



Evaluating instruction quality across narrative modality using measures of real-time cognitive load

S. P. Smith^a , K. Blackmore^a , S. Wark^b and M. Nowina-Krowicki^b

^a*School of Information and Physical Sciences, University of Newcastle, Callaghan, NSW 2308, Australia*

^b*AI for Decision Analytics, Joint Warfare & Operations, Joint & Operations Analysis Division, Defence Science & Technology Group, South Australia*

Email: shamus.smith@newcastle.edu.au

Abstract: Planning and operations tasks in complex environments place high information processing demands on decision makers. Within these environments, decision makers must simultaneously execute an operational plan, developed and enacted under pre-existing contextual information, whilst directing tasks and responding to real-time events. Receiving tasking and updated scenario information, in situ, can place significant cognitive load on decision makers.

In the research presented here, we consider the intersection between information delivery modalities and real-time cognitive load. Three information delivery conditions, namely a control condition containing only text and image-based media (see Figure 1:Right), a virtual human condition, using the Defence Science & Technology Group Virtual Human Storytelling system, combining multiple forms of visual prompts (virtual human avatar, as well as images) as well as audio cues (see Figure 1:Left), and an audio condition combining both image and audio cues only were considered.



Figure 1. Example information delivery modalities. Left: Virtual Human condition. Right: Text condition.

The results are promising for use of the Virtual Human Storytelling system. Across navigation, map annotation and resource allocation tasks, participants with instruction via the virtual human avatar tended to have higher performance scores. These participants also demonstrated better cognitive load profiles when compared to participants who had text or audio-based instruction. The research provides novel and interesting insights into the potential benefits of task instruction delivery for enhanced human performance in complex operating environments.

Keywords: *Decision support, virtual humans, simulation, cognitive load, information delivery*

1 INTRODUCTION

Situational awareness is a key requirement for operators, decision makers, and analysts. In the normal course of their roles this is achieved, in part, by immersion, exploration and manipulation of the information space in order to produce the evidence or products needed to support their actions, decisions, or analysis. This helps establish the context, and determine what is known, what is not known, what is important and what is not important to a particular situation. When automation is introduced to handle the ‘big data’ problems of volume, velocity, and variety, this pathway to situational awareness and understanding is largely lost. Narrative provides a way of establishing context and immersion, and hence re-establish a pathway for understanding.

The work described here extends previous work (Blackmore et al. 2019) and aligns with the Defence Science & Technology Group’s (DSTG) Agile C2 STaR Shot¹. Blackmore et al. (2019) investigated the delivery of narrative content from the DSTG Virtual Human Storytelling system, and measured information recall as a measure of post-task performance. This paper considers the intersection between information delivery modalities and real-time cognitive load. A new study was conducted to explore the performance and cognitive load of participants exposed to different forms of multimedia communication. In contrast to the earlier study where both the instruction and recall measures were passive, the current study required participants to achieve specific tasks in an interactive virtual environment. Instructions were provided at scenario start and followed with updated instructions triggered by events in the virtual environment. In addition to performance metrics gathered during task engagement, participants had their cognitive load measured in real-time.

2 BACKGROUND

2.1 Information Modality and Situation Awareness

Operational planning and command and control (C2) in complex environments places high information processing demands on decision makers. Within these environments, decision makers must simultaneously execute an operational plan, developed and enacted under pre-existing contextual information, whilst directing tasks and responding to real-time events. Within this operating environment, all information can be considered to filter through an Information Communication Layer (ICL).

In this context, the ICL consists of the type, modality, and temporality of information and data that operators use to develop the situation awareness that forms the basis of their decisions and resultant actions (Endsley 2000). Figure 2 provides a high-level conceptual overview of the role of the ICL in tasks, including specific human and task/system factors that may also impact on the overall operating environment.

2.2 Virtual Human Storytellers for Information Delivery

Virtual human storytellers mimic human information delivery and thus incorporate multiple perceptual modalities. While existing research has considered the influence of display type and narrative medium in the context of sensemaking (Hibbard et al. 2018), the potential benefits of virtual human storytellers for delivering contextual information in the form of a narrative requires investigation. In particular, identifying if context information delivered via a virtual human storyteller affects actions/performance in differing tasks (i.e. task dependence) and how real-time cognitive load varies during tasks is of importance.

