# LSVP: Towards Effective On-the-go Video Learning Using Optical Head-Mounted Displays

ASHWIN RAM and SHENGDONG ZHAO, NUS-HCI Lab, Department of Computer Science, National University of Singapore, Singapore

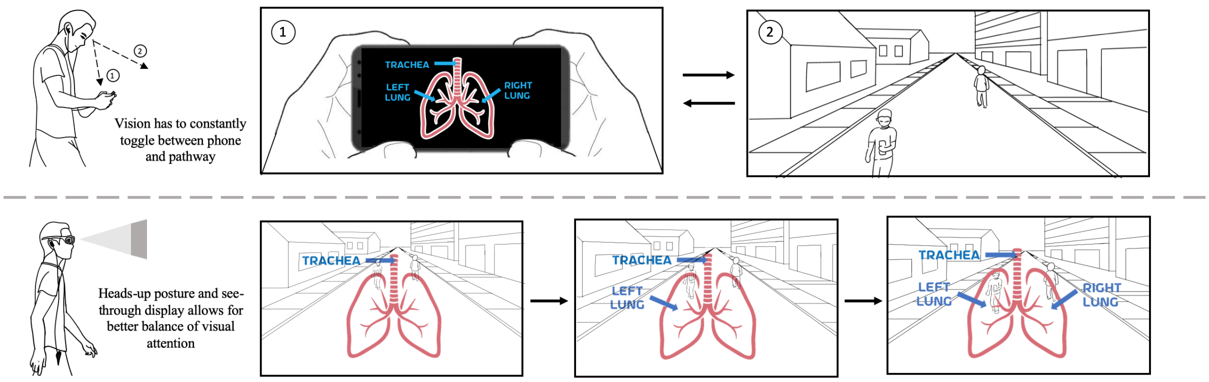



Fig. 1. On-the-go video learning situation in smartphones vs. smart glasses (or Optical Head-mounted Displays (OHMD)). On the phone, users need to switch their visual attention back and forth between the display and the environment when watching a Khan style video. On the OHMD, we transform Khan style videos into Layered Serial Visual Presentation (LSVP) with enhanced color contrast, transparent background, serial presentation of information, data persistence, and controlled information density to allow easy viewing of both the video content and the environment on the go.

The ubiquity of mobile phones allows video content to be watched on the go. However, users' current on-the-go video learning experience on phones is encumbered by issues of toggling and managing attention between the video and surroundings, as informed by our initial qualitative study. To alleviate this, we explore how combining the emergent smart glasses (Optical Head-Mounted Display or OHMD) platform with a redesigned video presentation style can better distribute users' attention between learning and walking tasks. We evaluated three presentation techniques: highlighting, sequentiality, and data persistence to find that combining sequentiality and data persistence is highly effective, yielding a 56% higher immediate recall score compared to a static video presentation. We also compared the OHMD against smartphones to delineate the advantages of either platform for on-the-go video learning in the context of everyday mobility tasks. We found that OHMDs improved users' 7-day delayed recall scores by 17% while still allowing 5.6% faster walking speed, especially during complex mobility tasks. Based on the findings, we introduce Layered Serial Visual Presentation (LSVP) style, which incorporates sequentiality, strict data persistence, and transparent background, among other properties, for future OHMD-based on-the-go video learning.

Authors' address: Ashwin Ram, ashwinram@u.nus.edu; Shengdong Zhao, zhaosd@comp.nus.edu.sg, NUS-HCI Lab, Department of Computer Science, National University of Singapore, 15 Computing Dr., Singapore, 117418.


2474-9567/2021/3-ART30 $15.00
https://doi.org/10.1145/3448118

CCS Concepts: • **Human-centered computing → Empirical studies in ubiquitous and mobile computing**.

Additional Key Words and Phrases: Smart glasses, dynamic information

**ACM Reference Format:**
Ashwin Ram and Shengdong Zhao. 2021. LSVP: Towards Effective On-the-go Video Learning Using Optical Head-Mounted Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 30 (March 2021), 27 pages. https://doi.org/10.1145/3448118

## 1 INTRODUCTION

Video is an effective tool for learning [29, 34]. It is widely utilized by platforms like Khan Academy [35], whose videos are watched every month by over 2 million users [53]. Research indicates that an increasing number of users utilize smartphones to learn from audio-visual content like flashcards during on-the-go scenarios [14, 52]. Dynamic content such as videos, however, can be much more challenging to watch on the go, as it requires users to engage with small screens for longer periods, causing a potential conflict with the need to attend to the environment, which is essential for safe navigation. Nevertheless, humans' need for information is ubiquitous, which leads us to an interesting research question: is it possible to make dynamic information such as videos accessible even during on-the-go scenarios (e.g., walking)?

One may wonder if users want to watch videos on the go in the first place. Our preliminary study results show that there is indeed such a need; however, several constraints make on-the-go video-watching a challenging task. First, users' attention gets fragmented between learning and attending to the demands of their on-the-go environment. Such fragmented attention is known to impact users' learning and walking negatively [10, 16]. Video as dynamic content requires more focused attention than other static forms of learning media [45], making it more challenging to learn during on-the-go scenarios.

Secondly, smartphones (henceforth, just phones) naturally induce a heads-down nature of interaction referred to as 'smartphone zombies' [2]. This heads-down interaction can lead to decreased situational awareness, thereby increasing the risk of accidents while walking [49]. Prolonged use of such interactions can also contribute to significant health problems such as turtleneck and spine injuries [26]. Apart from this, actively holding the phone while viewing videos for prolonged periods can be tiresome and tedious [4], which further hinders users' on-the-go video learning experience.

The emergent smart glasses (or Optical Head-Mounted Display henceforth, just OHMD) platform has shown potential in alleviating some of the aforementioned challenges. Recent research has employed OHMDs to perform basic mobile HCI tasks such as reading [48] and text editing [22] while walking, and found that the see-through nature and heads-up wearability of OHMDs allow higher task performance and better path navigation. Additionally, being wearable allows OHMDs to be operated hands-free (i.e., voice interaction), thus reducing fatigue caused by prolonged holding of the phone.

The benefits of OHMD show its potential as an alternative platform for on-the-go video learning. Yet, simultaneous viewing of digital information and the environment in OHMDs can also distract users, making it harder to concentrate on the video. Existing videos made with the assumption that users have dedicated attention while watching may not suit the OHMD platform. Hence, we need to investigate how video content should be presented to balance users' learning and navigational needs.

In order to find a video presentation style that is suitable for OHMD, we analyzed common video presentation techniques and narrowed down to three techniques that we regard as the most promising for OHMD-based video learning: *highlighting* and *sequentiality* [9, 31], which can guide users' attention, and *data persistence*, which can ease the temporal load on the working memory [38, 45, 47].

A controlled experiment evaluation of these techniques revealed that presenting information sequentially outperforms highlighting in terms of strengthening users' fundamental "remember" skills [36], resulting in 36%

higher immediate recall scores. This performance gain is achieved with minimal to no disruption in walking speed. Additionally, combining sequentiality and complete data persistence can enhance the immediate recall scores by 56% as compared with the baseline condition.

Based on the result, we developed *Layered Serial Visual Presentation (LSVP)*, a video presentation style specifically catered for on-the-go OHMD viewing. While LSVP may appear to be similar to the video presentation style of Khan Academy (Khan style video) where information is sequentially displayed, it has a number of important differences: 1) Data persistence is strictly preserved as once elements appear on the screen, they will not be removed until the end of the video (in contrast, Khan style video often refreshes the screen during the middle of the video). 2) To facilitate better path navigation, the background of the video is made transparent. 3) The color of the video content needs to be adjusted/enhanced to ensure relative clarity over the environmental background. 4) The information density needs to be controlled so that the number of elements, when all preserved on the screen, can fit the display.

To further validate this, we compared LSVP on OHMD with phones to delineate the benefits of either platform for on-the-go video learning in different realistic mobility scenarios. Our results showed that by using LSVP for video learning on OHMD, participants performed 17% better in 7-day delayed recall tests while walking 5.6% faster as compared with video learning using phones, especially in more complex navigation scenarios, thereby showing the potential for OHMD-based video learning.

The contribution of this paper is fourfold: 1) Enhance our understanding of on-the-go video learning through a qualitative study. 2) Empirically investigated the effectiveness of three presentation techniques on balancing user's learning performance and path-navigation ability in OHMD platforms. 3) Empirically compared two device platforms (phone vs. OHMD) and found that using OHMD can lead to better learning performance and faster walking speed in a number of mobility scenarios. 4) Design implications for existing and future videos to support on-the-go learning tasks on OHMD.

## 2 RELATED WORK

Our work covers three important research areas. First, we discuss prior works that have explored the use of videos for on-the-go learning using phones, especially while walking. Next, we look at how OHMD, as an emergent platform, can support learning applications including on-the-go video learning. Finally, we discuss literature related to enhancing video-based learning through the use of presentation techniques.

### 2.1 Multitasking during On-the-go Video Learning

Active developments in e-learning have led to a surge in the number of learning videos available online [53]. With the ability to access learning content on phones, users are increasingly engaging in on-the-go learning tasks [14, 23, 52]. However, previous work has found that walking affects learning on mobile devices [12, 16, 60]. Doolittle and Mariano [16] showed that an individual's ability to learn declined when performed in conjunction with walking. Similar results were found in the case of listening to educational podcasts while walking or jogging [12] and in other everyday scenarios such as navigating traffic and walking between buildings [60]. This decline in performance is attributed to the multitasking nature of on-the-go learning, which results in divided attention. For an effective on-the-go learning experience, it is necessary to balance this divided attention between the needs of both walking and learning. In this work, we explore how this trade-off can be achieved on the emergent OHMD platform.

### 2.2 OHMD as an Emergent Platform for Learning and On-the-go Tasks

OHMD, with its heads-up viewing and hands-free access, is a wearable technology that has the potential to support on-the-go information-seeking tasks. OHMDs have found increasing use in many domains such as medicine [1, 19], industry [56] and tourism [18]. Yet, its utility in education [6] has been mostly confined to Augmented Reality (AR) applications for learning. For example, OHMDs have been considered for teaching experimental physics in laboratories [30, 37] and anatomy [19]. However, these works do not focus on on-the-go situations where frequent context switching and distractions are prevalent. Recent works have explored how fundamental tasks such as reading [48] and text editing [22] can be re-designed to meet on-the-go needs, and their results encourage us to extend such efforts to video learning contexts. Being able to engage in dynamic content ubiquitously is an ambitious goal raised by Bret Victor [55]. Although the dynamic content we study is not the same as the one described by Bret, we hope that our investigation can contribute to the understanding of how dynamic content can be more effectively presented in on-the-go scenarios.

### 2.3 Effect of Presentation Techniques for Video Learning

Incorporating the right presentation techniques in videos can minimize the cognitive load on users' working memory and improve their learning [7]. In our discussion, we group these techniques into two main categories.

The first set of works focus on leveraging the modality effect of multimedia learning [51]. This can be done by highlighting relevant parts of the on-screen visuals, such as by varying their brightness [33], and color [5, 31] to guide the learner's attention and improve their recall ability. Similarly, studies in visual attention [28] have found that the sudden appearance of an object can also be an effective means of capturing attention. In the learning context, Jamet et al. tested this by sequentially presenting learning material on-screen [9, 31] and found that it led to faster processing [59] and higher levels of comprehension [9].

Other works have considered how the temporal load on the working memory can be eased during video learning [45]. The maintenance of information on-screen for the entirety of the video, which is known as the data persistence effect can be beneficial in this regard. This was identified by Lanir et al. [38] as an important aspect of traditional blackboard lectures that promoted self-paced active learning and also improved recall [47].

However, the benefits of both categories of presentation techniques discussed above are specific to traditional stationary learning scenarios. Such scenarios typically involve watching video lectures on a desktop or laptop while seated comfortably in an environment with minimal distractions. For on-the-go scenarios with more dynamic context switches, it is uncertain whether any of the above-mentioned strategies alone or in combination will be sufficient for on-the-go video learning, thereby motivating our investigation in Study 2.

### 2.4 Background

On-the-go video learning is an instance of multitasking where users have to focus on the learning and walking tasks simultaneously. To better understand and analyze such task interplay, we briefly discuss two existing theories; Resource competition Framework (RCF) [46] and Cognitive Load Theory (CLT) [50].

