Development of Cognitive Transfer Tasks for Virtual Environments and Applications for Adaptive Instructional Systems

Anne M. Sinatra[1(✉)], Ashley H. Oiknine[2,3], Debbie Patton[4], Mark Ericson[4], Antony D. Passaro[4], Benjamin T. Files[4], Bianca Dalangin[2], Peter Khooshabeh[4], and Kimberly A. Pollard[4]

[1] US Army Combat Capabilities Development Command Soldier Center – Simulation and Training Technology Center, Orlando, USA
anne.m.sinatra.civ@mail.mil

[2] Department of Psychological and Brain Sciences, University of California, Santa Barbara, USA
biancaydalangin@gmail.com

[3] DCS Corporation, Los Angeles, USA
aoiknine@dcscorp.com

[4] US Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground, Los Angeles, USA
{debra.j.patton4.civ, mark.a.ericson.civ, antony.d.passaro.civ, benjamin.t.files.civ, peter.khooshabehadeh2.civ, kimberly.a.pollard.civ}@mail.mil

Abstract. Spatial navigation and spatial learning are important skills with implications for many different fields, and can be trained in Virtual Environments (VEs). Research in these areas presents a critical challenge: how to best design tests of knowledge transfer to measure learning from the VEs. Additionally, spatial learning tasks and VEs can be implemented in Adaptive Instructional Systems (AISs), which could result in improved spatial navigation and learning. The current paper describes the development of transfer tasks to assess different levels of spatial learning from experience in a VE, including recall, recognition, route, bearings, and map drawing transfer tests. Details are provided in regard to how transfer questions were generated and the reasoning behind them. Implications for the use of similar tasks in AISs and Intelligent Tutoring Systems are discussed.

Keywords: Spatial navigation · Virtual environments · Transfer tasks · Adaptive Instructional Systems · Spatial learning

1 Introduction to Spatial Navigation in Virtual Environments

1.1 Introduction

Spatial navigation and spatial memory are important skills for a variety of occupations as well as for everyday living. Spatial learning is frequently studied in an experimental setting utilizing virtual environments (VEs) and immersive technologies, such as

R. A. Sottilare and J. Schwarz (Eds.): HCII 2019, LNCS 11597, pp. 340–351, 2019.
https://doi.org/10.1007/978-3-030-22341-0_27

Virtual Reality (VR) head mounted displays (HMDs), to display spatial tasks. An important challenge in spatial navigation research is how to design tests of learning transfer from the immersive VEs. By learning to navigate a VE there is the assumption that the navigational skills learned in the real world environments will be improved [3, 4].

Measures of transfer of learning are especially important for automated systems. Many Intelligent Tutoring Systems (ITSs) and Adaptive Instructional Systems (AISs) are in well-defined domains (e.g., math and physics), and few have been integrated with real-time feedback in external games and virtual environments [14]. However, with the development of domain-independent ITS frameworks such as the Generalized Intelligent Framework for Tutoring (GIFT), there are opportunities for integrating custom VEs with adaptive tutoring and after-the-fact assessment. Since spatial learning can involve interacting with a VE and meeting measurable goals, it is a natural fit for intelligent tutoring and incorporation into AISs.

1.2 The Current Environment and Applications for AISs

The current paper describes and details the development of experimental transfer tasks and assessment materials associated with an experiment to investigate the relationship among immersion, individual differences, and training transfer in VEs. The primary method used for determining transfer performance in the described experiment was based on traditional cognitive psychology approaches and measures. This paper describes the methods that were used to develop the transfer tasks that were associated with the VEs in order to examine learning as a result of engaging with the VEs, and provides an initial explanation of approaches that can be used to grade them. The implications for adapting these types of materials and approaches for AISs are discussed, as well as lessons learned from working as a distributed team to generate these materials.