Cognitive Load Theory (Chandler & Sweller 1991) and Cognitive Theory of Multimedia Learning (Mayer 2002) are used to provide guidelines for effective multimedia tools. These theories suggest that the use of multimedia materials in conjunction with audio visual mediums allows for integration of new information with prior knowledge to construct new knowledge, with effective materials reducing the cognitive load. However, redundant information, or representation of two similar types of information such as two forms of visual cues can result in split-attention and an increase in cognitive load (Craig et al. 2015).

2.3 Cognitive Load

The cognitive architecture of an individual is made up of both long-term memory and short-term memory, known as the working memory. Cognitive load describes how people process information in their working memory (Alharthi et al. 2018). Miller (1956) concluded that an individual can only save a limited number of items (around 5 – 9 items) in their working memory. Thus, the working memory can be overloaded to the

¹ Science, Technology and Research (STaR) Shots are challenging, inspirational and aspirational Science and Technology missions that align strategic research to Defence Force structure priorities. See <https://www.dst.defence.gov.au/strategy/star-shots>.

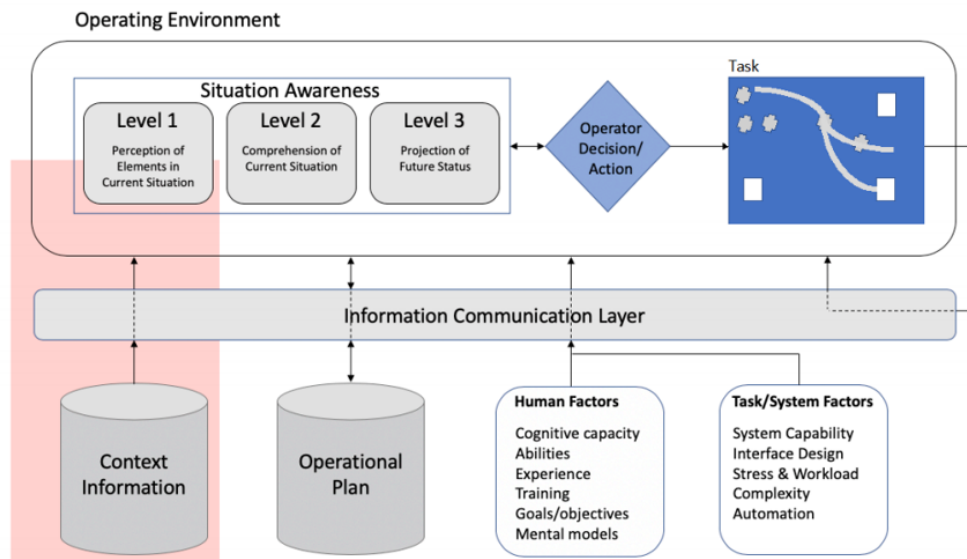


Figure 2. Conceptual model of the Information Communication Layer (ICL) in tasks, showing the focus of this research in pink (Adapted from (Endsley 2000)).

point where an individual can no longer effectively process new information. This is also known as cognitive overload and occurs when too many elements are being stored in the working memory (Chung et al. 2017).

Cognitive Load Theory proposes that people all have different thresholds in relation to the amount of ‘load’ their working memory can take (i.e. variability in the ratio between available and allocated resources). Based on this, some people need to exert more or less effort to complete the same task, compared to someone else (Thorpe et al. 2019). Cognitive load is therefore an important aspect that is used to understand the way an individual’s working memory is being utilised, and how difficult a person may find a given task.

3 METHOD

A user study was conducted following a randomised/control design. Human ethics approval was provided by the University of Newcastle (#H-2015-0163). Written informed consent was obtained from all participants. Unfortunately, COVID-19 significantly impacted the research with user studies periodically suspended during 2020 and 2021. Participants were randomly assigned to either one of three (3) conditions: the control/text (task information delivered by images and text); virtual human condition (task information delivered using an animated virtual person/storyteller with synthesized audio narration; images/text/audio condition (task information delivered via text and synthesized audio narration).

Information was delivered in bursts throughout the tasks. These information bursts provided participants with instructions for how to complete the given tasks, thus requiring participants to integrate new information into current situation awareness and update decisions/actions as required. Real-time cognitive load was measured using the Detection Response Task (DRT) (ISO 17488:2016) which has proved an effective measure for cognitive load in applied tasks (Stojmenova & Sodnik 2018).