The Resource Competition Framework (RCF) is useful for analyzing how users interact and multitask with HCI tasks (e.g., watching videos, learning) and mobility tasks (e.g., walking, avoiding collisions, wayfinding) in mobile contexts. It builds upon the important assumption that an individual's cognitive resources are limited in capacity, and these resources can be concurrently utilized for multiple tasks until the capacity limit is reached. When performing an HCI task in a mobile scenario, there is a constant competition between the mobile HCI tasks and ongoing mobility tasks for these cognitive resources. RCF further posits that with such competition, resources will be shared hierarchically based on the intrinsic motivational needs and goals of the individual. Those tasks which do not receive sufficient resources will be slowed-down, postponed, or terminated, leading to resource depletion penalty. Yet, it is also possible to avoid resource depletion by adopting compensatory strategies that can help manage cognitive resources efficiently.

Cognitive Learning Theory (CLT) states that, for effective learning, the cognitive load on the learners' working memory has to be minimal. This cognitive load is comprised of three parts: 1) intrinsic load, which is the load inherent to the topic being studied, 2) germane load, which is the level of cognitive activity required to consolidate the learned information, and 3) extrinsic load, that often arises from a poorly designed lesson. In the context of learning from multimedia content such as videos, the Cognitive Multimedia Learning Theory (CMLT) further elaborates upon these constructs of CLT using the modality effect [42], which posits that presenting learning material using both illustrations and spoken text can improve learning outcomes.

In this work, we use RCF in Study 1 for the deductive analysis of users' on-the-go video learning behavior. We rely on CLT and CMLT to narrow down on presentation techniques that can enhance on-the-go video learning on OHMD and to explicate the findings that arise from Study 2.

## 3 STUDY 1: CURRENT LANDSCAPE OF ON-THE-GO VIDEO LEARNING

While previous works have shown that on-the-go learning is commonplace, these findings are mostly associated with microlearning tasks (e.g., language learning) that occur in on-the-go contexts where the user is mostly stationary, such as while waiting or sitting during commutes [14]. Less is known about users' video learning while walking, a non-stationary on-the-go scenario that forms a significant part of a user's daily routine. Hence, in our initial investigation, we aim to understand the current landscape of video watching and learning during on-the-go scenarios involving walking. The insights from this study can shed light on the challenges that need to be tackled to improve the video-based learning experience of users in on-the-go scenarios. To this end, we aim to answer the following research questions:

**RQ1:** *What are the motivations and preferences of people regarding watching videos on the go (i.e., walking)?*

**RQ2:** *What are the factors that make watching videos while walking more difficult than in a stationary setting?*

The study was conducted in two stages. In the first stage, qualitative open-ended elicitation interviews were conducted to understand the motivation of users and the challenges they face when video-watching while walking. In the second stage, based on the insights from the interviews, we developed an online survey to understand how relevant our interview findings are in a larger population base.

### 3.1 Participants and Procedure

*3.1.1 Interviews.* Sixteen volunteers (7 females; M = 23.4, SD = 2.82 years) were recruited through open calls on chat groups. Half of the volunteers were working professionals, and the rest were from the university community. They were selected if they possessed and identified themselves as regular users of phones during "on-the-go" situations. Being on-the-go was described as "a situation where the user is in a mobile state" and examples were provided for clarity.

Semi-structured interviews were conducted either remotely using video conferencing (14) or in person (2). Remote interviews were favored due to the COVID-19 social distancing restrictions. The interview questions sought to elicit information on the frequency of mobile device usage while walking, type of use, motivation for watching videos (educational or entertainment) while walking and difficulties associated with it, media preferences while learning, and perceived limitations of learning using videos while walking.

*3.1.2 Surveys.* The interviews revealed that users were divided in their preference for watching videos while walking, for learning or otherwise. To further validate the generalizability of these results to the larger population, we developed a questionnaire with a total of 18 questions split into four main sections which gather information on: (1) The frequency and use cases of mobile devices in a variety of on-the-go situations. Sample questions include, "Which of the following on-the-go situations do you encounter on a regular basis ?", and "For what purposes do you use any of your handheld mobile devices during on-the-go situations?" (2) The prevalence of video watching while walking and the difficulties involved through a set of open and close-ended questions, such as, “How frequently do you find yourself watching videos (e.g., YouTube, social media, online courses) in the given situations?” (3) The opinions on learning from videos while walking and related difficulties, using close-ended questions (4) The demographics including age, gender, and educational background.

Pilot tests of the surveys were conducted with 5 users, leading to some questions being reworded or removed to ensure that understandable language was used in the survey and the length was appropriate (around 10 minutes). The final survey was created using Google Forms and distributed through e-mails and chat groups (Whatsapp, Telegram, etc.). Participants completed the survey at a time and location of their choice after indicating their agreement to complete the survey.

80 volunteers responded to the survey (37 females, M=23.55, SD=3.31 years). Participants were from a diverse range of technological (e.g., engineering, computer science) and non-technological (e.g., architecture, design) backgrounds. All respondents possessed and regularly used phones during common on-the-go situations such as while sitting and standing on public transport, and walking for commutes.

## 3.2 Data Analysis

Interviews were audio-recorded, transcribed, and finally coded using Thematic Analysis following both an inductive and deductive approach as laid out by Braun and Clarke [8]. For the deductive analysis, participants' statements relevant to RCF (Resource Competition Framework) [46] that describe user behavior while multitasking in on-the-go scenarios were identified. An inductive approach was then used to craft themes related to the perspectives on video-based learning while walking. Themes were derived from data based on both their frequency of occurrence and perceived substantive significance. Since the survey was meant to demonstrate the relevance of our interview findings in the larger population, the survey results were analyzed in conjunction with that of the interview.

# 4 STUDY 1: RESULTS

In this section, we first report the results of our deductive analysis using the relevant constructs of RCF. We then provide an overview of the inductive analysis with a focus on two major themes: usability barriers and attitude towards learning from videos while walking. For each statement, both interview and survey statistics will be presented, if available, in the format of interview stats followed by survey results (i.e., 15/16, 62.5% means 15 out of 16 interviewees, and 62.5% of the survey respondents agree with this statement).

## 4.1 Deductive Analysis

The most common on-the-go scenarios involving walking include commuting (15/16, 62.5%), short walks to nearby destinations such as cafeterias, meeting rooms, etc. (16/16, 70%), and recreational walks (6/16, 43.8%). Except for one interviewee who experienced motion sickness when using the phone while walking, all others multitasked during these walking scenarios to make use of idle time.

In general, users multitasked with mobile HCI tasks (e.g., texting, watching videos) that did not compete with ongoing mobility tasks of walking for attentional and cognitive resources. This finding is in agreement with the multitask interference property predicted by RCF. For example, auditory-only HCI tasks such as listening to music or news podcasts were preferred while walking, as these did not compete for visual resources that were already under heavy use by the mobility tasks of walking. Performing visually involved mobile HCI tasks such as watching videos while walking was challenging, and many users avoided it (7/16; 66.25%) or did so very rarely (4/16; 20%). This was mainly because watching videos required focused attention, and this could not be afforded while walking: “[While walking in the office] I need to be focused when watching a video. [But] I can't do that as looking at my phone makes me unaware of my surroundings" (P6). This highlights the need for devices that can offer better path navigation while facilitating visually intensive HCI tasks.

However, a number of users (5/16; 13.75%) frequently watched educational videos while walking as they had intrinsic motivational needs or personal time-critical goals to achieve: "I like to keep myself updated on the technology and automobile market" (P10), "My exam was coming soon and I couldn't waste any time" (P2). The educational videos watched by these users were generally visually engaging in nature and of diverse styles, ranging from slides with an instructor to complete animations. According to RCF, these users were able to watch and learn from videos while walking because resources for tasks are allocated hierarchically based on personal needs and goals. Consequently, regular video watchers were able to divert sufficient resources to accommodate video-based learning into their routine.

Apart from strong motivation, these users also adopted compensatory strategies to help them multitask. Such strategies are integral in RCF for avoiding the resource depletion penalties that can be incurred while multitasking. For instance, P10 remarked, "As I have been doing it [watching videos] for a long time, I know which parts [of the video] to skip and which parts to focus". This perceptual sampling strategy allowed users to effectively manage their limited resources between mobility and mobile HCI tasks. Nevertheless, the experience of watching videos while walking was far from ideal even for these users, with many stating that they often had to slow down their walking or pause the video when looking out for potential obstacles. This implies the need to adapt video content to the needs of on-the-go scenarios, allowing users to watch videos with less impact on mobility tasks.

It is also worth noting that several users who rarely watched videos on the go also expressed the desire to do so. For example, P1 remarked, "[while walking in traffic] I needed to respond to a video sent by some colleagues as soon as possible... I couldn't give my full attention to the video as I was walking on the road". Similarly, P14 expressed the wish to watch videos during commute: "My commute is quite long and I sometimes try to get work done... I tried watching some Udemy videos but because of the environment, you miss a few seconds [of the video] and then you need to keep rewinding it". Although they want to multitask, such unpleasant experiences lead these users to discontinue watching videos while walking.

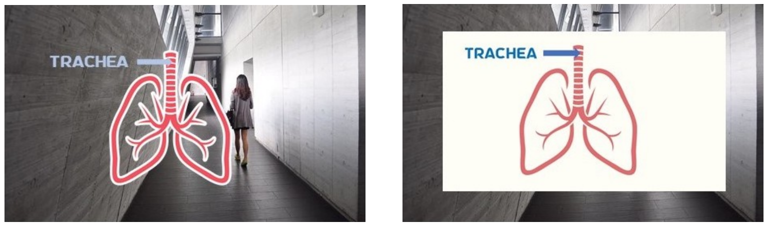
In summary, video is an effective media that users often need to use on a wide range of occasions. However, due to the situational impairments present while walking [24], users are often forced to give up watching videos on the go. Nevertheless, quite a few people try to adapt to the situation and frequently watch videos while walking due to personal goals, even though the experience of doing so is less desirable.

## 4.2 Inductive Analysis

We report findings from our inductive analysis, which revealed users' perspectives on watching and learning from videos while walking. We crafted two sub-themes from the data: usability barriers of phones and attitudes towards learning from videos while walking.

*4.2.1 Usability Barriers of Phones.* As expected, users (10/16; 70%) did not prefer watching videos while walking primarily because they found it difficult to manage their attention effectively between the video and the surrounding environment. Furthermore, users (11/16, 63.7%) mentioned that they feared bumping into people or vehicles: "I don't want to hit someone or slip and fall but I also want to focus on the video" (P12).

Several participants (6/16; 66.3%) mentioned that active holding and heads-down use of the phone was difficult and caused postural discomfort: "I need to take my phone out of my pocket, look down and keep the phone in my hand to watch. This whole thing is itself a process which makes me lose my concentration. This should be changed." (P5). Apart from this, a few users (3/16; 37.5%) found the screen size of phones to be limiting, especially for longer videos: "If the videos are longer than 5 minutes, I prefer to watch on a bigger screen like a laptop" (P1).




(a) (b)

Fig. 2. Difference between viewing videos with a black transparent background (left) and a colored opaque background (right) on OHMD

It serves to highlight here that much of the attention management difficulties and fears while walking arises due to the inherent limitations of the phone device. This calls for the consideration of alternative mobile platforms that could better manage on-the-go multitasking.

*4.2.2 Attitudes towards Learning from Videos while Walking.* In general, videos were highly favored by users (14/16, 71.25%) for learning in stationary situations. However, due to the situational impairments of watching videos while walking, users found themselves unable to engage in meaningful learning, although they needed to do so. Nevertheless, as previously described, some users had modified their learning expectations and found ways (compensatory viewing strategies) to make such learning useful. This indicates that there are potential benefits that can be gained from on-the-go video learning if the process of watching videos while walking can be facilitated in some manner.

To summarize, users' current on-the-go video learning experience using phones is far from ideal, primarily due to heads-down usage and the need to constantly toggle and manage their attention between the video and path navigation as depicted in Figure 1. This motivated us to conduct a follow-up study to investigate the OHMD as an alternative on-the-go video learning platform, as it allows a heads-up viewing experience that can potentially improve path navigation and alleviate postural discomfort.

