Received: 30 November 2021 | Revised: 15 April 2022 | Accepted: 11 June 2022
DOI: 10.1002/acp.3977

RESEARCH ARTICLE

WILEY

The effect of multi-tasking training on performance, situation awareness, and workload in simulated air traffic control

Stephanie C. Black[1] | Angela D. Bender[1,2] | Susannah J. Whitney[2] | Shayne Loft[1] | Troy A. W. Visser[1]

[1]University of Western Australia, Perth, Western Australia, Australia

[2]Department of Defence, Defence Science and Technology Group, South Australia, Australia

Correspondence
Stephanie C. Black, University of Western Australia, Perth, Western Australia, Australia.
Email: stephanie.black@research.uwa.edu.au

Funding information
Australian Government Research Training Program Scholarship Project, Grant/Award Number: 10306062; Defence Science and Technology Group, Grant/Award Number: 7120

Abstract

Increasingly higher demands are being made on the capacity-limited cognitive capabilities of human operators as they strive to maintain situation awareness (i.e., understanding "what is going on") and performance in complex tasks. In the current study we asked whether: (a) training administered via a mobile phone-based app could improve multitasking and (b) improved multitasking in the app would generalize to improved performance and situation awareness in a simulated air traffic control task (ATC). Participants completed the ATC task before and after multiple sessions of app-based multitasking training or control training. Multitasking on the app improved across training sessions. However, this did not lead to improved performance or situation awareness, or workload reduction, relative to control training on the ATC task. These outcomes indicate that app-based multitasking training based on repetition of a single training task will not necessarily yield generalizable benefits to human performance in other complex dynamic tasks.

KEYWORDS
multi-tasking, situation awareness, training

1 | INTRODUCTION

The last several decades have been characterized by rapid increases in information load and technological advancement (e.g., automation to aid tasks) across everyday life and work contexts such as aviation, defense, and healthcare. In order to successfully navigate these multitasking-based environments, it is vital that human operators have adequate situation awareness (SA). SA can be defined as understanding "what is going on" (Endsley, 1995b, p. 36), and although its exact nature remains a subject of theoretical debate (Pritchett, 2015), requires perception of relevant information, comprehension of its meaning, and prediction of future outcomes (Endsley & Jones, 2012).

Achieving SA requires simultaneously paying attention to events in one's surroundings, integrating this information with relevant background knowledge, and dynamically updating and prioritizing a situation model to inform decision-making (Vu & Chiappe, 2015). Such a conceptualization, in turn, strongly implicates a role for multitasking (see a meta-review by Pan & Rickard, 2018)—broadly defined as "the strategic allocation of resources among multiple tasks" (Gutzwiller et al., 2019)—in the development and maintenance of SA (Dehais et al., 2012; Gugerty, 2011).

Multitasking can be subdivided into three key component processes: performing more than one task simultaneously, alternating

Angela D. Bender is now at the Australian Taxation Office, Brisbane, Australia

between tasks, and task switching/prioritization (Chérif et al., 2018). These high-level processes, in turn, depend on more primary cognitive abilities (see reviews by Koch et al., 2018; Schuch et al., 2019). These primary abilities are chiefly working memory (a capacity-limited short-term memory storage; Baddeley & Logie, 1999), and attentional control (a mechanism that deliberately directs the focus of attention to relevant information and avoids irrelevant information; Kane et al., 2006).

One of the chief obstacles to successful multitasking is that human information processing is capacity limited (Kahneman, 1973; Norman & Bobrow, 1975; Wickens, 2002). For instance, Wickens (2002) theorizes in his four-dimensional multiple resource model that performance is adversely affected more when a task draws resources from the same dimension as another task than when the two tasks draw from independent resources. For this reason, complex tasks with competing demands for resources often exceed the processing capabilities of human operators (Wickens, 2008), which in turn, may degrade SA (Endsley, 1995b, 2006; Woods & Sarter, 2010) and consequently increase the risk of erroneous decisions (Adams et al., 1995).

One potential solution to overcome human capacity limitations that has received significant recent attention is cognitive training. Previous research has found that task-specific multitasking ability can be improved with training under controlled conditions (Dux et al., 2009; Liepelt et al., 2011; Verghese et al., 2018). For example, Dux et al. (2009) trained seven participants over two weeks in the so-called "dual-task" paradigm (Pashler, 1994; Schumacher et al., 2001; Tombu & Jolicoeur, 2004). On each trial, participants had to either: (a) make a vocal response to a spoken word, (b) make a key press in response to a visual image, or (c) do both tasks simultaneously (i.e., multitasking). Dux et al. (2009) found training across thousands of trials reduced reaction times (RTs) on all trial types, but this reduction was significantly larger on trials that required responding to both targets simultaneously.

While outcomes such as these point to the potential of cognitive training, many questions remain about its broader application. For one, it is unclear whether training benefits obtained in research labs with careful experimental control could also be obtained under less rigorous conditions. To test this possibility, we administered training using mobile-phone apps which offer opportunities for on-demand training, minimized costs, and wide availability, but allow less control over the training schedule and stimulus presentation.

Another issue concerns training generalizability. It has been suggested that multitasking training can lead to acquisition of elemental cognitive skills that transfer to other, potentially more complex, tasks (Garner et al., 2014). This assumption is consistent with evidence obtained by Liepelt et al. (2011) who trained participants to complete two simple tasks concurrently (hybrid training). They found performance not only improved over the course of hybrid training, but participants also showed benefits on two "near-transfer" tasks that had similar processing requirements to hybrid training but used different stimuli or response mappings.

Crucially, the evidence in favor of near-transfer is contentious as several studies have failed to show such benefits (Bender et al., 2017; Garner et al., 2015; Horne et al., 2020). Moreover, to our knowledge, there has been no examination of "far-transfer" of training benefits to tasks that share common underlying cognitive mechanisms but are substantially different on multiple other dimensions. Such far transfer is critical if multitasking training is to be helpful in improving human performance in complex real-world tasks. The present study investigates this issue by measuring the impact of app-based multitasking training on performance, SA, and subjective workload in a simulated air-traffic control (ATC) task.

1.1 | The present study

The primary goals of this research were to assess: (a) the effectiveness of mobile-app-based dual-task training paradigm similar that used by Filmer, Lyons, et al. (2017), (b) whether training leads to near-transfer on a dual-task paradigm similar to the training task but with different stimuli (as in Liepelt et al., 2011), and (c) whether training leads to far transfer by improving SA, performance, and subjective workload in a simulated ATC task. As detailed further below, the ATC task taps multitasking abilities by requiring participants to manage a sector of airspace by accepting incoming aircraft, handing off outgoing aircraft, and detecting and preventing aircraft conflicts (violations of minimum aircraft separation standards).

Our study incorporated methodological best practice for training (see reviews by Boot et al., 2011; Simons et al., 2016; Salas & Cannon-Bowers, 2001). In particular, we used an "active" control group that completed an alternative training procedure of similar length and engagement that did not require multitasking (in this case, a non-speeded visual search task; Kane et al., 2006). We also used an adaptive training procedure to avoid ceiling effects on training or testing measures.[1] To avoid differential expectancy effects across groups, we ensured participants did not know whether they were in the experimental or control group. Finally, we also emphasized that the experiment was designed to investigate human cognition and task performance, rather than framing it as a "cognitive training study," in order to avoid generating an expectation of training effects.

In addition to employing methodological best practice, a final innovation in our approach was to implement the training program on mobile (cellular) phones. This participant-centered scheduling and delivery of training on a handheld device meant that we had less control over stimulus delivery conditions, timing, and other aspects of practice, but it offered considerably easier access, portability, and flexibility.

2 | METHOD

2.1 | Participants

Ethics approval for this study was granted by the Human Research Ethics Committee of the University of Western Australia. The study conforms to the National Statement on Ethical Conduct in Human Research (National Health and Medical Research Council, 2007). Informed consent was given by all participants. One-hundred and twenty-eight participants were recruited via community and undergraduate student research participation pools in the School of Psychological Science at the University of Western Australia. Participants were awarded $10 (AUD) per research participation hour to cover incidental expenses such as travel and parking. Alternatively, if they were eligible psychology undergraduates, they could choose to instead be awarded one credit per hour towards a class participation component of their undergraduate unit up to the unit's credit maximum. In addition, all participants were offered a monetary performance-based bonus (max = $20 AUD) to increase their motivation. The sample size was chosen based on an a priori power analysis that indicated 90 participants were required to detect a small effect ($f = 0.15$) with $\alpha = 0.05$ and 80% power in a 2 × 2 analysis of variance, plus expected attrition due to participant withdrawal, incomplete data, technical errors, and other issues.

Our final sample consisted of 94 individuals (age: $M = 21.61$, $SD = 6.36$, range = 17–56; females = 52; males = 42) who completed both the pre- and post-training testing and required mobile training sessions. Forty-eight participants remained in the experimental condition after three participants failed to reach the required accuracy threshold during the pre-training near-transfer task (see below), three did not return for the post-training session, six failed to complete all required eight training sessions, and data from four was rendered unusable due to computer problems. Forty-six participants remained in the control condition after five participants failed to reach the required accuracy threshold during the pre-training near-transfer task, five did not return for the post-training session, four failed to complete all required eight training sessions, and data from four was rendered unusable due to computer problems.

