A Comparison of the Effects of Augmented Reality N-Back Training and Traditional Two-Dimensional N-Back Training on Working Memory

Bo Zhang and Nigel Robb

Abstract

We compared two versions of an n-back training program, differing from the graphical perspective, on the effects of working memory (WM) training and transfer. Sixty participants were trained on a traditional n-back task (2D perspective) or an augmented reality (AR) version of the same program. The AR version was rated more engaging and graphically stimulating. Pre- and post-performance on a 2D spatial working memory (SWM) test showed that while both groups improved, the distributions of improvement differed significantly between the groups; the group using traditional training showed generally more improvement on the easier levels of the SWM test. These results may be explained by the fact that the traditional (2D) version of the n-back training was more similar than the AR version to the SWM outcome measure (in virtue of its 2D presentation). This may support the common demands theory of training transfer, which claims that shared demands between training and transfer tasks accounts for near-transfer improvements.

Keywords

working memory, traditional n-back task, spatial working memory test, augmented reality n-back task, training transfer

Introduction

Working Memory and Cognitive Training

Working memory (WM) is an essential cognitive ability involved in many daily activities, such as managing tasks, problem-solving, and interacting with others (Tulbure & Siberescu, 2013). WM is the ability to temporarily store, process, and manipulate information during mental activities (Au et al., 2015; Jaeggi et al., 2010b). Studies have found that WM has a close relationship with fluid intelligence (Gf) (Engle et al., 1999) and that the improvement of WM can enhance Gf (Jaeggi et al., 2008; Jaeggi et al., 2010b; Wiley et al., 2011). Gf is the ability to think logically, reason, and solve problems in novel situations without relying on previously acquired knowledge (Au et al., 2015). Therefore, working memory improvement is closely related to learning and daily life, such as school preparation (Bierman et al., 2008), the development of academic skills (Allan et al., 2014), and mental health (Short et al., 2016).

Many studies demonstrate the potential of cognitive training to improve cognitive functions in healthy young adults (Tulbure & Siberescu, 2013), older adults (Anguera et al., 2013), children (Van Dongen-Boomsma et al., 2014), and even in clinical populations (Bahar-Fuchs et al., 2013; Robb et al., 2018). Even relatively short periods of computerized training can lead to significant improvements in the working memory and attention capacity of young healthy adults (Akter et al., 2015; Tulbure & Siberescu, 2013). This evidence indicates that working memory is malleable and can be improved by repeatedly engaging in cognitive training (Shipstead et al., 2012; Zhang et al., 2019), even in participants with no previous deficits (Tulbure & Siberescu, 2013). However, other previous research has found limited or no improvements in cognitive function after WM training (Melby-Lervåg & Hulme, 2013; Simons et al., 2016). In terms of the cognitive task used to measure improvement, it is common in the literature to distinguish between near- and far-transfer tasks. Transfer tasks that share many components with the trained task are used to illustrate near transfer, whereas tasks that share fewer components are seen as far...
transfer (Oei & Patterson, 2015). Simons et al. (2016) have found that training on one WM task can transfer to other similar tasks (near transfer), while producing no transfer to tasks that are dissimilar to the trained task (far transfer). In addition, research suggests that the near-transfer effects were more likely to sustain in visuospatial WM training than verbal WM training (Melby-Lervåg & Hulme, 2013). Also, stronger motivation and engagement may lead to better improvements following cognitive training, as participants are more engaged in the training (Katz et al., 2014). Hence, through innovating the visuospatial form of Virtual Reality [VR] training task (such as using augmented reality [AR]), this study attempts to explore the way of how to make the WM training more engaging and effective.

**Traditional Computerized Task for Working Memory Improvement**

A variety of computerized tasks have been used in WM training, including the n-back task which is used extensively (Forns et al., 2014; Jones et al., 2018). Studies find that training in performing n-back tasks can result in improvements in WM capability (Jaeggi et al., 2008; Jaeggi et al., 2010a). During n-back training, participants must deal with multiple processes such as observation, decision-making, selection, inhibition, interference resolution, and more (Jaeggi et al., 2010a). This process involves several aspects of cognition, including attention, memory updating, and working memory (Cormack et al., 2016). When completing an n-back task, the participant is presented with one or more series of stimuli (e.g., the location of the stimulus in a grid). A response is required whenever the current stimulus matches the stimulus in n positions back in the sequence (e.g., 1, 2, or 3) (Jaeggi, Buschkuehl, et al., 2010).

A typical example (Figure 1) from the work of Jaeggi et al. shows that an n-back task can be set with a single task with only one factor, such as visuospatial-nonverbal material in each trial (Jaeggi et al., 2010a; Jaeggi et al., 2008). Moreover, each trial generally consisted of a stimulus that was presented for 500 ms, followed by an interstimulus interval (ISI) of 2,500 ms, after which the next stimulus was presented.

The difficulty of n-back task training can be increased by rising the value of n. When the processing load increases systematically by manipulating the value of n, one of the most important dimensions in measuring improvement is participant selection accuracy (Jonides et al., 1997).

Traditionally, the n-back task needs to be installed and run on a computer screen; however, in recent years, it is available on the internet or mobile devices (e.g., smartphones, IPad; Tulbure & Siberscru, 2013). As technology rapidly advances, many researchers around the world are continuously exploring, trying to find new interventions that are more effective than traditional methods (Au et al., 2015; Katz et al., 2014). Some studies modified the features or forms of the traditional n-back task to make it more like a game and to investigate the effects on working memory (Katz et al., 2014; Nagle et al., 2015). These studies also proposed that future research might investigate whether AR technology can provide a more efficient way of cognitive training.