## 5 STUDY 2: IDENTIFYING SUITABLE PRESENTATION TECHNIQUES FOR OHMD-BASED VIDEO LEARNING

Although OHMDs have several advantages, its see-through display naturally leads users to share their visual bandwidth between the digital screen and the environment, making it harder for users to concentrate on the video content. Thus, existing videos, made with an assumption that users have dedicated attention while watching may not suit the OHMD platform. To find out, we began by conducting pilot studies to test existing learning videos on the OHMD platform.

Through a literature review, we identified six common lecture video styles [11] that differed in the form of instructional media and the level of human embodiment as shown in Figure 3. We conducted pilots with 4 participants by asking them to walk along a straight corridor and learn from lecture videos corresponding to each style using an OHMD. For all pilots and studies, we used an Epson BT-300 binocular optical see-through OHMD, which places the display in the center of the user's line of sight [61]. It has a 1280x720 px resolution display, 23° FoV with a projected distance of 80 inches at 5m.

During the pilot, participants were not allowed to interact with the video in any manner except to start playing it. After watching each video, participants filled a post-questionnaire about how different visual aspects (e.g., animations, text, background, diagrams, and instructor presence) in the video affected their learning experience.

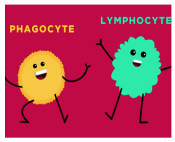
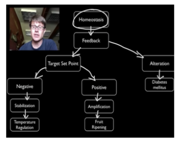
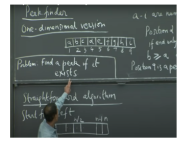
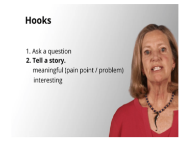
| | Crashcourse | TED-Ed | Bozeman Science | Khan Academy | MIT OCW | Coursera |
|---|---|---|---|---|---|---|
| Sample screenshot |  |  |  |  |  |  |
| Instructional media | Animation | Animation | Digital blackboard | Digital blackboard | Physical blackboard | Slides |
| Human embodiment | Instructor | Nil | Talking head | Pentip | Instructor | Talking head |

Fig. 3. Six common lecture video styles used in the Pilot study

We found that most existing video styles were not designed for on-the-go OHMD-based viewing due to several reasons. First, users preferred videos with a transparent background as they offered an unobstructed view of the environment (see Figure 2(a)) and improved path navigation. Many existing lecture video styles (e.g., 2D animation), however, utilize non-black backgrounds that appear opaque and obstruct a significant portion of the visual field when viewed on an OHMD (see Figure 2(b)). Instead, the digital blackboard style videos were more suitable for on-the-go viewing because the fully saturated black color became transparent in OHMDs, reducing interference to the primary navigational task.

Second, certain video styles did not fully align with the principles of CMLT [42]. For example, 2D animation-style videos usually showed a weak adherence to the modality principle by presenting certain information only as visuals, without any supporting auditory explanation. Lastly, many existing video styles presented content with relatively high information density. Although this is fine when users have dedicated attention, in the case of walking, we found it necessary to control the information density of the video so that it is easier to follow. Our pilots indicate that users preferred videos that had 15 or fewer visual elements[1] evenly displayed on the screen of the OHMD.

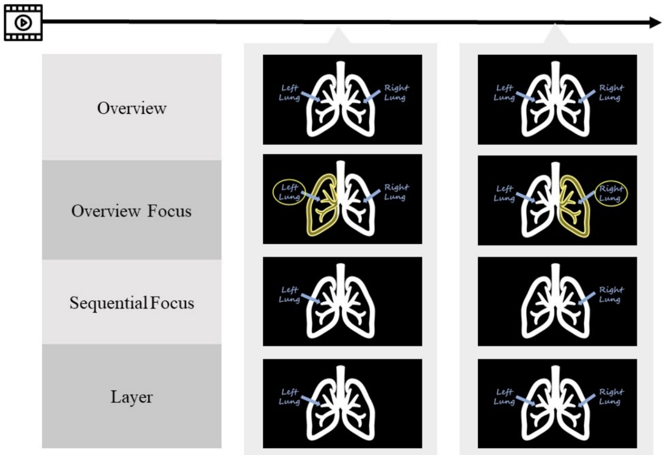
We then dug into existing literature for video presentation techniques that can possibly reduce the cognitive load and improve the viewing experience. We found three presentation techniques that seemed most promising for improving learning [9, 31, 47]: highlighting [31], sequentiality [9], and data persistence [47].

*Highlighting* refers to varying the salience of on-screen information along one or more perceptual dimensions such as color, orientation, or brightness. This technique is useful for guiding the visual attention of users to the relevant on-screen information, thereby facilitating visual search and enhancing the modality effect in video learning [31].

*Sequentiality* refers to the progressive disclosure of information on the screen [9]. The sudden appearance of information can guide the user's attention [28] to the appropriate content, making it easier to link the added information with orally spoken elements in the video. This property can reduce the user's cognitive load and improve the learning outcome [31]. Although both highlighting and sequentiality focus on content in a sequential manner, highlighting not only highlights information, but also makes information visible to users at all times. Conversely, in sequentiality, information is displayed in a just-in-time manner, i.e., information yet to be discussed will not be displayed on the screen until they are mentioned.

*Data Persistence* refers to the property whereby information added to the screen is maintained throughout the video. This property can reduce the temporal demand on the working memory, providing more time to encode information into the long term memory, and thus boost the learning outcome [45, 47]. It is also an important

[1]Visual elements refer to logically related chunks of text/diagrams that are spatially contiguous in the video.




| | Highlighting | Sequentiality | Data Persistence |
|---|---|---|---|
| Overview | x | x | ✓ |
| Overview Focus | ✓ | x | ✓ |
| Sequential Focus | x | ✓ | x |
| Layer | x | ✓ | ✓ |

Fig. 4. The four video styles compared in Study 2 to understand the individual impact of each presentation technique. The table on the right shows the presentation techniques incorporated in each video style.

aspect of traditional blackboard lectures that allows learners to refer back to earlier content and relate it with ongoing discussions [38].

While all three techniques look promising, what remains unknown is how each of them impacts on-the-go video learning on OHMD. Thus, we conducted a controlled experiment with the following research questions and hypotheses.

## 5.1 Research Questions and Hypotheses

**RQ1:** How does highlighting relevant visuals affect users' learning performance and walking ability in on-the-go situations?

While highlighting has been shown to improve users' learning performance in traditional stationary learning scenarios [31], the additional visuals used for highlighting may present added distractions while walking. Hence, we hypothesize that

**H1.1:** Highlighting will improve the user's learning performance

**H1.2:** Highlighting may result in lower walking speed

**RQ2:** How does a progressive presentation of video content affect user's learning performance and walking ability during on-the-go situations?

As sequential presentation guides focus without the need for additional visuals, we expect users' learning performance to improve while leaving their walking ability largely unaffected.

**H2.1:** Progressive presentation will improve the user's learning performance

**H2.2:** Progressive presentation will not cause a decline in walking speed

**RQ3:** What is the impact of data persistence on the learning performance and walking ability of users in on-the-go situations?

Whereas previous research suggests that persistence can improve learning [47], the increased on-screen content may hinder users' walking ability.

**H3.1:** Data persistence will improve the user's learning performance

**H3.2:** Data persistence can decrease walking speed

## 5.2 Measures

We collected three types of measures to compare the different presentation techniques: immediate recognition and recall scores to assess the learning performance based on the fundamental "remember" level of Bloom's taxonomy [36], Percentage of Preferred Walking Speed (PPWS) to measure the decline in users' walking speed, and a survey item to understand users' perceptions about individual presentation techniques.

**Immediate Recognition.** This tests users' ability to recognize whether a given piece of information is familiar. This is an easier "remember" level task than recall due to the availability of more cues for memory retrieval [54]. The recognition test consisted of assertion and multiple-choice questions. The participant's score (0-6) was used as a measure of their recognition ability.

**Immediate Recall.** This tests users' ability to retrieve information from memory with little to no cues provided to aid the memory retrieval process. The recall test consisted of filling in the blank or listing questions. The participant's score (0-9) was used as a measure of their recall ability.

**Percentage of Preferred Walking Speed.** This indicates how much slower a user walks as compared to his/her preferred walking speed. To get PPWS, we first obtain users' preferred walking speed (in m/s) by recording the time taken by them to walk a fixed distance without wearing any device. The PPWS (0-100%) is then obtained by dividing the walking speed during the task by the preferred walking speed.

**Preference for presentation techniques.** A post-experiment questionnaire item was provided after each video asking participants to rate aspects of each presentation technique that impacted them positively or negatively. This included questions like "Having information on the screen till the end of the video affected your on-the-go video learning experience (positively/negatively)."

### 5.2.1 *Scoring of Quiz.*

For both recognition and recall, correct answers received 1 point. If the word spellings in the recall task were incorrect, we calculated the error on a participant's answer using the Levenshtein [40] distance, which counts the minimum number of insertions, deletions, and substitutions needed to correct the spelling. If the error was greater than half the length of the word, it was awarded 0 points, otherwise, it was awarded 1 point. For instance, if the word 'sinoatrial' (word length = 10) was misspelled as 'cyanoareeal', no points were provided as the words differed by a Levenshtein distance of 6 units, which is greater than half the word length.

## 5.3 Lecture Videos: Style and Content

We used digital blackboard style videos from the Khan Academy e-learning platform in our study. We focused on this video style as previous research has shown that the characteristic digital handwriting feature in this video style is beneficial for learning [20]. For the study, we used videos from an introductory level medical course created by the *khanacademymedicine* YouTube channel. We chose a biology-related domain as it contains more factual knowledge [44], making it well-suited for assessing the learning measures of interest in our study, i.e., users' fundamental recognition and recall skills.

## 5.4 Participants & Apparatus

We recruited 16 participants P1-16 (6 female) from the university community. Their average age was $(25.38 \pm 0.79)$ years. Three of the users mentioned that they had prior experience using OHMD. Participants were ensured to be from a non-biology background to minimize any potential bias due to prior knowledge.

The videos were pre-loaded and viewed in an offline manner on the default video player application in Epson. We selected four videos from the *khanacademymedicine* Youtube channel (V1-V4). All videos were chosen to be of similar lengths (V1: 8:25; V2: 8:29; V3: 8:44; V4: 8:25) and were pilot tested to ensure a similar level of difficulty.

The recall/recognition questions were created based on the terms and facts discussed in each video and piloted to have a similar difficulty level across videos.

## 5.5 Design

The study sought to compare the effects of incorporating three presentation techniques ~ Highlighting (Yes, No) x Sequentiality (Yes, No) x Data Persistence (Yes, No). While a full factorial design has 8 conditions, only 4 of these were considered as shown in the table in Figure 4. Conditions where both sequentiality and data persistence are absent are invalid because the information discussed at the end of the video persists for the entire video, thus contradicting the data persistence property. Similarly, conditions involving both sequentiality and highlighting were dropped as both techniques serve to guide users' attention. Hence, we adopted a within-subject reduced factorial design [13] for the study, where conditions were counterbalanced using a balanced Latin square.

*5.5.1 Video Design.* The four videos used for the study were modified to include the corresponding presentation technique as shown in Figure 4. For the Overview and Overview Focus conditions, the video frame contains all the information for the entire video duration. To incorporate highlighting in the Overview Focus condition, the information associated with the ongoing audio explanation was enclosed within a fully saturated green boundary. The green hue for highlighting was chosen based on previous research that suggested the global effectiveness of green color for viewing on OHMD [21]. In the Sequential Focus and Layer conditions, the visuals were progressively disclosed on the screen, coinciding with the relevant audio explanation. While this information was removed immediately after the explanation was over in the Sequential Focus condition, the information stayed on the screen till the end of the lecture in the Layer condition.

*5.5.2 Walking Path.* Participants watched and learned from the video as they walked back and forth on a 54m long straight path consisting of objects such as dustbins placed along the sides. This setting was chosen to be representative of common environments traversed by people on a daily basis.

## 5.6 Procedure

The experiment started by providing participants with instructions regarding the various tasks in the study and recording their preferred walking speed by instructing them to walk at their normal pace along the straight path. Next, they were provided with an initial warm-up session where they watched a demo lecture video to familiarize themselves with the task.