2.2 | Materials

Training tasks were carried out on participants' mobile phones running a custom app. Participants completed all other tasks in a laboratory using a PC running Windows 7 attached to a display refreshing at 60 Hz (ATC task) or 100 Hz (all other tasks).

2.3 | Experimental design and procedure

The experiment consisted of a pre-training session conducted in the laboratory, five consecutive days of training (either multitasking or visual search) conducted at the participants' leisure on a mobile app, and a post-training session conducted in the laboratory (see Table A1 in the Appendix); with the pre- and post-laboratory sessions totaling approximately 3 h duration. Two ATC scenarios of approximately equivalent difficulty were used in the pre- and post-testing sessions (counterbalanced). Participants were randomly assigned to the multitasking or visual search training conditions.

2.4 | Training

2.4.1 | Multitasking training

The training program was implemented using the Unity 2018 game engine and deployed on Android and iOS mobile devices. Device refresh rates were set to 60 Hz, which is common for current mobile devices. The application was designed for screens with pixel densities of 326 to 640 ppi, and resolutions ranging from 750 × 1334 to 1440 × 2560 pixels. Display backgrounds were dark gray (RGB 149, 149, 149). Text was black (RGB 0, 0, 0) and presented in Arial 61 point (instruction text) or Din Alternate Bold (in-task text).

This multitasking training task was modeled after dual-task paradigms employed by Filmer, Lyons, et al. (2017). Trials were divided into three types; single visual, single auditory, and simultaneous visual–auditory. Similar to the near transfer task (see Figure 1), all trials commenced with a fixation cross at the center of the display for 200–600 ms (randomized in 100 ms steps). This was followed by the presentation of the target(s) for 200 ms, and a blank screen for 3000 ms until response. On single-target trials, the target was chosen randomly from a selection of three colored circles (visual trials) or three complex tones (auditory trials). On simultaneous trials, visual and auditory stimuli were randomly chosen such that each type of visual and auditory stimulus appeared equally often across single-target and simultaneous trial types. The stimuli were each associated with a different touch key. Key mapping was lateralized, with lateralization counterbalanced based on participant ID. For left lateralization, the visual stimuli were mapped to the touch keys A, B, and C on the

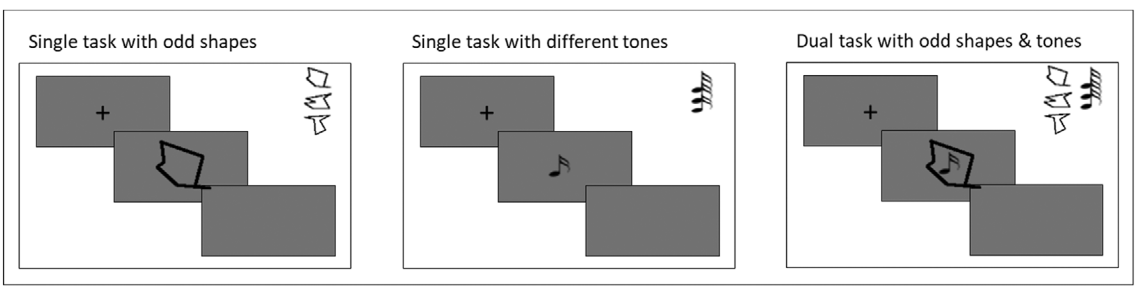

FIGURE 1 The near transfer task: A "dual-task" paradigm adapted from Bender et al. (2018)

left side of the screen; whereas the auditory stimuli were mapped to the touch keys D, E, and F on the right side on the screen. For right lateralization, the visual stimuli were mapped to the touch keys D, E, and F on the right side of the screen; whereas the auditory stimuli were mapped to the touch keys A, B, and C on the left side of the screen. Participants were instructed to identify the target(s) as quickly and accurately as possible but to respond sequentially to targets on simultaneous trials (i.e., to avoid simultaneous or grouped responses; however, participants could respond to stimuli in either order, visual followed by auditory stimulus or vice versa).

Practice consisted of three blocks of 32 trials for each trial type. Participants were instructed to make their response(s) within an 1800 ms response window. To ensure the response mappings were encoded, participants who failed to achieve 80% accuracy on a practice block were presented with the task instructions again and required to repeat the block with a longer response time window.

Training sessions consisted of four blocks of 96 trials equally divided among trial types (randomly intermixed). Training lasted for eight sessions (3072 total trials), in line with previous experiments (e.g., Bender et al., 2017; Filmer, Varghese, et al., 2017; Filmer, Lyons et al., 2017; Garner et al., 2014).

To encourage peak performance in the training sessions, response windows in each block were varied such that the length of the window corresponded to the RT at the 75th percentile of the distribution from the previous block (see Garner et al., 2014). In addition, points were awarded as part of a bonus scheme for achieving 95% accuracy and responding within the response window on 80% of the trials. These could be converted into a monetary bonus of up to $10 (AUD) at the end of training. To help participants track their performance, cumulative average RT and point totals were displayed after each block.

2.4.2 | Active control (visual search) training

Training was delivered via an app like that used for multitasking training in which participants completed a non-speeded visual search task similar to that used by Di Lollo et al. (2000). On each trial, a central fixation cross was followed by the presentation of an equidistant matrix of circles, each with a 60° portion of their perimeter removed. These portions could be in one of six positions in 60° increments around the perimeter. Circles were presented for 33–500 ms (see below). Participants were asked to report the location of the missing portion of a target circle (designated by a surrounding circular frame that remained on the display for 400 ms) by pressing a designated area on the touchscreen.

Practice consisted of two blocks of 32 trials, each starting with a matrix display of 300 ms. Participants who failed to reach 80% mean accuracy during the last practice block were given an additional block of practice trials in which the matrix display duration was increased by 33 ms (to a maximum of 500 ms).

Training sessions consisted of four blocks of 96 trials each, which were repeated eight times (3072 total trials). To encourage peak performance, matrix display duration on each block was decreased by 33 ms when at least 80% accuracy was achieved on the previous block and increased by 33 ms if this minimum was not met. To encourage peak performance, one bonus point was awarded if a participant achieved a correct response, if the response was incorrect, then one bonus point was subtracted. Whenever the matrix display time decreased or the target display time increased, an additional two bonus points were awarded (adapted from Garner et al., 2014). Bonus points were exchanged into a monetary bonus of up to $10 (AUD) at the end of the mobile app training. To help participants track their performance, the total number of bonus points awarded for that block and the total number of bonus points earned overall were displayed after each block.

2.5 | Outcome measures

2.5.1 | Near transfer task

The task was similar to the multitasking training task described above, except that visual targets were complex shapes (see Figure 1; Filmer, Lyons, et al., 2017), the auditory targets were novel complex tones, and responses were made using the keyboard.

Practice consisted of four practice blocks of 18 trials separately for each trial type, with feedback on accuracy after each trial. Participants who failed to reach 80% mean accuracy during the last practice block of each trial type were given an additional block of practice trials.

The main task consisted of six blocks of 27 trials consisting of equal numbers of each trial type presented in random order. In addition, points were awarded as part of a bonus scheme for fast and accurate performance to encourage effort and vigilance. These could be converted into a monetary bonus of up to $10 (AUD) at the end of training. To help participants track their performance, cumulative average RT and point totals were displayed after each block.

2.5.2 | Far transfer task

The ATC lab task (Fothergill et al., 2009) requires participants to assume the role of air traffic controllers responsible for the safe passage of aircraft through their assigned sector. It has been used extensively to examine complex task performance, prospective memory, human-automation teaming, and other human factors issues with trained novices (e.g., Bowden & Loft, 2016; Loft, 2014) and expert air controllers (e.g., Loft et al., 2009).

During the task, participants monitored a display (see Figure 2) depicting aircraft traveling at various speeds and altitudes (indicated by aircraft data blocks) along pre-determined flightpaths. Participants were required to accept aircraft entering the sector and hand off aircraft leaving the sector by pressing A or H, respectively, then left clicking on the circle representing the aircraft. Aircraft needed to get accepted when they approached within 5 nautical miles of entering the sector boundary and handed-off when they approached within 5 nautical miles of exiting the sector boundary.