**AR for Cognitive Training**

AR, as a new way to understand and learn about the real world, allows users to learn virtual information in a real context, extending the depth of learning about the real environment (Akçayır et al., 2016; Sungkur et al., 2016). AR is an
interactive experience in which a real-world environment is enhanced with perceptual information generated using computers (Schueffel, 2017). AR applications present virtual objects which appear to be situated in the real world (e.g., by augmenting the feed from a smartphone camera so that a 3D cube appears to be placed on the floor that the user is viewing through the camera, see Figure 2), to create the sensation of immersion in a contextual learning environment which combines real and virtual elements (Sungkur et al., 2016). This simple means of interaction creates a new mode of learning, which is easily used even by students who have no prior experience of using computers (Lu & Liu, 2015).

Recently, AR has been used to create novel and mobile tasks or to improve cognitive health. Studies have found that AR technologies have the capability to effectively induce psychological reactions in users (Brito & Stoyanova, 2018). A trend has emerged in which researchers have begun to explore the feasibility of using AR technology to create cognitive training tasks (Boletsis & McCallum, 2014, 2016). In addition, AR technology is strongly connected to a user’s cognitive and physical functionality, as it has a positive impact on mental processes and supports spatial cognition (Scanlon et al., 2016).

Moreover, some studies have found evidence that AR-based cognitive training is effective in improving a participant’s ability to perform. For example, AR-based cognitive training can enhance the behavioral and cognitive functions of children with attention deficit (Shema-Shiratzky et al., 2018) and can improve spatial visualization in older people (Boletsis & McCallum, 2016; Hoe et al., 2019). Furthermore, studies suggest that AR plays a significant role in increasing students’ motivation and attention throughout the learning process (Bacca et al., 2014). When AR provides users with an interesting and pleasant virtual scientific experience, they may enjoy the learning process more, which may shorten their perception of time spent learning new information (Brito & Stoyanova, 2018). Some studies argue that the application of AR in educational settings can help students

![Figure 2. Traditional n-back task and AR n-back task. Note. AR = augmented reality.](image-url)
learn more effectively and increase knowledge retention relative to traditional 2D desktop interfaces (Billinghurst et al., 2015; Mitrovic et al., 2009).

To the best of our knowledge, no studies have investigated the differences between AR and traditional interfaces in the context of WM training. Investigating this difference is important, as the positive experience associated with AR may lead to greater engagement with cognitive training. However, it is also possible that AR training may be less effective in terms of transfer to laboratory measures of WM due to dissimilarities between the training and laboratory tasks (e.g., Oei & Patterson, 2015, argue that training transfer depends on common demands between the trained task and the transfer task). Thus, in this study, we compared the effects of traditional WM training with AR WM training in a controlled training experiment.

Method

A Traditional N-back Task and an AR N-back Task

In this study, we compared the effects of training with traditional and AR versions of a simple visuospatial n-back task. The visuospatial stimulus consisted of green squares appearing in one of nine different locations spaced on a grid on a black screen. Figure 2 (the figures of the traditional n-back task and the AR n-back task) indicates the main differences between the two versions of n-back tasks. In the AR version, the player perceives the stimulus as located in real-world surroundings, while in the traditional version, the stimulus is perceived as being located on the screen of the device. Both versions were built on a game engine called Unity (unity.com). The AR n-back task was developed using a combination of the Unity game engine and Apple’s ARKit AR Framework.

According to the principles described in the previous studies (Jaeggi et al., 2007, 2008, 2009), both versions of the n-back tasks were developed with the same specific rules. The tasks were used with three load levels (1-back, 2-back, and 3-back). On the grid, a response is needed whenever a stimulus is presented onscreen that matches the one presented in n positions back in the sequence (Jaeggi et al., 2007, 2008). As shown in Figure 2, participants were required to press the “Match” button (located at the bottom right of the screen) for correct targets during performing the n-back task and no responses were required for incorrect targets. After several user tests, we determined the speed of the visuospatial stimulus in this study. Each trial consisted of a stimulus that was presented for 2,000 ms followed by an ISI of 2,500 ms, after which the next stimulus was presented (Jaeggi et al., 2010a).

Two groups of participants performed separate experimental sessions of traditional n-back tasks and AR n-back tasks. Each session included a 1-back task, 2-back task, or 3-back task in each day of training. Each task trial lasted 3 min. Within this time frame, the visuospatial stimulus in different trials (n = 1, 2, or 3) was arranged in a pseudo-randomized order, which was matched for the number of targets (33%) and non-targets (67%) (Jaeggi et al., 2010a). Participants were encouraged to respond to the correct targets as quickly and accurately as possible. The processing load can be varied by manipulating the value of n, which can be shown by changes in accuracy (Jonides et al., 1997). Hence, in this study, the performances on both n-back tasks were assessed in terms of participant accuracy, which includes scoring and false clicks.

Participants

A total of 60 participants took part in the experiment (average age = 19.33 years, SD = 1.15 years, 29 females). They were undergraduate students from a university in China who were fulfilling a course credit, without any specified selection criteria. All participants received a payment of 15 dollars to encourage them to take the trial seriously. Written informed consent was collected from the students before they began the working memory training and participants were informed that they could drop out of the study at any time. Prior to beginning the study, ethics approval was obtained from the institutional ethics committee.

Materials

Measure of spatial working memory. To measure WM changes, we used the spatial working memory (SWM) task from the Cambridge Neuropsychological Test Automated Battery (CANTAB) (Owen et al., 1990; Robbins et al., 1994; Saults et al., 1999).