To avoid fatigue, the experiment was conducted in two sessions lasting approximately 40 minutes each. During each session, participants watched and learned from two videos on the OHMD, while walking back and forth along the straight path. They were allowed to passively watch the video only once, but they were allowed to adjust the OHMD to align with their view angle during the experiment if required.

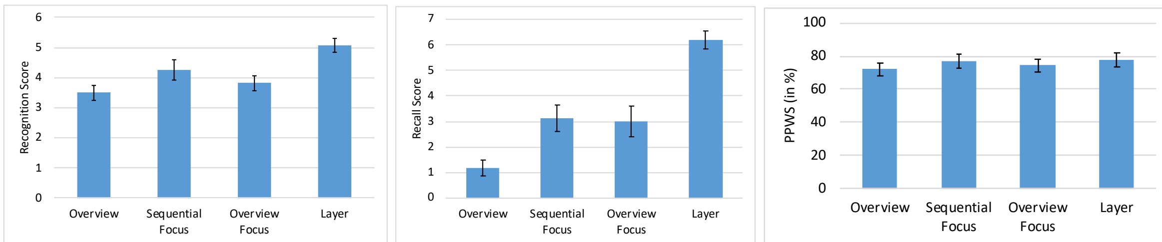
After watching each video, participants took a quiz that tested their recognition and recall based on content learned in that video and subsequently completed a post-questionnaire. A short break was provided before moving on to the next video to avoid fatigue.

# 6 STUDY 2 - RESULTS

Immediate recognition/recall and PPWS scores did not meet the normality assumption of ANOVA (Shapiro-Wilk tests, $p < .05$). Therefore, the non-parametric Friedman's test was used, with Wilcoxon signed rank tests and Bonferroni correction for post hoc analysis.

## 6.1 Learning Performance

The Friedman's test indicated a significant main effect in both immediate recall ($\chi^2_{(3,N=16)} = 34.76, p < 0.001$) and recognition ($\chi^2_{(3,N=16)} = 20.81, p < 0.001$) as shown in Figure 5(a) and 5(b). Further post hoc analysis revealed the following:




(a) Recognition (b) Recall (c) PPWS

Fig. 5. (a) Immediate recognition scores (b) Immediate Recall scores and (c) PPWS comparison for 16 participants.

*Effects of Highlighting*: Immediate recall was better in the Overview Focus condition ($3.00 \pm 0.59$) than in the Overview condition ($1.19 \pm 0.31$) ($p_{bonf} = 0.024, d = 0.94$) indicating that highlighting relevant visual elements improves learning.

*Effects of Sequentiality*: We found that participants scored significantly higher in the Layer condition ($6.25 \pm 0.35$) as opposed to the Overview condition in the recall tasks ($p_{bonf} < 0.001, d = 1.00$). A significant improvement was also seen in participants' recognition scores when using the Layer condition ($5.06 \pm 0.23$) instead of the Overview condition ($3.50 \pm 0.24$) ($p_{bonf} < 0.001, d = 0.93$) . This demonstrates the importance of progressively presenting information for on-the-go learning.

*Effects of Data Persistence*: Performance in both recall ($p_{bonf} < 0.001, d = 0.93$) and recognition ($p_{bonf} = 0.01, d = 1.00$) tasks significanlty declined when data persistence was removed. This decline in performance was much higher in recall tasks, with Sequential Focus condition having an average score of ($3.12 \pm 0.51$) as opposed to Layer condition ($6.25 \pm 0.35$), providing clear evidence of the importance of data persistence for learning.

## 6.2 Percentage of Preferred Walking Speed

A significant main effect ($\chi^2_{(3,N=16)} = 13.65, p = 0.003$) was found in the PPWS between conditions shown in Figure 5(c). Post hoc analysis showed that:

*Effects of Highlighting*: No significant difference between the Overview Focus and Overview condition in walking speed was found.

*Effects of Sequentiality*: There was a significant difference in the PPWS between Layer ($77.5 \pm 4.2$) and Overview condition ($72.10 \pm 4.0$)($p_{bonf} = 0.01, d = 0.81$).

*Effects of Data Persistence*: No significant difference was observed between the Sequential Focus and Layer condition, implying that data persistence does not impact the user's walking speed.

## 6.3 Post-questionnaire

Overall, more than half the participants (10/16) preferred having a transparent background for the video while walking. We also found that most participants (12/16) prefer not to have all elements always presented on-screen, as in the Overview and Overview Focus conditions. Most participants (12/16) however, mentioned that their learning experience was better when the visual content was highlighted, as in the Overview Focus condition. In both the Layer and Sequential Focus conditions, the content was progressively added to the screen. Yet, while only a few participants (8/16) preferred this sequential feature in the Sequential Focus condition, most participants (14/16) found it beneficial in the Layer condition. Moreover, the retention of discussed content until the end of the video was also found to be beneficial by most participants (14/16) in the Layer condition. Apart from this, several participants also mentioned issues with the clarity of the video content when viewed against a clear background, which hindered their ability to read certain words in the video.

### 6.4 Discussion

We organized the discussion below into parts that answer each of the research questions that we put forth earlier.

**RQ1:** How does highlighting relevant information affect users' learning performance and walking ability in on-the-go situations?

In our hypothesis H1.1, we had predicted that highlighting would improve both the recognition and recall performance of users. Our results support this, showing significant gains in both learning tasks. This improvement can be attributed to the appropriate guidance of users' attention to relevant on-screen information, which helps reduce the cognitive load on users' working memory. These benefits of highlighting were also evident from participants' comments in the post-questionnaire; while participants in the Overview condition found it "difficult to know what exactly he [speaker] was talking about" (P10), this issue was largely resolved in the Overview Focus condition where participants felt it to be "helpful when the speaker highlighted what he was talking about on the diagram" (P11).

However, our expectation from H1.2 that the walking speed would decrease in the Overview Focus condition in comparison to the Overview condition did not hold. This could be due to the increased amount of cognitive resources spent on visual search in the Overview condition, which in turn reduces the cognitive resources available for walking.

**RQ2:** How does a progressive presentation of video content affect users' learning performance and walking ability during on-the-go situations?

Our prediction in H2.1 that sequentially presented content would improve both the recognition and recall scores of users has proved true. This is because the sudden appearance of on-screen content evokes visual attention, which in turn eases visual searching and reduces the user's cognitive load.

Interestingly, the benefits of using sequentiality outweighed that of highlighting, with users' scoring much higher when the content is progressively added on-screen. This finding can be adequately interpreted from the point of view of Perceptual Load Theory (PLT) [39], which is useful for explaining the selective attention processes that occur in perceptual tasks. According to PLT, the attentional resources get utilized such that task-relevant (target) stimuli are processed before task-irrelevant (distractor) stimuli. Moreover, if more attentional resources are needed to focus on the target, i.e., the perceptual load of the target task is high; PLT states that there is less spillover of attention to the distractors. In the context of video learning, the target can be interpreted as information being explained in the video, and distractors refer to the remaining information on screen.

Applying PLT to our findings, we see that since both highlighting and sequentiality operate to guide users' attention and ease the visual selection process, incorporating either technique will reduce the overall perceptual load of the target task. In such a low perceptual load situation, according to PLT, the leftover attentional resources would spill over to the distractors, i.e., the remaining on-screen content. The crucial difference, however, is that in the Layer condition, this distractor content has already been discussed whereas in the Overview Focus condition most of the distractors have not yet been discussed. Therefore, the attention spill over positively reinforces learning in the former condition, but hinders learning in the latter condition.

Furthermore, users were able to achieve a 3.9 % faster walking speed when information was presented sequentially as opposed to being highlighted. This could be due to the presence of additional visuals required for highlighting in the Overview Focus condition, which can cause distractions, whereas progressively presented content inherently evokes attention without the need for additional visuals that can distract the user.

**RQ3:** What is the impact of data persistence on the learning performance and walking ability of users in on-the-go situations?

The learning performance of participants was better in the Layer condition as opposed to the Sequential Focus condition, implying the importance of data persistence. This was also evident in the user's post-questionnaire comments. For instance, participant P7 remarked, "I could not remember the words while attempting the quiz because the words weren't retained on-screen after explanation." This could be in large due to the temporal demand imposed on the learner's working memory by the transient video content [45]. Research has shown that unattended information in an individual's working memory lasts for about 10-15 seconds before being degraded or replaced with new information [25]. For successful recall, this information must be encoded and consolidated into the long term memory within this duration [32, 43]. In the case of the Sequential Focus condition, the duration available for this encoding process is much shorter than in the Layer condition, thereby increasing the demand on users' working memory. Additionally, evidence also indicates that persistence can encourage active learning where the user decides how to interlink the discussed contents leading to better recall [38, 47].

This controlled study revealed the significance of each of the three presentation techniques. However, due to the unbalanced design of the study, it is difficult to distill the exact effect of each technique on the user's learning and walking ability. For example, to identify the effect of sequentiality, we compare Sequential Focus with the Layer condition. However, since both conditions include data persistence, a pairwise comparison between the conditions would show the effect of both sequentiality and its interaction with data persistence.

Nevertheless, the results inform us on the relative significance of the techniques. They indicate that both sequentiality and complete data persistence are crucial, and should be combined for on-the-go video learning on OHMDs. Also, as informed by our pilots, it is necessary to ensure the right level of information density in the video (at most 15 visual elements evenly spaced out on-screen) and use a transparent video background to improve the viewability of the surroundings. In addition, the original video content must be recolored using shadow techniques in order to improve its color contrast with the environment.

We consolidate these requirements into the Layered Serial Visual Presentation (LSVP) style for OHMD-based on-the-go video learning. Notably, although we move forward with the LSVP style in its current form for future evaluation, we further modify LSVP based on an additional design aspect that arose in Study 3. Thus, the final LSVP style, which we propose for future on-the-go video learning is further elaborated in Section 9.

## 7 STUDY 3: COMPARING OHMD VS. PHONE

Our previous study suggests that learning videos presented using LSVP style is best suited for OHMD-based on-the-go video learning. Yet, it is unclear how our proposed solution compares against the current phone-based video learning approach. While the advantage of OHMDs lies in its heads-up, hands-free usage, and its ability to facilitate seamless switching of visual attention between the environment and video content, its see-through display also makes it difficult for users to concentrate on the video content. Despite its heads-down usage, phones offer a more exclusive viewing experience. Also, OHMDs today typically have a lower resolution than phones and users are more familiar with phones than with OHMDs. These differences make it difficult to predict which platform is more suitable for on-the-go video learning. To find out how the OHMD platform compares with the de facto phone platform for on the go video learning, we conducted a controlled experiment. For our investigation, we compared the platforms in different mobility scenarios of varying complexity.

### 7.1 Research Questions and Hypotheses

To better understand how phones and OHMDs differ in their ability to balance users' learning performance and walking ability in various mobility scenarios, we investigated the following two main research questions:

**RQ1:** How do the platforms compare in enabling users to learn from videos and walk simultaneously?

While the focused attention afforded by phones may be beneficial for on-the-go video-based learning, this viewing behavior could degrade users' navigation performance, especially in visually challenging mobility scenarios. Contrarily, since the OHMD displays video content alongside the surroundings, it will enable quicker attention-switches between the video content and surroundings. This led us to hypothesize that,
**H1.1:** Phones will improve users' learning performance and path navigation in simple mobility scenarios.
**H1.2:** OHMDs will improve users' learning performance and path navigation in complex mobility scenarios.

**RQ2:** What role does video design play in facilitating on-the-go video learning?
Although both LSVP and Khan style present data sequentially, LSVP uses complete data persistence as opposed to Khan, where the data only persists partially. As a result, LSVP would naturally provide users with more time to encode the learning content than Khan. Thus, we hypothesized that,
**H2:** LSVP style will outperform Khan style in terms of learning performance, irrespective of the platform.
It is unclear, however, as to what effect the video design will have on users' walking speed.