All participants performed three logistics tasks including (i) a navigation/routing task, (ii) a map annotation task, and (iii) a construction/resource allocation task. Instructions for each task was provided by pre-recorded videos, in MP4 format. Images for the videos were captured from a custom virtual environment that was built in the video game Minecraft. This virtual environment was also the task environment.

3.1 Task Environment and Participants

Minecraft (Java version 1.5.2. Mojang, www.minecraft.net) was used as the base environment for the study tasks. The aim was to provide a virtual environment that (i) had a low entry level for the participants, so they could spend more time doing tasks rather than learning the environment and (ii) was an environment with active interaction. Minecraft is an open-world 3D video game that can be played in first-person perspective where



Figure 3. Example Minecraft views. Left: A landscape to navigate across (Task 1). Right: A harbour environment for the resource allocation task (Task 3).

players can travel over a 3D cartoon like landscape and utilize tools to both create and destroy environmental components and thus modify the virtual environment. Interaction in the user study was via a wired Xbox controller (gamepad). Figure 3 shows example views of the task environment.

Minecraft allows customization of both the environment terrain and plugins to gather user data from sessions. For this project a Minecraft server was hosted in Sydney (via Shockbyte – game server provider, <https://shockbyte.com/>). Analytic data on user performance was downloaded at the end of each session. Tasks 1 and 3 were completed in the Minecraft environment. Task 1 involved participants navigating from a deployment area to an extraction point in the environment while collecting medical supplies, represented as boxes in the environment. Task 3 required participants to allocate boxes to different collection points in an area with three loading zones (see Figure 3: Right) environment. Task 2 used a map view of the Minecraft environment and required participants to annotate a physical copy (see Figure 4: Right) with elements that they had encountered while navigating the environment in Task 1.

Participants were recruited from undergraduate and postgraduate students at University of Newcastle. Upon recruitment, participants were sequentially allocated to a study condition (Control/Text, Virtual Human, or Audio, as described in Section 3.2), with the sequence adjusted to ensure equal condition allocation to all groups. Participants were asked to provide written, informed consent and completed a demographic survey at the commencement of the experimental session. The demographic survey included basic demographic questions, as well as their previous interaction with virtual humans and virtual worlds.

3.2 Instruction Stimuli

For each of the three tasks, the participants were shown a video briefing. This briefing presented an artificial dispute over a region of land and the need to collect and distribute medical supplies. The briefing also discussed the names and supporters of fictitious countries involved in the conflict, and noted the goals and sub-goals of the tasks. Following the conclusion of a video, participants either engaged with the task environment (i.e. Minecraft in Tasks 1 and 3) or with annotating a physical map (Task 2). During Tasks 1 and 3, a second video was shown after a trigger event, for example after two minutes, to provide updated goals to the participants.

Control Condition (Text). The control condition presents the participants with a video that includes text overlay on a briefing report style of images. All task information is provided within the text overlay, and is supplemented by the images; no audio is present. A screenshot of this condition is shown in Figure 1: Right.

Virtual Human Condition. In addition to the series of images shown in the control condition, an animated virtual human character is overlaid on the briefings, akin to a newsreader on a news report. Text presented in the control condition is instead provided using audio. The virtual human mimics human facial expressions and mouth movements when speaking. The speech is generated based on the control condition text using text-to-speech conversion software. Minimal text is provided via a news report style headline banner at the bottom of the screen. A screenshot of this condition is shown in Figure 1: Left.

Audio Condition. The audio condition (sound and text) presents the audio from the virtual human condition instead of the text from the control condition, and without the presence of the virtual human. Minimal text is

