Cognitive and perceptual skills in game-like training tools: transfer of training from static to dynamic contexts

Lisa Wise¹, Gregor McLean², Jason Skues¹ and Lisa Johnston¹

¹ Swinburne University of Technology, Victoria, Australia
² Defence Science and Technology Organisation, Australia

Abstract

Background: This paper explores the robustness of the cognitive and perceptual skills developed within a game-like instrument scanning training task. The task uses static, stylised instruments and explores transfer of training to perceptual manipulation and to a dynamic training task. Method: Participants (n=50) completed a static instrument scanning task (including static transfer), presented via a laptop under supervised conditions. They then completed a dynamic instrument scanning task to test for transfer of static training to the more realistic dynamic environment. A small number of experienced pilots (n=3) completed the dynamic task alone to ensure that the task tapped skills relevant to the real world flying domain. Results: Participants could perform reasonably accurately (median 80% correct compared with a chance level of 10%) on the static instrument scanning task after minimal training, and response times improved significantly across training trials, including static transfer trials. Performance on the dynamic task was uniformly poor even for the subset of high-performing participants (n=19) who were the most accurate on the static task. However, the experienced pilots were able to perform well on the dynamic task suggesting that this task draws on skills required in the aviation context. Conclusion: Despite superficially good performance on a game-like static instrument scanning task, there is little evidence of transfer of training to a dynamic training environment with more ecological validity. The gap between game-like training environments and real world operational tasks is likely to be substantially greater, such that the potential for training transfer is further reduced.

©Wise et al: Licensee HFESA Inc.

Background

Young people who have grown up in a digital world may potentially spend as much time interacting with digital technologies and virtual worlds as they do with the real world around them [1]. As a result of this trend, researchers and practitioners are now considering how best to leverage new technologies for educational and training purposes to better engage with the assumed learning style of recruits from the digital generation [1]. This paper explores the nature of the cognitive and perceptual skills developed within a game-like instrument scanning training task using static, stylised instruments representing the six-instrument cockpit display of a light aircraft (Figure 1).

Previous aviation-focused research [2] suggests that novices can perform at the same level as experienced pilots after minimal training on a Perceptual Learning Module incorporating an instrument training task. McLean et al. [3-4] replicated the original study of Kellman and Kaiser [2] using a sample of undergraduate psychology students, and found that even participants with minimal intrinsic interest in aviation appeared to be capable of rapidly developing instrument scanning skills in this game-like training environment. While such findings superficially appear to offer exciting training opportunities using new technologies, the task of identifying the aircraft situation from the instrument panel was a relatively simple one and the response times (from 4 – 10 secs) were of an order of magnitude suggesting cognitive, rather than purely perceptual processing.

McLean et al. [3-4] included an additional condition in their study to see whether skills learned in the Perceptual Learning Module were based on true perceptual learning or would transfer to a new instrument display representing the same static instrument information, but in a perceptually distinct way. Participants were initially trained on a highly stylised instrument panel similar to the panel used by Kellman and Kaiser [2]. This training was shown to transfer to an instrument display constructed from more realistic instrument images, supporting the idea that the learning process included a cognitive dimension rather than being based on purely perceptual display elements.