## 7.2 Measures

We reused the immediate recognition/recall tests and PPWS measures from Study 2. We also incorporated two additional measures:

**Delayed Recognition & Recall.** Given that language learning literature uses delayed tests as a strong indicator that learning has happened in long term memory, we included 7-day delayed tests for a more comprehensive understanding of the differences between the platforms, in terms of users' fundamental "remember" skills [36]. The quiz was sent out to the participants via mail and consisted of recognition and recall questions. The question formats were similar to those in the immediate tests. The participant's score (0-10) was used as a measure of their recognition and recall abilities. This added measure served to strengthen our learning performance results.

**Platform preference.** Participants were asked to indicate the platform they preferred for each mobility task and their overall preference in a post-experiment questionnaire.

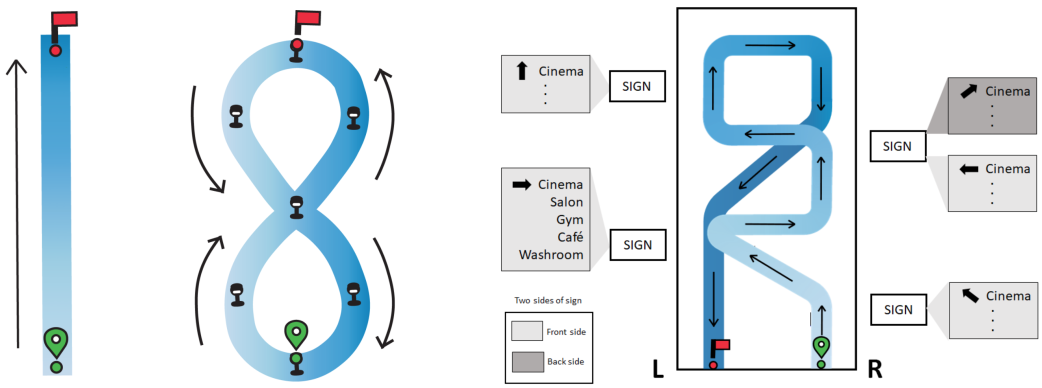
## 7.3 Design

A within-subject design with 2 Platforms (OHMD, Phone) x 2 Video Designs (Khan, LSVP) x 3 Mobility Tasks (Sim, Obs, Nav) was used (details below), which resulted in 12 conditions per participant. The order of Platform and Video Design variables were counterbalanced using a balanced Latin square design while the order of Mobility Tasks was presented in sequential order with increasing difficulty.

With the number of conditions and tasks involved, the experiment was estimated to take 3 hours to complete. Based on pilot feedback, we split the experiment into two 80-90 minute sessions with an inter-session break of one hour to balance between continuity of the experiment and possible fatigue. Further details on the design of the independent variable and their levels are given below.

*7.3.1 Mobility Tasks.* We included three mobility tasks with increasing levels of difficulty/complexity as shown in Figure 6. These tasks simulate realistic everyday walking situations. Given that video visibility on the OHMD is affected by outdoor lighting [21], all scenarios were designed in indoor lighting conditions to minimize any platform-specific bias due to video visibility.

The simplest task: *Sim* used a straight path with no obstacles (shown in Fig 6(a)). The slightly more difficult Obs task was designed as a curved 8-figure path with obstacles along the 8-figure outline (shown in Fig 6(b)), which is representative of situations that require obstacle avoidance and more complex motor skills than a straight path [27]. The Nav task was more challenging. It simulated wayfinding and required more cognitive faculties than those necessary for Obs and Sim tasks [46] (more details below). The increasing difficulty of the mobility tasks




(a) Sim (b) Obs (c) Nav

Fig. 6. Mobility tasks of varying complexity. In the Sim and Obs task, the paths measure 22.8m from start (green pin) to finish (red pin). For Obs, one round measures 16.8m, taking 1.35 rounds to complete the path. For the Nav task, the distance is measured to be ≈ 50m from start to finish.

was further confirmed by pilots, which showed a decrease in users' Preferred Walking Speed (PWS) (Sim - 1.05 m/s, Obs - 0.9 m/s, Nav - 0.83 m/s).

*Design of Nav Task.* This task simulated a real-world navigation scenario where users needed to decide between possible paths (represented by directions) at a junction (represented by signboards) to reach the correct location (see Fig 3(c)). At the starting point (green circle in Fig 6(c)), users were initially assigned to one of five possible destinations (Salon, Cinema, Gym, Cafe, Washroom). Consider the 'Cinema' as an example destination. The signboard at the starting point shows an up-left arrow, which directs the user to walk diagonally to the left with a 45° angle. Once the second signboard is reached, new instructions for the 'Cinema' appear, which instructs the user to walk horizontally to the right. Following that instruction, users will see the third signboard, which provides additional instructions, and so on until s/he reaches the destination (red flag in Fig 6(c)). The signboards were placed at an optimal viewing height of 1.5 meters [3]. Note that although 'Cinema' in Fig 6(c) always appears as the first item, the order in which the destination appears on the signboard was randomized in the experiment to avoid shortcut behavior.

*7.3.2 Video Design.* We considered two video designs: the LSVP style from Study 2 adapted for the OHMD and Khan style, which is an unmodified Khan Academy video suited for the phone.

## 7.4 Participants & Apparatus

Sixteen participants (8 females) aged between 20 to 26 years (22.81 ± 2.09) with high school or higher degrees were recruited for the study. Participants were selected if they regularly watched videos on their phone while walking and if they had not taken any course whose curriculum overlapped with the biology-related video content used for the study. The latter condition was necessary to minimize any potential bias due to prior knowledge that may aid users' learning. Each participant was compensated in monetary form ($7.20 per hour). No participants from Study 2 or any of our pilot studies participated in this study.

The setup for the OHMD platform is identical to that in Study 2. For the Phone platform, we let participants use their own phones (instead of providing a standard phone) to maximize device familiarity [22]. The phones

Table 1. Descriptive results for all conditions in Study 3

| | | Sim | | | | Obs | | | | Nav | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | OHMD | | Phone | | OHMD | | Phone | | OHMD | | Phone | |
| | | M | SE | M | SE | M | SE | M | SE | M | SE | M | SE |
| **Immediate Recognition** | **LSVP** | 4.25 | 0.25 | 3.62 | 0.30 | 4.12 | 0.22 | 3.25 | 0.33 | 3.94 | 0.25 | 3.19 | 0.39 |
| | **Khan** | 3.50 | 0.20 | 4.25 | 0.25 | 3.43 | 0.26 | 3.81 | 0.16 | 2.81 | 0.21 | 3.37 | 0.30 |
| **Immediate Recall** | **LSVP** | 4.06 | 0.21 | 4.06 | 0.25 | 3.81 | 0.23 | 1.75 | 0.31 | 3.12 | 0.18 | 0.87 | 0.27 |
| | **Khan** | 1.18 | 0.34 | 2.31 | 0.45 | 1.75 | 0.32 | 1.44 | 0.27 | 1.00 | 0.26 | 1.87 | 0.34 |
| **Delayed Recognition** | **LSVP** | 4.00 | 0.24 | 2.68 | 0.22 | 3.62 | 0.38 | 3.25 | 0.25 | 3.37 | 0.27 | 2.44 | 0.33 |
| | **Khan** | 3.06 | 0.26 | 2.87 | 0.30 | 2.87 | 0.36 | 2.56 | 0.33 | 2.00 | 0.27 | 2.00 | 0.26 |
| **Delayed Recall** | **LSVP** | 3.00 | 0.48 | 0.37 | 0.18 | 2.50 | 0.39 | 0.62 | 0.27 | 1.19 | 0.27 | 0.18 | 0.13 |
| | **Khan** | 0.12 | 0.12 | 0.31 | 0.19 | 0.37 | 0.15 | 0.62 | 0.15 | 0.44 | 0.13 | 0.37 | 0.15 |
| **PPWS (%)** | **LSVP** | 84.94 | 1.93 | 82.38 | 2.13 | 79.13 | 2.21 | 78.94 | 2.37 | 92.69 | 1.68 | 82.69 | 2.64 |
| | **Khan** | 81.13 | 2.52 | 81.44 | 2.24 | 77.19 | 2.52 | 77.56 | 2.42 | 87.75 | 2.17 | 86.63 | 2.02 |

had a screen resolution of 1280 x 720 or higher to support the 720p quality of the videos. To have a more realistic comparison with OHMD, we did not constrain the participants' mobile video watching habits such as the choice of hand(s), orientation, and the viewing distance for holding the phone. All videos had the same playback speed and were received offline.

The video style and content domain remain unchanged from Study 2. We selected 12 videos related to different concepts from the *khanacademymedicine* YouTube channel (V1-V12). All videos were trimmed to be of similar lengths (M=6.03 min, SD=0.067 min) and were pilot tested to ensure a similar level of difficulty. Questions for the recall/recognition tests were also designed in a manner similar to that in Study 2.

## 7.5 Procedure

Participants were initially briefed about the tasks in the experiment. Next, as a warm-up exercise, participants walked along an alternate straight path while learning from a six-minute demo video, watching the first half of the video on the OHMD and the rest on their phones.

Before starting the conditions in a mobility task, participants' Preferred Walking Speed (PWS) in that mobility task was recorded. During the experiment, participants watched videos on the platforms and walked at their natural pace as they learned the video content and engaged in the mobility task. The resulting walking speed was also monitored for each condition. After each video, participants took a quiz that assessed their immediate recall and recognition abilities for the topic learned in that condition. A short break was provided between each condition to reduce fatigue.

After the experiment, participants completed a questionnaire and a short interview where they described their preference for a platform in each mobility task, along with the reasons for their choice. The experiment took 4 hours to complete (inclusive of the inter-session break).

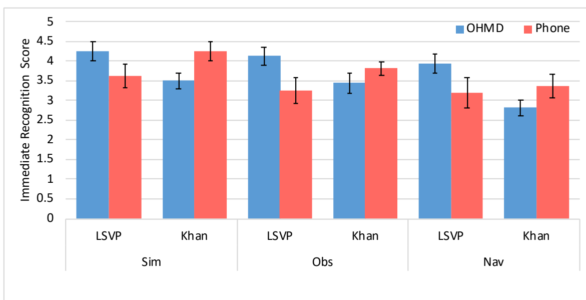
After 7 days, a quiz was sent to the participants to test their delayed recall and recognition abilities. The quiz contained a total of 120 questions, with 10 questions (5 recognition, 5 recall) on each video they watched in a condition. Users were instructed to complete the test solely based on what they remembered and were asked not to refer to any online/offline content.

## 8 STUDY 3: RESULTS

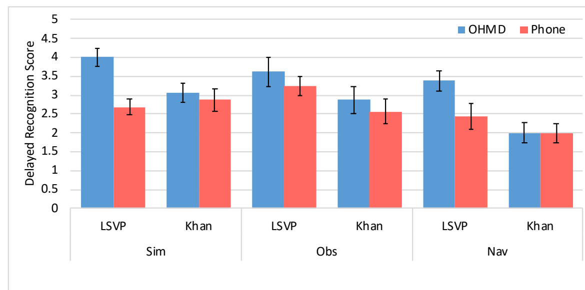
All results violated the normality assumption of ANOVA (Shapiro- Wilk tests, p<.05). Therefore, repeated measures ANOVA was performed after applying aligned rank transformation [58], with Bonferroni corrected post hoc analysis. Table 1 summarizes the descriptive results.

### 8.1 Learning Performance

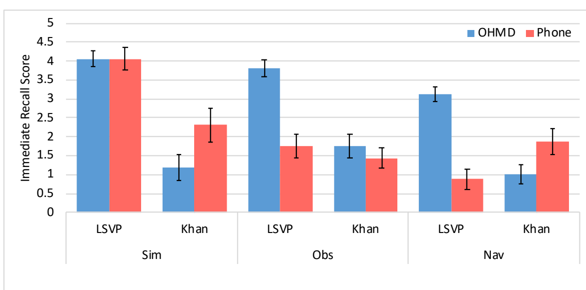
A repeated measures analysis of variance (ANOVA) was performed individually on the immediate recognition, immediate recall, delayed recognition, and delayed recall scores after ART ~ Platform x Video Design x Mobility Task. The results are shown in Figure 7 and Figure 8.



