Between- and Within-Subjects Experiences with Desktop Simulations

Robert Hubal
RTI International
3040 Cornwallis Road, RTP, NC, 27709-2194 USA
rhubal@rti.org

ABSTRACT
This paper presents observations made during studies of usability, user acceptance, and effectiveness of a range of desktop simulations. Among the topics covered are users' buy-in into the context of the simulations (acceptance and engagement), the way that users act within simulations (their means of interactions), how multiple users *inter*act while using simulations (their interactive behaviors), and how these behaviors change over time.

Author Keywords
Acceptance, desktop simulation, embodied conversational agents, engagement, usability.

ACM Classification Keywords
H.5.1: Multimedia Information Systems; H.5.2: User Interfaces – Graphical user interfaces (GUI), Interaction styles (e.g., commands, menus, forms, direct manipulation).

INTRODUCTION
During the past dozen years the author and colleagues at the author's institution developed several series of desktop training, assessment, and marketing simulations. All of these simulations were interactive 3D, requiring active participation by the user that influenced the simulation flow, hence attaining Level IV interactivity as defined by [12]. By *desktop* simulation is meant that all of these simulations ran on a personal computer with no peripheral devices (projection or head mounted display, special controller) demanded.

Most of these simulations were fielded and their use was observed. Among the many variables of interest during the observations were the experiences with these desktop simulations across users as well as for given users across time. Lessons learned from such observations form the content of this paper.

DESKTOP SIMULATIONS
Descriptions of many of these desktop simulations are given in published reports of usability, acceptance, and effectiveness analyses [1,4,6-10]. Very briefly (but also see the Acknowledgements section below), the simulations were of three types, dealing either with procedural skills or with interaction skills, or having a marketing focus. For instance, one series of procedural skills training simulations focused on U.S. Army maintenance technicians and their need to learn to troubleshoot malfunctions in ground tracked vehicles. Desktop simulations were fielded covering a number of vehicle variants and upgrades to equipment on the vehicles. Meanwhile, a second series of simulations addressed interpersonal skills such as interviewing, de-escalating, and social competency. All of these kinds of simulations employed embodied conversational agents (ECA) [3]. Lastly, several simulations were developed for (mostly) commercial clients to market innovation, to run conjoint analyses, or to disseminate product information.

The remainder of this paper is organized into two sections, with some overlap, addressing the two focus areas of the workshop: Differences among users (*"between-subjects"*) in how the desktop simulations were experienced, and differences for individual users across time (*"within-subjects"*) in their experiences. Each section is organized by experiential measures of acceptance, engagement, means of interaction, and interaction behaviors.

BETWEEN-SUBJECTS
The following are observations of users' general experiences with these desktop simulations.

User Characteristics
The range of users was tremendous. For those simulations that were developed for the U.S. Army, users tended to be younger than twenty years and male, ethnically diverse, generally but not always having some computing or gaming experience. For the other simulations users ranged in age from fifteen to fifty, represented both genders, were doctors and law enforcement officers and at-risk youth and prisoners and college students and tradeshow attendees, experienced the simulation in a classroom or clinical setting or public space, and stayed within the experience for anywhere from one to thirty minutes. Their goals were to learn about items or respond appropriately to provocative situations or seek information or answer questions or demonstrate other specific skills; but for a

IUI Workshop on Users' Preferences Regarding Intelligent User Interfaces: Differences among Users and Changes over Time, February 8, 2009, Sanibel Island, FL, USA.

given user experiencing a given simulation in a given usage situation, that user's goals did not change.

