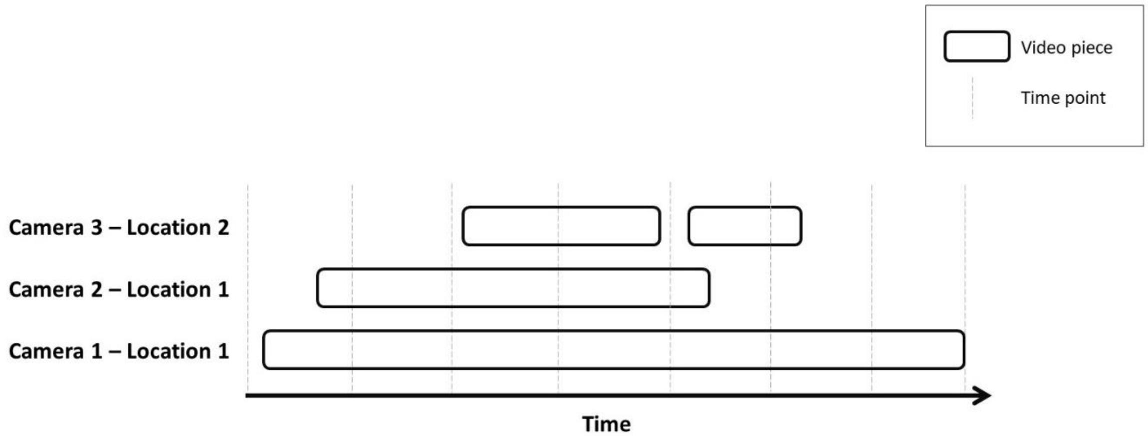


**Figure 1.** An abstract depiction of an example of the output of the Recording Performance phase. Video pieces from different cameras are synchronized relative to the performance time. Each video piece is used to show the task environment from a unique perspective or location.

### *What to Record*

Except for trainees, the elements that were identified in the previous phase (i.e., task, experts, tools) as well as additional relevant factors (e.g., time) should be recorded. The specific focus of recordings depends on the task and the goal of the expert in performance. For example, in marksmanship, it is important to record the complete body posture of the marksman, but in driving this might not matter. Also, we might need to record minute details of experts' actions or use an eye-tracker to understand their visual attention during performance; however, we need to consider that none of our recordings should interfere with experts' performance or distract them from their performance settings. Moreover, it is important to include all relevant task environments. For example, airplane maintenance technicians might work in more than one location where they should be recorded.

***How to Record***

We need to set up our recording equipment to capture important aspects of experts' performance. To do this, we should pay attention to the key technical details of the recording such as cameras' point of view, focal point, and audio recording devices (if needed). Such technical considerations can also inform us of what type of recording devices we need. In setting up our recording equipment, we should minimize the chance of interfering with experts' performance (see Aiello & Kolb, 1995).

## Analyzing the Recordings

We now should have one or multiple pieces of recording from the performance of each expert. The goal of this phase is to produce a single video piece of performance for each expert, with highlights over specific behaviors (as shown in Figure 2). To do so, we should first compile all video pieces into one, and later, highlight specific behaviors in that piece.

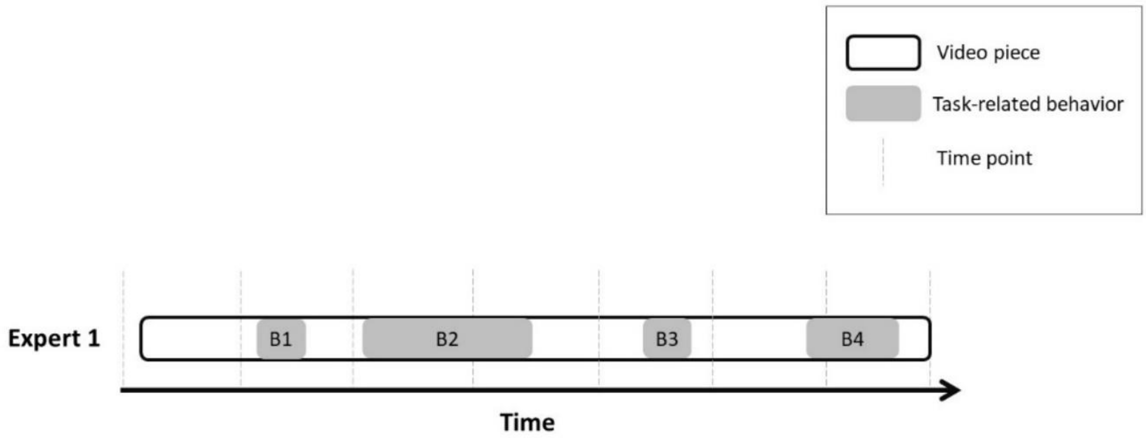


**Figure 2.** An abstract depiction of the output of Analyzing the Recordings phase. Each expert's performance should be represented by a single video piece with specified behaviors (highlighted areas in grey), and Bs refer to behaviors.

***Creating A Single Piece of Video***

Suppose we have more than one video piece for each expert. It is possible that one of the multiple video streams has the best performance-related content, and so, we might need to have a better focus on that piece. On the other hand, some behaviors need more than one video input, each input with equal importance in performance. So, we need to decide how we should combine multiple video feeds into one output display. This is necessary for the next step which is highlighting behaviors during the performance of experts. To create a single video piece, we often need to see multiple video inputs in a single display during analysis. Figure 3 shows two examples of the compilation of multiple inputs in a single display (see Advance Driving School, 2016).