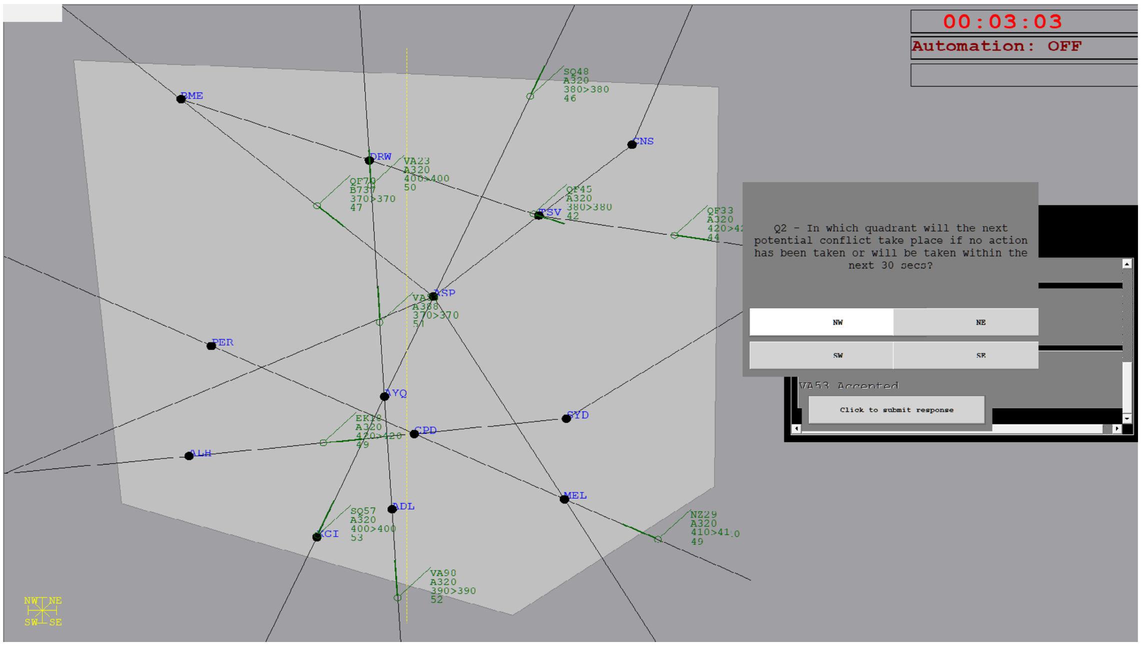

FIGURE 2 Illustration of the ATC task adapted from ATC-lab (Fothergill et al., 2009). The light gray polygon marks the participant's sector, and the lines denote aircraft flight paths. Aircraft speed is presented in knots, and altitude in feet. The aircrafts are represented by small green circles with a line indicating the direction the aircraft is flying in and a text block specifying the call sign, speed, type, and altitude. In the top left quadrant, aircraft QF 73 and VA53 are in future conflict (i.e., they will violate separation if the participant does not intervene by increasing the altitude of one of the aircraft from 370 to 380). In this example the screen is frozen whilst an SA question has appeared asking about the potential conflict; the participant is expected to highlight the correct answer and click the button below.

Participants were also required to resolve conflicts. Aircraft were in conflict if they would, given their respective flight levels, velocities, and headings, simultaneously violate vertical and lateral separation in the future. To resolve a detected conflict participants were told to instruct one of the aircraft involved in the conflict to climb 10,000 ft, by left clicking on the aircraft, selecting which aircraft it is in conflict with and then clicking the ok button. Participants completed three scenarios (a 20-minute practice and two 30-minute testing scenarios). During each of the testing scenarios 10 scripted conflicts, 6 near misses and approximately 100 aircraft acceptances and handoffs needed to be performed. Throughout the simulation aircraft acceptance, handoff and conflict resolution accuracies and RTs were recorded.

2.5.3 | Situation awareness and workload measures

A modified Situation Present Assessment Method (SPAM; Durso & Dattel, 2004) probe technique was employed during the ATC task to assess objective situation awareness. The first SPAM query was presented 2–3 min into each ATC scenario with additional queries presented every 1.5–2 min thereafter. SPAM distinguishes workload from SA by first presenting a "Ready for Question?" prompt and an auditory alert that participants were instructed to respond to as quickly as possible. The time taken for the individual to accept the SPAM query often correlates with subjective workload (Strybel et al., 2008; Vu et al., 2012), with a longer response time indicating greater workload.

If there was no response within 30 s, the prompt was removed, and the scenario continued. Otherwise, the scenario was paused, and an SA query appeared that asked participants about the past, current and future state of the ATC scenario (in line with the three levels of SA; Endsley, 1995a), along with four possible response options (see Table A2 in the Appendix). Participants were encouraged to respond as quickly and accurately as possible by mouse click, with lower accuracy or longer RTs indicating poorer SA (Durso & Dattel, 2004). The assumption in using SPAM is that operators who are maintaining better SA should know, or know where to find, the appropriate information to address the query and are thus able to respond faster to SPAM queries (see Chiappe et al., 2016).

Note that we used SPAM rather than the Situation-Awareness Global Assessment Technique (SAGAT; Endsley, 1995b) because the latter method requires the individual to recall information based on a detailed stable mental representation of their task environment. However, past studies with the current ATC task have indicated that participants tend to rely more on frequent interactions with information displays to maintain SA.

At the end of each ATC scenario, participants also completed the Situational Awareness Rating Technique (SART; Salmon et al., 2009; Taylor, 1990) to assess subjective SA and the NASA-TLX (Hart & Staveland, 1988) to assess subjective workload. The SART consisted of 10 ratings on a 7-point Likert scale (1 = low, 7 = high) that measured overall task understanding, attentional demand, and supply. The NASA-TLX consisted of six rating and 15 weighting questions evaluating ten factors: overall workload, task difficulty, time pressure, own performance, physical effort, mental effort, frustration, stress, fatigue, and activity type.

2.5.4 | Other measures

After the final experimental session, participants were asked to rate their training experience (whether "training was fun") on a scale of 1 (minimal) to 10 (maximal) as per Anguera et al. (2013).

3 | RESULTS

3.1 | Data cleaning

3.1.1 | Outliers

For the multitasking training and near-transfer tasks, the inter-target response interval (RT_2–RT_1, see Tombu & Jolicoeur, 2004) was calculated on each dual-target trial, and trials with intervals of less than 50 ms were excluded from further analysis. Additionally, for each task, both accuracy and median RT data for participants were omitted from further analyses if either were more than 1.5× the interquartile range away from the overall mean (Jones, 2019). Together, this resulted in a total of 4.07% of data being omitted across the tasks.

3.2 | Analysis procedure

Graphs are shown with 95% confidence intervals (CI) to allow intuitive visualization of the precision of the study (Krzywinski & Altman, 2013). We used a Bayesian approach to analyze the data. To allow more intuitive understanding of the strength of the empirical evidence, for *t*-tests we report BF_{01} when evidence is stronger for the null hypothesis, and we report BF_{10} when evidence is stronger for the alternative hypothesis. Similarly for ANOVA, we report BF_{excl} when evidence is stronger for the null hypothesis, and we report BF_{incl} when evidence is stronger for the alternative hypothesis. Thus, the reported Bayes Factors will always be greater than 1. We follow Lee and Wagenmakers (2013); adjusted from Jeffreys' (1961) standard of interpreting Bayes Factors of 1–3 as "anecdotal," 3–10 as "moderate," 10–30 as "strong," 30–100 as "very strong," and above 100 as "extreme." We used default Cauchy priors (Bayesian *t*-tests: $r = 0.707$; Gronau et al., 2020; Bayesian Repeated Measures ANOVA: $r_{\text{fixed effects}} = 0.5$, $r_{\text{random effects}} = 1$, $r_{\text{covariates}} = 0.354$; Rouder et al., 2012; see also van Doorn et al., 2021). For ANOVA we compared all models to the null model, for which $BF_{01} = BF_{10} = 1.0$.

3.3 | Training experience

Participants rated their training experience near the middle of the scale ($M = 5.03$, $SD = 2.22$). Importantly, an independent groups *t*-test comparing training experience across the groups indicated moderate evidence in favor of the null hypothesis ($BF_{01} = 4.26$), indicating there were no differences between the groups.

3.4 | Outcomes of training

3.4.1 | Multitasking training

Table 1 depicts mean accuracy and RTs for auditory and visual targets on single and dual-target trials as a function of training session. Dual-cost scores were calculated by subtracting accuracy and RTs for visual and auditory targets on single target trials from those on dual-target trials, and then summing the scores for auditory and visual targets (Bender et al., 2017). Dual-cost RTs were then divided by accuracy to create a composite performance estimate (adjusted RTs; Figure 3, left panel; Chambers et al., 2006; Townsend & Ashby, 1983), with smaller numbers indicating better multitasking ability. A paired samples *t*-test comparing the first and last sessions indicated extreme evidence for a reduction in adjusted RTs ($BF_{10} = 4.71\text{E}{+}06$) consistent with task-specific improvement in multitasking across sessions.

3.4.2 | Visual search training

Figure 3 (right panel) depicts search matrix display times as a function of session. A paired samples *t*-test comparing the first and last sessions indicated extreme evidence for a reduction in matrix display times ($BF_{10} = 2.024\text{E}{+}19$), consistent with task-specific improvement in visual search across sessions.

3.5 | Near transfer

Table 2 depicts mean accuracy and RTs for auditory and visual targets on single and dual target trials as a function of testing session. To test if multitasking training yielded benefits on a similar task with unpracticed stimuli, we computed adjusted RT dual-cost scores (using the method and dual-cost calculations as outlined above) on the near transfer task separately as a function of testing session (pre- vs. post-training) and training group (multitasking vs. active control). These adjusted RTs are shown in Figure 4 and were submitted to a 2 (testing session) × 2 (training group) mixed-design analysis of variance (ANOVA). This yielded anecdotal to moderate evidence in favor of the null hypothesis for both main effects and the interaction ($1.000 \le BF_{excl} \le 3.165$), indicating that multitasking training did not yield improvements on a similar task despite evidence for improvements in the training task itself.