The CANTAB SWM test begins with a number of colored boxes (4, 6, 8, and 12 boxes in this study) being shown on the screen of an iPad (see Figure 3). The aim of the SWM test is that, by process of elimination, the participant should find a yellow “token” within a number of boxes by hand touching and using them to fill up an empty column on the right-hand side of the screen. The key point is that once a token has been located from a particular box, it should not appear there again during that particular trial. The number of boxes is gradually increased from 4 to 12 (4, 6, 8, 12), with 12 being the most difficult. To discourage the use of stereotyped search strategies, the colors and positions of the boxes change from trial to trial.

Currently, CANTAB SWM offers two difficulty level tests: Recommended Standard 2.0 (SWM Standard) and Recommended Standard 2.0 Extended (SWM Extended). Two significant terms are used in the CANTAB SWM tests to measure WM changes:

1. SWM between errors: The number of times the subject incorrectly revisits a box in which a token has previously been found.
2. SWM strategy: The number of times a subject begins a new search pattern from the same box they started with previously.

SWM Standard test collects data of five key variables: between errors across trials with 4 boxes, between errors across trials with 6 boxes, between errors across trials with 8 boxes, between errors across trials with 4, 6, and 8 boxes (SWMBE468) and strategy for 6 and 8 boxes trials (SWMS). Compared with the SWM Standard test, the key variables of the SWM Extended test are expanded with two more variables: between errors on trials with 12 boxes (SWMBE12) and strategy for 12 boxes trials (SWMSX). Neither CANTAB SWM tests collect the data of strategy for 4 box trials because the 4 box trials is not difficult enough for participants to apply complex strategy on.

In the past decades, many studies used two key variables of CANTAB SWM: total between errors (SWMBE468) and strategy on 6 and 8 boxes (SWMS) to predict the cognitive function on orders (e.g., Csipo et al., 2019; Wu et al., 2020) and to investigate the effects of WM training on the cognitive development of impaired populations (e.g., Cacciamani et al., 2018; Cocchi et al., 2009). Based on these similar study aims, therefore, this study decided to use SWMBE468 and SWMS as the key variables to investigate the effects of WM training. However, as the participants of this study were young healthy adults, the SWM Extended test (with two more expanded key variables: SWMBE12 and SWMSX) was selected as the WM measurement tool. SWMBE12 and SWMSX can be used in impaired populations, but also healthy controls, due to the more difficult stages mitigating ceiling effects. Hence, this study separately analyzes the data of SWMBE468 and SWMBE12 (between errors), along with the analysis of SWMS and SWMSX (strategy) to measure healthy participants’ WM changes.

In addition, the SWM Extended model also collected basic demographic information, including age, sex, and the highest level of education per participant.

### Table 1. User Experience Questionnaire (Nine Items).

| Questions                                                                 |
|---------------------------------------------------------------------------|
| The n-back task was easy to play                                           |
| The text within the n-back task was easy to understand                     |
| The design of the n-back task screen made sense to me                      |
| The graphics of the n-back task were stimulating                          |
| When I was playing the n-back task, I was totally absorbed in it           |
| When I was playing the n-back task, I felt just the right amount of challenge |
| The game was too fast                                                     |
| The game was too slow                                                     |
| Overall, I enjoyed using this software                                     |

**User experience questionnaires.** Following training, participants completed a user experience questionnaire (see Table 1) for the respective training programs (traditional and AR). The questionnaire was written in Chinese and English. Nine questions within the questionnaire were developed based on usability and user experience surveys from Nielsen’s Heuristic Evaluation (Nielsen, 1993) and the Computer System Usability Questionnaire (CSUQ) (Lewis, 1995). In addition, two questions were based on the Flow Short Scale (Engeser & Rheinberg, 2008), which is designed to determine a person’s level of flow experience (Csikszentmihalyi & Csikszentmihalyi, 1992), which is recognized as an enjoyable state of absorption and optimal challenge that is associated with positive outcomes in e-learning (Rodríguez-Ardura & Meseguer-Artola, 2017). All items were 7-point Likert-type-style scales with possible response ranging from 1 (strongly disagree) to 7 (strongly agree).

### Training Procedure

A flowchart of the study procedure is illustrated in Figure 4. At the beginning of the study, participants were randomly divided into two groups, with 30 participants in each group (see Table 2). Before training, all participants completed the CANTAB SWM for the first time on the first day for the purpose of baseline measures. The instructions for the SWM test were explained to the students before the initiation of the test.

Two groups of participants then received separate training in two versions of n-back tasks for 4 days. Each session lasted 3 min, and with two sessions conducted on each of the 4 days (eight sessions in total). The instructions for the n-back task were explained to the participants prior to the training, followed by three practice trials (when \( n = 1 \) or \( n = 2 \)) over the first 2 days. On the remaining five sessions, all participants received training in performing 3-back tasks. As the participants of this study were healthy undergraduates, 3-back tasks (the hardest level of n-back task in this study) had been set as the main training task. The participants’ performance was assessed in terms of n-back task accuracy (scoring and false clicks) and their data were recorded throughout the duration of the training.
After the training had finished, all participants completed the CANTAB SWM a second time on the last day to obtain an indication of improvement in their SWM. Following the test, all the participants immediately completed the usability questionnaire.

All above tests and trainings were conducted on six new iPad Pros (IOS 13.1). The entire experiment was conducted in a well-equipped laboratory during the practice class of the course of New Media Application. Participants were asked to perform all cognitive tests and trainings on their own, in a quiet space and to the best of their ability.

**Statistical Analysis**

Data were analyzed using IBM SPSS version 26.0 and JASP version 0.11.1. Differences in mean performance within the groups from baseline to posttreatment were analyzed by using paired-sample *t*-tests. Differences between the groups were analyzed using independent sample *t*-tests. In addition to the outcome measures from the CANTAB SWM mentioned above, we also analyzed the performance of the participants in the n-back training program. All data were measured at 95% confidence intervals (CIs), and the threshold for statistical significance was set at *p* < .05.