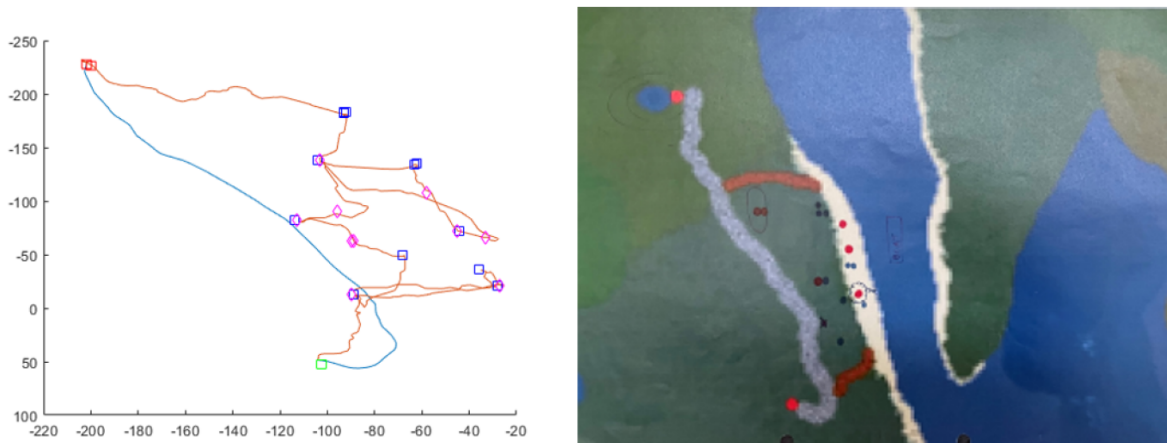


Figure 4. Left: Task 1 visualisation of one participant showing author generated minimum path (blue line), participant path (orange line), start point (green square), exit point (red square), box pickups (blue squares/pink diamonds). Right: Example annotated map from Task 2.

provided via a news report style headline banner at the bottom of the screen. A screenshot of this condition is shown in Figure 3: Right.

3.3 Measurement of Cognitive Load

Cognitive load was measured throughout each task using the Detection Response Task (DRT). The DRT is a form of secondary task which has been shown to be an effective measure of cognitive load (Conti et al. 2012) and is registered in ISO 17488:2016 as a proven and effective measure (Stojmenova & Sodnik 2018). Under this standard, a haptic stimuli, or vibration buzzer, is attached to a participant's shoulder and reaction time and hit rate to haptic events is recorded with a response range of 100-2500ms.

The sensory stimulus provided through the haptic buzzer occurred randomly every 3-5 seconds, with participants instructed to respond to these stimuli by pressing a button attached to their non-dominant index finger. The time difference (in ms) between the stimuli presentation and the button press is captured as the "response time", with longer response times, or the absence of a response, indicative of higher mental load. A hit rate (%), which is indicative of accuracy, can be determined for responses outside the 100-2500ms range.

3.4 Performance Measures

Correctly following instructions in the virtual environment was the main metric for participants' understanding of each task. Each task required both engagement with the task and dynamic situational awareness in the context of task goals. A score system of goals and sub-goals in each task was defined from the instructions provided in the videos. Using a standardised scoring system allowed for consistent marking of participant performance across each task and for a "quality" score, based on task objectives, of performance to be determined. During the Minecraft tasks (Tasks 1 and 3), data analytics were collected on participant movement and environment events/interactions. These were then visualised (see Figure 4: Left) to support scoring the tasks. Task 2 required participants to manually annotate a map of the virtual environment they had just navigated in Task 1 (see Figure 4: Right).

4 RESULTS

This section presents high-level results from the data collected during the user study. As noted earlier, COVID-19 significantly impacted participant recruitment. Due to low participant numbers, the results in the following sections are indicative rather than statistically significant. Therefore, we have opted for graphical representation of trends. Seventeen participants completed sessions with the three tasks. Participants were split evenly across the three conditions (Control/Text $n=5$, Virtual Human $n=6$, Audio $n=6$) and balanced for gender at time of recruitment. Participants identified as 82% male and 18% female. Participant age ranges were 18-25 year (76%), 26-40 years (18%) and 41-55 years (6%). Also all participants identified as video gamers.