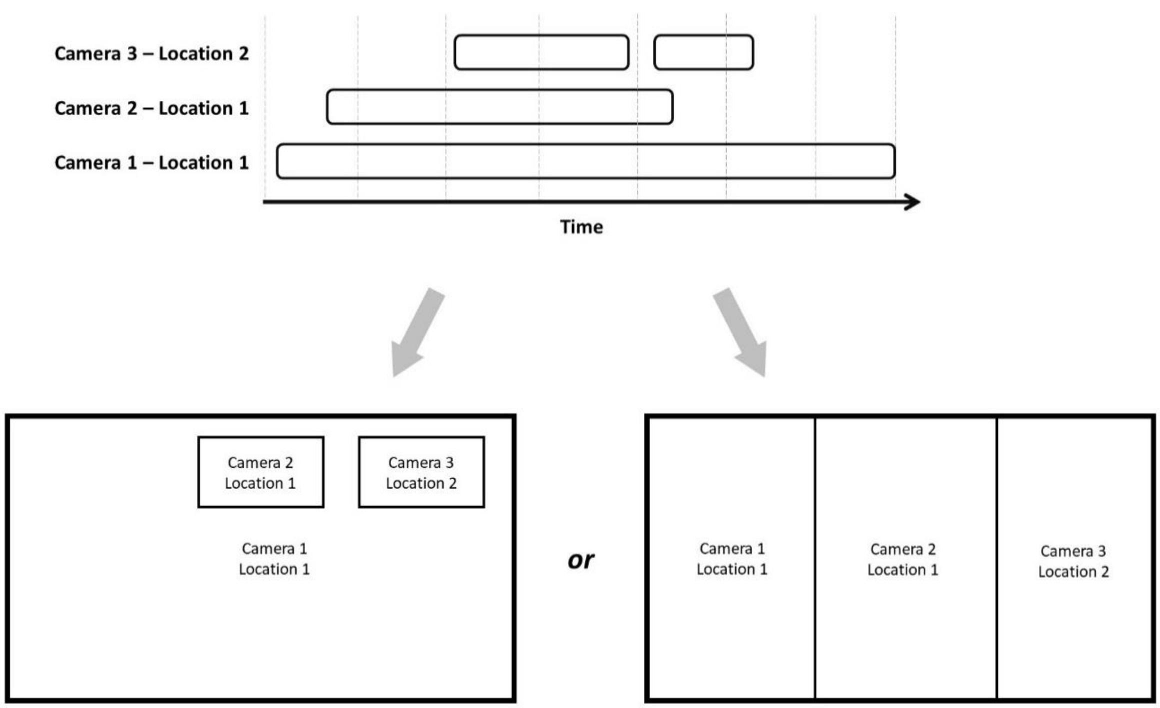

**Figure 3.** Two different compilations of a display to show the video pieces from multiple cameras. In the primary-secondary example (left) one video piece has the primary performance-related content, and in the equal-importance example (right) all video pieces have equal importance in performance.

### *Specifying Behaviors*

Each behavior has a duration within an expert's entire performance (grey boxes in Figure 2). For example, in landing an airplane, checking if the wheels are fully open can constitute one behavior. By consulting with SMEs, we can ensure that the specified behaviors are relevant to task performance. And, to increase the validity of our analysis, we can ask multiple individuals (preferably SMEs) to analyze the videos independently and in parallel (e.g., Mathiassen et al., 2013). The aggregation of their analysis can reduce the subjectivity in behavior specification.

Behaviors can be defined and specified in various levels of abstraction. For example, a macro-level behavior can be parallel parking of an automobile as it contains many details, and a micro-level behavior can be pressing the brake pedal in the same task. Choosing the level of abstraction depends on the task and what skills should be trained. In a relevant note, some aspects of performance might be difficult to record via video. For example, how much force an expert is physically applying to a controller or pedal is difficult to capture via video, and so, we might need special devices to measure these factors. The results of such measurements can supplement the videos to maximize the task-related information that will be available to designers.

To highlight behaviors, we can observe the video pieces and manually specify single behaviors. Another option is to use available computational video-processing tools to help us in behavior specification. Such tools are currently used for human performance analysis in various domains (e.g., Loukas, 2018; Shih, 2017). Considering the use of computational resources and tools can be particularly useful when we have many experts with long durations of performance and numerous behaviors that would take too long for humans to analyze.

**Validation of Behaviors**

By now we should have one piece of video for each expert. In this phase, we need to review the recordings to see how experts perform the same behavior to average their performance in that behavior. This averaging leads to a model of *average expert behavior (AEB)*. This model represents the expertise in this context and will be used in design. Figure 4 shows an abstract depiction of the process in this phase.

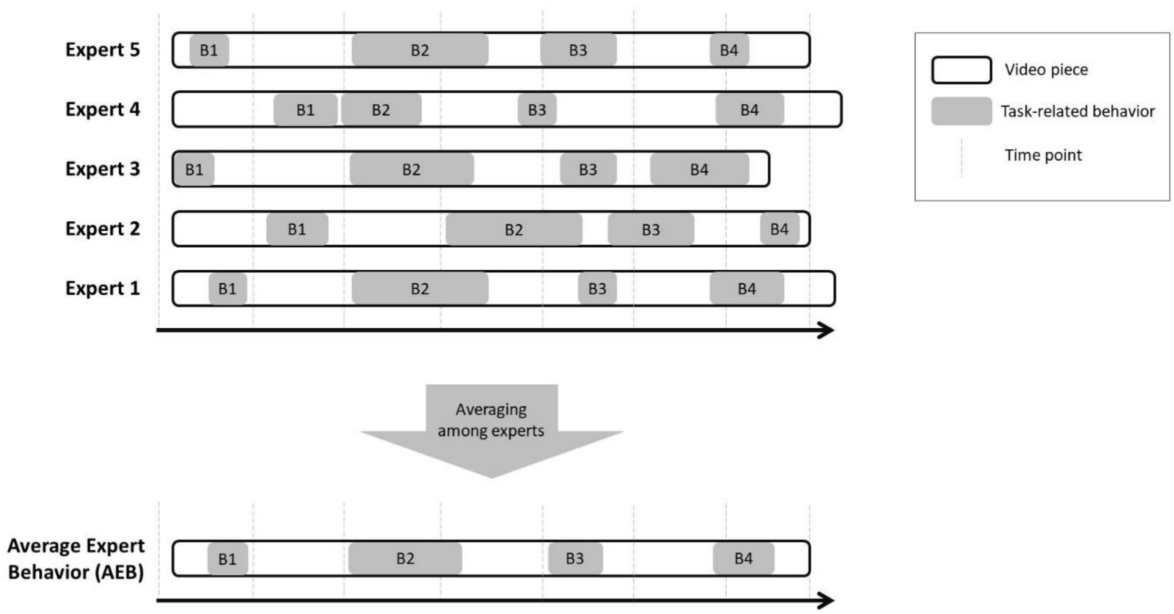


**Figure 4.** Producing the model of average expert behavior.

The first step in averaging is to choose the criteria of comparison between behaviors. This is an important advantage of this method compared to the traditional task-analysis methods, as we would have the freedom to review the recordings and choose and test the criteria of performance; if one criterion was not satisfactory (e.g., not shared among experts), we can choose another. An example of a quantitative criterion that often distinguishes experts from novices is timing in performance (e.g., Beek & Turvey, 1992). The temporal analysis can include measuring the duration of performing a behavior (the width of gray boxes in each expert's performance in Figure 4), the delay between two behaviors (the distance between two gray boxes in Figure 4), and the entire duration of performance. This will also help us in creating a temporal dimension for the AEB (the bottom of Figure 4) that can be used for further quantitative or qualitative comparisons between behaviors.

In addition to timing, we can use other quantitative criteria in performance such as degree, length, and force that experts represent in their performance. We might also need to use qualitative criteria in comparisons. For example, one qualitative criterion that can be used for a variety of tasks is the order of behaviors in task performance (e.g., Carroll & Olson, 1988). In this case, averaging consists of knowing what sequence of behaviors were followed by the majority of experts.

Once we averaged each behavior, the result can be in the form of numbers and statistics, shapes, prototypes, figures, animations, or any other form that can inform us of how, on average, experts performed a behavior. Having this information for multiple or all behaviors creates the general model of AEB (Figure 4, bottom). By now, it is clear that this approach is in stark contrast to conventional task-analysis, as the model of expert behavior comes from reviewing the videos of experts, not from the logical steps needed for a goal.

**Implementation**

To use the AEB in practice, we first need to have a minimal simulation of the target task environment (i.e., a low-fidelity simulation) that would provide the possibility for trainees to practice the task in similar environment in which the experts were recorded. We will later supply this minimal simulation with the information that we have in the AEB as well as the videos.