**Results**

**Pre-Training Performance on CANTAB SWM**

In terms of baseline performance on the CANTAB SWM, the traditional and AR training groups had different distributions in terms of between errors (the number of times the subject incorrectly revisits a box in which a token has previously been found; see the “Measure of spatial working memory” section) on 4, 6, and 8 box trials, and different distributions in terms of strategy (the number of times a subject begins a new search pattern from the same box they started with previously; see the “Measure of spatial working memory” section). However, the difference in the pre-training strategy on 6 and 8 box trials (SWMS) was the only significant difference (Mann–Whitney *U* test, *W* = 283.50, *p* = .012). The traditional training group had a generally less effective strategy than the AR training group before training (Figure 5).

**Accuracy on N-Back Training Programs**

With regard to scores on the traditional n-back and AR n-back training programs, there were no significant differences in terms of average score on the 3-back trials, nor in terms of improvement in scores on the 3-back trials over the five 3-back training sessions. Paired-sample *t*-tests showed that both groups significantly improved their training tasks score between the first and final 3-back session (Table 3).

We also compared the mean number of false clicks across the five 3-back training sessions, and the mean improvement in incorrect clicks over five 3-back training sessions. In this case, the data were normally distributed, and the AR training group had a mean incorrect score of 5.873 (SD = 3.301). The traditional training group had a significantly lower mean of 4.000 (SD = 2.257) (Student’s *t*-test, *t* = −2.566, *p* = .013).

**Table 2. Demographics for Each Training Group.**

| Demographics         | AR training group (n = 30; 17 female) | Traditional training group (n = 30; 12 female) |
|----------------------|--------------------------------------|-----------------------------------------------|
|                      | *M* | *SD* | *M* | *SD* |
| Age (years)          | 19.700 | 1.088 | 18.967 | 1.098 |
| Education (years)    | 13.000 | 1.000 | 13.000 | 1.000 |

Note. AR = augmented reality.
The difference in improvement of incorrect clicks over five 3-back training sessions was not significant. The training performance data of AR training group and traditional training group is attached in the online appendix of this study.

**Improvements to CANTAB SWM**

We used paired-sample tests to examine the hypotheses that each group would show an improved performance in terms of between errors on the CANTAB SWM after training. As improvement is shown by a reduction in errors, the alternative hypothesis was that the number of pre-training errors would be higher than the number of post-training errors. As the post-training data were not all normally distributed (Figure 6), we used non-parametric Wilcoxon signed-rank tests. Both groups significantly reduced the number of between errors across the 4, 6, and 8 box trials (SWMBE468) and the 12 box trials (SWMBE12) (Table 4).

Next, we compared between errors (SWMBE468 and SWMBE12) improvement between the two groups. Improvement was calculated by subtracting the number of post-training between errors from the pre-training between errors, where a higher number indicates greater improvement. There were two outliers in the traditional training group: one participant made 21 additional between errors after training, while another participant made a total of 68 fewer between errors after training (Figure 7). In each of the following analyses, we indicate whether or not these outliers are included in the analysis.

Including both outliers, an independent sample t-test showed no significant differences between the two groups in terms of improvement in between errors after training (Table 5). However, the data were not normally distributed (Table 6), which is an assumption of the Student’s t-test.

As such, there were two possible approaches, and both are reported here. The first option was to exclude the outliers from the traditional group. Doing so meant that the improvement scores of both groups were normally distributed and, with the outliers excluded, a Student’s t-test showed that the traditional training group improved more significantly than the AR training group in terms of between errors in 4, 6, and 8 box trials after training (Table 7). The AR group improved more than the traditional training group in terms of between errors on 12 box trials, although this difference was not significant (Table 7).

However, given the relatively large spread of both datasets (see standard deviations in Table 7), it is perhaps more relevant to consider a second option: investigating the differences in distributions using a non-parametric test. Therefore, we used a Mann–Whitney test, which does not assume normally

---

**Figure 5.** Distribution of the pre-training strategy on 6 and 8 box trials for both groups.

*Note. A higher score indicates a less effective strategy. AR = augmented reality.*

**Table 3.** Paired-Sample t-Tests Showing Significant Improvement in Score on Both Types of Training Over Five Sessions.

| Training groups | Score on first 3-back session | Score on final 3-back session | Student’s t-test |
|-----------------|-------------------------------|-------------------------------|-----------------|
|                 | M    | SD   | M    | SD   | t    | df  | p    |
| Traditional training | 12.467 | 5.444 | 16.100 | 4.908 | −3.338 | 29 | .001 |
| AR training     | 14.367 | 6.228 | 17.700 | 5.844 | −3.276 | 29 | .001 |

*Note. The alternative hypothesis is that first score is lower than final score. AR = augmented reality.*
Table 4. Pre- and Post-Training Between Errors Showing Significant Improvement on the CANTAB SWM After Both Types of Training.

| Training groups variables | Pre-training between errors | Post-training between errors | Wilcoxon signed-rank test |
|---------------------------|-----------------------------|-----------------------------|--------------------------|
|                           | M   | SD  | M   | SD  | W   | p   |
| Traditional n-back training |    |     |     |     |     |     |
| SWMBE12                   | 29.067 | 15.402 | 20.700 | 12.801 | 356.500 | .001 |
| SWMBE468                  | 14.833 | 8.956  | 6.900  | 7.640  | 325.500 | <.001 |
| AR n-back training        |    |     |     |     |     |     |
| SWMBE12                   | 28.533 | 11.602 | 17.200 | 14.554 | 378.500 | <.001 |
| SWMBE468                  | 10.567 | 8.955  | 5.900  | 7.680  | 337.000 | .005 |

Note. The alternative hypothesis is that the number of errors post training is less than the number of errors pre-training. CANTAB SWM = Cambridge Neuropsychological Test Automated Battery Spatial Working Memory; AR = augmented reality.
distributed data. In this case, we found that the distributions were significantly different in the case of improvement on between errors for 4, 6, and 8 trials only (outliers included, Figure 8, Table 8). Most noticeably, in the traditional training group, 22 participants (73.3%) reduced their number of between errors on 4, 6, and 8 box trials by at least 5. The corresponding figure in the AR training group was 15 (50%) participants.