Acceptance & Engagement

Users nearly universally enjoyed the experiences, so it has been instructive to understand when that was not the case. It turns out a user's expectations prior to experiencing a simulation influenced his or her acceptance and engagement. When it was not clear to a user what his or her role was to be during an experience, or when the purpose of the usage situation was not clearly explained, then that user tended to find the experience less favorable. To a lesser extent, and somewhat conflated with prior expectations, a user's computing or gaming experience also influenced acceptance and engagement, in both directions. When a user with computing or gaming experience was led to believe that the simulation would meet his or her fidelity expectations, and the simulation did not meet them, then that user tended to find the experience unfavorable. On the other hand, when a user with computer or gaming experience understood constraints and limitations imposed both within games and within these simulations, and the simulation still attained a reasonable (defined by that user) fidelity, then that user tended to find the experience quite favorable.

There were instances when users' acceptance and engagement strongly changed, namely when entities within the simulation behaved unexpectedly. For instance, in an application where a law enforcement officer was learning to manage an encounter with a schizophrenic individual (represented by an ECA), if the ECA was animated into a strange position (e.g., walking through a bench – an inadvertent graphical error within the simulation) then this behavior broke the user's sense of presence and adversely influenced measures of engagement. In contrast, in an application where a tradeshow attendee asked booth-specific questions of a virtual tradeshow attendant (also represented by an ECA), when the user posed what s/he considered a difficult question and when that question was still answered correctly by the ECA, the user's acceptance of the technology greatly increased.

Means of Interaction

Given that there were many types of tasks for users to accomplish across the simulations, it follows that there were many forms of interaction to enable those tasks. This section describes some of those forms as they cut across simulations.

Navigation

Several of the simulations required the user to navigate about a virtual environment. For instance, in learning troubleshooting skills on a vehicle, it is usually necessary to check status displays of a variety of different pieces of equipment. The troubleshooting simulations provided multiple navigation methods, including free navigation using the mouse and/or keyboard (employed by users who had already run through simulation lessons or who had gaming experience in navigating 3D worlds), controlled navigation using icons or item lists having preset viewpoints (useful for users who might have known what procedure to perform next but were not sure where to go to perform the procedure), and forced navigation during lessons where the user was still acquiring knowledge and skills. All of these methods were used as well in other desktop simulations, as appropriate. So in a pair of driving simulations, users manipulated either arrow keys or an appended gaming steering wheel to stay within the lanes of a road, while in a pair of simulations meant for training inspection skills (i.e., to learn to identify what kinds of discrepant objects or object components might indicate problems to be reported), users navigated between set viewpoints by means of mouse clicking and within 360° views at each viewpoint by means of mouse dragging.

Users nearly universally had little difficulty with the means of navigation. Indeed, even early studies [5] with relatively low-fidelity virtual environments suggested that users learned spatial and relational distances in the virtual environments that transferred readily to live environments. What users could not learn from the virtual environments was scale; that was best learned in the live environment.

Interaction with objects

The simulations involved different types of interactions with objects and entities in the virtual environment. Again for the maintenance and operations trainers, where the objects of interest are pieces of equipment having knobs, switches, dials, and pushbuttons, user interaction was mouse controlled, with changeable cursors used to represent different tools the user might employ (e.g., a wrench, a probe, or a hand). Not dissimilarly for the series of applications involving dialog and basic (virtual) physical interactions between a clinician (user) and virtual patient, user interaction was mouse controlled, with changeable cursors used to represent the different tools (e.g., an otoscope, a stethoscope, and a hand). Both series of applications required that the user employ the tool correctly, so that, for instance, a hand could not loosen a bolt or, similarly, a stethoscope placed on the shoulder rather than chest produced no sound. Not any of these simulations, it is important to point out, was meant to teach the actual physical skills involved in the use of these tools; the belief is those skills are best acquired in mock-up or real-life situations [5]. Rather, the simulations were intended to teach the procedures to follow so that time would not need to be taken in the mock-up or real-life environment to learn those procedures.