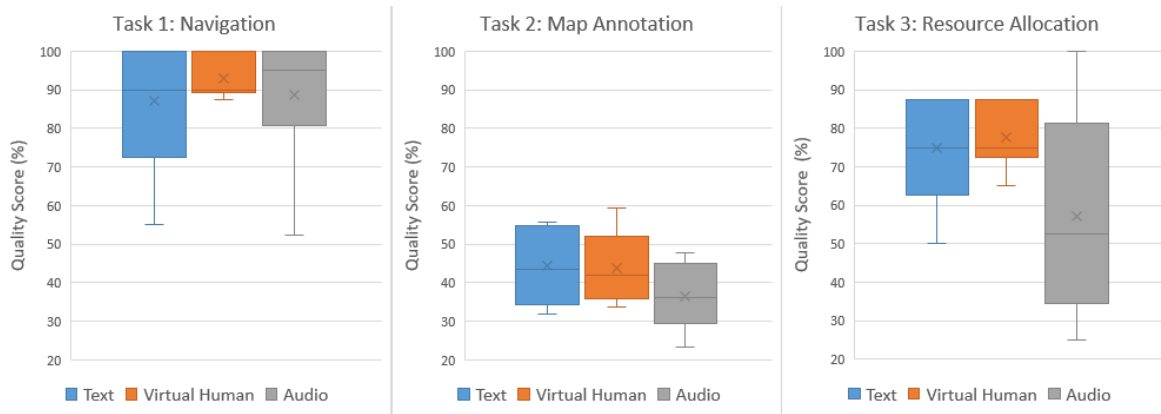


Figure 5. Results of participant performance based on following task instructions across Tasks 1-3. Performance measured as Quality (%) of goals and sub-goals completed. Means = X.

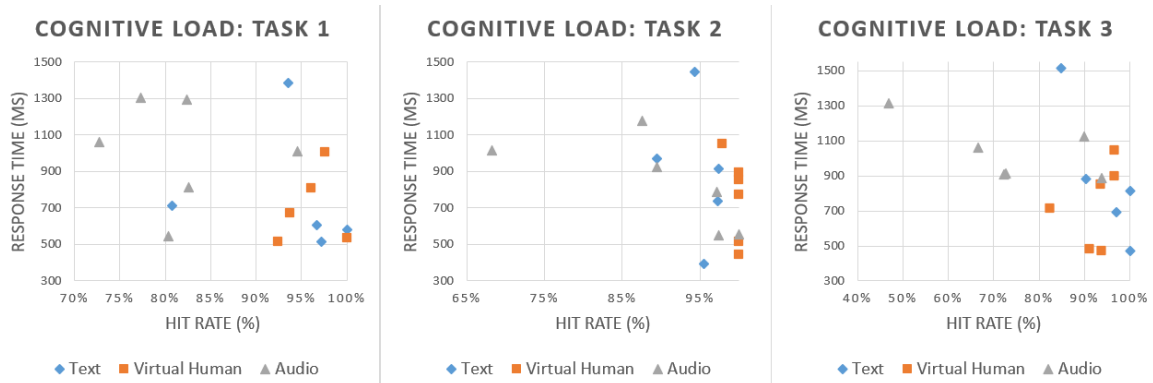


Figure 6. Individual participant results of cognitive load measured via DRT response times (ms) and response hit rate (%).

4.1 Performance

Figure 5 shows the results of participants' quality scores in the three tasks over the different instruction conditions. Task 1 was easiest task with the main goal being to navigate around the virtual environment and collect boxes of different colours. This is evident by high scores across all three conditions. However, the Virtual Human condition has a slightly higher mean, smaller interquartile spread and most participants with scores greater than 90%. Task 2 was the most difficult task, requiring participants to recall their actions from Task 1 and update a physical map. All participants generally scored poorly in this task and there was little difference between the three conditions. Task 3 shows a significant difference with the Audio condition. The Virtual Human condition is slightly better and again has a smaller interquartile spread.

4.2 Cognitive Load

Response time (ms) and hit rate (%) to the DRT haptic stimuli across Tasks 1-3 is shown in Figure 6. For response time, lower scores are preferred as they indicate faster reactions. For hit rate, a higher percentage indicates higher accuracy. Thus in Figure 6 data points in the bottom right corner are the best as they minimise reaction time and maximise hit rate (i.e. accuracy)². With the clustering across Tasks 1-3, the Virtual Human condition had the best overall cognitive load profile. The Audio condition performed the worst across all three tasks. The Text condition and the Virtual Human conditions were comparable in Task 3 and the Virtual Human condition was the best in Task 2. The Text condition had a single poorly performing response time data point

²NB. Review of an outlier data point in the Virtual Human condition (Task 1) indicated equipment failure and was thus removed.

across all three tasks and on review of the data this was determined to be the same participant. The low scoring data points in the Audio condition in Tasks 2 and 3 were different participants. When combined with the performance results (Figure 5), the Audio condition performed the worst across performance and cognitive load. Although the Text and Virtual Human had similar performance scores, the Virtual Human performance results had smaller interquartile ranges and a better profile across the cognitive load measures.