As the AR group utilized a significantly stronger strategy on 6 and 8 box trials before training, we also used independent sample $t$-tests to compare post-training strategy scores and improvement. With all data included (SWMS and SWMSX), there were no significant differences between the two groups post-training in terms of strategy score, nor in terms of improvement in strategy score (i.e., post-training strategy subtracted from pre-training strategy).

**User Experience Questionnaire Results**

Table 9 shows that the AR version of the n-back task was found to have graphics that were significantly more motivating as compared with the traditional version, and the participants playing the AR version felt significantly more absorbed in the n-back task than the traditional group. There were no significant differences between the groups on any of the other user experience questionnaire items.

**Discussion**

As the recent advances in AR technology, AR applications have been seen as a potential means of improving cognitive training outcomes by many researchers (Boletsis & McCallum, 2014, 2016). Study found that transfer effectiveness increases are closely attributable to the transfer tasks’ similarity to the practiced tasks, as well as the category of WM tasks on which improvements are all seen as near transfer (Simons et al., 2016). Therefore, this study focuses on comparing the efficacy of an AR n-back task and a traditional 2D n-back task in leading to the training and transfer improvements in WM.

Results show that after just a short period of AR-based n-back training, participants showed significantly improved CANTAB SWM test scores. As such, our findings support previous research suggesting that AR technologies have the capability to positively influence cognitive functions (Boletsis & McCallum, 2016; Brito & Stoyanova, 2018; Scanlon et al., 2016). Furthermore, we found that the AR version of the n-back test was rated as more engaging and having more stimulating graphics than the traditional version. Given that engagement with cognitive training is recognized as a key factor contributing to the effectiveness of such training (Robb et al., 2018), these findings suggest that AR and other related technologies could play a role in the effectiveness of cognitive training software.

However, we also found differences in the distributions of performance improvement between the AR and traditional training groups. Specifically, we found that the traditional training group had a distribution of improvement in (easier) 4, 6, and 8 box trials, which suggested better levels of improvement, while the AR group had a higher mean improvement in (more difficult) 12 box trials, although only the first of these differences was statistically significant. This suggests that a minor change in the graphical context of an n-back training program can have effects on the outcomes, even after only a short period of training. This finding makes a timely contribution to our understanding of the effects of cognitive training, as researchers in this field have recently begun to highlight how research on cognitive training must acknowledge the many specific aspects of a training environment. This is because differences in content may lead to different effects, which should be specified and controlled for in research (Dale & Green, 2017; Green et al., 2019). Our research provides empirical evidence to support this point, which undoubtedly warrants further research using carefully controlled versions of training programs to investigate these effects.

We suggest that some of our findings may contribute to our understanding of transfer in cognitive training. Specifically, our finding that improvement in the traditional training group in 4, 6 and 8 box trials was significantly different (generally, more improved) than the AR group may be of particular importance. In virtue of the fact that the stimuli

![Figure 7. Between errors improvement showing two outliers in the traditional training group.](image-url)

| Table 5. Independent Student Sample $t$-Test, Comparing Improvement in Between Errors Between the Two Groups After Training. |
| Variables                     | t    | df  | p   |
|--------------------------------|------|-----|-----|
| Improvement (SWMBE12)         | 0.818| 58.000 | .417 |
| Improvement (SWMBE468)       | –1.592| 58.000 | .117 |

Note. All data included.
are presented in a two-dimensional display, the traditional n-back training program is clearly more similar to the CANTAB SWM than the AR training program. One plausible explanation of our results is that training on the traditional n-back task produced improvements on the similar (to the training) CANTAB SWM (i.e., near transfer) while training on the AR n-back tasks produced less improvement on the dissimilar (to the training) CANTAB SWM (i.e., far transfer). It is therefore possible that our findings can be explained in terms of the common demands theory of training transfer (Oei & Patterson, 2015; Simons et al., 2016), which holds that transfer of training to different tasks occurs when the trained task and transfer task share some common demands, such that a previously acquired task set (e.g., a representation of the task stored in working memory), can be drawn on to facilitate performance of the transfer task. Furthermore, the fact that the traditional training group did not further improve in the (more difficult) 12 box trials may also be explained in terms of shared demands. That is, if we assume that the level of difficulty of the task is one such demand that can be shared.

It may be argued that our findings are limited by the fact that practice effects have been found in the CANTAB SWM (Levaux et al., 2007; Lowe & Rabbitt, 1998). Although this does limit our findings regarding pre and post improvement in both groups, it is important to note that our two main findings—specifically, (a) that the AR training program was found to be more engaging and have more appealing graphics and (b) that the distributions of improvement on between errors was significantly different between the two groups—are not limited by any such practice effects. This is emphasized by noting that Lowe and Rabbitt (1998) claimed that practice effects were particularly prevalent in tests of executive function (which include working memory) due to the fact that the discovery of a strategy can dramatically affect participants’ subsequent performance on the tests. In our sample, while both groups showed a significant improvement in strategy after training, there were no significant differences between the groups in terms of strategy improvement. Therefore, the difference in the distributions of improvement in terms of error reduction must be explained by something other than strategy. As such, our main findings do not depend on the absence of the practice effects in the CANTAB SWM test.