Dialog

Some of the simulations involved dialog between the user and one or more ECA. Different simulations used different means for enabling this dialog. One simulation, intended for first responders to demonstrate skills at trauma patient assessment, used pull-down menus for the user to select dialog inputs to the ECA, who responded with computer-generated speech. Another simulation, meanwhile, the one intended for law enforcement officers managing an encounter with a schizophrenic individual, used natural language input that was interpreted on-the-fly by the ECA. Other simulations used typed free text that was also interpreted dynamically. Finally, a trio of simulations intended to provide realistic situations to assess social competency skills used a wizard-of-Oz approach where the user spoke naturally with the ECA but an experimenter immediately coded the input and sent that code to the ECA to drive its next responses.

Only limited data from these observations suggest how to implement dialog for different users, although the focus for these simulations was more on tailoring the application content rather than tailoring to different types of users. Some users immediately began conversing with the ECA, suggesting an immediate buy-in into the simulation, whereas other users showed considerable wariness that dissipated only as they became convinced of the reasonableness of the ECA responses. In general non-realistic means of dialog input (selection from a pull-down and typing in text) were less accepted and required longer to achieve engagement. When the natural language input functioned well, users quickly became engaged. For this reason, the situated assessment simulations (those intended to gauge social competency) employed a human-in-the-loop, obviously necessitating extra effort on the part of researchers using the simulations but ensuring the perceived (by the user) valid functioning of natural language input.

Interaction Behaviors

Users' actions while engaged in the experiences with the simulations were measured. Two types of actions demonstrated engagement across users. First, as is the case with most forms of media that encourage a *presence* [11], when experiencing most of these desktop simulations users exhibited gestures and verbal and facial expressions that indicated engagement and emotional involvement. This was true of driving simulators where users leaned while steering to stay within lanes, of interaction skills simulators where users sometimes yelled or expressed exasperation at the ECA, and of signal systems operations simulators where users became frustrated when the simulated equipment did not perform as they expected (that is, at learning moments when the equipment performed correctly but the users' expectations of performance were incorrect).

Second, when allowed, users grouped together to jointly experience situations or point out previously noticed interesting cases of simulation flow to each other. Whenever simulations were experienced in classroom environments, for instance, users having trouble learning or demonstrating skills often received assistance from classmates, who pointed to discrepant or important objects, redirected simulation flow, and identified key decision points. Similarly, when having completed required lessons and when given time, groups of users often re-ran the simulations and attempted to push the boundaries to find where the simulations break down. These actions were perceived as positive, as they all suggest that users became engaged in the simulations, and in some cases reflected users' beliefs about how the simulations could best be used.

WITHIN-SUBJECTS

The following are observations of individual users' continued experiences with these desktop simulations.

User Characteristics

In far fewer instances, compared to one-time usages, were users able to use the simulations repeatedly. The observations in this section, then, rely on a smaller sample of users than do the between-subjects observations. Overall, the observations do not reflect any changes in users' goals, needs, or usage situations, though they do reflect changes in users' capabilities and attitudes.

For the situated assessments of social competency, users experienced three different situations each at two times (a pre- and post-intervention design). These users were both at-risk youth and a prison population. Similarly for one of the driving simulations, users were asked to run all five levels of the simulation twice. These users were law enforcement officers. Finally, the users who experienced the maintenance training and signal systems operations simulations were soldiers required to run through all lessons to pass a course. Because the simulations could have scores of lessons, the experiences of these users is perhaps most relevant to a within-subjects discussion.

Acceptance & Engagement

The assessments for social competency and driving were conducted separated in time, hence the main observation that can be made regarding users' cross-time experiences is that users remained engaged in the simulations. Performance as well as other measures of user behavior (e.g., gestures) generally improved between assessments.

The users running through the many lessons of their respective maintenance training or signal systems operations simulations had distinctive characteristics. Specifically, as soldiers, they were required to follow orders, and since one order was to complete the lessons, their acceptance of the simulations in a sense was not important. Further, any novelty of and unfamiliarity with using the simulations wore off relatively quickly; even users with little computing or gaming experience became adept at the basics of navigation and interaction within a short time. Still, these users' actions while using the simulations suggested that most, though not all, found the simulations continually engaging. Post-hoc surveys of these users revealed beliefs that the simulations were realistic, were well-designed, and had merit for training.