For example, the information about each behavior in the AEB can be used as a condition in the system that would monitor trainees' practice and inform them of proper performance regarding timing, movements, and other performance criteria. This monitoring and feedback would then require additional tools in the simulator (e.g., eye-tracker, video-monitoring systems). Also, reviewing the videos informs us of what tools were used by experts, and this can inform the decisions of hardware and software configurations of the simulator.

It is important to remember that designing a new simulator for a task might not only need an extensive consultation with SMEs, but would also require comprehensive information about many aspects of the target environment, tools, and the task-relevant behaviors of experts that need to be included in our recordings. A simpler way that the AEB can be used is to bring small-scale modifications to the existing simulators. This is particularly useful if we have a set of recordings that are incomprehensive, unstandardized, and may violate many of the conditions in the method. And even if we have comprehensive recordings, it might not always be efficient to design new systems if there are existing systems in use. For example, if we can verify and average only one behavior from the recording of experts, we can use the information as a condition in the system that can monitor and guide trainees' performance. Additionally, we can first review the working of an existing simulator, derive a scenario of task, and observe how experts would perform that scenario through videos (e.g., Cannon-Bowers, Burns, Salas, & Pruitt, 1998). We can use this question to inform our method, initiate the recordings, and use the recordings for improving simulators.

Bringing small-scale improvements to the existing systems might also be more convincing from an organizational perspective, as organizations are often more cautious about using novel simulators than using older ones with improvements. This can pave the way to a smoother transition toward adopting the method by organizations (for further discussion on this topic, see Salas et al., 2006, 20).

**Summary**

The relationship between the four elements and the five phases of the method is depicted in Figure 5, and the phases of the method are summarized in Table 1. Reviewing the phases of Validation and Implementation tells us an important advantages of this method versus traditional task-analysis: because the video provides a direct picture of performance, it gives us the freedom to choose the performance criteria of comparisons among experts. As such, hidden aspects of the expertise that might go unnoticed by experts can be discovered by analyzing the videos.

This section described the method in an ideal scenario in which all requirements were provided and all detailed were followed. The next section presents an example that shows how the method works in a real-world situation with unstructured data.

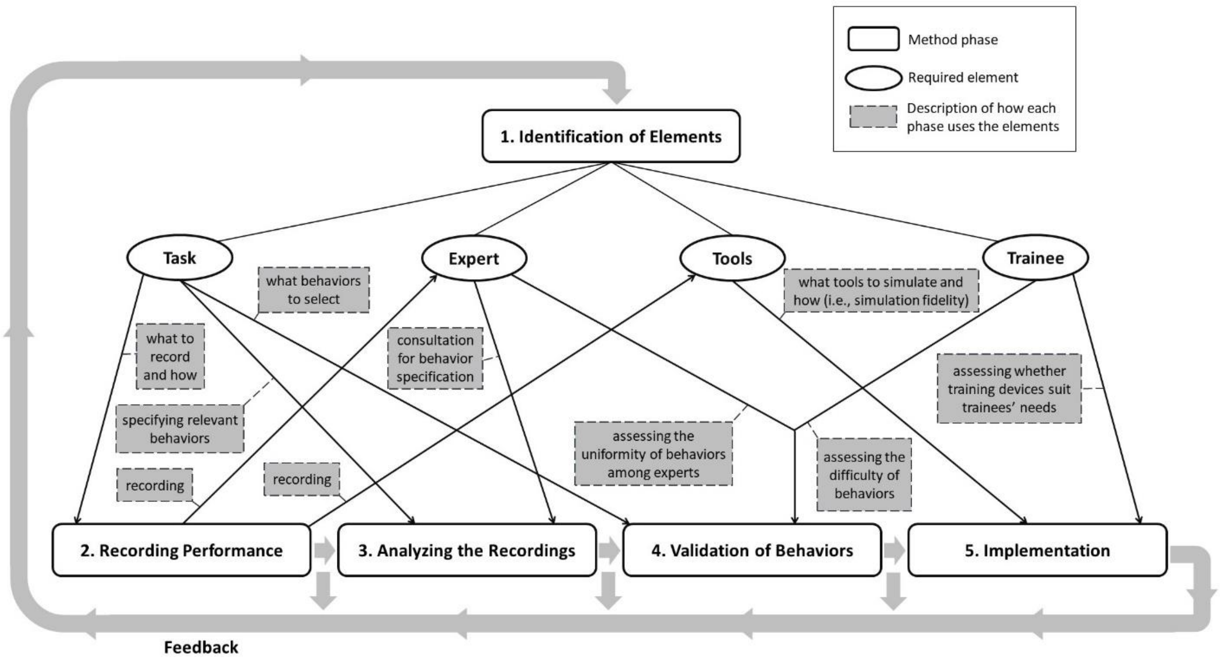


**Figure 5.** An abstract depiction of the videorecording method. The feedback is considered so that at any step, if further information or elements were needed, designers add that information in the design cycle.

**Table 1**

*A Summary of the Input and Output of Each Design Phase*

| | Phase | Input | Output |
|---|---|---|---|
| 1 | Identification of Elements | Task-analysis, interview, observation, etc. | Specification of task, experts, tools, and trainees. |
| 2 | Recording Performance | Observation of performance including the specified elements. | Piece(s) of video of the expert and tools during performance (Figure 1). |
| 3 | Analyzing the Recordings | (Output of Phase 2) | For each expert, a single piece of video with highlights over single behaviors (Figure 2). |
| 4 | Validation of Behaviors | Video pieces of the performances of all experts (Figure 4, top). | A validated model of average expert behavior (Figure 4, bottom). |
| 5 | Implementation | The average expert behavior. | Hardware and software specifications for the design and improvements of simulators to guide trainees in practice. |

## An Experimental Example

In this example, the videorecording method is used to provide design suggestions for heavy-truck driving simulators. Such simulators are used to train the essentials of what is needed to drive a truck and pass the driver's license test (see Figure 6). This example is provided to show how important phases of the method (i.e., analyzing and validating behaviors) can be used.


**Figure 6.** Two examples of truck-driving simulators used for training. Derived from Veterans Health Administration (2009).

### Identification of Elements (Example)

The *task* in this example is heavy-truck driving—i.e., moving the truck from one location to another in a city or highway. After conducting a comprehensive search to identify *experts*, it was fortunate to find three experts who uploaded the videos of their performance online for public access. The videos satisfied the basic requirements to be used in the method (i.e., similar recordings, road, visibility, traffic, time-of-day). There were four videos from the three experts; general information of the experts and videos is summarized in Table 2 and sample snapshots of videos are shown in Figure 7.

**Table 2**

*Information of Experts and Their Videos*

| Expert | Estimated age of the expert | Total video length in minutes | Reference |
|---|---|---|---|
| 1 | 60 | 17 | Logan (2016) |
| 2 | 60 | 28 | West Side Transport (2016) |
| 3 | 40 | 18 | CDLTestTruck (2012) |
| | | 15 | CDLTestTruck (2011) |


**Figure 7.** Sample snapshots of videos used in the study. The video of Expert 1 is shown on the left, Expert 2 is shown in the middle, and Expert 3 is shown on the right.