2 Generating Transfer Tasks for a Virtual Environment

2.1 General Experiment Design and Task Description

In order to examine the relationship between immersion, individual differences and training transfer, a study was designed that had three levels of immersion: (i) low: desktop monitor with desktop speakers, (ii) medium: partially occluded HMD with headphones, and (iii) high: fully occluded Oculus Rift HMD with circumaural headphones. The visual fidelity of the devices used in each level increased as the associated immersion level went up (e.g., the Oculus had higher visual fidelity than the partially occluded HMD). Additionally, the fidelity of the associated audio increased with the level of immersion. The low immersion audio included only distance-based intensity cues played over loudspeakers. The medium immersion audio condition used directional audio cues based on free-field head-related transfer functions without head motion cues. The high immersion condition used head motion tracked directional cues and room acoustic cues using Oculus' 3D audio spatialization effects from their audio Software Development Kit.

Three different mini and three different larger Unity3D-based environments were created for use in the experiment. The objects used within the environments included Unity "free assets", and objects developed by the research team. The mini-environments were intended to serve as an introduction and practice environment before engaging in the full task and were highly tied to schemas. The mini-environments were: (i) Holiday environment (items included a pot of gold at the end of a rainbow, four leaf clover, and a Christmas tree), (ii) Museum environment (items included a dinosaur skeleton and a knight's armor), and (iii) Recreation Room environment (items included a dart board and a chess board). Each of these mini-environments included items that were appropriate for the theme assigned to the environment and could be used for assessment in the transfer tasks. Three larger main environments were also created and had more traditional items of interest in them. The main environments were (i) House (items included an alarm clock and a bed), (ii) Office (items included a conference table and a vending machine), and (iii) School (items included a chalkboard and an abacus). Each environment featured a scavenger hunt; there were 8 target scavenger hunt items in each main environment. Additionally, there were a number of incidental non-target items that were included in each environment, with the numbers being as consistent as possible between environments. Environments and immersion levels were counterbalanced. The study had a within subjects design, with participants returning after no less than 14 days for each additional session to minimize for carry-over effects. Therefore, all participants who completed each of the three sessions experienced all environments and immersive technologies. As an example of an environment layout, an overhead view of the Office environment is shown in Fig. 1.

Fig. 1. Overhead screenshot of the Office environment with researcher-generated room labels (labels not shown to participants). Different items can be viewed in each room, such as a coat rack, treadmill, conference table, and copy machine.

2.2 Task Description

In the experiment, the participants were tasked with navigating through the environment and conducting a scavenger hunt for 8 target objects that each had numbered flags on top of them. In every condition, the participants navigated with an Xbox One controller. They were asked to look serially for a target item (e.g., "Find item number 1.") with the given numbered flag instead of the item name so as to not introduce any semantic information about the item to the participant. This was to ensure that when participants engaged with the transfer tests they were remembering the environment that they navigated and not information or instructions that they were told. When a target item was found, the participant would indicate it by pressing a button on the controller, and the system would record it and ask them to find the next item number. Each main environment had an equal number of rooms (13), and equal number of target items (8) and a similar number of incidental (non-target) items. The participants had 15 min to engage in the scavenger hunt, and if they completed the task prior to this they were asked to explore the environment until the time completed. Careful consideration was given to which items would be considered target items and ensuring that there was no duplication of target items between the environments. For instance, if a sink was a target object in one environment, it was ensured that (a) no other environment had the same sink and (b) no sinks served as target items in other environments. In order to ensure that distinct (i.e. no identical objects) questions could be asked about each environment, the selection of the target items for the scavenger hunt was tightly coupled with the creation of the transfer tasks that would be used for later assessment.

In order to ensure that the environments were equally complex, the number of rooms, size of rooms, and number of doors were the same, and the number of items in each room were counted similarly with high inter-rater agreement, as this would influence the number of items that it would be possible to recall. Further details about the creation of the environments, metrics supporting equal complexity, and the determination of the order of the scavenger hunt is documented by Files et al. [5]. This equivalency ensured that transfer questions were derived from equally complex environments, could be balanced, and that no environment had an obvious advantage or disadvantage.

2.3 Transfer Tasks and Development

The transfer tasks were divided into five types with increasing complexity, based on the level of spatial knowledge processing required, from recall and recognition to more complex and integrative survey knowledge [10]. Types included: free recall, recognition, route description, bearings, and map drawing. All the tasks except the map drawing were completed in a computer-based survey system presented on a Surface Pro tablet. The map drawing was completed using a pen and paper. After the completion of the interaction with the environment, the participants engaged in these activities in the aforementioned order as to not "give away" or activate a schema for the environment. For example, because map drawing requires the most complete processing and integration of the participant's spatial memory, we wanted the participant to complete this last. If they completed the map drawing before the route descriptions, they may be pulling from their memory of their own drawn map to answer the question rather than the memory of their actual experience in the environment. Similarly, because participants were shown photos of objects during the recognition portion, we wanted this to occur after the free recall. Otherwise, participants may have remembered an object from its photo during the recognition transfer task rather than from their memory of their experience in the environment.