Means of Interaction and Interaction Behaviors

As would be expected, users' ability to navigate within the virtual environments and their interactions with objects in the environments improved over time. No formal measures were taken of users' means of interaction, but informal observations suggest first, that once a user became familiar with useful techniques, such as using keyboard shortcuts or preset viewpoints, then he or she stuck with those techniques, and second, if a user was shown a more efficient technique, then that user incorporated the new technique into his or her experience with the simulation.

Users' interaction behaviors, too, tended to become less informative. Unlike a game where different levels present challenges that force the player into continued involvement [2], these simulations involved slightly different content from lesson to lesson but not overtly different challenges. Thus, in a way, the users' less overt actions suggested continued acceptance of the simulation; they grew to view the simulation and its capabilities and limitations as a given, just not as engaging as upon the first experience.

Dialog interactions were not present in the maintenance training and signal systems operations simulations. The only observations of cross-time dialog interactions came from users who were testers of some of the ECA-based applications. In general, those users' dialog interactions became more concise and clearly worded, regardless of the means used (e.g., spoken vs. typed text). However, much of this conciseness could be attributed to these users' familiarity with the simulation, hence their understanding of exactly what dialog was needed to appropriately influence the flow of the simulation. Furthermore, given that they were testers, these users' acceptance of the simulation as realistic and useful became unremarkable, so that their dialog interactions would be observed to be of interest only when they were pushing the boundaries of the simulation to try to make it break down.

IMPLICATIONS

The goal of the workshop is to address designers' and evaluators' understanding of between-subjects and within-subjects factors that can influence use and acceptance of intelligent user interfaces. Though far from complete, the observations made in this paper from a wide range of desktop simulations suggest specific lessons learned that could be of use to designers and evaluators.

The simulation should be designed, first, with the range of capabilities of users in mind. Aspects of the simulation like navigation and interaction, when possible, should handle multiple methods, so that users can choose how they wish to engage with the simulation initially and as they learn its affordances. The simulation should also present realistic displays and use of tools, perhaps by using changeable cursors, to more accurately reflect what actions would be taken in the real world. Next, the simulation should break down gracefully, since obvious mistakes can immediately and permanently affect users' sense of presence. It is okay for the simulation to break down, and for users to find the simulation's limitations, so long as the limitations are understood and require active seeking on the part of users rather than normal use. Next, the simulation should if possible turn users into proponents, for instance by advocating its use or by providing a means for some users to assist other users. (This recommendation, it should be noted, necessarily depends not only on the simulation but also on the environment in which the simulation is used, such as a work environment where initial users can promote its adoption or in a classroom environment where proficient students can help less proficient students.) Last, the simulation should be built on aspects of user interface design that other researchers and developers have shown to work. In that regard, the simulations presented here have had the benefit of following each other, and lessons learned in developing each simulation have informed later developments.

ACKNOWLEDGMENTS

This paper focuses on desktop simulations that were developed over the past some twelve years, and represents user studies that have occurred over that time. As such the work has depended on very many former and current colleagues, both at RTI and at partner institutions.

Brief outlines of the variety of simulations follow:

- *Driving simulations*. One simulation, tested with law enforcement officers and others, measured field-of-view and ability to respond to increasingly complex levels of stimulus demands within a driving task. Another simulation was geared towards pre-teens (and thus future drivers) and demonstrated effects of driving while impaired.
- *Conjoint analysis*. Early simulations were developed for commercial clients to enable designers to gauge consumers' interests by having consumers compare versions (i.e., simulated prototypes) of their products. A recent simulation simulates shopping to collect information on consumer purchase decisions.
- *Interactive video*. Simulations built for military and commercial clients used interactive videos to enable ship personnel and aircraft baggage handlers to learn to navigate a given environment, to identify discrepant features of objects in the environment, and to learn the steps of standard procedures (e.g., inspection, loading) done in that environment.
- *Maintenance trainers*. A series of diagnostic trainers for maintenance training, developed for the U.S. Army, were intended for maintenance personnel to learn to diagnose and troubleshoot ground tracked vehicle faults and failures.
- *Distributed signal systems operations trainers*. A separate series of setup/operations trainers developed for the U.S. Army were for signal operations and maintenance personnel for assignment-oriented and

sustainment training involving signal systems and shelters.