Other findings also suggest caution in interpreting performances of experts versus novices, given that these two groups bring very different domain knowledge to the “training task”, and may actually be performing very different cognitive tasks despite the apparent objective similarity of the stimuli [5-7]. Novices have only the information provided by the stimuli to generate classifications of aircraft situation, whereas experts not only need to select task-relevant cues from the experimental stimuli, but also need to identify and ignore task-irrelevant cues that might be meaningful in a real-world situation, and may also spend time searching for cues that they think should be task-relevant but have not been provided to them.
The current study further examines the robustness of skills learned in the static instrument scanning task to perceptual and cognitive manipulations. This study explores performance on an instrument scanning task using video-based stimuli incorporating dynamic instrument motion (perceptual cues for motion) rather than static stimuli using arrows to denote instrument motion (a cognitive representation of motion used in the static tasks discussed above). Rather than judging the aircraft situation directly, a situation that is changing dynamically during the time course of the video, the task was to identify if any of the instruments became incongruent with the aircraft situation as would happen in an instrument failure situation. In order to complete this task successfully, the participant needed to understand the relationship between the aircraft instruments and the out-the-window representation of the aircraft’s situation. Preliminary data from a pilot version of this experiment reported only minimal transfer of training from static to dynamic conditions [8]. If transfer of training from the static to dynamic situation occurs, it suggests that the perceptual and cognitive skills developed within the stand-alone static training task have the potential to be of use in more realistic aviation training environments. Conversely, if no transfer of training occurs, there are implications for the way in which training tasks using digital technologies and virtual environments are developed for real-world operational tasks, in aviation, but also in other skilled performance domains. The face validity of using domain-relevant stimuli and responses (e.g., using a task loosely based on aircraft instrument scanning in an aviation context) may mask issues with construct validity in terms of the perceptual and cognitive skills underlying training and performance.

In this study, we used a brief version of the static instrument scanning task [3-4], followed by the dynamic task [8] under supervised conditions to test for transfer of training between the static and dynamic conditions, as would be required if the static training was leading to meaningful learning. We included two control groups. A small sample of experienced pilots undertook the dynamic task alone to ensure that it had sufficient ecological validity to draw on their aviation experience. A second control group of people with no aviation background undertook the dynamic task alone so we could assess the degree to which prior static instrument scanning training would improve performance on the dynamic task. If the dynamic task taps the instrument scanning skills of experienced pilots, we would expect experienced pilots to perform this task at a higher degree of accuracy than non-pilots. Furthermore, if the static training task provides meaningful instrument scanning training, we would expect participants trained on the static task to perform the dynamic task at a higher degree of accuracy than untrained non-pilots. However, as argued above, if the static training task is actually a different task for novices and pilots due to their different understanding of the stimuli, we would not expect transfer from static to dynamic training tasks.

Method

Participants

The sample consisted of a total of 62 participants (21 Males and 41 Females) of which 74% were 20 years of age or younger, 20% were between 21 and 30 years of age, and 6% were over 30 years of age. The sample was organised into 3 groups with 50 participants randomly allocated to the Experimental Group and 9 participants allocated to a Non-Pilot Control Group. A further 3 participants were recruited into an Experienced Pilot Group in order to confirm a degree of task validity regarding the relevance of the experiment to the aviation context. Each of these experienced pilots reported logging at least 200 hours of flying experience.

Materials

The experiment consisted of two separate computer-based instrument scanning tasks; a Static Instrument Scanning (SIS) training task based on the design used by McLean et al. [3] and a dynamic Instrument Failure Video (IFV) task based on the design used by McLean et al. [8]. Both tasks were programmed in Inquisit v4 (Millisecond Software, http://www.millisecond.com), which allows precise timing (with ms resolution) of stimuli and responses.

Static Instrument Scanning (SIS) Task

The SIS task presented participants with a static image of a standard six instrument panel (Figure 1, top-left panel), including an Airspeed Indicator, Attitude Indicator also known as the Artificial Horizon, Altitude Indicator, Turn Coordinator, Heading Indicator, and the Vertical Speed Indicator. Prior to any experimental conditions, participants were given a detailed explanation of each instrument, how to interpret instrument changes, and shown how to interpret them in combination. When presented with the instrument panel, the participants’ objective was to determine the aircraft’s situation from ten possible scenarios as quickly as possible (via button press). The ten scenarios comprised nine aircraft situations: “Straight and Level”, “Level Left Turn”, “Level Right Turn”, “Level Climb”, “Level Descent”, “Climbing Left Turn”, “Climbing Right Turn”, “Descending Left Turn”, or “Descending Right Turn”, and as in the previous studies [3-4], an additional aircraft situation, where the instrument panel represented “Incongruent” information. An example of “Incongruent” is shown in Figure 1 (lower-middle panel), where five of the six instruments display readings consistent with a level climb, yet the altitude indicator is incongruent suggesting a descending aircraft. The incongruent condition was included to ensure that participants engaged with each individual instrument in order to confirm their response, rather than relying on a single instrument.
As was noted in previous studies [2-4], participants reached high levels of accuracy after only a small number of trials and many participants lost motivation to complete the training trials. So for this study, the SIS task comprised only a small number of trials compared (70) with previous studies (330) to maintain motivation. The 50 participants in the Experimental Group first completed a short series of 10 practice trials before completing a training block of 30 trials in which they were instructed to identify the aircraft situation as quickly as possible. Following these blocks, participants then completed a ‘Static Transfer Condition’ block (30 trials). This block consisted of an identical task, however utilized a set of stimuli generated using more realistic instrument images (Figure 1, lower-right panel. See McLean et al.[2,8] for further details.