The main *tool* in this task is the truck and its performance-related devices (e.g., steering wheel, mirrors). The trucks that the experts used were regular 18-wheelers (i.e., Class 8 trucks in the US Gross Vehicle Weight Rating classification). The streets and the incoming traffic were performance-related environmental details included in the videos. *Trainees* are individuals whose goal is to pass the Class A Commercial Driver's License (CDL) Skills Test that is officially required to drive heavy trucks (Class 7 and Class 8 trucks, more than 26,001 pounds of weight) in the US. The minimum age requirement for this type of license is 18 years in most states in the US.

**Recording Performance (Example)**

The driving performance was recorded with a camera that was installed inside drivers' cabin, showing the front of the vehicle and the road ahead through the windshield (Figure 7). The driving of one of the experts is recorded with a wide-angle lens that helped including more details of the road (i.e., Expert 1), and the other two experts were recorded with normal lenses. The side mirrors can also be seen in the recordings of two of the experts that show the two sides of the truck (i.e., Expert 1 and Expert 2). The videos were recorded in similar road, traffic, visibility, and climate conditions, and the mean length of the videos was about 20 minutes. All videos underwent minor editing as they simply showed a continuous driving session.

**Analyzing the Recordings (Example)**

Because for each expert there was one camera that produced one video stream of performance, there was no need to combine different video sources and create a single piece of recording. So, each performance-related behavior was specified during the performance of each expert. To do this, each video was replayed and each behavior was manually identified and temporally highlighted. Table 3 shows the result of the analysis and the specified behaviors. To facilitate the validation, labels for behaviors (e.g., "Turn left") were standardized between the three experts. It should be noted that because this is an example, the analysis of the recordings was conducted manually. In real-world cases, because there might be more experts to analyze with longer duration of performance and more behaviors, we can consider using computational and machine-learning techniques for the analysis.

**Table 3**

*The Specified Behaviors During the Performance of the Three Experts*

| Expert 1 | | Expert 2 | | Expert 3 | |
|---|---|---|---|---|---|
| 03:16 - 03:46 | Turn left | 00:42 - 01:11 | Turn left | First video | |
| 04:20 - 04:40 | Turn right inters. | 01:48 - 02:05 | Turn right inters. | 01:38 - 01:54 | Turn right inters. |
| 04:50 - 05:09 | Curve right | 07:00 - 07:39 | Curve right | 02:47 - 02:59 | Lane change |
| 05:55 - 06:16 | Curve right | 08:50 - 09:05 | Curve right | 04:00 - 04:20 | Turn left |
| 06:29 - 06:45 | Merge to highway | 09:45 - 10:00 | Merge to highway | 04:56 - 05:16 | Turn left |
| 07:10 - 07:20 | Lane change | 12:45 - 13:00 | Exit highway | 05:28 - 05:54 | Roadside stop |
| 07:39 - 07:48 | Lane change | 13:38 - 13:52 | Merge to highway | 08:21 - 09:18 | Parallel parking |
| 08:37 - 08:56 | Exit highway | 14:05 - 14:13 | Lane change | 11:30 - 11:45 | Turn left |
| 09:05 - 09:32 | Pass traffic circle | 14:43 - 15:00 | Exit highway | 12:33 - 12:50 | Turn right inters. |
| 10:39 - 11:07 | Turn right inters. | 15:15 - 15:37 | Turn left | 12:51 - 13:11 | Curve right |
| 11:40 - 12:00 | Turn left | 16:02 - 16:16 | Turn left | 13:32 - 13:48 | Merge to highway |
| 12:09 - 12:38 | Turn Left | 16:30 - 16:41 | Merge to highway | 14:35 - 14:48 | Exit highway |
| 13:00 - 13:13 | Curve right | 17:05 - 17:18 | Exit highway | 15:58 - 16:18 | Turn left |
| 13:47 - 14:06 | Turn left | 18:03 - 18:16 | Merge to highway | 17:16 - 17:38 | Turn left |
| 14:38 - 15:04 | Curve left | 21:30 - 21:39 | Exit highway | 18:11 - 18:26 | Roadside stop |
| 15:34 - 15:45 | Lane change | 22:39 - 22:53 | Turn right inters. | Second video | |
| 15:45 - 15:55 | Lane change | 25:15 - 25:30 | Turn left | 07:33 - 07:50 | Turn right inters. |
| 16:05 - 16:16 | Exit highway | 27:54 - 28:04 | Turn left | 09:19 - 09:37 | Turn left |
| 16:30 - 17:14 | Pass traffic circle | | | 10:39 - 10:49 | Lane change |
| 17:20 - 17:38 | Pass traffic circle | | | 12:32 - 12:47 | Turn right inters. |
| 18:27 - 18:52 | Turn left | | | 13:43 - 13:56 | Merge to highway |
| 18:53 - 19:09 | Roadside stop | | | 14:37 - 14:47 | Exit highway |
| 20:12 - 20:36 | Turn right inters. | | | 14:53 - 15:06 | Turn left |
| | | | | 15:06 - 15:19 | Turn right |
| | | | | 15:38 - 15:53 | Turn right inters. |
| | | | | 17:10 - 17:31 | Turn right |
| | | | | 19:50 - 20:07 | Turn left |

## Validation of Behaviors (Example)

For the sake of brevity, two behaviors were chosen for validation. The first is Turning Right at an Intersection (i.e., "turn right inters." in Table 3). This behavior is a 90-degree turning of the truck to the right at an intersection. As shown in Table 3, Expert 1 has three, Expert 2 has two, and Expert 3 has five instances of this behavior which makes ten instances in total. The second behavior is Lane Change that moves the truck from one lane to another in a multi-lane road. There are seven instances of this behavior in total: four in Expert 1, one in Expert 2, and two in Expert 3.

### *Quantitative Averaging*

One important advantage of using the video is in giving us the freedom to choose the performance criteria in averaging the behaviors. Here, because it is a familiar measure in performance-analysis, the timing in performance was chosen as a quantitative criterion of comparison in this example. But it should be noted that in real-world cases, we might need to start with different performance criteria such as accuracy or safety measures. For the turning-right behavior, the videos showed that the mean time for all ten instances of this behavior among the three experts was 18.30 seconds with the standard deviation of 4.47 seconds. In other words, the data shows that it takes 18.30 seconds for experts to turn right at an intersection. This finding will be implemented in the simulator to guide trainees in practice. For the lane change, averaging all seven instances of the behavior resulted in the mean time of 10.00 seconds and the standard deviation of 1.29 seconds. In averaging of the two behaviors, the timings in Table 3 were used, and Figure 8 summarizes the results. It is important to note the advantage of the method compared to conventional task-analysis is in directly using experts' data in tabulating Table 3, without any rational analysis of their performance or relying on their own reports.