- *Marketing usage of ECA.* Rocky and Roxanne, virtual tradeshow attendants, were fielded at several tradeshows and conference exhibits to answer basic questions that a booth attendant would answer.
- *Clinical usages of ECA.* Simulations involving ECA were designed to enable survey and medical researchers to learn standard responses to questions of informed consent; to provide virtual role plays for law enforcement officers to learn to manage encounters with mentally ill individuals; to provide an environment where clinicians could learn good practices for eliciting information from and establishing rapport with pediatric patients and learn proper techniques to make accurate differential diagnoses of rare illnesses; to assess adolescents' and prisoners' social competency skills in risky situations; and for recovering pediatric patients to demonstrate their perceptions of emotional expressions.

REFERENCES

1. Armsby, P.P., Link, M.W., Hubal, R., Guinn, C.I., Flicker, L., & Caspar, R. (2003). Accessibility and acceptance of a virtual respondent-based interviewer training application. *Survey and Statistical Computing IV. The Impact of Technology on the Survey Process*. Chesham, UK: ASC.
2. (2005). Engagement and motivation in games development processes. British Educational Communications and Technology Agency (BECTA): Coventry, U.K.
3. Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.) (2000). *Embodied conversational agents*. Cambridge, MA: MIT Press.
4. Evens, N., Whiteford, B., Frank, G., & Hubal, R. (2006). User interface lessons learned from distributed simulations. *Proc. I/ITSEC* (pp. 1276-1285). Arlington, VA: NDIA.
5. Frank, G.A., Helms, R.F., & Voor, D. (2000). Determining the right mix of live, virtual, and constructive training. *Proc. I/ITSEC* (pp. 1268-1277). Arlington, VA: NDIA.
6. Guinn, C., & Hubal, R. (2004). An evaluation of virtual human technology in informational kiosks. *Proc. ICMI* (pp. 297-302). New York, NY: ACM Press.
7. Guinn, C., Hubal, R., Frank, G., Schwetzke, H., Zimmer, J., Backus, S., Deterding, R., Link, M., Armsby, P., Caspar, R., Flicker, L., Visscher, W., Meehan, A., & Zelon, H. (2004). Usability and acceptability studies of conversational virtual human technology. *Proc. SIGdial Workshop on Discourse and Dialogue* (pp. 1-8). East Stroudsburg, PA: ACL.
8. Hubal, R. (2005). Design and usability of military maintenance skills simulation training systems. *Proc. HFES* (pp. 2110-2114). Santa Monica, CA: HFES.
9. Hubal, R., & Frank, G. (2007). Lessons learned in embodying tutoring for interactive skills. *Proc. Int'l Workshop on Virtual Instructors*. Washington, DC: IEEE VIPRG.
10. Hubal, R.C., Fishbein, D.H., Sheppard, M.S, Paschall, M.J., Eldreth, D.L., & Hyde, C.T. (2008). How do varied populations interact with embodied conversational agents? Findings from inner-city adolescents and prisoners. *Comp. Hum. Beh., 24*(3), 1104-1138.
11. Lee, K.M. (2006). Presence, explicated. *Comm. Theory, 14*(1), 27-50.
12. U.S. Army Training and Doctrine Command. (2001). TRADOC Pamphlet 350-70-2: Multimedia courseware development guide. Ft. Monroe, VA.