Free Recall. First, participants were asked to freely recall any and all objects that were present in the environment. The phrasing of the prompt was carefully worded as “Please list all of the objects you encountered in the virtual environment”, in order to encourage participants to list everything they recalled even if it was not specifically part of the numbered search task. For the mini environment participants had 1.5 min and the opportunity to type up to 25 items. In the main environments participants had 3 min and were able to type in as many as 74 items. These numbers were deemed sufficient based on pre-pilot runs where non-naïve participants were allowed to recall an unlimited number of items during the time limit. After the initial recall, the participants were prompted to indicate, of the objects recalled, which had been target objects in the scavenger hunt, and which objects had included any audio components. These measures indicate how much target and incidental (non-target) knowledge was gained from navigating the environment without providing any cues for recall. It is expected that participants are likely to recall items that were “targets” in the scavenger hunt task, but also that they will have noticed other objects throughout the exploration of the VE. Initial planned approaches for analysis of the recall questions include compiling the number of items listed by the participant; determining if there is accuracy in identifying the characteristics of the objects (e.g., sound; target), and looking for any frequently missed or added items.

Recognition. Next, participants were given a recognition test that was meant to examine the impact of schemas on their knowledge of the environment. Participants were presented with actual images from the environment as well as foil images, and needed to determine if they were in the environment or not. There were two forms of these questions that had different overall intentions. The first type presented an image and asked the participant to indicate with a “yes” or a “no” if the specific object was in the environment. This type of question presented items that could reasonably be in the environment among objects that were actually in the environment. This was to see if participants were relying on their schema for the specific environment rather than direct memory of the environment. This method has been previously used in route learning research in order to determine what information was gleaned from the interaction [15]. The second type of question presented images of four objects and asked the participants to indicate which of the objects was present in the environment. All images in both question types featured an isolated image of the object on a white background to ensure no contextual information would impact response to transfer and allow for a seamless introduction to foils. We included four levels of cases in which these objects were schematically similar (e.g. objects found in a baby’s room), functionally similar (e.g., different objects that tell time), identical but varied in color (e.g. same power drill but in four distinct colors), and cases in which objects differed in small details (e.g., all four were clocks of the same color, shape, and function but varied in number font, artistry, etc.). These questions were designed to determine if specific types of errors were being made, and if they varied systematically across conditions.

In the case of both types of questions (yes vs. no and multiple choice), there were a mix of target and non-target objects included in the questions. This decision was made to distinguish if there is a difference in memory for objects from the task that were being engaged with (target) vs. incidental objects from the environments (non-target). It should be noted that there were no time constraints on the recognition items. Examples of these two types of questions are in Fig. 2.