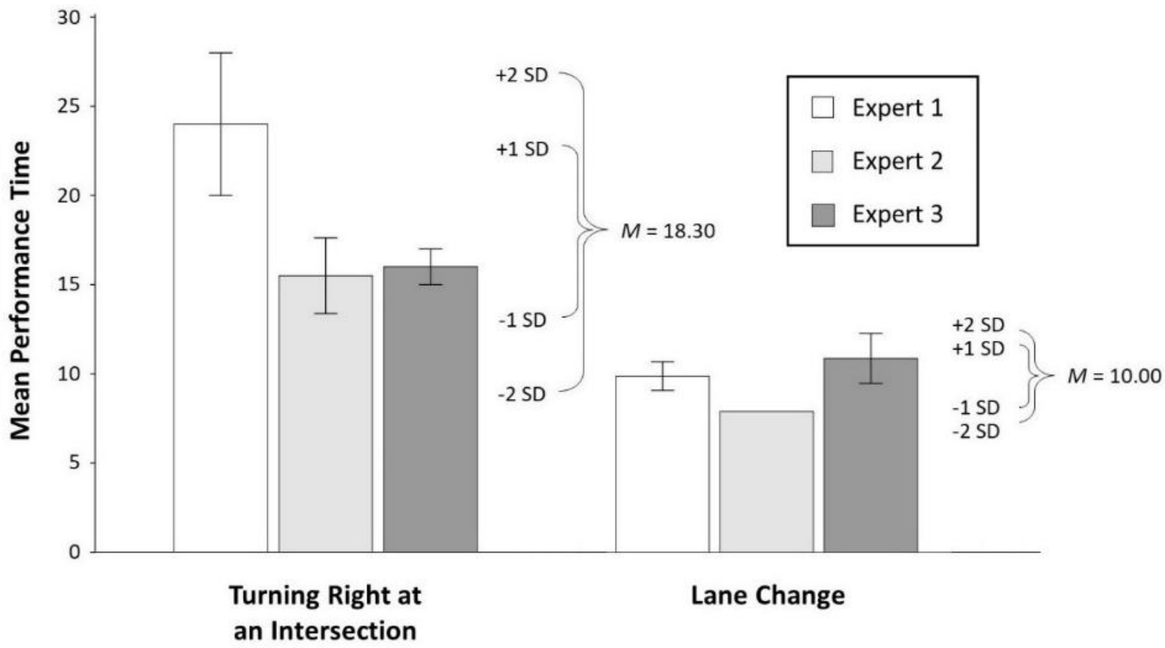


**Figure 8.** The results of the temporal averaging for the two behaviors. M is mean and SD is standard deviation.

### *Qualitative Averaging*

The advantage of video in providing freedom for analysis is greater regarding qualitative averaging as most qualitative aspects should be chosen after reviewing the videos. Task-analysis methods do not provide such a freedom as performance criteria must be decided beforehand. Here, the order of the sub-task performance was chosen as a qualitative criterion, but in real-world cases, different qualitative criteria might be used. The question is what specific steps did experts follow in performance for each of the two behaviors? For the turning-right behavior, Expert 1 suggested in his video that the performance can be broken into four steps that should be completed in order: Approach, Stop, Turn Through, and Completion. Reviewing the videos from the other two experts showed that this suggestion helped to analyze this behavior, and this sub-task classification was used here to make the qualitative averaging. Table 4 describes how each expert turned right according to each of these four steps.

By aggregating the results in Table 4 it was found that that the following steps were followed by at least two of the three experts:

- Approach: Activate the turn signal and reduce speed 100 feet ahead. Detect the stop sign.
- Stop: Complete stop two feet before the stop sign.
- Turn Through: First, move forward a bit to have a better view on the traffic. While turning slightly, drive to the extended space. Look at the right-side rear mirror; when the trailer wheels are cleared, make a sharp turn.
- Completion: Return to the right lane quickly. Cancel the signal. Continue.

The same qualitative averaging can be applied to the lane change behavior with the following results:

- Initiate: Road should be straight and not curving. Turn on the turn signal well in advance of lane change (at least 400 feet ahead).
- Check traffic: Check the traffic behind before changing lane by looking at the side mirrors.
- Change: Slowly turn the steering wheel (can be quantified). Position the truck and trailer properly in the new lane.
- Completion: Cancel the signal.

**Table 4**

*A General Description of How Each Expert Followed the Four Steps When Turning Right*

| | Approach | Stop | Turn Through | Completion |
|---|---|---|---|---|
| Expert 1 | Activate the turn signal 100 feet before the intersection, reduce speed. | Completely stop two feet before the sign. | Before the main turn, move the truck forward a bit to have a better view. If no traffic, drive to the extended amount of space (second 12-foot lane from the right), sharp turn after the tires are cleared on the right. Check the side rear mirrors. | Check the side mirrors to ensure there is no car coming from the back, return your vehicle to the correct lane, continue, cancel the turn signal. |
| Expert 2 | Detect the stop sign, scan the intersection, reduce speed, activate the turn signal. | Completely stop, look at the incoming traffic, wait for a big gap in traffic. | First, move forward a bit to have a better view. Sharp turn after the tires are cleared on the right rear mirror. | Return to the right lane quickly to prevent the following vehicles to enter the right lane during the turn. |
| Expert 3 | From 100-125 feet ahead turn on the signal, reduce speed, detect the stop sign. | Completely stop two feet before the stop sign. | Move forward a bit. While turning, look at the side rear mirrors to see if no car entered the gap while you were turning. | Return to the right lane quickly, cancel the signal. |

*Note.* The text in the table is derived from the experts' verbal descriptions as well as the author's analysis of the experts' performance.

**Implementation (Example)**

Now, the quantitative and qualitative information of experts' performance should be used in simulators to guide trainees in practicing the two behaviors. The information should be presented to trainees in the form of feedbacks, and there are many decisions that should be made in providing feedbacks (e.g., concurrent feedback vs. terminal feedback, visual vs. auditory). The details of these decisions highly depend on the task and would go beyond the scope of our discussion (for a review, see Sigrist, Rauter, Riener, & Wolf, 2013). So, here a simple suggestion is presented of how to use the results of experts' performance in design.

We can adopt a concurrent feedback (i.e., providing feedbacks to trainees during practicing with the simulator) provided through visual cues on the screen of the simulator. Before practicing with the simulator, trainees are instructed that their performance will be monitored and matched against the performance of experts that is embedded in the simulator, and that they will receive respective feedback during practice to improve their performance.

Now, we need to implement the performance results in the simulator. Regarding the quantitative analysis, the curly brackets in Figure 8 shows how the timing windows for the two behaviors were defined. Specifically, if during a practice session in the simulator, a trainee's performance time was close to the mean (i.e., less than 1 standard deviation from the mean, or within the smaller brackets in Figure 8), the system would inform the trainee of the good performance timing. If the trainee's performance took between 1 and 2 standard deviations from the mean (i.e., between the small and big brackets in Figure 8), the system would indicate acceptable performance with guidance to further improve the timing. And if the trainee's timing was far away from the experts' mean (i.e., more than 2 standard deviations from the mean, or outside of the larger brackets in Figure 8), the system would inform the trainees that they were too fast (or slow) to perform the behavior and they need to accelerate (or decelerate) their performance.