This study demonstrates the feasibility of future work using a controlled version of training programs to investigate the influence of specific features (e.g., 2D vs. 3D perspective) on the effectiveness of cognitive training programs. Furthermore, our work shows that such features may be associated with positive improvement on one outcome measure (as the AR training was more engaging and graphically appealing), while simultaneously leading to less or no improvement on another outcome measure (as the traditional training had a significantly better distribution of improvement in terms of error reduction on the CANTAB tests). This illustrates the complex interaction of various features of cognitive training programs and should encourage future research to understand these interactions.

Limitation

Although this study has a number of important implications, several limitations also need to be considered. First, the number of training sessions was few. We suggest further research with longer training time to further verify the transfer results, which may reveal more scientifically significant finding on
cognitive development. Second, this study did not include a non-WM training control group, thereby lack of the difference assessment between non-WM training and WM training. Finally, this study was conducted with a single task to measure the training and transfer effects. However, for the purposes of this study, there are several reasons on why we conducted such an experimental design. As many studies suggest that traditional n-back task is an effective instrument to improve participants’ WM (Jaeggi et al., 2010a; Jaeggi et al., 2007, 2008, 2009; Jones et al., 2018), this study, therefore, did not include a non-WM training control group to compare whether the trainings were effective in either n-back group. Moreover, previous study suggested that a simpler single n-back task makes the investigation of the process in training and transfer more accessible, as the dual n-back task is too complex to be understood by the participants (Jaeggi et al., 2010b).

In addition, the near-transfer effects were more intended to maintain in visuospatial WM training than the verbal WM training (Melby-Lervåg & Hulme, 2013). From these considerations, we used a single n-back task with visuospatial-nonverbal material, in which n was set with 3 to increase the difficulty of WM task in this study. Furthermore, previous studies found that short-term (e.g., 1–2 weeks) cognitive training can result in cognitive improvements (Akter et al., 2015; Au et al., 2015), and especially produce significant improvements in young healthy adults (Tulbure & Siberescu, 2013). Therefore, to explore the possibility of short WM training effects, we conducted a short training with eight sessions in 4 days, as the participants in the present investigation were healthy undergraduates.

Conclusion and Future Study

As WM is a significant cognitive skill that affects a wide range of life functions, even small improvements can have deep societal implications (Au et al., 2015; Tulbure & Siberescu, 2013). It is becoming very clear to us that training on WM with the AR n-back task holds much promise. It would be interesting for future research to create an AR-specific transfer measure for the same type of the present work to investigate the transfer in an AR WM task. Our study
nevertheless provides some further limited evidence for the effectiveness of both traditional and AR n-back training, while highlighting how differences in the visual mode of presentation of a WM training task may have important effects on both transfer and engagement with the training. As these two factors are important in considering the efficacy of WM training, our study provides evidence that there may be crucial considerations concerning the correct balance of engagement and training effectiveness. It is clear from our results that making simple alterations to the visual presentation of a WM training task can affect participant’s engagement with the training, while simultaneously affecting the effectiveness of the training. In addition, we show that such simple alterations in visual presentation of WM training tasks may have important implications in terms of near and far transfer of training effects, and we suggest that this may provide support for the common demands theory of training transfer (Oei & Patterson, 2015). Our findings suggest that the level of improvement on a transfer task after training may be closely related to how similar or dissimilar the training and transfer tasks are in terms of, among other factors, whether they are both presented in a traditional, 2D format or in an AR and/or 3D format.

Authors’ Note
The quantitative data collected for this study can be accessed by request, by contacting authors.

Acknowledgments
The authors would like to thank the reviewers for providing valuable comments which helped us to greatly improve this manuscript. Also, the authors would like to acknowledge Professor Xinyi Wu and lecturer Yichun Liu for their contributions to organize the participants for this study.

Author Contributions
B.Z. designed the study, did statistical analyses, collected the data, and wrote the manuscript. N.R. performed the statistical analyses and wrote the manuscript.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was sponsored by Peak Discipline Construction Project of Education at East China Normal University, China Postdoctoral Science Foundation under Grant 2018M641953 and Education Youth Program of Shanghai Philosophy and Social Science Planning Project under grant B2021001.

Ethics Statement
Prior to beginning the study, ethics approval was obtained from the institutional ethics committee in East China Normal University under approval number: HR 055-2019.

ORCID iD
Bo Zhang https://orcid.org/0000-0002-6529-8720

Supplemental Material
Supplemental material for this article is available online.