Figure 1. Experimental stimuli. Top-left panel shows the six main instruments of a Cessna cockpit as stylized static stimuli. The top-right panel shows the timeline of stimulus presentation for each trial within a standard trial block. While the top-left panel is congruent with a straight and level aircraft situation, the lower left panel is compatible with a climbing left turn. The middle lower panel conversely shows an “Incongruent” display, with most instruments compatible with a level climb, yet the altitude indicator suggests a descending situation. The rightmost lower panel shows the transfer condition using a more realistic instrument display. Arrows on some instruments indicate the direction of motion (outside of the instrument where the dial moves and inside the instrument where the needle moves).

Figure 2. Example of a screenshot from the IFV task showing the aircraft flying straight and level. In this example, the Turn Coordinator (highlighted in red) has failed indicating a left turn while the remaining instruments suggest straight and level flight.
Instrument Failure Video (IFV) Task

Participants viewed 32 short 30-second videos of flight simulation footage captured using flight simulation software (Laminar Research, X-Plane – see Figure 2 for a screenshot). Each trial commenced with a straight and level aircraft, before a maneuver from the situations depicted in SIS task was initiated (e.g., “Climbing Left Turn”, “Descending Right Turn”), and then finally the aircraft would return to straight and level flight. In 50% of video segments, at approximately the 10, 15, or 20 second mark, one of the instruments would “fail” (i.e. cease to respond) and as such would no longer be congruent with the other instruments (akin to the “Incongruent” aircraft situation in the SIS task). The participant’s objective was to observe the video and determine whether any instrument had failed, and identify that instrument as quickly as possible via a mouse click on the failed instrument.

Data Analyses

Descriptive statistics were computed to identify overall patterns of responding (percent correct and response latency) across the conditions and sample. Two Friedman’s analysis of variance tests for dependent measures were used to test for transfer of training. One compared accuracy of performance for participants under static SIS task and the dynamic IFV task and the other compared speed of responses across static SIS conditions. A Kruskal Wallis test for independent groups was used to compare accuracy of performance across groups of participants grouped by their performance on the static SIS task, and the two control groups.

Results

In order to test for transfer of training from static to dynamic conditions, it was first needed to be established that participants could learn the static SIS task with the reduced number of trials in the static training phase.

Response accuracy

Consistent with previous studies [2-4], the participants could perform reasonably accurately on the static SIS task after minimal training (median accuracy of 80% compared with a chance level of 10% in the Practice condition, which were performed in the absence of time pressure), and median accuracy remained around 70% as participants attempted to improve the speed of performance in the Training and Transfer conditions (Figure 3, Panel A). This accuracy was lower than in previous studies but still well above chance level (1/10 response options).

However accuracy on the dynamic IFV task was considerably lower than on the static SIS task (median accuracy of 48%). A Friedman’s ANOVA showed these differences in accuracy to be significant ($\chi^2=58.30, \text{df}=3, p<0.0001$). Accuracy on the dynamic IFV trials was significantly lower than for all other conditions. The small improvement in accuracy from SIS Training (Median=56%) to SIS Transfer (Median=68%) did not reach significance, although coupled with the improvement in response time described below, suggests that participants were still learning the task. In all cases, the critical difference (alpha = 0.05 corrected for the number of tests) was 34.06.