The same conditional checking applies to qualitative averaging. For example, in the Turn Through step of the turning-right behavior, the system can monitor trainees to see if they visually check the clearance of the trailers' wheels with side mirrors before making a sharp turn. Similarly, the system can check if trainees change lane in curvy or straight roads.

To check trainees' performance, monitoring systems should be included in the simulator (e.g., eye-tracker). Additionally, only the tools and gauges that experts used in performance are necessary to be included in the simulator (e.g., odometer, tire pressure gauge, break valves). As a result, the videos can provide hardware and software suggestions for the design. With a similar but more comprehensive analysis of all behaviors, the method can shape the hardware and software characteristics of the simulator.

The above suggestion was one way to provide feedback to trainees; in reality, researchers and designers of truck-driving simulators should determine how to administer the feedbacks. What was presented in the example was respecting the limited space, and the analysis of experts' behaviors could be more sophisticated. For example, averaging the performance time for all instances of a behavior among the three experts might have oversimplified the details of their performance because, for example, different experts might perform their behaviors with different speeds, and this might be a part of their style in performance. Or, performance criteria other than timing should be used for the design of truck-driving simulators. Nonetheless, the direction for a more sophisticated use of the method is obvious by now.

# Discussions

The following paragraphs outline the strengths and limitations of the videorecording method.

## Contributions and Strengths

The presented method brings the following contributions to the HFE in general and training-simulator design community in particular:

- Relying on the performance of one expert can sway the training because one expert's approach to performance might be different from those of other experts. So, the method used a *society-of-expert* approach to minimize the problem of individual differences between experts' performance.
- The method standardized the guidance from experts to trainees. The guidance does not involve the presence of experts and their explicit judgments (e.g., as Figure 6 shows, see Hoffman et al., 1998) as doing so can distract trainees from practicing the task and might be inaccurate (e.g., Collins, 2006; Polanyi, 1966; Robinson, 1974). Recording experts during performance and using it to provide indirect guidance through feedback is presented as a solution.
- The method is generic as it can be applied to various domains that use training-simulators. Nonetheless, this genericity did not prevent the method from providing specific and low-level guidance for designers (as shown in the experimental example). This combination of genericity and practicality can help designers to use the method as a ready-to-use tool in a variety of domains.

The method provided normative information of performance—how the task *should* be performed. This “should” was derived from reviewing the performance of multiple experts through videos. Traditional task-analysis, however, is logically descriptive by telling us how the task *can* be performed. This highlights the difference between studying experts in this method versus “users” or “operators” as occasionally practiced in task-analysis. The presented method received performance information directly from experts via video, but task-analysis breaks down a task into steps. There is an important distinction between focusing on experts’ skills rather than steps, as insufficient attention to experts shifted the focus of design to simulating target task environments (e.g., simulation fidelity) and other trivial aspects of a task. This negligence resulted in confusion for HFE practitioners, researchers, and designers (e.g., Fowlkes et al., 2009; Hamstra et al.,2014; Roberts et al., 2020; Wickens et al., 2013) and prevented trainees to practice the skills. Using expert performance in the presented method provided a direction that helped resolving this confusion.

The videos of expert truck drivers in the experimental example provided an informative case because they all contained the verbal reports of drivers. Although the verbal reports were useful in analyzing the videos, the performance-related information that was derived from the videos was largely independent of verbal reports. Rather, the information came from observing drivers’ behaviors in the videos. This shows how relying on verbal reports and other text-based media as commonly used in task-analysis cannot reveal certain important aspects of experts’ behaviors.

In a relevant note, it has been shown that experts might be unaware and unconscious of their own approach and details to performance (e.g., Collins, 2006; Polanyi, 1966; Robinson, 1974). In other words, experts can perform implicitly, without the explicit knowledge of how they perform. As a result, there is skepticism toward asking for experts' explicit judgments to guide trainees and designing training-simulators. Rather, we need to see how experts perform the task in natural settings and without interference, and use their performance to characterize the skills. This was the central goal of the presented method, and unlike other research that use experts' explicit knowledge outside of the target environment (e.g., Hoffman et al., 1998; Jansson et al., 2015; Klein & Borders, 2016), the method prioritizes the natural performance of experts.

Aside from providing an accurate and direct picture of skills and other benefits of using the video as a medium (see Heath et al., 2010; Petersen et al., 2011), the advantage of video is in being a dynamic source of information. Often, it is difficult to predict what should be the focus of task-analysis in performance. The video can resolve this difficulty by providing the opportunity of unpredictable analyses with different criteria and goals that a designer might need at any step of the design, including environmental factors (e.g., tools, locations). On the contrary, text-based task-analysis should determine what should be known before conducting the analysis, and as such, the interpretation of performance is static and cannot provide new information after the analysis is completed.

One might argue that the results gained in the provided experimental example could also be gained without using video or by conventional task-analysis methods. Indeed. However, we should remember that the same recordings in the example could be used to gain different results. If we changed the performance criteria in analysis, we would get new information. Generally, the videorecording method, on its own, has the capacity to model experts' performance and provide valuable design information. Unfortunately, there were not many well-formulated design methods in the literature that could be used for comparisons, and so, we can only think of conventional task-analysis methods (e.g., verbal protocol, hierarchical task-analysis). From this perspective, if we put the videorecording method side-by-side each of conventional task-analysis methods, it becomes clear that no other single method has the capacity of the videorecording method in terms of sufficiency and independence to guide the design in various tasks. This capacity comes from relying on the two key factors of experts and videorecording that provide necessary—if not sufficient—information for designing training simulators.

And the experimental example showed how the method can be flexible in working with unstandardized videos. In fact, the videos came from different sources, were recorded with different equipment, had different lengths, and experts practiced different orders of behaviors during their performance. Nonetheless, the method could still be used to present a model of their individual behaviors that proved to be useful in design.

In short, the video can resolve various problems and shortcomings of the traditional task-analysis. The arguments against the costs associated with using videorecording and observations (e.g., Stanton et al., 2005/2018, p. 43) are counterbalanced by the recent availability of recording devices and computational tools for video-analysis (e.g., Loukas, 2018; Shih, 2017).

## Limitations

The premise of the method is that the information of skills is representable through video. However, the skills in many tasks cannot be easily observed, and so, the sole reliance on the method is limited to those tasks for which some or most skills can be externally observed. Table 5 shows examples of tasks and domains for which the method may or may not be applicable.

One proposed solution is that for tasks that have unobservable aspects, the videos should be supplied with other modes of information such as verbal reports, information of physical force, description of goals and subgoals, and any other mode of data that can provide skill-related information regarding the actions and decisions in performance. Therefore, the method is not intended to replace all other existing methods of data collection and design, but can be used together with other methods and techniques such as verbal protocol.

**Table 5**

*Examples of Tasks and Domains for Which the Method May or May Not Be Applicable for Training-Simulator Design*

| May be applicable | May not be applicable |
|---|---|
| Transportation (e.g., flying, sailing, driving) | Air traffic control |
| Surgery | Troubleshooting |
| Maintenance operations | Professional design |
| Sport | Computer programming |
| Drawing | Recognition tasks |

*Note*. The classification is based on three criteria: observability, specifiability, and measurability of behaviors. This is not a definitive classification but an example of tasks and domains. The classification is partially derived from Ericsson, Hoffman, Kozbelt, and Williams (2018).