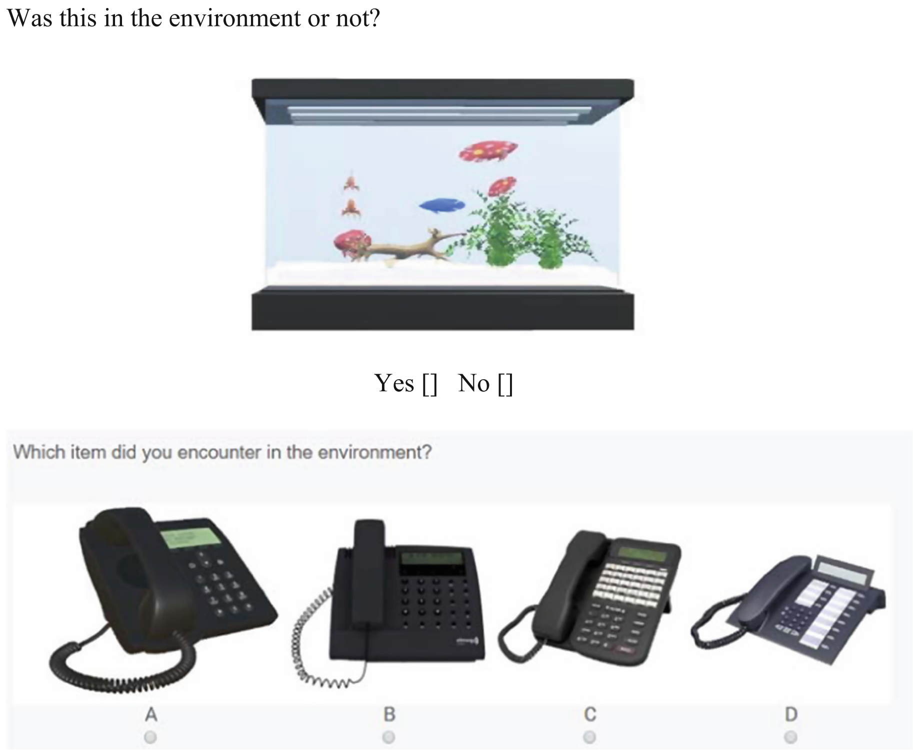

Fig. 2. The top question is an example of recognition question type 1, which was a "yes" or "no" response. The bottom question is an example of recognition question type 2, which presents items that differed in small details and asks for the one in the environment to be identified.

Initial analysis approaches for the first type of recognition questions are expected to involve a count of how many accurate and inaccurate "yes" or "no" responses were given and if there were any differences in accuracy for target or non-target items. For the second type of recognition questions the overall accuracy will be examined, as well as which type of characteristic question results in the greatest number of errors. This will provide additional information about the level of information that was encoded about the objects in the environment and if a particular level of immersion was conducive to a given "resolution" of recall. For example, if some subjects performed better on the low-level specific detail questions for the most immersive condition (HMD) then that would be a case in favor of that level of immersion being most effective for that subset of subjects.

Route Description. After the recall and recognition questions, participants were asked to describe the route between two specified objects. More specifically, participants were asked to imagine giving directions to a person who had never been in the environment before and type in those directions. Below the route prompt, an example of a route description of a commonly known route (to the current testing area from the building lobby) was displayed. The example included the initial orientation in the environment at the starting object (e.g. elevator), directives (e.g. turn right), landmarks (e.g. couch), and goal location (e.g. lab). Once participants reviewed the instructions, they received either 60 or 90 s (mini and main environment respectively) to describe the route in a free response textbox. In choosing which objects to use in these questions, careful consideration was given to the number of rooms that would need to be traversed, and if the participant was being asked to describe a route from target to target, target to non-target, non-target to target, or non-target to non-target. The number of questions about objects of the aforementioned types were kept equal between all conditions. This allows us to make comparisons between memory for scavenger hunt target items and incidental items.

Our preliminary approach to evaluating these descriptions was largely borrowed from the work of Lovelace and colleagues [9]. This scoring protocol includes route quality, landmarks, elaboration, errors, and segments completed, among others. In addition to the route quality ratings from Lovelace et al. [9], we added categories that included interpretations of landmarks, vagueness, and whether participants reported they did not remember. Upon completion of the scoring protocol, we plan to establish interrater reliability among coders on a sub-set of the route descriptions. Once there is agreement, the remaining descriptions will be distributed amongst coders for evaluation.

Bearings. Next, participants were given a test to identify the bearing directions of an object given a specific viewpoint. This was designed to evaluate their spatial understanding of the relationship between objects and the environment. Easy (within the same room), medium (1 or 2 rooms over), and hard (3 rooms over or more) questions were developed, again with careful consideration to the characteristics of the objects that were being considered (e.g., target to non-target; non-target to non-target) and ensuring they were equal across conditions. Participants were told to indicate the direction of a goal object given a starting viewpoint as though they were pointing to an object with a penetrating laser. The analogy of a "penetrating laser" was used to convey that the direction reported should not be influenced by potential obstructions such as walls or other objects. Below these instructions, participants were provided with an example question to familiarize themselves with the question type. The example used the spatial relationship between California, Washington, and Oregon, which should be common knowledge to our participants (living in California). The example prompt asked participants to imagine standing on Oregon facing towards Washington and to indicate at what direction they can find California. Underneath the prompt, the question contained a palm-sized circle with mouse-clickable tick marks every three degrees along the inner circumference. In the middle of the circle was a photo with the object they are to imagine “standing on” and above the circle is the object they are to imagine “facing.” See Fig. 3 for an example of one of the bearings questions.