3.6 | Far transfer (ATC performance, SA, and workload)

3.6.1 | ATC performance

Raw accuracy and RTs on the ATC tasks can be seen in Table 3. As in other tasks, we used these scores to compute adjusted RT scores for each task separately as a function of testing session and training group (see Figure 5, left panel). Separate 2 (testing session) × 2 (training group) mixed-design ANOVAs were conducted to analyze performance on each ATC task.

Analysis of the aircraft acceptance task yielded anecdotal to moderate evidence in favor of the null hypothesis for both main effects and the interaction ($2.123 \le BF_{excl} \le 3.226$), suggesting multitasking training did not improve task performance. Analysis of the aircraft handoff task similarly yielded anecdotal to moderate evidence in favor of the null hypothesis for both main effects and the interaction ($1.748 \le BF_{excl} \le 3.731$), suggesting multitasking training did not improve performance.

Analysis of the aircraft conflict detection task yielded extreme evidence for a main effect of testing session ($BF_{incl} = 20474.569$), indicating conflict detection performance improved across sessions. However, there was anecdotal to moderate evidence in favor of the null hypothesis for the main effect of training group and the interaction ($2.392 <= BF_{excl} <= 3.802$), suggesting that multitasking training did not improve performance.

3.6.2 | SA (SPAM)

Raw SPAM accuracy and RTs are presented in Table 4. We used these scores to compute adjusted RT scores separately as a function of testing session and training group (Figure 6, left panel). A 2 (testing session) × 2 (training group) ANOVA yielded very strong evidence for a main effect of testing session ($BF_{incl} = 33.752$), indicating objective SA improved across sessions. However, there was anecdotal to moderate evidence in favor of the null hypothesis for the other effects ($2.278 <= BF_{excl} <= 4.082$), suggesting multitasking training did not improve objective SA during the ATC task.

3.6.3 | SA (SART)

Scores were calculated separately as a function of testing session and training group using the formula $SA = U - (D - S)$, where U = summed understanding ratings, D = summed demand ratings, S = summed supply ratings, and higher scores indicate greater SA (Figure 6, right panel). A 2 (testing session) × 2 (training group) ANOVA yielded anecdotal evidence for the main effect of testing session ($BF_{incl} = 1.489$), indicating subjective SA improved across sessions. However, as with the SPAM measures, there was moderate evidence in favor of the null hypothesis for the remaining effects ($3.717 \le BF_{excl} \le 4.525$), suggesting multitasking training also failed to improve subjective SA during the ATC task.

3.6.4 | Workload (SPAM ready prompt latency)

The median time to accept SPAM queries was calculated separately as a function of testing session and training group (Figure 7, left panel) and submitted to a 2 × 2 ANOVA. This yielded moderate evidence

TABLE 1 Mobile app training performance: RTs (in ms) and accuracy (proportion correct) for single-auditory, single-visual, dual-auditory, and dual-visual trials

Session	1	2	3	4	5	6	7	8
Raw RTs (ms)	*Median (SD)*							
Single visual	902 (103.65)	834 (111.41)	801 (85.56)	800 (98.88)	799 (90.49)	800 (98.26)	788 (74.19)	768 (87.59)
Single auditory	1167 (139.06)	1069 (125.83)	1017 (102.33)	967 (103.91)	983 (114.52)	968 (119.91)	967 (124.57)	936 (116.15)
Dual visual	1178 (171.15)	1067 (185.29)	1067 (150.74)	1068 (178.37)	1067 (172.87)	1033 (148.11)	1050 (182.32)	1025 (144.73)
Dual auditory	1333 (177.64)	1168 (206.04)	1150 (184.02)	1100 (176.88)	1099 (185.94)	1101 (194.79)	1067 (201.39)	1067 (181.24)
Raw accuracies	*Mean (SD)*							
Single visual	0.87 (0.08)	0.94 (0.03)	0.94 (0.04)	0.94 (0.04)	0.94 (0.05)	0.92 (0.06)	0.93 (0.05)	0.93 (0.05)
Single auditory	0.8 (0.12)	0.9 (0.07)	0.91 (0.07)	0.94 (0.04)	0.91 (0.06)	0.93 (0.04)	0.92 (0.06)	0.92 (0.06)
dual visual	0.65 (0.18)	0.79 (0.12)	0.81 (0.13)	0.85 (0.10)	0.85 (0.08)	0.84 (0.11)	0.84 (0.12)	0.86 (0.08)
Dual auditory	0.65 (0.18)	0.79 (0.12)	0.81 (0.13)	0.85 (0.10)	0.85 (0.08)	0.84 (0.11)	0.84 (0.12)	0.86 (0.08)

Note: The RTs are all median RTs.
Abbreviation: RT, reaction time.

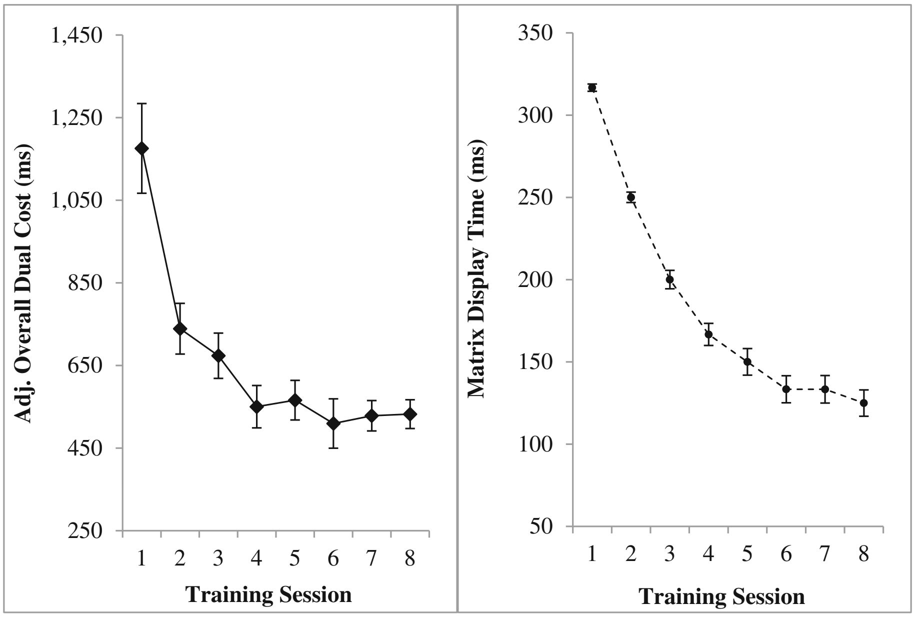

FIGURE 3 Mobile app training performance. Left panel shows adjusted dual costs for the experimental training. Right panel shows matrix display times. Lower values indicate improved performance. 95% CI error bars

TABLE 2 Near transfer task performance: RTs (in ms) and accuracy (proportion correct) for single-auditory, single-visual, dual-auditory, and dual-visual trials

Session	Pre-training	Post-training
Raw RTs	*Median (SD)*	
Single visual	754 (106.67)	620 (62.54)
Single auditory	914 (100.83)	772 (121.40)
Dual visual	1021 (187.50)	844 (155.21)
Dual auditory	1257 (122.84)	1063 (124.42)
Raw accuracies	*Mean (SD)*	
Single visual	0.92 (0.05)	0.95 (0.03)
Single auditory	0.9 (0.05)	0.94 (0.04)
Dual visual	0.88 (0.06)	0.89 (0.07)
Dual auditory	0.88 (0.06)	0.89 (0.07)

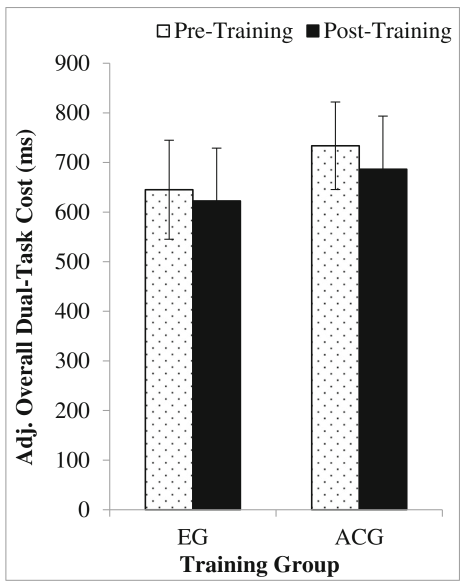

FIGURE 4 Near transfer task performance. Lower adjusted overall dual cost indicates improved performance. 95% CI error bars. ACG, active control group; EG, multitasking group

for the main effect of training group ($BF_{incl} = 4.156$), suggesting the experimental group showed reduced workload across both pre and post training sessions in the ATC task. However, there was moderate support for the null hypothesis for all other effects ($4.444 <= BF_{excl} <= 4.785$), suggesting that multitasking training did not reduce workload during the ATC task.

3.6.5 | Workload (weighted NASA TLX)

Weighted NASA TLX Indices (Hart & Staveland, 1988) were calculated separately as a function of testing session and training group (Figure 7, right panel) and submitted to a 2 × 2 ANOVA. This yielded extreme evidence for the main effect of time ($BF_{incl} = 3267.637$), but moderate evidence for the null hypothesis for all other effects ($3.021 <= BF_{excl} <= 3.876$). Thus, the evidence from the NASA TLX converged with the SPAM measure of workload to indicate that multitasking training did not reduce workload during the ATC task.