References
Akçayır, M., Akçayır, G., Pektaş, H. M., & Ocak, M. A. (2016). Augmented reality in science laboratories: The effects of augmented reality on university students’ laboratory skills and attitudes toward science laboratories. Computers in Human Behavior, 57, 334–342. https://doi.org/10.1016/j.chb.2015.12.054
Akter, S., Hassan, M. R., Shahrir, M., Akter, N., Abbas, M. G., & Bhuiyan, M. A. (2015). Cognitive impact after short-term exposure to different proton pump inhibitors: Assessment using CANTAB software. Alzheimer’s Research & Therapy, 7(1), Article 79. https://doi.org/10.1186/s13195-015-0164-8
Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. Developmental Psychology, 50(10), 2368–2379. https://doi.org/10.1037/a0037493
Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., & Johnston, E. (2013). Video game training enhances cognitive control in older adults. Nature, 501(7465), 97–101. https://doi.org/10.1038/nature12486
Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. Psychonomic Bulletin & Review, 22(2), 366–377. https://doi.org/10.3758/s13423-014-0699-x
Bacca, J., Baldiris, S., Fabregat, R., & Graf, S. (2014). Augmented reality trends in education: A systematic review of research and applications. Journal of Educational Technology & Society, 17(4), 133–149.
Bahr-Fuchs, A., Clare, L., & Woods, B. (2013). Cognitive training and cognitive rehabilitation for mild to moderate Alzheimer’s disease and vascular dementia. Cochrane Database of Systematic Reviews, 6(6), 103. https://doi.org/10.1002/14651858.CD003260.pub2
Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. Development and Psychopathology, 20(3), 821–843. https://doi.org/10.1017/S095457940800394
Billinghurst, M., Clark, A., & Lee, G. (2015). A survey of augmented reality. Foundations and Trends in Human-Computer Interaction, 8(2–3), 73–272. https://doi.org/10.1561/1100000049
Boletis, C., & McCallum, S. (2014). Augmented reality cube game for cognitive training: An interaction study. Studies in
Hoe, Z.-Y., Lee, I.-J., Chen, C.-H., & Chang, K.-P. (2019). Using
Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge,
Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A.
Engeser, S., & Rheinberg, F. (2008). Flow, performance and mod-
Boletsis, C., & McCallum, S. (2016). Augmented reality cubes for
Csipo, T., Lipecz, A., Fulop, G. A., Hand, R. A., Ngo, B. T. N.,
Cacciamani, F., Salvadori, N., Luchetti, E., Calabresi, P., & Parnetti, L. (2018). Evidence of practice effect in CANTAB spatial working memory test in a cohort of patients with mild cognitive impairment. *Applied Neuropsychology: Adult*, 25(3), 237–248.
Cocchi, L., Walterfang, M., Testa, R., Wood, S. J., Seal, M. L.,
Cortez, J., . . .Soulsby, B. (2009). Grey and white matter
abnormalities are associated with impaired spatial working memory ability in first-episode schizophrenia. *Schizophrenia Research*, 115(2–3), 163–172.
Cormack, F. K., Taptiklis, N., Barnett, J. H., King, J., & Fenhert, B. (2016). High-frequency monitoring of cognition, mood and
behaviour using commercially available wearable devices. *Alzheimer’s & Dementia, 12*, P159. https://doi.org/10.1016/j.jalz.2016.06.263
Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (Eds.). (1992). *Optimal experience: Psychological studies of flow in consciousness*. Cambridge university press.
Cspio, T., Lipécz, A., Fulop, G. A., Hand, R. A., Ngo, B. T. N.,
Dzialendzik, M., . . .Silva-Palacios, F. (2019). Age-related decline in peripheral vascular health predicts cognitive impairment. *Geroscience*, 41(2), 125–136.
Dale, G., & Green, C. S. (2017). The changing face of video games and video gamers: Future directions in the scientific study of video game play and cognitive performance. *Journal of Cognitive Enhancement*, 1(3), 280–294.
Engeser, S., & Rheinberg, F. (2008). Flow, performance and mod-
erators of challenge-skill balance. *Motivation and Emotion*, 32(3), 158–172. https://doi.org/10.1007/s11031-008-9102-4
Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309
Forns, J., Esgaola, M., López-Vicente, M., Suades-González, E.,
Alvarez-Pedrero, M., Julvez, J., . . .Sunyer, J. (2014). The n-back Test and the Attentional Network Task as measures of childhood neuropsychological development in epidemiological studies. *Neuropsychology*, 28(4), 519–529. https://doi.org/10.1037/neu0000085
Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge,
Ball, K. C., . . .Faccoetti, A. (2019). Improving method-
ological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement*, 3(1), 2–29.
Hoe, Z.-Y., Lee, I.-J., Chen, C.-H., & Chang, K.-P. (2019). Using an augmented reality-based training system to promote spatial visualization ability for the elderly. *Universal Access in the Information Society*, 18(2), 327–342. https://doi.org/10.1007/s10209-017-0597-x
Jaeggi, S. M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig,
W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, Affective, & Behavioral Neuroscience*, 7(2), 75–89. https://doi.org/10.3758/CABN.7.2.75
Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105
Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010a). The concurrent validity of the N-back task as a working memory
measure. *Memory, 18*(4), 394–412. https://doi.org/10.1080/09658211003702171
Jaeggi, S. M., Schmid, C., Buschkuehl, M., & Perrig, W. J. (2009). Differential age effects in load-dependent memory processing. *Aging, Neuropsychology, and Cognition*, 16, 80–102. https://doi.org/10.1080/13825580802233426
Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F.,
Jonides, J., & Perrig, W. J. (2010b). The relationship between n-back performance and matrix reasoning—Implications for training and transfer. *Intelligence*, 38(6), 625–635. https://doi.org/10.1016/j.intell.2010.09.001
Jones, M. R., Katz, B., Buschkuehl, M., Jaeggi, S. M., & Shah, P. (2018). Exploring N-Back Cognitive Training for Children with ADHD. *Journal of Attention Disorders*, 24(5), 704–719. https://doi.org/10.1177/1087076817797293
Jonides, J., Schumacher, E. H., Smith, E. M., Lauber, E. J., Awh, E.,
Minoshima, S., & Koepe, R. A. (1997). Verbal working memory
load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience*, 9(4), 462–475. https://doi.org/10.1162/jocn.1997.9.4.462
Katz, B., Jaeggi, S., Buschkuehl, M., Stegman, A., & Shah, P. (2014). Differential effect of motivational features on training improvements in school-based cognitive training. *Frontiers in Human Neuroscience*, 8(1), 242–252. https://doi.org/10.3389/fnhum.2014.00242
Levaux, M. N., Potvin, S., Sepehry, A. A., Sablier, J., Mendrek, A.,
& Stip, E. (2007). Computerized assessment of cognition in schizophrenia: Promises and pitfalls of CANTAB. *European Psychiatry*, 22(2), 104–115. https://doi.org/10.1016/j.eurpsy .2006.11.004
Lewis, J. R. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-computer Interaction*, 7(1), 57–78. https://doi.org/10.1080/10447319509526110
Lowe, C., & Rabbitt, P. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theore-
tical and practical issues. *Neuropsychologia*, 36(9), 915–923. https://doi.org/10.1016/S0028-3932(98)00036-0
Lu, S. J., & Liu, Y. C. (2015). Integrating augmented reality tech-
nology to enhance children’s learning in marine education. *Environmental Education Research*, 21(4), 525–541. https://doi.org/10.1080/13504622.2014.911247
Melby-Lervåg, M., & Hulme, C. (2013). Is working memory
training effective? A meta-analytic review. *Developmental Psychology*, 49(2), 270–291. 
Mitrovic, A., Martin, B., Suraweera, P., Zakharov, K., Milik, N.,
Holland, J., & McGuigan, N. (2009). ASPIRE: An authoring system and deployment environment for constraint-based tutors.
Nagle, A., Riener, R., & Wolf, P. (2015). High user control in game design elements increases compliance and in-game performance in a memory training game. *Frontiers in Psychology*, 6, 1774–1789.