Response Latencies

Response latencies improved across static training trials blocks for both correct and incorrect responses (Figure 3, Panel B). Overall response times decreased from Practice to Training (difference of 2.56 sec), and again from Training to Transfer (difference of 0.5 sec). A Friedman’s ANOVA
test for related samples showed that these differences were all significant ($\chi^2=40.44$, df=2, $p<0.0001$). In all cases, the critical difference (alpha = 0.05 corrected for the number of tests) was 23.94.

**Training progression and transfer of training**

In order to examine the progression of learning throughout the experiment, accuracy data from each participant across all experimental trials was compiled into a single plot (Figure 3, Panel C). Shaded cells represent correct responses, and open cells represent incorrect responses. Participants from the Experimental Group were divided into three sub-groups based on accuracy of performance in the final phase of SIS training (the SIS transfer condition). The high performing group were participants performing at over 80% accuracy, the medium performing group were participants performing at between 50%-80% accuracy (well above chance) and the poorly performing group were participants who scored less than 50% correct for all phases of the experiment. These participants were sorted into descending order of accuracy such that the most densely shaded area of the graph depicts highest accuracy of performance (the upper right quadrant of SIS trials comprising the later trials of the highest performing participants). Two control groups, a pilot group (n=3) and a non-pilot group (n=8) performed only the dynamic IFV task.

Although experienced pilots were able to perform well on the dynamic IFV task (mean accuracy of 72% correct), performance on the dynamic task was uniformly poor for non-pilots pre-trained on the static task, reaching only 50% and 48% accuracy for the high and medium performing groups respectively, evidenced by the patchy shading across the dynamic phase. The group that performed poorly on the static task also performed poorly on the dynamic task. The control group of non-pilots who performed only the dynamic IFV task surprisingly performed slightly better (mean accuracy of 56%) than the best performing group pre-trained on the static SIS task, suggesting a trend towards negative transfer from the static to dynamic training situation.

To summarise, it can be seen that the SIS low group (Median=38%) have the worst performance on the dynamic IFV task and the pilots have the best performance (Median=72%). There is a trend for improvement in SIS medium group (Median=47%) and SIS high group (Median=50%). The performance of the control group (Median=56%) is actually slightly higher than that of the SIS high group. A Kruskal Wallis test for independent groups showed that performance differed significantly across the groups, $H(4)=16.76$, $p=0.002$. Post-hoc comparisons across all groups showed that the significant difference was driven by the poor relative performance of the SIS low group, and to a lesser extent by the better performance of the pilots (n=3). Comparison between the participants who learned the SIS task (medium and high performing groups) was similar to, or slightly worse than that of the control group, suggesting that there was no transfer (or slightly negative transfer) from static to dynamic tasks.

![Figure 4. Spread of responses for each type of instrument failure in the dynamic task. Note that the Air Speed Indicator did not fail in any video trials.](image)

**Response accuracy for individual instruments**

In order to explore possible reasons for poor performance in the dynamic context, we looked at the spread of responses across different instrument failure conditions in the dynamic IFV task. Figure 4 shows the response spread for each type of instrument failure. “NONE” indicates a situation where no instrument failed (50% of trials). The selection of “NONE” as a response indicated that the participant did not detect any instrument failure. The notation “Bad Click” indicates that the participant’s response (clicking the mouse on the instrument that failed) did not correspond to an instrument in the panel.

It can be seen from these data that the most common error was failing to detect an instrument failure (NONE). However it can also be seen that the two most intuitive instruments in terms of their (dynamic) movement, the Turn Coordinator and Heading Indicator, which both move for the duration of an aircraft turn, were the easiest instrument failures for participants to detect and were...
least often incorrectly perceived to have failed. Failure of the Vertical Speed Indicator was also commonly detected, although the Vertical Speed Indicator was incorrectly assumed to have failed at a high rate in other conditions. It was clear that participants had difficulty interpreting the relationship between the Artificial Horizon and the Altitude Indicator, with the pattern of responses almost identical for the two types of instrument failures. So while participants may have been able to identify when the two instruments were in conflict from their SIS training, they were unable to identify which instrument was conveying the correct information with respect to the aircraft situation.