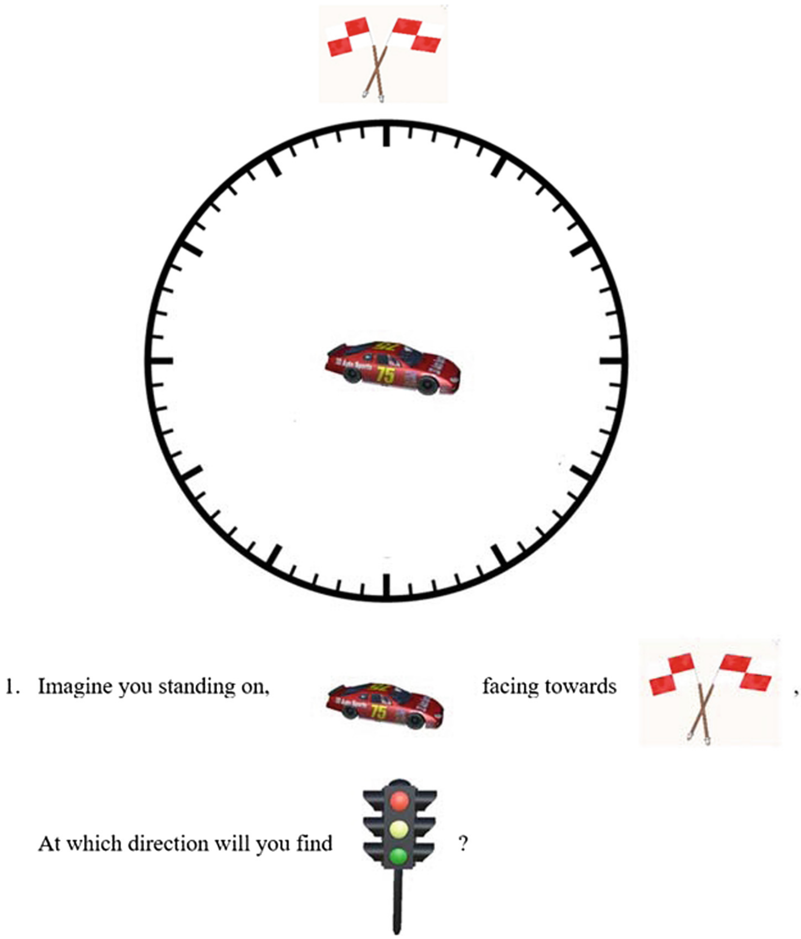

Fig. 3. An example of a Bearings question.

Map Drawing. Finally, participants were given a piece of 8.5 × 11 in. graph paper and an erasable pen. Participants were instructed to draw a floor plan of the environment. They were asked to label each room, refrain from using any label twice, and to indicate doorways with an “X”. This overarching task was developed in order to measure the survey knowledge of the participants and assess how accurate their mental map of the rooms and configurations were as they relate to each other. Further, the participants were asked to label the rooms in order to determine if they had schematized the information. An erasable pen was provided to ensure that participants were able to generate a map that was legible for input into the Gardony Map Drawing Analyzer (GMDA) [6], while also retaining the ability to erase as needed. Participants had either 2 or 5 min to draw the floor plan when drawing the mini or main environment respectively. We plan on using an open source and validated graph analyzer, GMDA, to evaluate how accurate participants' map drawings were [6]. First we will draw boundaries for each room on top of a true map of the environment within the software to serve as the template. Once an accurate template has been created, we will develop a protocol for scanning the participant maps and upload them to the GMDA. Using the same boundary drawing protocol we used for template creation, we will analyze the scanned map drawing for accuracy. The metrics we may use to compare maps include the number of rooms drawn, room position relative to others, scale, rotational bias, and angle bias among others.