3.6.6 | Summary of results

Please see Table 5 for a summary of our results.

TABLE 3 ATC task performance: Acceptance, handoff, and conflict resolution RTs and accuracy (i.e., hit rate)

	Pre-training		Post-training	
Session	EG	ACG	EG	ACG
RTs	*Median (SD)*			
Acceptance (ms)	3202 (840.11)	3220 (1005.10)	3077 (722.26)	3182 (772.85)
Handoff (ms)	2407 (907.62)	2576 (646.91)	2410 (672.59)	2661 (659.00)
Conflict resolution (s)	124 (52.35)	114 (29.13)	95 (45.87)	99 (33.19)
Accuracy	*Mean (SD)*			
Acceptance	0.96 (0.03)	0.96 (0.03)	0.97 (0.03)	0.97 (0.02)
Handoff	0.95 (0.04)	0.96 (0.03)	0.97 (0.03)	0.97 (0.03)
Conflict resolution	0.88 (0.11)	0.91 (0.11)	0.94 (0.07)	0.95 (0.07)

Note: Means (SDs).
Abbreviations: ACG, active control group; EG, multitasking group.

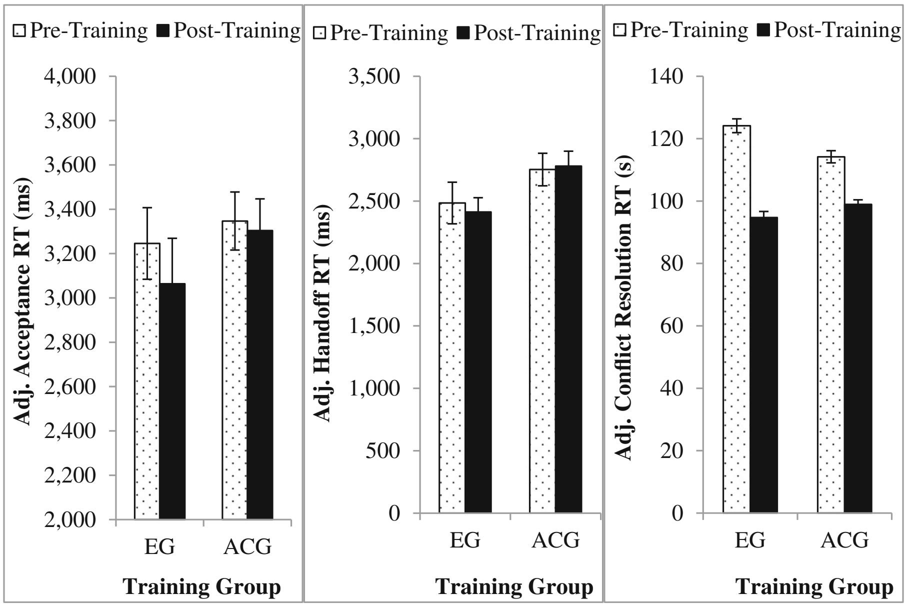

FIGURE 5 ATC task performance: Hit rate adjusted acceptance (left), handoff (middle) and conflict resolution RTs (right). 95% CI error bars. ACG, Active control group; EG, multitasking group

TABLE 4 RTs and accuracy (proportion correct) for SPAM past, present, future, and overall

	Pre-training		Post-training	
Session	EG	ACG	EG	ACG
Raw SPAM RTs	*Median (SD)*			
Past	11.75 (5.37)	13.25 (5.00)	9.00 (3.29)	10.04 (3.64)
Present	11.73 (3.82)	12.27 (3.46)	11.41 (2.95)	12.02 (3.06)
Future	13.54 (3.78)	13.11 (4.01)	11.61 (3.87)	13.08 (3.25)
Overall	11.87 (3.47)	12.57 (3.70)	11.05 (2.63)	11.34 (3.00)
Raw SPAM accuracies	*Mean (SD)*			
Past	0.86 (0.13)	0.88 (0.14)	0.85 (0.15)	0.90 (0.11)
Present	0.95 (0.08)	1.00 (0.00)	0.91 (0.13)	0.93 (0.10)
Future	0.90 (0.10)	0.89 (0.12)	0.93 (0.09)	0.94 (0.09)
Overall	0.90 (0.06)	0.90 (0.08)	0.90 (0.08)	0.93 (0.05)

Note: 95% CI error bars.
Abbreviations: ACG, active control group; EG, experimental group; PCG, passive control group.

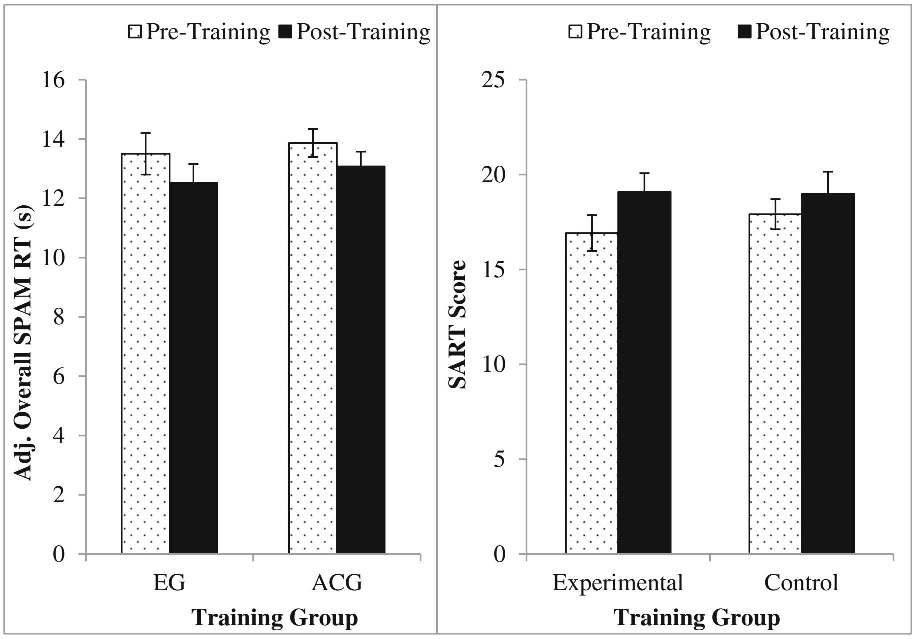

FIGURE 6 Objective and subjective SA: Adjusted SPAM RT (left panel) and SART scores (right panel). 95% CI error bars. ACG, active control group; EG, multitasking group

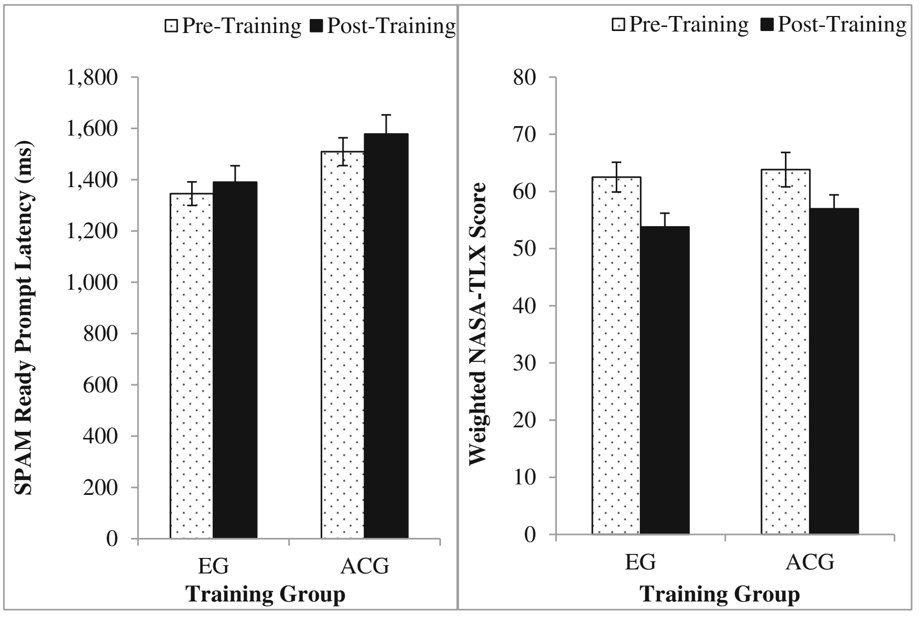

FIGURE 7 Far transfer task objective and subjective workload: SPAM ready prompt latency (ms, left panel) and weighted NASA TLX scores (right panel). 95% CI error bars. ACG, active control group; EG, multitasking group

4 | GENERAL DISCUSSION

Human operators are at substantial risk of being cognitively overloaded because of increasing demands imposed by task complexity and new technologies. One potential way to ameliorate this problem could be to increase the cognitive capacity of the human operator. Some promise in this regard has been shown by dual-task training paradigms in which participants complete several hours of a simple computer-based task designed to develop cognitive processes underlying multitasking (e.g., Liepelt et al., 2011). While such training has been shown to improve performance on the trained task over time in controlled laboratory conditions, it is less clear whether similar benefits could be achieved using a mobile app-based version of a training protocol that is less controlled, and, if so, whether training benefits could transfer to similar tasks (near transfer) or dissimilar tasks (far transfer) that rely on the same cognitive process that was trained.