Discussion

The results confirmed findings from previous studies [2-4] that many participants with no background in aviation can learn the static instrument scanning task relatively quickly, such that they can accurately identify an aircraft situation from these instruments. Although accuracy reaches high levels very quickly, improvement in response times occurs over subsequent training trials. As has been noted in previous studies, if the task is not learned in the first ten trials, participants do not appear to improve during the course of the experiment. Wise et al [4] attributed failure to learn, and failure to improve performance to differences in motivation, particularly in the context of their web-based study in which participants completed the experimental task in unsupervised conditions. In the current study, all data were collected under direct supervision of one of the investigators, and all participants completed all conditions. However it is still not clear from our data whether some participants lose motivation due to lack of skill, rather than fail to develop skill due to lack of motivation. The low performing group in this study were not identifiable by their overt behaviour while undertaking the study, and all participants appeared to be sufficiently motivated to complete the task.

The data from this study showed that, in contrast to our previous studies [3-4], participants were still improving in response time during the static transfer condition, and did not reduce their accuracy. It appeared that they were still consolidating the static task, due to the reduced number of training trials. These findings give support to the idea that the instrument scanning was overlearned on the stylised stimuli in our previous studies [3-4], such that a change in static stimulus display (static transfer) disrupted performance accuracy. This finding suggests that it is important to include variety in initial training sets to ensure that training does not become reliant on cues associated with the specific training stimuli rather than the generalizable cues that are relevant to the specific task itself.

The main focus of this study was to examine the transfer of training from static training conditions to more dynamic conditions with greater ecological validity. The ecological validity of the task was supported by the fact that the pilot group were able to perform the task accurately with no training. There was, however, minimal transfer from the static SIS training task to the dynamic IFV task, as evidenced by the poor performance of even the high-performing participants in the static SIS training task. In fact there was a trend towards negative transfer from static to dynamic conditions, with the control group with no aviation experience and no static SIS training performing slightly better than the high and medium performing static SIS training groups. We cannot rule out that the slightly better performance of the control group may also have been due to the fact that the other participants were fatigued from undertaking the SIS training first. The lack of transfer from static to dynamic training conditions suggests that perceptual exposure to instrument motion is a critical factor in understanding the relationship between instrument information and the aircraft situation. It is likely that cognitive markers of motion (arrows used in the static SIS stimuli to denote instrument motion) are not processed in a way that transfers to the dynamic more realistic condition where motion cues are provided perceptually rather than cognitively.

Conclusion

Despite good performance on a game-like static instrument scanning task, there is little evidence of transfer of training to a dynamic training environment with more ecological validity. The gap between game-like training environments and real world operational tasks is likely to be substantially greater, and the potential for training transfer is further reduced. This study suggests that when attempting to develop perceptual or cognitive skills, it is not sufficient to train either of these skills in isolation. Training of component skills has clear benefits, but training design must go beyond superficial face validity to create training tasks within a meaningful dynamic environment that acknowledges the interaction between perceptual and cognitive factors. This is particularly important for game-like training tools [1] where it is must be acknowledged that psychomotor and perceptual skills are not always well-understood or fully represented, but the importance of developing cognitive and decision-making skills is promoted [9].

Acknowledgements

We acknowledge the technical support provided by Mr John Cassell, Mr David Yamouni and Dr Benedict Williams in preparing the experimental stimuli for use in this study. This research was supported by funding from the Defence Science and Technology Organisation Research Agreement with Swinburne University of Technology (RA 2010/112925/1).

References

1. Gee JP. What video games have to teach us about learning and literacy. 2007, Palgrave Macmillan, New York
4. Wise L, Williams B, McLean GMT, Skues J. Online training: Sustainability in the face of the unskilled, unsupervised, unmotivated “long tail”. Future Challenges,