5 CONCLUSIONS

The results indicate that participants receiving instruction by the virtual human generally performed better. Across navigation, map annotation and resource allocation tasks, participants with instruction via the virtual human avatar tended to have higher quality performance scores. These participants also demonstrated better cognitive load profiles when compared to participants who had text or audio-based instruction. Participants in the Audio condition performed the worst across task performance and cognitive load impact. However, the low participant count makes the results indicative rather than robust. Much of the current results are based on manual performance review, i.e. comparing in-session activity logs to task objectives. However, it is possible that participants were completing tasks reactively in the environment (i.e. just collecting all the boxes they saw) rather than following the instructions from the videos. Mapping attention to the videos and any in-session focus would be useful here. However, this cannot be verified without tracking the participant eye movements (i.e. which area of the screen they were paying attention to) with the use of eye tracking technology.

Therefore, in addition to increasing participant numbers, future work will consider a new study with participant gaze tracking to determine features participants are focused on during instruction and while completing in-situ tasks. Also it would be worth considering the short and long term impact of information delivery; the current study only considers delivery immediate before and mid-task and does not consider potential for longer term retention. Long term retention may be important in tasks where timely instruction updates may vary.

REFERENCES

- Alharthi, S. A., Raptis, G. E., Katsini, C., Dolgov, I., Nacke, L. E. & Toups, Z. O. (2018), Toward understanding the effects of cognitive styles on collaboration in multiplayer games, in 'Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing', CSCW '18, Association for Computing Machinery, New York, NY, USA, p. 169–172.
- Blackmore, K. L., Smith, S. P., Nesbitt, K. V., North, L., Wark, S. & Nowina-Krowicki, M. (2019), Evaluating a virtual human storyteller for improved decision support, in S. Elsworth, ed., 'MODSIM2019, 23rd International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand', pp. 207–213.
- Chandler, P. & Sweller, J. (1991), 'Cognitive load theory and the format of instruction', *Cognition and Instruction* **8**(4), 293–332.
- Chung, C., Kadan, A., Yang, Y., Matsuoka, A., Rubin, J. & Chechik, M. (2017), The impact of visual load on performance in a human-computation game, in 'Proceedings of the 12th International Conference on the Foundations of Digital Games', FDG '17, Association for Computing Machinery, New York, NY, USA.
- Conti, A. S., Dlugosch, C., Vilimek, R., Keinath, A. & Bengler, K. (2012), An assessment of cognitive workload using detection response tasks, in 'Advances in Human Aspects of Road and Rail Transportation', CRC Press, pp. 735–743.
- Craig, S. D., Twyford, J., Irigoyen, N. & Zipp, S. A. (2015), 'A test of spatial contiguity for virtual human's gestures in multimedia learning environments', *Journal of Educational Computing Research* **53**(1), 3–14.
- Endsley, M. R. (2000), Theoretical underpinnings of situation awareness: A critical review, in M. R. Endsley & D. J. Garland, eds, 'Situation Awareness Analysis and Measurement', CRC Press, pp. 1–24.
- Hibbard, S. J., Whitney, S. J., Carter, L., Fidock, J. J. T., Temby, P. & Thiele, L. (2018), Making virtual sense: Display type and narrative medium influence sensemaking in virtual environments, in A. Naweed, M. Wardaszko, E. Leigh & S. Meijer, eds, 'Intersections in Simulation and Gaming', Springer International Publishing, Cham, pp. 222–236.
- Mayer, R. E. (2002), 'Multimedia learning', *Psychology of Learning and Motivation* **41**, 85–139.
- Miller, G. A. (1956), 'The magical number seven, plus or minus two: Some limits on our capacity for processing information', *Psychological Review* **63**(2), 81–97.
- Stojmenova, K. & Sodnik, J. (2018), 'Detection-response task—uses and limitations', *Sensors* **18**(2).
- Thorpe, A., Nesbitt, K. & Eidels, A. (2019), Assessing game interface workload and usability: A cognitive science perspective, in 'Proceedings of the Australasian Computer Science Week Multiconference', ACSW 2019, Association for Computing Machinery, New York, NY, USA.