2.4 Distinction of Transfer Task Levels and Cognitive Psychology

These five levels of tasks described in detail above were developed to measure spatial knowledge acquisition including spatial-working memory (remembering objects, shapes, and colors in recall and recognition) and survey knowledge (e.g. route descriptions, bearing directions, map drawing). Survey knowledge includes visual, geometrics, relational, emotional, and descriptive information [7], and requires information to be processed more deeply than with simple route or recognition memory [15]. Having these different levels of transfer tasks allows for nuanced comparison of spatial learning that occurred under conditions of different levels of immersion and different individual traits. From a cognitive psychology perspective, the initial spatial-memory questions (item recall, item identification) are representative of a shallow automatic encoding that can be acquired with minimal processing and attention [2, 8]. Whereas, survey knowledge as represented by being able to explain the relative location of items to each other and drawing a map requires more in depth and effortful processing to achieve [2]. Additionally, by asking questions about target items as well as non-target items it can lend insight into the encoding processes that are occurring during the interaction. Comparisons can be done to determine if the same level and type of knowledge was achieved in all conditions, or if certain levels of immersion were more conducive to the development of knowledge of the environment. For instance, if the high immersion condition results in higher performance on the survey knowledge measures it could have important implications for the design and display of VEs.

3 Application of Methodology and Implications for AISs

The described experiment and process used for the development of transfer tasks is highly relevant to the development of AISs. There are multiple relevant lessons learned and processes that were used during the development of the tasks. The team which developed these questions was highly distributed, which resulted in the materials needing to be very organized and highly annotated. Many AISs have teams that are often working in different locations and that need to work collaboratively on the development of the system. Additionally, environments and transfer evaluations, such as the project described, could be incorporated within an intelligent tutoring framework, such as the GIFT [14] to engage in continuous assessment, adaptive tutoring, and individual remediation based on performance. GIFT has many functions that are useful to enhance and conduct psychology experiments [1, 11–13].

3.1 Implications of Examining Learning Transfer in an AIS

The learning transfer tasks that were described can be updated and used in similar training environments to assess the level and types of knowledge that are being gained from an interaction. By considering the distinction between shallow and deep encoding it can lend insight into how much is being learned from interaction with an ITS or system. If it is determined that certain characteristics of ITSs result in deeper knowledge (e.g. the ability to identify bearings or generate a map) then they can be utilized in future ITSs and AISs in general. The specific types of questions that were used in our study are generalizable and can be updated or changed based on the topic or goals of a tutor that is being created. By determining if there is learning transfer after engaging with a VE it can help to improve the design of ITS interventions and provide a trainer with valuable information about what type of knowledge their learners are getting out of engaging with the ITS.

3.2 Application of VEs and Transfer to the Generalized Intelligent Framework for Tutoring

Spatial navigation and learning is an ideal task to incorporate into an AIS using a framework such as GIFT. The VEs that were created for the described experiment were created using Unity3D. GIFT has the ability to interface with external training applications, and a gateway already exists between Unity and GIFT. Therefore, if desired, it would be fairly straightforward to connect a Unity environment such as the one described with GIFT. It would be necessary to ensure that the desired performance and state information (e.g., timing, finding target items during the scavenger hunt) was being communicated from Unity to GIFT. This could then be used to provide adaptive feedback that could be authored based on the desired training outcomes. If an ITS were to be created to assess spatial learning, the approaches that were used for the currently described transfer tasks would be useful. An ITS framework such as GIFT includes a survey authoring system that can be used to create nearly all of the types of questions that were described (with the exception of map drawing). While it would be difficult to leverage the assessment of recall questions in real time, accuracy on recognition and bearings questions could be determined in real time and used to provide recommendation as well as feedback on ways that learners could improve their performance. There would be many opportunities to provide adaptive feedback in real time in a VE, and to additionally store information for post-processing after the fact to inform future training.