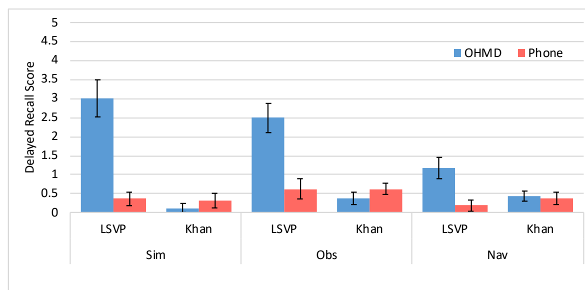

(a) Immediate Recognition




(b) Delayed Recognition




(c) Immediate Recall




(d) Delayed Recall

Fig. 7. (a)-(d) Measured outcomes for learning performance comparing OHMD with phone

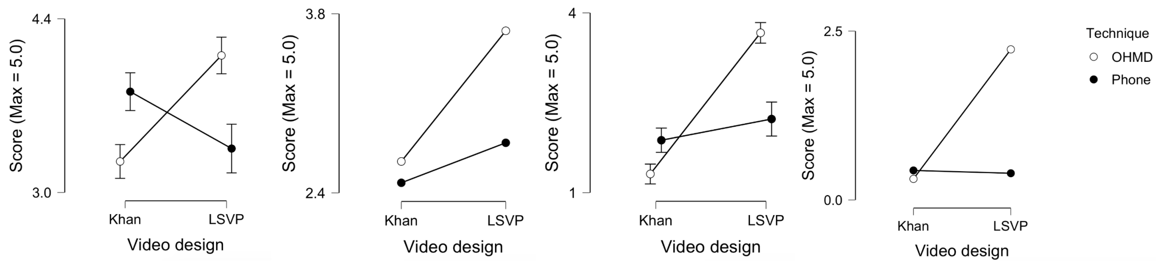
#### 8.1.1 *Immediate Recognition.*

The ANOVA revealed a significant main effect for *Mobility Task* ($F_{2,30} = 8.30, p = 0.001, \omega^2 = 0.132$). The scores were significantly lower in the *Nav* task as compared to *Obs* ($p_{bonf} = 0.008, d = 0.90$) and *Sim* ($p_{bonf} = 0.010, d = 0.87$) however, no significant difference between *Obs* and *Sim* tasks were found.

Additionally, there was a significant *Platform x Video Design* ($F_{1,15} = 25.16, p < 0.001, \omega^2 = 0.234$) interaction effect on the immediate recognition performance. Post hoc analysis showed that overall, OHMD with LSVP style ($4.10 \pm 0.14$) significantly outperformed OHMD with Khan ($3.25 \pm 0.13$)($p_{bonf} < 0.001, d = 0.91$) and phone with LSVP ($3.35 \pm 0.19$)($p_{bonf} = 0.024, d = 0.84$).

#### 8.1.2 *Immediate Recall.*

There was a significant main effect for *Platform* ($F_{1,15} = 6.31, p = 0.024, \omega^2 = 0.101$), *Mobility Task* ($F_{2,30} = 15.57, p < 0.001, \omega^2 = 0.292$) and *Video Design* ($F_{1,15} = 57.08, p < 0.001, \omega^2 = 0.524$).




(a) Immediate Recognition (b) Delayed Recognition (c) Immediate Recall (d) Delayed Recall

Fig. 8. Interaction effects between Platform and Video design for Learning performance

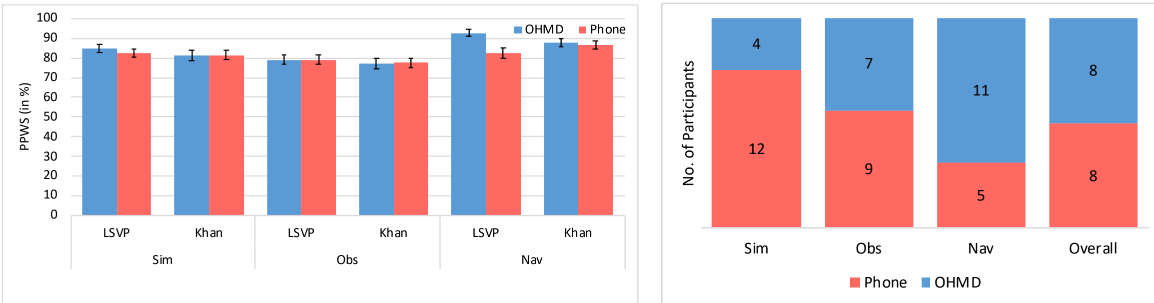
Participants' score was better overall using the OHMD ($2.49 \pm 0.16$) than with the phone ($2.05 \pm 0.16$)($p_{bonf} = 0.024, d = 0.63$). Additionally, the LSVP style ($2.94 \pm 0.16$) outperformed the Khan design ($1.60 \pm 0.14$)($p_{bonf} < 0.001, d = 1.89$).

There were also significant interaction effects observed on immediate recall scores: *Platform x Mobility Task* ($F_{2,30} = 15.18, p < 0.001, \omega^2 = 0.240$), *Platform x Video Design* ($F_{1,15} = 74.31, p < 0.001, \omega^2 = 0.420$), and *Mobility Task x Video Design* ($F_{2,30} = 7.84, p = 0.002, \omega^2 = 0.159$). Post hoc analysis showed that overall, OHMD has comparable performance to the phone in case of the *Sim* task but outperformed the phone on the *Obs* ($p_{bonf} < 0.001, d = 1.00$) task. In terms of video design, the LSVP style in OHMD ($3.67 \pm 0.13$) offered a significant improvement in immediate recall scores across all mobility tasks, as compared to the Khan design in both the phone ($1.87 \pm 0.21$)($p_{bonf} < 0.001, d = 0.91$) and OHMD ($1.31 \pm 0.18$)($p_{bonf} < 0.001, d = 1.00$). Furthermore, while the use of LSVP style showed no significant difference in scores for either platform in the case of the *Sim* task, this difference became significant with increasing mobility task complexity. In case of the *Obs* task, the use of LSVP style on OHMD outperformed that on the phone significantly by ($2.06 \pm 0.31$)($p_{bonf} < 0.001, d = 1.00$) whereas in the *Nav* task, the difference was higher ($2.25 \pm 0.23$)($p_{bonf} < 0.001, d = 1.00$).

*8.1.3 Delayed Recognition.* A significant main effect was seen in the *Platform* ($F_{1,15} = 12.61, p = 0.003, \omega^2 = 0.144$), *Mobility Task* ($F_{2,30} = 16.28, p < 0.001, \omega^2 = 0.211$)and *Video Design* ($F_{1,15} = 13.71, p = 0.002, \omega^2 = 0.188$). In general, participants were able to perform better using the OHMD ($3.16 \pm 0.14$) than with the phone ($2.63 \pm 0.12$)($p_{bonf} = 0.004, d = 0.88$) in 7 day delayed recognition test. In addition to this, the LSVP style ($3.23 \pm 0.13$) outperformed the Khan design ($2.56 \pm 0.13$)($p_{bonf} = 0.002, d = 0.92$). The scores also decreased as the mobility task became more complex, and was significantly lower in the *Nav* task than the *Obs* ($p_{bonf} < 0.001, d = 1.28$) and *Sim* ($p_{bonf} < 0.001, d = 1.28$) task.

Furthermore, a significant interaction effect was observed between *Platform x Video Design* ($F_{1,15} = 25.16, p < 0.001, \omega^2 = 0.234$) on the delayed recognition performance. Post hoc tests showed that overall, OHMD with LSVP style ($3.67 \pm 0.18$) significantly outperformed OHMD ($2.65 \pm 0.18$)($p_{bonf} = 0.012, d = 0.83$) and phone ($2.48 \pm 0.17$)($p_{bonf} = 0.018, d = 0.79$) with Khan style.

*8.1.4 Delayed Recall.* A significant main effect for *Platform* ($F_{1,15} = 145.38, p < 0.001, \omega^2 = 0.622$), *Mobility Task* ($F_{2,30} = 15.57, p < 0.001, \omega^2 = 0.292$) and *Video Design* ($F_{1,15} = 79.19, p < 0.001, \omega^2 = 0.616$). Overall, using the OHMD improved participants' delayed recall scores than when they used the phone by ($0.85 \pm 0.06$)($p_{bonf} < 0.001, d = 3.01$). The LSVP style also led to significantly better performance than the Khan design by ($0.94 \pm 0.12$)($p_{bonf} < 0.001, d = 2.22$).




(a) PPWS (b) Subjective Platform Preference

Fig. 9. Comparison of phones with OHMD in terms of (a) PPWS (b) Preferred Technique measured for 16 participants.

The ANOVA also revealed the following significant interaction effects: *Platform x Video Design* ($F_{1,15} = 124.16, p < 0.001, \omega^2 = 0.654$), *Platform x Mobility Task* ($F_{2,30} = 4.43, p = 0.021, \omega^2 = 0.111$) and *Mobility Task x Video Design* ($F_{2,30} = 20.54, p < 0.001, \omega^2 = 0.341$). Post hoc analysis showed that the use of LSVP style on the OHMD ($2.23 \pm 0.25$) offered a significant improvement in delayed recall scores, as compared to using the Khan design on the phone ($0.39 \pm 0.19$)($p_{bonf} < 0.001, d = 1.00$) and the OHMD ($0.31 \pm 0.08$)($p_{bonf} < 0.001, d = 1.00$).

## 8.2 Percentage of Preferred Walking Speed

We performed repeated measures ANOVA on PPWS scores ~ *Platform x Video Design x Mobility Task*, shown in Figure 9. There was a significant main effect for *Platform* ($F_{1,15} = 7.52, p = 0.015, \omega^2 = 0.013$), *Mobility Task* ($F_{2,30} = 38.66, p < 0.001, \omega^2 = 0.173$), and *Video Design* ($F_{1,15} = 26.85, p < 0.001, \omega^2 = 0.009$). PPWS with the phone ($81.60 \pm 0.96$) was significantly lower than with the OHMD ($83.80 \pm 1.01$)($p_{bonf} = 0.015, d = 0.68$). PPWS was also higher while using the LSVP style ($83.46 \pm 0.93$) than the Khan design ($81.95 \pm 0.99$)($p_{bonf} < 0.001, d = 1.29$). Furthermore, there was a significant *Platform x Mobility Task* ($F_{2,30} = 5.63, p = 0.008, \omega^2 = 0.016$), and *Platform x Video Design* ($F_{1,15} = 9.92, p = 0.007, \omega^2 = 0.009$) interaction effect on PPWS.

Post hoc comparison showed that the OHMD ($90.22 \pm 1.42$) allowed participants to maintain significantly higher PPWS than the phone ($84.66 \pm 1.67$) in the *Nav* task ($p_{bonf} < 0.001, d = 1.00$). The improved PPWS of OHMD on the *Nav* task, however, was observed only when the OHMD used the LSVP design as opposed to other possible platform-video design combinations (all at the $p_{bonf} < 0.001$ level). For the other mobility tasks and video designs, there was no significant difference between the OHMD and the phone.

## 8.3 Discussion

**RQ1:** How do the platforms compare in managing the user's visual/cognitive needs during on-the-go video learning?

Overall, participants were able to walk faster and learn better with the OHMD as compared with the phone. These results were unexpected given that most (except one) of the participants were first time users of the OHMD. In particular, the OHMD had a significant advantage over the phone when the mobility task complexity was higher, as seen in the case of the *Obs* and *Nav* tasks. Thus, our hypothesis that the OHMD will enable better on-the-go video learning in complex mobility tasks was validated.

In terms of learning ability, there was no significant difference in both the immediate and delayed recognition scores achieved by participants using either platform. This indicates that both platforms can sufficiently handle the visual/cognitive needs necessary for easier learning tasks like recognition. On the other hand, in the more challenging immediate and delayed recall tasks, the OHMD outperformed the phone in complex mobility situations where users needed to switch their attention between the video and the surroundings. In such situations, the OHMD offered lower attention-switching latency as compared to the phone, thereby allowing users to utilize their visual/cognitive resources for learning effectively. This was also verified by participants (68.8%), who felt that switching attention between the content and the surroundings was more effortless on the OHMD display as compared to moving their head up and down while using the phone.