Additionally, the decision of how to implement the AEB in simulators can itself open contentious debates. For example, some authors supported the effectiveness of providing feedback during practice, but others supported the use of feedbacks after practice (for discussions, see Sülzenbrück, & Heuer, 2011; Walsh, Ling, Wang, & Carnahan, 2009). This decision should be made by reviewing previous research and considering the properties of the task and the training program (see Sigrist et al., 2013).

In a relevant note, it is one of the general assumptions of the current work that practicing expert performance through feedback is an effective strategy in training; an assumption that is well-supported in the literature (e.g., Huegel, Celik, Israr, & O'Malley, 2009; Klein & Borders, 2016). Therefore, the effectiveness of the simulators that would be built by using the method was not tested in this work. Nonetheless, before being adopted in practice, design methods need to be tested in practice and with real simulators within training programs. Therefore, for conducting a rigorous test, a longitudinal study may need to be conducted to measure the long-term training benefits of the simulators built upon the presented videorecording method.

Another difficulty can be in finding shared behaviors among multiple experts and averaging the performance among them to create the model of AEB. The difficulty is in situations in which we cannot find shared behaviors, or when averaging the behaviors does not yield a meaningful result. For example, suppose we have two expert troubleshooting technicians with similar levels of expertise. When their order of actions and decisions in the same troubleshooting problem is entirely different from each other, we cannot easily model their skills in the AEB. This problem is more likely to emerge for more complex tasks or if the number of available experts for recording is limited as different experts might have different approaches to performance which makes it challenging to derive a single performance model. In such cases, we either need to enroll more experts—that is not always feasible—or use a different approach for training the skills in those tasks (for further reading, see Duncan, 1985; Scardamalia & Bereiter, 1991).

And there were limitations in the extent and details of the experimental example. For example, it could have increased the validity of our model if experts recorded their driving under different scenarios and operational conditions (e.g., weather, environmental lighting). As a result, the conclusions in the provided example can only cover a narrow range of operational conditions in performance (e.g., day, light traffic, sunny, etc.). And as another example, timing is used as one of the main performance criteria in performance modelling of the example. However, in real-world cases, we will need to use other performance criteria (e.g., accuracy, movements, etc.). Indeed, to be used for designing real-world truck driving simulators, a more comprehensive study is needed, and as such, the example was simply used to show the practicality of the method.

Due to the nature of the method that works with qualitative data, additional challenges might also appear (e.g., Heath et al., 2010). Examples of these challenges are determining the required number of experts, finding and enrolling experts for recordings, efforts needed to record the videos, and subjectivity in analyzing videos and modeling the expertise (see Goldman et al., 2014). Although some of the challenges can be addressed by consulting with SMEs, the final decision of whether to use the method should depend on cost-benefit analyses and available resources.

## Conclusion

The design of training-simulators is a difficult and ill-defined problem to solve. The lack of a unified approach, theory, or method to address the design problem is indicative of its difficulty as solutions have often been tailored to specific problems. Although the problem-specific approach in providing solutions was effective in practice, it prevented the creation of generic design methods. Such methods can bring researchers and designers under the same design problem so that experiences can be shared, and solutions can be scientifically tested and verified. This can accelerate the progress of training-simulators in future. The need for generic design methods is particularly important these days with the advance and availability of modern devices (e.g., virtual-reality tools) that can lead to growing confusions for designers. The current research proposed such a method in which it is argued that the performance of experts should be considered as the main informational resource for design. And it is shown how the videorecording is the medium that can supply a rich and direct qualitative information of skills in performance.

The videorecording method as standardized in the article can be used for various goals in HFE. For example, the model of expert performance can be used for evaluating existing training-simulators. More importantly, we can use a similar videorecording method in studying the experience of trainees, as the video allows us to see what elements of a simulator are guiding trainees' attention, and what can help or distract them during practice. As such, future methods and techniques derived from the same approach can lead to addressing lingering design problems such as the fidelity requirements in the interface of training-simulators. The hope is to see more examples of design methods in future that would improve upon the experiences of the existing knowledge in the field.

**References**

Advance Driving School. (2016, June 14). *Perfect hill starts in a manual car | Learn to drive: Basic skills* [Video]. YouTube. https://www.youtube.com/watch?v=bSPFbZm-oiw&list=PLWLUNYkDgcb842smCw6k-Csa7j6TCxLzb&index=11

Aiello, J. R., & Kolb, K. J. (1995). Electronic performance monitoring and social context: Impact on productivity and stress. *Journal of Applied Psychology, 80*(3), 339–353.

Alliger, G. M., Beard, R., Bennett, Q., & Colegrove, C. (2012). Understanding mission essential competencies as a job analysis method. In M. A. Wilson, W. Bennett, S. Gibson, and G. M. Alliger (eds), *The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations*. New York, NY: Routledge.

Annett, J. (2004). Hierarchical task analysis. In N.A. Stanton, A. Hedge, K. Brookhuis, E. Salas & H. Hendrick (Eds), *Handbook of Human Factors and Ergonomics Methods*. Boca Raton, FL: CRC Press.

Annett, J., & Duncan, K. D. (1967). Task analysis and training design. *Occupational Psychology, 41*, 211–221.

Beek, P. J., & Turvey, M. T. (1992). Temporal patterning in cascade juggling. *Journal of Experimental Psychology: Human Perception and Performance, 18*(4), 934–947.

Blomberg, J., Giacomi, J., Mosher, A., & Swenton-Wall, P. (1993). Ethnographic field methods and their relation to design. In D. Schuler & A. Namioka (Eds.), *Participatory design: Principles and practices* (pp. 123–155). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cannon-Bowers, J. A., Burns, J. J., Salas, E., & Pruitt, J. S. (1998). Advanced technology in scenario-based training. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 365–374). American Psychological Association.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Boca Raton, FL: CRC Press.

Carroll, J. M., & Olson, J. R. (1988). Mental models in human-computer interaction: Research issues about what the user of software knows. In M. Helander (Ed.), *The Handbook of Human-Computer Interaction* (pp. 45–65). Amsterdam, Netherland: North Holland.

CDLTestTruck. (2011, December 19). *Class A CDL road test Texas 469-332-7188* [Video]. YouTube. https://www.youtube.com/watch?v=MqK8jKSDCGs

CDLTestTruck. (2012, January 15). *Class A CDL road test Garland Texas 469-332-7188* [Video]. YouTube. https://www.youtube.com/watch?v=In6tR9Zm5Ig

Clancey, W. J. (2006). Observation of work practices in natural settings. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 127–145). Cambridge, UK: Cambridge University Press.

Collins, H. (2010). *Tacit and explicit knowledge.* Chicago, IL: University of Chicago Press.

Cream, B. W., Eggemeier, F. T., & Klein, G. A. (1978). A strategy for the development of training devices. *Human Factors, 20*(2), 145–158.