With these questions in mind, we first created a mobile phone-based app that implemented a multitasking training program similar to Filmer, Lyons, et al. (2017). Following best-practice guidelines (Boot et al., 2011; Simons et al., 2016) we also implemented a visual search training program similar to Di Lollo et al. (2000) to be used by an active control group. Additionally, both apps used adaptive training procedures that changed in difficulty depending on participant

TABLE 5 Summary of Bayesian repeated measures ANOVA test results for each measure

Experimental component	Measure	Terms/effects	BF >1	Lee & Wagenmakers (2013); adjusted from Jeffreys (1961)	Support for
Near transfer task performance	Adj RT dual cost	Time	$BF_{incl} = 1.000$	Anecdotal	H1
		Group	$BF_{excl} = 3.040$	Moderate	H0
		Group*Time	$BF_{excl} = 3.165$	Moderate	H0
Far transfer task performance	Acceptance Adj RT	Time	$BF_{excl} = 2.809$	Anecdotal	H0
		Group	$BF_{excl} = 2.123$	Anecdotal	H0
		Group*Time	$BF_{excl} = 3.226$	Moderate	H0
	Handoff Adj RT	Time	$BF_{excl} = 3.731$	Moderate	H0
		Group	$BF_{excl} = 1.748$	Anecdotal	H0
		Group*Time	$BF_{excl} = 1.883$	Anecdotal	H0
	Conflict resolution Adj RT	Time	$BF_{incl} = 20474.569$	Extreme	H1
		Group	$BF_{excl} = 3.802$	Moderate	H0
		Group*Time	$BF_{excl} = 2.392$	Anecdotal	H0
Far transfer task SA	SPAM Adj RT	Time	$BF_{incl} = 33.752$	Very Strong	H1
		Group	$BF_{excl} = 2.278$	Anecdotal	H0
		Group*Time	$BF_{excl} = 4.082$	Moderate	H0
	SART	Time	$BF_{incl} = 1.489$	Anecdotal	H1
		Group	$BF_{excl} = 4.525$	Moderate	H0
		Group*Time	$BF_{excl} = 3.717$	Moderate	H0
Far transfer task workload	SPAM ready prompt all	Time	$BF_{excl} = 4.785$	Moderate	H0
		Group	$BF_{incl} = 4.156$	Moderate	H1
		Group*Time	$BF_{excl} = 4.444$	Moderate	H0
	NASA-TLX	Time	$BF_{incl} = 3267.637$	Extreme	H1
		Group	$BF_{excl} = 3.021$	Moderate	H0
		Group*Time	$BF_{excl} = 3.876$	Moderate	H0

Note: Showing BF_{incl} if evidence for alternative hypothesis, BF_{excl} if evidence for null-hypothesis.

performance to avoid ceiling effects and to maintain engagement. We then examined the impact of both multitasking and control training on performance, SA, and subjective workload in a simulated ATC task.

Addressing our first question, we found strong evidence that multitasking improved over the course of the mobile-app based training sessions, replicating earlier laboratory-based training outcomes (e.g., Bender et al., 2017; Filmer, Lyons, et al., 2017; Strobach et al., 2013). For instance, Bender et al. (2017) found 44.58% improvement in overall dual-cost and we found improvements of 11.59% for raw overall dual RT cost, 62.74% for raw overall dual accuracy cost and 54.73% for adjusted overall RT cost. Our results are also broadly similar to those obtained in training studies with other perceptual paradigms. For example, Enns et al. (2017) found performance improvements on an attentional blink task implemented on a mobile-based app similar to those seen in more controlled laboratory-based training regimens (Braun, 1998; Choi et al., 2012; Taatgen et al., 2009; Tang et al., 2014).

Addressing our second research question, multitasking training did not lead to near transfer of benefits to the same task with different stimuli. This adds to accumulating evidence in the literature that near transfer is the exception rather than the norm, at least with respect to multitasking (Bender et al., 2017; Garner et al., 2015; Horne et al., 2020). Additionally, multitasking training also failed to yield far transfer benefits on ATC performance, SA, or workload, suggesting using a simple dual-task paradigm is unlikely to benefit more complex tasks. To our knowledge, this represents the first study that has tested for far transfer arising from a multitasking training paradigm.

The evidence for near transfer in some past studies notwithstanding, our results align closely with the literature on expertise acquisition (Farrington-Darby & Wilson, 2006; Healy et al., 2014; Lesgold, 1983; Samuels & Flor, 1997), which suggests that training narrowly defined tasks leads to benefits as a result of automatization of underlying processes specific to the trained task. This automatization is thought to reduce the impact of limitations in cognitive capacity by fostering information chunking and encapsulation, the development of which relies on constancy of stimuli and procedure (Brown & Carr, 1989; Logan et al., 1996). From this argument, it follows that automaticity of task-specific processes is unlikely to lead to general improvements in other tasks that have different stimuli and/or procedures.

Of course, alternative explanations for our failure to obtain near- or far-transfer effects could be made based on particular elements of our methodology. For example, it could be possible that this failure reflects the app-based mode of training. However, we consider this unlikely given our success at showing improvement on the training task, analogous to many studies in the literature. Additionally, it could be suggested that the improvements in conflict detection and subjective SA across testing sessions might reflect beneficial effects of both dual-task and visual search training. Although we think it is more likely that practice effects are responsible for these improvements, one way to address this question more directly would be to replicate the current study with the addition of a passive control group that completed no training. If the impact of testing session reflects improvement arising from both multitasking and visual search training, then the passive control group should not show similar changes across testing session. On the other hand, if the improvement reflects practice effects, then it should also occur in the passive control group.

Another possibility is that with further multitasking training, we might see greater benefits to underlying cognitive skills and that this would be sufficient to transfer to other tasks. We think that insufficient training is unlikely to explain our failure to find near or far transfer for the following reason. The length of our training regimen was comparable to those in previous studies, which ranged from only a sixth of our number of trials (i.e., 465 trials; Bender et al., 2017) to twice the amount (i.e., 5694 trials; Liepelt et al., 2011).

4.1 | Limitations and future research

It is well known that training transfer can be affected by motivation levels and individual differences such as cognitive ability, self-efficacy, and goal orientation (Salas & Cannon-Bowers, 2001; also see review by Baldwin & Ford, 1988). This may be relevant because our sample consisted of student and community participants recruited from a university setting. Though this sample is likely highly representative of new entrants into the workforce who will be confronted by significant cognitive challenges on the job, they are not broadly representative of the population. In particular, our participants are likely to have had above-average cognitive ability. Moreover, students are closer to the age when cognitive ability peaks than the general population (Ardila, 2007). These characteristics may have limited our ability to detect benefits of multitasking training due to ceiling effects. In addition, university students might not see the benefits of cognitive training, as opposed to participants from a more general pool for whom this might be more job relevant. With these points in mind, attempting to replicate our training effects in older and more diverse samples may lead to different outcomes.

Another potential limit to the effectiveness of our training regimen stems from the relatively modest engagement participants reported with the app (average "fun" rating of 5.03/10). It could be reasonably expected that improving participants' enjoyment while using the app might improve training engagement and therefore the magnitude of training benefits. Although, as noted above, training effects were highly significant and similar in magnitude to other studies that did find near transfer, improved engagement could nonetheless theoretically lead to greater benefit. This aligns with expectations of improved training engagement and subsequent training results arising from "gamification" (see Anguera et al., 2013; Anguera & Gazzaley, 2015; Bediou et al., 2018), and suggests that future efforts to make cognitive training apps more akin to video games may yield beneficial outcomes.

There is also some reason to believe that attempting to train a general cognitive process using a single task may itself be a flawed approach. Studies of individual differences in cognitive ability often use a latent factors approach to assess ability, combining performance from several paradigms to tap a common cognitive ability to derive a reliable measure. This approach acknowledges that performance on any one task is likely determined to a relatively small extent by an underlying cognitive ability and to a much larger extent by situational variables, participant factors, and measurement error (Redick et al., 2016). A reasonable corollary of this logic is that repeatedly performing a single training task should, at best, lead to limited improvement to the underlying cognitive ability associated with the task because much of task performance reflects the action of other factors.

An additional consideration is that successful multitasking depends upon other underlying cognitive skills such as working memory and attentional control (e.g., Redick et al., 2016), and these more fundamental skills may be differentially tapped depending on the type of multitasking being performed. This raises two issues. First, trying to improve multitasking broadly through training on a single task is likely to be unsuccessful unless the training task strongly taps into all of the relevant underlying cognitive skills (Harrison et al., 2013). Second, efforts to show near- or far-transfer of training will be differentially successful depending on whether the training task taps into the same fundamental cognitive skills as the near- or far-transfer task. For example, multitasking training that improves working memory may benefit multitasking in other tasks that depend heavily on working memory but be less effective if the other tasks depend heavily on attentional control. This analysis suggests that simultaneous training on several types of tasks associated with a cognitive process (particularly a complex one) would be much more likely to be effective than the current single-task training approach.

4.2 | Practical implications

In terms of practical implications, first, this study has added to the evidence that it is possible to achieve similar improvements with participant-administered and -scheduled mobile app-based multitasking training as it is with the more controlled laboratory-based computer training. Given the flexibility and ease of deployment as well as the maturity of the existing mobile app technology, our results suggest that implementation of training programs on mobile platform is as likely to be successful and more easily accessed than using traditional digital educational platforms.