Participants were also able to reach significantly higher overall walking speeds while using the OHMD than with the phone with two exceptions. 1) In the Sim task, no significant increase in walking speed was observed. This is expected as the *Sim* task does not require much attention switching, so that walking speeds are not as affected. 2) In the *Obs* task, the walking speed was equivalent for both platforms. From the interview comments, we learned that obstacles in the *Obs* task could easily be detected from a natural heads-down position merely by "looking sideways from the phone screen". In such a scenario, participants had mixed feelings about the benefits of the OHMD, with 7 of 16 participants finding the phone just as comfortable as the OHMD.

The advantages offered by the OHMD, however, were prominent in the more complex *Nav* task, with 11 of 16 participants finding it easier and faster (5.7% higher PPWS) to walk using the OHMD as opposed to the phone. This is because with heads-up viewing on OHMD, participants could easily switch their attention to the signboards using peripheral vision. However, on the phone, they had to toggle their view from heads-down to heads-up, resulting in discomfort and delay.

**RQ2:** How does the LSVP video design help facilitate on-the-go video learning in the different mobility tasks?

Validating our hypothesis, the LSVP style was found to be more suitable for on-the-go video learning in general, improving users' learning and walking ability. In particular, there was a significant increase in participants' immediate and delayed recall scores, further confirming our findings from Study 2 that the complete retention of on-screen information can help reduce the temporal demand on users' working memory, which in turn allows for better encoding of learning content.

Yet, these benefits were more apparent only in the more complex *Obs* and *Nav* tasks when the LSVP style was coupled with the OHMD. This is because when using the OHMD, users had constant access to the video content on the display as they walked, a feature that complements and allows users to utilize the information retained on-screen by the LSVP style effectively. On the other hand, using the LSVP style on the phone did not produce a similar effect as users constantly needed to lift their heads up from the phone during complex mobility tasks, and so were unable to make effective use of the retained on-screen information. Consequently, we see no significant difference in using either the LSVP or Khan style on phones.

## 9 OVERALL DISCUSSION

Our interviews and surveys revealed a gap that exists between users' needs and their current on-the-go video learning experience using phones. In particular, due to the heads-down nature of interaction, users found it difficult to balance their attention between the video and path-navigation. Despite these disadvantages, watching videos on the go is still frequently practiced by a handful of users who use compensatory strategies such as perceptual sampling to help them efficiently manage their attentional resources between the mobility and HCI tasks. To resolve this gap and facilitate users' on-the-go video learning experience, in Study 2, we experimented with how combining the OHMD platform with a redesigned video presentation style can help balance video learning and navigational needs. Our results showed that the Layer presentation style, which combines sequential presentation and complete data persistence effectively improves users' immediate recognition scores by 26% and recall by 56%. Based on the study results, we propose LSVP, a modified Layer presentation style, as the most suitable presentation style for OHMD-based video learning.

The above findings were further validated in Study 3, where we investigated how our proposed solution (combining OHMD with LSVP video design) compared with the phone-based approach in accommodating the needs of video learning in mobility tasks with varying difficulties. With LSVP, we found it possible to achieve comparable or even better video learning using OHMD, as our participants scored better in both immediate and 7-day delayed recognition/recall tests. In addition, OHMD allowed users to walk faster on paths with greater navigational difficulty. The heads-up viewing and see-through properties of OHMD were crucial to this improvement of navigational performance, as evidenced by participants' open-ended comments: "I could easily look out to see the signboards without having to look up and down" (P3). "I felt safer learning with the OHMD, knowing that I can also see what is outside at all times" (P8).

These results indicate the need to choose a mobile platform that allows users to achieve meaningful learning while providing them with a feeling of safety while walking. In simple on-the-go scenarios where users can pay more attention to the learning content, phones may be a better option as they provide users with an exclusive focus on the video content. On the other hand, in on-the-go multitasking situations where the mobility task is less intensive, the properties of OHMD helped them navigate the path better, making it potentially safer for on-the-go multitasking. However, in highly complex navigation scenarios (i.e., driving fast on a curvy road), a cognitive bottleneck may occur. In such cases, OHMD may not be able to help either. Hence, we recommend using OHMD for on-the-go video learning while performing less intensive complex mobility tasks, and phones for simpler on-the-go scenarios.

While this work takes a first step towards establishing OHMD with LSVP as a promising approach for on-the-go video learning, one aspect of LSVP needs further research. Several participants in Study 3 found the digital handwriting feature in LSVP to be unsuitable for reading text in videos while walking. "Seeing some words on the OHMD was quite tiring because the handwriting was not very legible" (P6). Surprisingly, this issue did not arise in Study 2, even though the same style of digital handwriting was used. One possible explanation for this is that Study 2 users did not encounter the more complex mobility tasks tested in Study 3. However, when faced with more complex navigational tasks, their cognitive bandwidth reduced and magnified the legibility issue. Thus, we conclude that legibility and clarity needs further consideration for the LSVP style of video presentation.

Looking forward, we envision a future where on-the-go video learning can be more seamlessly integrated into our daily lives, with on-the-go video designs such as LSVP that support OHMD as an output platform. Below we delineate the LSVP design based on the cumulative findings from our studies.

## 9.1 Layered Serial Visual Presentation (LSVP)

In the pilots and multiple controlled studies conducted in this paper, we encountered several desirable properties that educational videos must possess in order to be better suited for on-the-go video learning. Here, we consolidate these properties into two sets, which taken together form the LSVP style .

The first set of properties primarily focuses on how the video content must be presented to improve users' recall/recognition abilities. In particular, the information should appear sequentially so that users' visual attention is guided effectively, and it must be strictly preserved on screen after it appears to reduce the temporal load on the working memory. In preserving the content, however, the information density needs to be controlled (by limiting the number of elements introduced in the video) so that all information can be retained on the screen without causing visual clutter.

The second set of properties concern the characteristics of the video content that complement the see-through properties of the OHMD, facilitating path navigation. For this, the background of the video needs to be made transparent. Such transparency, however, can affect the color contrast of the video content with the environment. Consequently, the contrast of the video content needs to be adjusted and enhanced to ensure relative clarity over the environmental background. To further enhance clarity, it is also advisable to use legible fonts that make reading text easier.

To convert an existing learning video designed for viewing on the desktop or phone platform into the above LSVP style suitable for on-the-go viewing on OHMD, we recommend the following steps for designers:

*Step 1:* Use short learning videos. In the case of longer videos, consider logically chunking the video to durations of 6-8 minutes or less.

*Step 2:* Add each piece of information (text/diagram) sequentially on the screen and allow it to persist on the screen for the entirety of the video.

*Step 3:* Modify the background of the video to make it transparent and discard any fill colors in diagrams.

*Step 4:* Recolor text/diagrams by increasing the contrast and shadow properties using color correction tools.

*Step 5:* Use more legible text and symbols, and avoid digital handwriting that can be difficult to read at a glance.

## 10 LIMITATIONS & FUTURE WORK

While the heads-up and see-through nature of an OHMD can be beneficial for path navigation and learning, heads-up displays, in general, have been reported to reduce the user's situational awareness in several scenarios due to attentional tunneling [15, 17, 41, 57]. This issue is also relevant to OHMDs, which require users to switch their attention between different depth planes in order to focus on the content and environment [57]. Taking this into consideration, we hope to use measures such as event detection to explicitly quantify the detraction of situational awareness in different mobility scenarios in the future.

Our current analysis assessed the efficacy of on-the-go video learning on acquiring fundamental recognition and recall skills suggested in Bloom's taxonomy [36]. In the future, however, we are interested in understanding how well our findings can be translated when higher levels of learning objectives such as "understand" or "apply" are being considered. Moreover, it would also be exciting to know whether our results are extendable to other subject domains such as mathematics and physics.

Another limitation of our study is its restriction to indoor environments. This design choice was intended to avoid visibility issues on the OHMD that occur due to natural lighting [21]. Although this limitation was partially mitigated by designing the mobility tasks to reflect realistic mobility scenarios, we hope to overcome this limitation in the future by considering how our proposed approach can help users learn in real commute and transit situations. This can also shed light on how users prefer to pace their learning in these scenarios, which can help the design of intelligent context-aware on-the-go video learning systems.

Lastly, although our results are significant, the sample size of participants considered in our study is moderate. Furthermore, our studies only included participants between the age of 18 and 29. While this is a representative age group for the activity under consideration, it limits the generalizability of our findings. To further strengthen our claims, we hope that future research can replicate our results with more diverse user groups, thereby preserving the scientific integrity of our findings.

## 11 CONCLUSION

In this work, we developed an OHMD-based system that enables video learning during users' idle time, such as while walking during commute or transits. Understanding the limitations and need to improve users' on-the-go video learning experience based on a thorough qualitative study, we conducted empirical studies to identify presentation techniques that can be used to redesign and adapt existing videos to meet the needs of on-the-go learning on OHMDs. Moreover, we compared our proposed solution against the current phone platform in a variety of realistic mobility scenarios, and found that OHMD can be a promising tool for on-the-go video learning. Our results suggest that OHMD, with an adapted video design that sequentially presents and persists information after it appears, can help improve users' fundamental recognition and recall skills as well as their walking abilities, better than phones. Our findings contribute to literature on mobile learning and on-the-go multitasking, and we discuss how future videos can be designed to support on-the-go video learning on OHMD.

## ACKNOWLEDGMENTS

This project is funded by the NUS Advanced Robotics Centre. We thank Zhang Yue for her generous help with designing some of the figures in the paper. We also thank the reviewers for their insightful comments that helped to improve the paper.

## REFERENCES

[1] Urs-Vito Albrecht, Ute von Jan, Joachim Kuebler, Christoph Zoeller, Martin Lacher, Oliver J Muensterer, Max Ettinger, Michael Klintschar, and Lars Hagemeier. 2014. Google Glass for Documentation of Medical Findings: Evaluation in Forensic Medicine. *J Med Internet Res* 16, 2 (2014), e53. https://doi.org/10.2196/jmir.3225

[2] Markus Appel, Nina Krisch, Jan-Philipp Stein, and Silvana Weber. 2019. Smartphone zombies! Pedestrians' distracted walking as a function of their fear of missing out. *Journal of Environmental Psychology* 63 (2019), 130 – 133. https://doi.org/10.1016/j.jenvp.2019.04.003

[3] Ahmad Qadri Basri and Ruhaizin Sulaiman. 2013. Ergonomics Study of Public Hospital Signage. In *Current Trends in Ergonomics (Advanced Engineering Forum, Vol. 10)*. Trans Tech Publications Ltd, Kuala Lumpur, Malaysia, 263–271. https://doi.org/10.4028/www.scientific.net/AEF.10.263

[4] Sophia Berolo, Richard P. Wells, and Benjamin C. Amick. 2011. Musculoskeletal symptoms among mobile hand-held device users and their relationship to device use: A preliminary study in a Canadian university population. *Applied Ergonomics* 42, 2 (2011), 371 – 378. https://doi.org/10.1016/j.apergo.2010.08.010

[5] Mireille Betrancourt. 2005. *The Animation and Interactivity Principles in Multimedia Learning*. Cambridge University Press, Cambridge, 287–296. https://doi.org/10.1017/CBO9780511816819.019

[6] Matt Bower and Daniel Sturman. 2015. What are the educational affordances of wearable technologies? *Computers & Education* 88 (2015), 343 – 353. https://doi.org/10.1016/j.compedu.2015.07.013

[7] Cynthia J. Brame. 2016. Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE—Life Sciences Education* 15, 4 (2016), es6. https://doi.org/10.1187/cbe.16-03-0125

[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[9] M. Bétrancourt, P. Dillenbourg, and C. Montarnal. 2003. *Computer Technologies in Powerful Learning Environments: The Case of Using Animated and Interactive Graphics for Teaching Financial Concepts.* Elsevier, Amsterdam, 143–157. http://infoscience.epfl.ch/record/33770

[10] Soledad Castellano and Inmaculada Arnedillo-Sánchez. 2016. Sensorimotor Distractions When Learning with Mobile Phones On-the-Move., In 12th International Conference on Mobile Learning. *International Association for Development of the Information Society*, 89–93. https://eric.ed.gov/?id=ED571434

[11] Konstantinos Chorianopoulos. 2018. A Taxonomy of Asynchronous Instructional Video Styles. *The International Review of Research in Open and Distributed Learning* 19, 1 (Feb. 2018), 296–311. https://doi.org/10.19173/irrodl.v19i1.2920

[12] Joke Coens, Ellen Degryse, Marie-Paul Senecaut, Jorge Cottyn, and Geraldine Clarebout. 2011. Listening to an educational podcast while walking or jogging: Can students really multitask? *International Journal of Mobile and Blended Learning (IJMBL)* 3, 3 (2011), 23–33. https://doi.org/10.4018/jmbl.2011070102

[13] Linda M Collins, John J Dziak, and Runze Li. 2009. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. *Psychological methods* 14, 3 (2009), 202–224. https://doi.org/10.1037/a0015826

[14] Tilman Dingler, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. 2017. Language Learning On-the-Go: Opportune Moments and Design of Mobile Microlearning Sessions. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 28, 12 pages. https://doi.org/10.1145/3098279.3098565

[15] Benjamin J. Dixon, Michael J. Daly, Harley H.L. Chan, Allan Vescan, Ian J. Witterick, and Jonathan C. Irish. 2014. Inattentional Blindness Increased with Augmented Reality Surgical Navigation. *American Journal of Rhinology & Allergy* 28, 5 (2014), 433–437. https://doi.org/10.2500/ajra.2014.28.4067

[16] Peter E Doolittle and Gina J Mariano. 2008. Working memory capacity and mobile multimedia learning environments: Individual differences in learning while mobile. *Journal of Educational Multimedia and Hypermedia* 17, 4 (2008), 511–530. https://psycnet.apa.org/record/2008-16147-003

[17] Matt Dunleavy, Chris Dede, and Rebecca Mitchell. 2009. Affordances and Limitations of Immersive Participatory Augmented Reality Simulations for Teaching and Learning. *Journal of Science Education and Technology* 18, 1 (February 2009), 7–22. https://doi.org/10.