Davids, K., Button, C., & Bennett, S. (2008). *Dynamics of skill acquisition: A constraints-led approach*. Champaign, IL: Human Kinetics.

de Groot, A. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.

Diaper, D. (2004). Understanding task analysis for human-computer interaction. In D. Diaper & N. A. Stanton (eds), *The handbook of task analysis for human-computer interaction* (pp. 5–47). Mahwah, NJ: Lawrence Erlbaum Associates.

Duncan, K. D. (1985). Representation of fault-finding problems and development of fault-finding strategies. *Programmed Learning and Educational Technology, 22*(2), 125-131.

Engström, T., & Medbo, P. (1997). Data collection and analysis of manual work using video recording and personal computer techniques. *International Journal of Industrial Ergonomics, 19*(4), 291–298.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8), 725–747.

Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams A. M. (Eds.) (2018). *Cambridge handbook of expertise and expert performance*. Cambridge, UK: Cambridge University Press.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327–358.

Fowlkes, J. E., Neville, K. J., Owens, J. M., & Hafich, A. J. (2009). Challenges to the development of pedagogically driven engineering requirements for complex training systems. *Theoretical Issues in Ergonomics Science, 10*(3), 217–229.

Gilbreth, F. B. (1911). *Motion studies*. New York, NY: Van Nostrand.

Goldman, R., Pea, R., Barron, B., & Derry, S. J. (Eds.). (2014). *Video research in the learning sciences*. New York, NY: Routledge.

Goode, N., Salmon, P. M., & Lenné, M. G. (2013). Simulation-based driver and vehicle crew training: Applications, efficacy and future directions. *Applied Ergonomics, 44*(3), 435–444.

Hall, R. (2000). Videorecording as theory. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 647–664). Mahwah, NJ: Erlbaum.

Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & 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.

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research*. Thousand Oaks, California: Sage Publications.

Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors, 40*(2), 254–276.

Huegel, J. C., Celik, O., Israr, A., & O'Malley, M. K. (2009). Expertise-based performance measures in a virtual training environment. *Presence: Teleoperators and Virtual Environments, 18*(6), 449–467.

Jansson, A., Erlandsson, M., & Axelsson, A. (2015). Collegial verbalisation–the value of an independent observer: an ecological approach. *Theoretical Issues in Ergonomics Science, 16*(5), 474–494.

John, B. E., & Kieras, D. E. (1996). Using GOMS for user interface design and evaluation: Which technique? *ACM Transactions on Computer-Human Interaction, 3*(4), 287–319.

Jordan, B., & Henderson, H. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences 4*(1), 39–103.

Kirwan, B., & Ainsworth, L.K. (1992). *A guide to task analysis*. London, UK: Taylor and Francis.

Klein, G., & Borders, J. (2016). The ShadowBox approach to cognitive skills training: An empirical evaluation. *Journal of Cognitive Engineering and Decision Making, 10*(3), 268–280.

Logan, M. (2016, December 18). *Class A CDL skills test "Road Test" video #13* [Video]. YouTube. https://www.youtube.com/watch?v=LA6t4ys0MHk

Loukas, C. (2018). Video content analysis of surgical procedures. *Surgical Endoscopy, 32*(2), 553–568.

Mahmood, T., & Darzi, A. (2004). The learning curve for a colonoscopy simulator in the absence of any feedback: No feedback, no learning. *Surgical Endoscopy and Other Interventional Techniques, 18*(8), 1224–1230.

Mathiassen, S. E., Liv, P., & Wahlström, J. (2013). Cost-efficient measurement strategies for posture observations based on video recordings. *Applied Ergonomics, 44*(4), 609–617.

Militello, L. G., & Hutton, R. J. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*(11), 1618–1641.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259.

Petersen, E. S., Nyce, J. M., & Lützhöft, M. (2011). Ethnography re-engineered: the two tribes problem. *Theoretical Issues in Ergonomics Science*, *12*(6), 496–509.

Pink, S. (2013). *Doing visual ethnography*. Thousand Oaks, CA: Sage Publications.

Polanyi, M. (1966). *The tacit dimension*. Gloucester, MA: Peter Smith.

Proctor, R. W., & Vu, K. L. (2006). Laboratory studies of training, skill acquisition, and retention of performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The cambridge handbook of expertise and expert performance* (pp. 265–286). Cambridge, UK: Cambridge University Press.

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.

Robinson, W. L. (1974). Conscious competency—The mark of a competent instructor. *The Personnel Journal—Baltimore, 53*, 538–539.Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest, 13*(2), 74-101.

Salas, E., Wilson, K. A., Priest, H. A., & Guthrie, J. W. (2006). Design, delivery, and evaluation of training systems. In G. Salvendy (Ed), *Handbook of human factors and ergonomics* (pp. 472–512). Hoboken, NJ: Wiley and Sons.

Scardamalia, M., & Bereiter, C. (1991). Literate expertise. In K. A. Ericsson, J. Smith (Eds.), *Toward a General Theory of Expertise. Prospects and Limits* (172–194). Cambridge, UK: Cambridge University Press.

Schraagen, J. M. (2009). Designing training for professionals based on subject matter experts and cognitive task analysis. In K. Ericsson (Ed.), *Development of professional expertise* (pp. 157–179). New York, NY: Cambridge University Press.

Sigrist, R., Rauter, G., Riener, R., & Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review, 20*(1), 21–53.

Shih, H. C. (2017). A survey of content-aware video analysis for sports. *IEEE Transactions on Circuits and Systems for Video Technology, 28*(5), 1212–1231.

Smode, A. F. (1974). Recent developments in instructor station design and utilization for flight simulators. *Human Factors, 16*(1), 1–18.

Stanton, N. A. (2006). Hierarchical task analysis: Developments, applications, and extensions. *Applied Ergonomics, 37*(1), 55–79.

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2005/2018). *Human factors methods: A practical guide for engineering and design*. New York, NY: Routledge.

Sülzenbrück, S., & Heuer, H. (2011). Type of visual feedback during practice influences the precision of the acquired internal model of a complex visuo-motor transformation. *Ergonomics, 54*(1), 34–46.

Van Merrionboer, J. J. G., & Boot, E. W. (2009). Research on past and current training in professional domains: the emerging need for a paradigm shift. In K. A. Ericsson (Ed.), *Development of professional expertise* (pp. 131–156). Cambridge, UK: Cambridge University Press.

Veterans Health Administration. (2009, July 1). *Truck driver training* [Video]. YouTube. https://www.youtube.com/watch?v=M6-qITME420

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Walsh, C. M., Ling, S. C., Wang, C. S., & Carnahan, H. (2009). Concurrent versus terminal feedback: It may be better to wait. *Academic Medicine*, *84*(10), S54–S57.

West Side Transport. (2016, June 8). *Road test* [Video]. YouTube. https://www.youtube.com/watch?v=IKhXtrinLvo

Wickens, C. D., Hutchins, S., Carolan, T., & Cumming, J. (2013). Effectiveness of part-task training and increasing-difficulty training strategies: A meta-analysis approach. *Human Factors, 55*(2), 461–470.