Second, this study has added to the evidence that there is a lack of training-related near- and far transfer benefits arising from current cognitive training methodologies. With the increased evidence against these methodologies, the path is being cleared to start focusing on finding alternative approaches to implementing cognitive training, in order to improve its effectiveness.

4.3 | Conclusion

In summary, this study has shown that app-based cognitive training can produce similar improvements over the course of training as earlier studies that conducted training in more controlled environments. However, benefits arising over the course of training did not yield near- or far-transfer to novel tasks thought to tap the same underlying cognitive process. We acknowledge that this conclusion rests on reporting evidence for the null in this article. However, we believe it is crucial that null findings are published in order to avoid the potential "file drawer" problem. Indeed, the importance of establishing the robustness and generality of psychological effects has received much attention (Pashler & Wagenmakers, 2012; Yong, 2012), and is particularly vital when the resulting knowledge could be used by practitioners in safety-critical work settings (Jones et al., 2010). It is also critical that meta-analysis of research topics like cognitive training include all results obtained, not just the positive cases.

ACKNOWLEDGMENTS

We gratefully acknowledge that the Commonwealth of Australia supported this research through an Australian Government Research Training Program Scholarship and a Defense Science and Technology Group (DSTG) Top-Up Scholarship to S.B., through the Australian Army and Defense Science Partnerships agreement of the Defense Science and Technology Group (ID 7120), as part of the Human Performance Research network (HPRnet). We thank Dr Ryan Li for programming the mobile apps, Dr Michael English for programming the PC version of the dual-task training paradigm, and Dr Vanessa Bowden for her help with the ATC task. Open access publishing facilitated by The University of Western Australia, as part of the Wiley - The University of Western Australia agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Stephanie C. Black https://orcid.org/0000-0001-9321-745X

ENDNOTE

[1] A passive control group was not used to avoid the introduction of differing motivation and expectancy effects (Shipstead et al., 2012)

REFERENCES

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors*, *37*(1), 85–104.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., & Gazzaley, A. (2013). Video game training enhances cognitive active control in older adults. *Nature*, *501*(7465), 97–101.

Anguera, J. A., & Gazzaley, A. (2015). Video games, cognitive exercises, and the enhancement of cognitive abilities. *Current Opinion in Behavioral Sciences*, *4*, 160–165.

Ardila, A. (2007). Normal aging increases cognitive heterogeneity: Analysis of dispersion in WAIS-III scores across age. *Archives of Clinical Neuropsychology*, *22*(8), 1003–1011.

Baddeley, A. D., & Logie, R. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge University Press.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, *41*(1), 63–105.

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, *144*, 77–110. https://doi.org/10.1037/bul0000130

Bender, A. D., Filmer, H. L., Naughtin, C. K., & Dux, P. E. (2017). Dynamic, continuous multitasking training leads to task-specific improvements but does not transfer across action selection tasks. *Nature Partner Journals: Science of Learning*, *2*(1), 14.

Bender, A. D., Loft, S. D., Lipp, O. V., Bowden, V. K., Whitney, S. J., & Visser, T. A. W. (2018, November). Situation awareness mediates the relationship between working memory ability and performance on the air traffic control task. *Paper presented at the Defence Human Sciences Symposium*, Perth, Australia.

Boot, W. R., Blakely, D. P., & Simons, D. J. (2011). Do action video games improve perception and cognition? *Frontiers in Psychology*, *2*, 226.

Bowden, V. K., & Loft, S. (2016). Using memory for prior aircraft events to detect conflicts under conditions of proactive air traffic control and with concurrent task requirements. *Journal of Experimental Psychology: Applied*, *22*(2), 211–224. https://doi.org/10.1037/xap0000085

Braun, J. (1998). Vision and attention: The role of training. *Nature*, *393*(6684), 424–425.

Brown, T. L., & Carr, T. H. (1989). Automaticity in skill acquisition: Mechanisms for reducing interference in concurrent performance. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 686.

Chambers, C. D., Stokes, M. G., Janko, N. E., & Mattingley, J. B. (2006). Enhancement of visual selection during transient disruption of parietal cortex. *Brain Research*, *1097*(1), 149–155. https://doi.org/10.1016/j.brainres.2006.04.084

Chérif, L., Wood, V., Marois, A., Labonté, K., & Vachon, F. (2018). Multitasking in the military: Cognitive consequences and potential solutions. *Applied Cognitive Psychology*, *32*(4), 429–439.

Chiappe, D., Morgan, C. A., Kraut, J., Ziccardi, J., Sturre, L., Strybel, T. Z., & Vu, K.-P. L. (2016). Evaluating probe techniques and a situated theory of situation awareness. *Journal of Experimental Psychology: Applied*, *22*(4), 436–454. https://doi.org/10.1037/xap0000097

Choi, H., Chang, L. H., Shibata, K., Sasaki, Y., & Watanabe, T. (2012). Resetting capacity limitations revealed by long-lasting elimination of attentional blink through training. *Proceedings of the National Academy of Sciences*, *109*(30), 12242–12247.

Dehais, F., Causse, M., Vachon, F., & Tremblay, S. (2012). Cognitive conflict in human-automation interactions: A psychophysiological study. *Applied Ergonomics*, *43*, 588–595.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, *129*(4), 481–507.
Durso, F., & Dattel, A. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Trembley (Eds.), *A cognitive approach to situation awareness: Theory, measures and application* (pp. 137–154). Aldershot.
Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., & Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron*, *63*(1), 127–138.
Endsley, M. R. (1995a). A taxonomy of situation awareness errors. In R. Fuller, N. Johnston, & N. McDonald (Eds.), *Human factors in aviation operations* (pp. 287–292). Avebury Aviation, Ashgate Publishing Ltd.
Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64.
Endsley, M. R. (2006). Situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 528–542). Wiley.
Endsley, M. R., & Jones, D. (2012). *Designing for situation awareness: An approach to user-centered design* (2nd ed.). CRC Press.
Enns, J., Kealong, P., Tichon, J., & Visser, T. (2017). Training and the attentional blink: Raising the ceiling does not remove the limits. *Attention, Perception, & Psychophysics*, *79*, 2257–2274. https://doi.org/10.3758/s13414-017-1391-9
Farrington-Darby, T., & Wilson, J. R. (2006). The nature of expertise: A review. *Applied Ergonomics*, *37*(1), 17–32.
Filmer, H. L., Lyons, M., Mattingley, J. B., & Dux, P. E. (2017). Anodal tDCS applied during multitasking training leads to transferable performance gains. *Scientific Reports*, *7*(1), 12988.
Filmer, H. L., Varghese, E., Hawkins, G. E., Mattingley, J. B., & Dux, P. E. (2017). Improvements in attention and decision-making following combined behavioral training and brain stimulation. *Cerebral Cortex*, *27*(7), 3675–3682.
Fothergill, S., Loft, S., & Neal, A. (2009). ATC-lab^Advanced: An air traffic control simulator with realism and control. *Behavior Research Methods*, *41*, 118–127. https://doi.org/10.3758/BRM.41.1.118
Garner, K. G., Matthews, N., Remington, R. W., & Dux, P. E. (2015). Transferability of training benefits differs across neural events: Evidence from ERPs. *Journal of Cognitive Neuroscience*, *27*(10), 2079–2094.
Garner, K. G., Tombu, M. N., & Dux, P. E. (2014). The influence of training on the attentional blink and psychological refractory period. *Attention, Perception, & Psychophysics*, *76*(4), 979–999.
Gronau, Q. F., Ly, A., & Wagenmakers, E. J. (2020). Informed Bayesian t-tests. *The American Statistician*, *74*(2), 137–143. https://doi.org/10.1080/00031305.2018.1562983
Gugerty, L. (2011). Situation awareness in driving. In J. Lee, M. Rizzo, D. Fisher, & J. Caird (Eds.), *Handbook for driving simulation in engineering, medicine and psychology* (pp. 265–268). CRC Press.
Gutzwiller, R. S., Wickens, C. D., & Clegg, B. A. (2019). The role of reward and effort over time in task switching. *Theoretical Issues in Ergonomics Science*, *20*(2), 196–214.
Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, *24*(12), 2409–2419.
Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in psychology* (Vol. 52, pp. 139–183). North-Holland Press.
Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2014). Training principles to advance expertise. *Frontiers in Psychology*, *5*, 131.
Horne, K. S., Filmer, H. L., Nott, Z. E., Hawi, Z., Pugsley, K., Mattingley, J. B., & Dux, P. E. (2020). Evidence against benefits from cognitive training and transcranial direct current stimulation in healthy older adults. *Nature Human Behaviour*, *5*, 146–158. https://doi.org/10.1038/s41562-020-00979-5
Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
Jones, K. S., Derby, P. L., & Schmidlin, E. A. (2010). An investigation of the prevalence of replication research in human factors. *Human Factors*, *52*, 586–595. https://doi.org/10.1177/0018720810384394
Jones, P. R. (2019). A note on detecting statistical outliers in psychophysical data. *Attention, Perception, & Psychophysics*, *81*, 1189–1196. https://doi.org/10.3758/s13414-019-01726-3
Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Prentice-Hall.
Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search exploring the boundaries of executive attention. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *32*, 749–777.
Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin*, *144*(6), 557–583.
Krzywinski, M., & Altman, N. (2013). Points of significance: Error bars. *Nature Methods*, *10*, 921–922. https://doi.org/10.1038/nmeth.2659
Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
Lesgold, A. M. (1983). *Acquiring expertise* (No. UPITT/LRDC/ONR/PDS-5). Pittsburg University, Pennsylvania. Learning Research and Development Centre.
Liepelt, R., Strobach, T., Frensch, P., & Schubert, T. (2011). Improved intertask coordination after extensive dual-task practice. *Quarterly Journal of Experimental Psychology*, *64*(7), 1251–1272.
Loft, S. (2014). Applying psychological science to examine prospective memory in simulated air traffic control. *Current Directions in Psychological Science*, *23*, 326–331.
Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: Incorporating environmental constraints. *Journal of Experimental Psychology: Applied*, *15*(2), 106–124. https://doi.org/10.1037/a001611
Logan, G. D., Taylor, S. E., & Etherton, J. L. (1996). Attention in the acquisition and expression of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 620–638.
National Health and Medical Research Council. (2007). *National statement on ethical conduct in human research*. National Health and Medical Research Council.
Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*(1), 44–64.
Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, *144*(7), 710–756. https://doi.org/10.1037/bul0000151
Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220–244.
Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*, 528–530.
Pritchett, A. (2015). Preface to the JCEDM special issue on situation awareness. *Journal of Cognitive Engineering and Decision Making*, *9*(1), 3. https://doi.org/10.1177/1555343415572807
Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., Kane, M. J., Hambrick, D. Z., & Engle, R. W. (2016). Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General*, *145*(11), 1473–1492.
Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001
Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, *52*(1), 471–499.
Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring situation awareness in

complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, *39*(3), 490–500.

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *13*(2), 107–121.

Schuch, S., Dignath, D., Steinhauser, M., & Janczyk, M. (2019). Monitoring and control in multitasking. *Psychonomic Bulletin and Review*, *26*, 222–240. https://doi.org/10.3758/s13423-018-1512-z

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., & Meyer, D. E. (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science*, *12*, 101–108.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–654.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*(3), 103–186.

Strobach, T., Liepelt, R., Pashler, H., Frensch, P. A., & Schubert, T. (2013). Effects of extensive dual-task practice on processing stages in simultaneous choice tasks. *Attention, Perception, & Psychophysics*, *75*(5), 900–920.

Strybel, T. Z., Vu, K. P. L., Kraft, J., & Minakata, K. (2008). Assessing the situation awareness of pilots engaged in self spacing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(1), 11–15.

Taatgen, N. A., Juvina, I., Schipper, M., Borst, J. P., & Martens, S. (2009). Too much control can hurt: A threaded cognition model of the attentional blink. *Cognitive Psychology*, *59*(1), 1–29.

Tang, M. F., Badcock, D. R., & Visser, T. A. W. (2014). Training and the attentional blink: Limits overcome or expectations raised? *Psychonomic Bulletin & Review*, *21*(2), 406–411.

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational awareness in aerospace operations (AGARD-CP-478)* (pp. 3/1–3/17). NATO-AGARD.

Tombu, M., & Jolicoeur, P. (2004). Virtually no evidence for virtually perfect time-sharing. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 795–810. https://doi.org/10.1037/0096-1523.30.5.795

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modelling of elementary psychological processes*. Cambridge Univ. Press.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Komarlu Narendra Gupta, A. R., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*(3), 813–826. https://doi.org/10.3758/s13423-020-01798-5

Verghese, A., Mattingley, J. B., Garner, K. G., & Dux, P. E. (2018). Decision-making training reduces the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 195–205.

Vu, K., & Chiappe, D. (2015). Situation awareness in human systems integration. In D. A. Boehm-Davis, F. T. Durso, & J. D. Lee (Eds.), *APA handbook of human systems integration* (pp. 293–308). American Psychological Association. https://doi.org/10.1037/14528-019

Vu, K. P. L., Strybel, T. Z., Battiste, V., Lachter, J., Dao, A. Q. V., Brandt, S., Ligda, S., & Johnson, W. (2012). Pilot performance in trajectory-based operations under concepts of operation that vary separation responsibility across pilots, air traffic controllers, and automation. *International Journal of Human-Computer Interaction*, *28*(2), 107–118.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455.

Woods, D. D., & Sarter, N. (2010). Capturing the dynamics of attention active control from individual to distributed systems: The shape of models to come. *Theoretical Issues in Ergonomics Science*, *11*(1), 7–28.

Yong, E. (2012). In the wake of high-profile controversies, psychologists are facing up to problems with replication. *Nature*, *483*, 298–300.

How to cite this article: Black, S. C., Bender, A. D., Whitney, S. J., Loft, S., & Visser, T. A. W. (2022). The effect of multi-tasking training on performance, situation awareness, and workload in simulated air traffic control. *Applied Cognitive Psychology*, *36*(4), 874–890. https://doi.org/10.1002/acp.3977

APPENDIX A

TABLE A1 Experiment schedule for participants on five-day training cycle

Day	Group: Multitasking	Group: Active visual search
1	Informed consent Pre-testing • Questionnaires • ATC introduction and practice • ATC simulation 1st Scenario incl. SPAM and SART • NASA-TLX questionnaire Training • 1 Session dual-task app	Informed consent Pre-testing • Questionnaires • ATC introduction and practice • ATC simulation 1st Scenario incl. SPAM and SART • NASA-TLX questionnaire Training • 1 Session visual search app
2	Training • 2 Sessions dual-task app	Training • 2 Sessions visual search app
3	Training • 2 Sessions dual-task app	Training • 2 Sessions visual search app
4	Training • 2 Sessions dual-task app	Training • 2 Sessions visual search app
5	Training • 1 Session dual-task app Post-testing • Training questionnaire • ATC simulation 2nd Scenario incl. SPAM and SART • NASA-TLX questionnaire Debrief	Training • 1 Session visual search app Post-testing • Training questionnaire • ATC Simulation 2nd Scenario incl. SPAM and SART • NASA-TLX questionnaire Debrief

TABLE A2 SPAM presented used during the ATC simulation

Scenario A
Q1—What aircraft is on the same flight level as aircraft QF45 in the NE quadrant?
Q2—In which quadrant will the next potential conflict take place if no action has been taken or will be taken within the next 30 s?
Q3—What is the flight level of the aircraft that you last accepted to the VA route?
Q4—What common waypoint will aircraft QF94 and aircraft AA36 both pass?
Q5—At what altitude is aircraft NZ20 traveling?
Q6—What is the flight level of the two aircraft that you last accepted to the QF and SQ route?
Q7—What waypoint are aircraft QF59 and SQ27 currently closest to?
Q8—Will aircraft AA58 and aircraft NZ17 cross path in the NW, NE, SW, or SE quadrant?
Q9—What waypoint do aircraft AA35 and QF52 both have to cross?
Q10—Are aircraft SQ47 and VA37 traveling at the same speed?
Q11—What is the next waypoint that QF87 has to cross?
Q12—What is the flight level of the aircraft that you last accepted to the EK route?
Q13—What is the flight level of the aircraft that you last accepted to the VA route?
Q14—Are aircraft AA63 and aircraft NZ19 currently located in the NW, NE, SW, or SE quadrant?
Q15—Which quadrant currently has the most number of aircraft inside the flight sector?
Q16—In which quadrant will the next potential conflict take place if no action has been taken or will be taken within the next 30 s?
Q17—What waypoint does the last accepted aircraft on the AA route have to cross next?
Q18—What aircraft did you last hand off on the QF route?
Scenario B
Q1—Are aircraft EK23 and aircraft AA31 currently located in the NW, NE, SW, or SE quadrant?
Q2—What aircraft is on the same flight level as aircraft SQ57?
Q3—In which quadrant will the next loss of separation take place within the next 30 s if no action will be taken?
Q4—What was the flight level of the aircraft that you last accepted into the sector in the NE quadrant?
Q5—What aircraft is on the same flight level as aircraft AA37 in the SW quadrant?
Q6—Will aircraft SQ79 and aircraft VA32 cross path in the NW, NE, SW, or SE quadrant?
Q7—How many aircraft needed accepting within the last 30 s in the SW quadrant?
Q8—Which quadrant currently has the most number of aircraft in the flight sector?
Q9—What were the flight level of the two aircraft that you last accepted in to the sector?
Q10—Closest to which waypoint will aircraft AA83 and aircraft NZ55 cross path?
Q11—What is the next waypoint that QF40 has to cross?
Q12—What was the flight level of the aircraft that you last accepted in to the sector?
Q13—Will aircraft QF59 and aircraft EK85 cross path in the NW, NE, SW, or SE quadrant?
Q14—Are aircraft VA28 and SQ56 traveling at the same speed?
Q15—Which quadrant currently has the most number of aircraft in the flight sector?
Q16—How many aircraft did need handing off within the last 30 s?
Q17—Will aircraft VA95 and aircraft NZ33 cross path in the NW, NE, SW, or SE quadrant?
Q18—What common waypoint will aircraft SQ51 and aircraft QF81 both pass?