3.3 Implications of Working with a Distributed Team

As is the case with many AISs, the research team involved in the described study were not all working in the same location. The group was distributed between the West coast, Southeast, and Mid-Atlantic of the United States, and was comprised of 10+ members. In order to support these many locations, planning meetings occurred over the phone, with some in-person synchronization meetings of the entire team. In order to engage in the actual development of the transfer tasks, a subset of the group met consistently through phone calls in order to come up with an overall plan for the task development. A full list of questions of each type for each environment were developed by the subgroup of researchers and then compiled and edited by one researcher in order to balance the characteristics of items and questions. All of the questions were carefully compared to ensure for consistency between the transfer questions for the different environments. Careful notes and descriptions were included with the final version of the questions, as they were implemented by an additional group of team members. These additional team members collected, modified, and prepared the required images, and one team member formatted and entered the questions into an electronic survey platform. Finally, the team members involved with the development of the transfer questions reviewed the versions that were implemented in the survey platform.

4 Conclusions

The current paper describes the development of transfer tasks to evaluate the types of knowledge that participants gained after navigating in Unity3D-based virtual environments with different levels of immersive technology. The types of questions that were generated and described offer insight into the types of questions that could be generated for use in AISs by instructors and researchers who wish to assess the type of information (shallow vs. deep) that was retained after engaging in a VE. The described environments and tasks could be used as a basis for a spatial navigation tutor using a framework such as GIFT. The approaches for developing the tasks in a distributed team, and for ensuring consistency between experimental conditions offer insight into processes that can be used for future AIS design and experimentation.

Acknowledgements. This work was funded by the US Army Research Laboratory's Human Research and Engineering Directorate. The authors would like to acknowledge Jason Moss who contributed to the conceptualization and development of the experiment. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. Boyce, M.W.: From concept to publication - a successful application of using GIFT from the ground up. In: Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym4), p. 125 (2016)
2. Craik, F.I., Lockhart, R.S.: Levels of processing: a framework for memory research. J. Verbal Learn. Verbal Behav. **11**(6), 671–684 (1972)
3. Darken, R.P., Banker, W.P.: Navigating in natural environments: A virtual environment training transfer study. In: Proceedings of VRAIS 1998, pp. 12–19 (1998)
4. Darken, R.P., Goerger, S.R.: The transfer of strategies from virtual to real environments: an explanation for performance differences? In: Proceedings of Virtual Worlds and Simulation 1999, pp. 159–164 (1999)
5. Files, B.T., Oiknine, A.H., Thomas, J., Khooshabeh, P., Sinatra, A.M., Pollard, K.A.: Same task, different place: Developing novel simulation environments with equivalent task difficulties. In: Applied Human Factors and Ergonomics 2019 Conference (in press)
6. Gardony, A.L., Taylor, H.A., Brunyé, T.T.: Gardony map drawing analyzer: software for quantitative analysis of sketch maps. Behav. Res. Methods **48**(1), 151–177 (2016)
7. Golledge, R.G., Dougherty, V., Bell, S.: Acquiring spatial knowledge: survey versus route-based knowledge in unfamiliar environments. Ann. Assoc. Am. Geogr. **85**(1), 134–158 (1995)
8. Hasher, L., Zacks, R.T.: Automatic and effortful processes in memory. J. Exp. Psychol. Gen. **108**(3), 356 (1979)
9. Lovelace, K.L., Hegarty, M., Montello, D.R.: Elements of Good Route Directions in Familiar and Unfamiliar Environments. In: Freksa, C., Mark, D.M. (eds.) COSIT 1999. LNCS, vol. 1661, pp. 65–82. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48384-5_5
10. Siegel, A.W., White, S.H.: The development of spatial representations of large-scale environments. Adv. Child Dev. Behav. **10**, 9–55 (1975)
11. Sinatra, A.M.: The 2018 research psychologist's guide to GIFT. In: The Proceedings of the Generalized Intelligent Framework for Tutoring (GIFT) User Symposium (GIFTSym6) (2018)
12. Sinatra, A.M.: The updated research psychologist's guide to GIFT. In: Proceedings of the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym4), p. 135 (2016)
13. Sinatra, A.M.: The research psychologist's guide to GIFT. In: Proceedings of the 2nd Annual GIFT Users Symposium, pp. 85–92 (2014)
14. Sottilare, R.A., Brawner, K.W., Sinatra, A.M., Johnston, J.H.: An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT) (2017). GIFTtutoring.org
15. van Asselen, M., Fritschy, E., Postma, A.: The influence of intentional and incidental learning on acquiring spatial knowledge during navigation. Psychol. Res. **70**(2), 151–156 (2006)