1007/s10956-008-9119-1
[18] Steven Feiner, Blair MacIntyre, Tobias Höllerer, and Anthony Webster. 1997. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Personal Technologies* 1, 4 (1997), 208–217. https://doi.org/10.1007/BF01682023
[19] J Ferrer-Torregrosa, J Torralba, MA Jimenez, S García, and JM Barcia. 2015. ARBOOK: Development and assessment of a tool based on augmented reality for anatomy. *Journal of Science Education and Technology* 24, 1 (2015), 119–124. https://doi.org/10.1007/s10956-014-9526-4
[20] Logan Fiorella, Andrew T Stull, Shelbi Kuhlmann, and Richard E Mayer. 2019. Instructor presence in video lectures: The role of dynamic drawings, eye contact, and instructor visibility. *Journal of Educational Psychology* 111, 7 (2019), 1162–1171. https://doi.org/10.1037/edu0000325
[21] J. L. Gabbard, J. E. Swan, and D. Hix. 2006. The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence* 15, 1 (2006), 16–32. https://doi.org/10.1162/pres.2006.15.1.16
[22] Debjyoti Ghosh, Pin Sym Foong, Shengdong Zhao, Can Liu, Nuwan Janaka, and Vinitha Erusu. 2020. EYEditor: Towards On-the-Go Heads-Up Text Editing Using Voice and Manual Input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376173
[23] Joanne Gikas and Michael M Grant. 2013. Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. *The Internet and Higher Education* 19 (2013), 18–26. https://doi.org/10.1016/j.iheduc.2013.06.002
[24] Mayank Goel, Leah Findlater, and Jacob Wobbrock. 2012. WalkType: Using Accelerometer Data to Accomodate Situational Impairments in Mobile Touch Screen Text Entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2687–2696. https://doi.org/10.1145/2207676.2208662
[25] E. Bruce Goldstein. 2007. *Cognitive Psychology: Connecting Mind, Research and Everyday Experience*. Wadsworth Cengage Learning, Australia. https://books.google.se/books?id=XkBoSZ0O1bwC
[26] Ewa Gustafsson, Sara Thomee, Anna Grimby-Ekman, and Mats Hagberg. 2017. Texting on mobile phones and musculoskeletal disorders in young adults: a five-year cohort study. *Applied ergonomics* 58 (2017), 208–214. https://doi.org/10.1016/j.apergo.2016.06.012
[27] Rebecca J. Hess, Jennifer S. Brach, Sara R. Piva, and Jessie M. VanSwearingen. 2010. Walking Skill Can Be Assessed in Older Adults: Validity of the Figure-of-8 Walk Test. *Physical Therapy* 90, 1 (01 2010), 89–99. https://doi.org/10.2522/ptj.20080121
[28] Anne P Hillstrom and Yu-Chin Chai. 2006. Factors that guide or disrupt attentive visual processing. *Computers in human behavior* 22, 4 (2006), 648–656. https://doi.org/10.1016/j.chb.2005.12.003
[29] Wen-Jung Hsin and John Cigas. 2013. Short Videos Improve Student Learning in Online Education. *J. Comput. Sci. Coll.* 28, 5 (2013), 253–259. https://dl.acm.org/doi/10.5555/2458569.2458622
[30] María Blanca Ibáñez, Ángela Di Serio, Diego Villarán, and Carlos Delgado Kloos. 2014. Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education* 71 (2014), 1–13. https://doi.org/10.1016/j.compedu.2013.09.004
[31] Eric Jamet, Monica Gavota, and Christophe Quaireau. 2008. Attention guiding in multimedia learning. *Learning and instruction* 18, 2 (2008), 135–145. https://doi.org/10.1016/j.learninstruc.2007.01.011
[32] Annette Jeneson and Larry R Squire. 2012. Working memory, long-term memory, and medial temporal lobe function. *Learning & memory* 19, 1 (2012), 15–25. https://doi.org/10.1101/lm.024018.111
[33] Hyun-Ju Jeung, Paul Chandler, and John Sweller. 1997. The Role of Visual Indicators in Dual Sensory Mode Instruction. *Educational Psychology* 17, 3 (1997), 329–345. https://doi.org/10.1080/0144341970170307
[34] Robin H Kay. 2012. Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior* 28, 3 (2012), 820–831. https://doi.org/10.1016/j.chb.2012.01.011
[35] Khan Academy 2020. *Khan Academy*. Khan Academy. Retrieved August 27, 2020 from http://www.khanacademy.org/
[36] David R. Krathwohl. 2002. A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice* 41, 4 (2002), 212–218. https://doi.org/10.1207/s15430421tip4104_2
[37] Jochen Kuhn, Paul Lukowicz, Michael Hirth, Andreas Poxrucker, Jens Weppner, and Junaid Younas. 2016. gPhysics—Using smart glasses for head-centered, context-aware learning in physics experiments. *IEEE Transactions on Learning Technologies* 9, 4 (2016), 304–317. https://doi.org/10.1109/TLT.2016.2554115
[38] Joel Lanir, Kellogg S. Booth, and Leah Findlater. 2008. Observing Presenters' Use of Visual Aids to Inform the Design of Classroom Presentation Software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 695–704. https://doi.org/10.1145/1357054.1357165
[39] Nilli Lavie, Aleksandra Hirst, Jan W De Fockert, and Essi Viding. 2004. Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General* 133, 3 (2004), 339–354. https://doi.org/10.1037/0096-3445.133.3.339
[40] V. I. Levenshtein. 1965. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Dokl. Akad. Nauk SSSR* 163 (1965), 845–848. Issue 4. http://mi.mathnet.ru/eng/dan31411
[41] Jonathan L. Levy, David C. Foyle, and Robert S. McCann. 1998. Performance Benefits with Scene-Linked Hud Symbology: An Attentional Phenomenon? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 42, 1 (1998), 11–15. https:

//doi.org/10.1177/154193129804200104

[42] Richard E. Mayer. 2002. Multimedia learning. Psychology of Learning and Motivation, Vol. 41. Academic Press, 85 – 139. https://doi.org/10.1016/S0079-7421(02)80005-6

[43] Martijn Meeter and Jaap MJ Murre. 2004. Consolidation of long-term memory: Evidence and alternatives. *Psychological Bulletin* 130, 6 (2004), 843–57. https://doi.org/10.1037/0033-2909.130.6.843

[44] Jennifer Momsen, Tammy Long, Sara Wyse, and Diane Ebert-May. 2010. Just the Facts? Introductory Undergraduate Biology Courses Focus on Low-Level Cognitive Skills. *CBE life sciences education* 9 (12 2010), 435–40. https://doi.org/10.1187/cbe.10-01-0001

[45] Hong Kok Ng, Slava Kalyuga, and John Sweller. 2013. Reducing transience during animation: a cognitive load perspective. *Educational Psychology* 33, 7 (2013), 755–772. https://doi.org/10.1080/01443410.2013.785050

[46] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. 2005. Interaction in 4-Second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Association for Computing Machinery, New York, NY, USA, 919–928. https://doi.org/10.1145/1054972.1055101

[47] Eric D. Ragan, Alex Endert, Doug A. Bowman, and Francis Quek. 2012. How Spatial Layout, Interactivity, and Persistent Visibility Affect Learning with Large Displays. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '12)*. Association for Computing Machinery, New York, NY, USA, 91–98. https://doi.org/10.1145/2254556.2254576

[48] Rufat Rzayev, Paweł W. Woundefinedniak, Tilman Dingler, and Niels Henze. 2018. Reading on Smart Glasses: The Effect of Text Position, Presentation Type and Walking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3173574.3173619

[49] HyeonHui Shin and KyeongMi Kim. 2014. Effects of cervical flexion on the flexion-relaxation ratio during smartphone use. *Journal of physical therapy science* 26, 12 (2014), 1899–1901. https://doi.org/10.1589/jpts.26.1899

[50] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12, 2 (1988), 257 – 285. https://doi.org/10.1016/0364-0213(88)90023-7

[51] Huib K Tabbers, Rob L Martens, and Jeroen JG Van Merriënboer. 2004. Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British journal of educational psychology* 74, 1 (2004), 71–81. https://doi.org/10.1348/000709904322848824

[52] Bernardo Tabuenca, Stefaan Ternier, and Marcus Specht. 2013. Supporting Lifelong Learners to Build Personal Learning Ecologies in Daily Physical Spaces. *Int. J. Mob. Learn. Organ.* 7, 3/4 (Oct. 2013), 177–196. https://doi.org/10.1504/IJMLO.2013.057160

[53] Clive Thompson. 2011. *How Khan Academy is changing the rules of education.* Retrieved January 23, 2021 from https://www.wired.com/2011/07/ff-khan/

[54] Endel Tulving and Zena Pearlstone. 1966. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior* 5, 4 (1966), 381 – 391. https://doi.org/10.1016/S0022-5371(66)80048-8

[55] Bret Victor. 2014. Humane Representation of Thought: A Trail Map for the 21st Century. In *Proceedings of the Companion Publication of the 2014 ACM SIGPLAN Conference on Systems, Programming, and Applications: Software for Humanity (SPLASH '14)*. Association for Computing Machinery, New York, NY, USA, 5. https://doi.org/10.1145/2660252.2661746

[56] Xiangyu Wang, Soh K Ong, and Andrew YC Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4, 1 (2016), 1–22. https://doi.org/10.1007/s40436-015-0131-4

[57] Christopher D. Wickens and Amy L. Alexander. 2009. Attentional Tunneling and Task Management in Synthetic Vision Displays. *The International Journal of Aviation Psychology* 19, 2 (2009), 182–199. https://doi.org/10.1080/10508410902766549 arXiv:https://doi.org/10.1080/10508410902766549

[58] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[59] Patricia Wright et al. 1990. Integrating Diagrams and Text. *Technical Writing Teacher* 17, 3 (1990), 244–54. https://eric.ed.gov/?id=EJ421261

[60] Yue Zhao, Tarmo Robal, Christoph Lofi, and Claudia Hauff. 2018. Stationary vs. Non-Stationary Mobile Learning in MOOCs. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization (UMAP '18)*. Association for Computing Machinery, New York, NY, USA, 299–303. https://doi.org/10.1145/3213586.3225241

[61] Xianjun Sam Zheng, Cedric Foucault, Patrik Matos da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. 2015. Eye-Wearable Technology for Machine Maintenance: Effects of Display Position and Hands-Free Operation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2125–2134. https://doi.org/10.1145/2702123.2702305