Nielsen, J. (1993). *Usability engineering*. Academic Press.

Oei, A. C., & Patterson, M. D. (2015). Enhancing perceptual and attentional skills requires common demands between the action video games and transfer tasks. *Frontiers in Psychology*, 6, Article 11. https://doi.org/10.3389/fpsyg.2015.00113

Owen, A., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, 28(10), 1021–1034. https://doi.org/10.1016/0028-3932(90)90137-D

Robb, N., Northridge, J., Politis, Y., & Zhang, B. (2018). Parental intention to support the use of computerized cognitive training in children with genetic neurodevelopmental disorders. *Frontiers in Public Health*, 6, 309–316. https://doi.org/10.3389/fpubh.2018.00309

Robbins, T. W., James, M., Owen, A., Sahakian, B. J., McInnes, L., & Rabbitt, P. M. (1994). Cambridge Neuropsychological Test Automated Battery (CANTAB): A factor analytic study of a large sample of normal elderly volunteers. *Dementia*, 5(5), 266–281.

Rodriguez-Ardura, I., & Meseguer-Artola, A. (2017). Flow in e-learning: What drives it and why it matters. *British Journal of Educational Technology*, 48(4), 899–915. https://doi.org/10.1111/bjet.12480

Sahakian, B. J., & Owen, A. (1992). Computerized assessment in neuropsychiatry using CANTAB: Discussion paper. *Journal of the Royal Society of Medicine*, 85(7), 399–402.

Scanlon, L., O’Shea, E., O’Caoimh, R., & Timmons, S. (2016). Usability and validity of a battery of computerised cognitive screening tests for detecting cognitive impairment. *Gerontology*, 62(2), 247–252. https://doi.org/10.1159/000433432

Schueffel, P. (2017). *The concise fintech compendium*. School of Management Fribourg.

Shema-Shiratzky, S., Brozgol, M., Cornejo-Thumm, P., Gevadayan, K., Rotstein, M., Leitner, Y., & Mirelman, A. (2018). Virtual reality training to enhance behavior and cognitive function among children with attention-deficit/hyperactivity disorder: Brief report. *Developmental Neurorehabilitation*, 22(6), 1–6. https://doi.org/10.1080/17518423.2018.1476602

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, 138(4), 628. https://doi.org/10.1037/a0027473

Short, M. M., Mazmanian, D., Onioso, N., & Mushquash, C. J. (2016). Executive function and self-regulation mediate dispositional mindfulness and well-being. *Personality and Individual Differences*, 93, 97–103. https://doi.org/10.1016/j.paid.2015.08.007

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.

Sungkur, R. K., Panchoo, A., & Bhoyroo, N. K. (2016). Augmented reality, the future of contextual mobile learning. *Interactive Technology and Smart Education*, 13(2), 123–146. https://doi.org/10.1108/ITSE-07-2015-0017

Tulbure, B. T., & Sibereci, I. (2013). Cognitive training enhances working memory capacity in healthy adults. A pilot study. *Procedia-Social and Behavioral Sciences*, 78, 175–179. https://doi.org/10.1016/j.sbspro.2013.04.274

Van Dongen-Boomsma, M., Vollebregt, M. A., Buiter, J. K., & Slats-Willems, D. (2014). Working memory training in young children with ADHD: A randomized placebo-controlled trial. *Journal of Child Psychology and Psychiatry*, 55(8), 886–896. https://doi.org/10.1111/jcpp.12218

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. (2011). New rule uses drive the relation between working memory capacity and Raven’s Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 256–263. https://doi.org/10.1037/a0021613

Wu, D., Yang, C. C., Chen, K. Y., Lin, Y. C., Wu, P. J., Hsieh, P. H., . . . Hu, C. J. (2020). Hydrolyzed chicken extract (Probeptigen®) on cognitive function in healthy middle-aged people: A randomized double-blind trial. *Nutrients*, 12(5), 1362.

Zhang, Q., Wang, C., Zhao, Q., Yang, L., Buschkuehl, M., & Jaeggi, S. M. (2019). The malleability of executive function in early childhood: Effects of schooling and targeted training. *Developmental Science*, 22(2), Article e12748. https://doi.org/10.1111/desc.12748