



Review article

A review of augmented reality systems and their effects on mental workload and task performance



Nor Farzana Syaza Jeffri *, Dayang Rohaya Awang Rambli

Computer Information Sciences Department, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia

ARTICLE INFO

Keywords:Augmented reality
Mental workload
Cognitive load
Task performance
Augmented reality systems

ABSTRACT

Augmented Reality (AR) systems have been shown to positively affect mental workload and task performance across a broad range of application contexts. Despite the interest in mental workload and the increasing number of studies evaluating AR use, an attempt has yet to be made to identify the relationship between the effects of AR on mental workload and task performance. This paper seeks to address this gap in AR technology literature. With a better understanding how AR affects mental workload and task performance, researchers and developers can design more effective AR systems. 34 articles investigating the effects of the use of AR systems were selected for the review. A positive correlation was found between effects on mental workload and effects on task performance: if the effect on mental workload is positive, then the effects on task performance are more likely to be positive as well, and vice versa. Effectiveness of AR systems were shown to be influenced by the type of AR display device used, relevance and timeliness of content, information presentation, user characteristics and task characteristics. Additionally, the paper addresses the use of the concept of mental workload and limitations in current literature.

1. Introduction

Augmented reality (AR) is an emergent technology that allows for the augmentation of interactive digital information upon the real-world. This ability to convey spatially and temporally meaningful information in real-time makes AR technology a promising “option to support knowledge-intensive works” [1]. AR applications have been adopted in various application domains such as education [2, 3], medicine [4], manual assembly and manufacturing [5, 6], online retail [7], maintenance [6, 8, 9], and in-vehicle map navigation [10, 11]. By using AR to provide visual guidance and task-relevant information where and when needed, cognitive workload could be reduced, users would be subjected to lower cognitive demands and thus improve task performance. Indeed, some studies reported positive effects of AR use on both mental workload and task performance [12, 13]. However, there are also studies that reported positive effects on mental workload yet negative effects on task performance [8, 14]. Whereas some studies reported insignificant differences in task performance and even increases in cognitive workload when using AR [1, 10].

Despite the growing number of research on AR technology use and its numerous reported effects on user's mental workload and task performance, little effort has been conducted to consolidate these studies in an attempt to understand the correlation, if any, between the effects of

AR use on mental workload and effects on task performance. Does a reduction in mental workload result in an increase in task performance? Are the effects on task performance resultant of the effect on mental workload? Furthermore, the concept of mental workload is a longstanding grey area. There is no single accepted definition of the concept and there exists many variations of what constitutes mental workload [15, 16, 17, 18]. Despite the heterogenous nature of the concept, it remains an important type of measurement for evaluating the usability of technologies across many application areas. How is the concept of mental workload adopted in AR literature? How is mental workload measured? This paper seeks to address this gap in AR literature concerning mental workload and task performance by answering the following questions:

1. How are mental workload and task performance measured in studies evaluating the use of an AR system?
2. What are the effects of AR use on mental workload and task performance?
3. Is there a relationship between the effects on mental workload and the effects on task performance?
4. How does AR affect mental workload and task performance?

* Corresponding author.

E-mail address: nor_17008594@utp.edu.my (N.F.S. Jeffri).<https://doi.org/10.1016/j.heliyon.2021.e06277>

Received 23 October 2019; Received in revised form 28 August 2020; Accepted 9 February 2021

By addressing these questions through a review of past relevant studies, the paper contributes to the current knowledge by understanding how AR systems influence human cognition. In doing so, researchers and developers can design more effective AR systems. This paper is organized as follows. We begin with defining the concepts of Augmented Reality and Mental Workload in Section 2. The literature review methodology is described in Section 3, followed by the results and discussions in Section 4. The paper is concluded in Section 5.

2. Background

2.1. Augmented reality

Augmented Reality (AR) allows for the augmentation of virtual information upon the real-world in real-time [19]. It is one of the ways in which virtual and real environments can be combined, based on the Mixed Reality Continuum introduced by Milgram and Kishino [20]. In contrast to Virtual Reality (VR), which immerses the user in a completely new virtual environment, AR aims to supplement the user's current reality by augmenting virtual content on top of the physical environment. Azuma [19] has outlined the three main features that define AR technologies:

- Registered in 3D
- Interactive
- Displayed in real-time

The potential of AR as an assistive technology is largely attributed to its ability to present relevant information where and when needed [1]. AR technologies have mostly been applied to support visual perception by presenting information in the forms of spatially-registered textual annotations, sound and 3D objects and animations. However, researchers have begun to expand this range by augmenting olfactory and gustatory senses as well [21]. AR is being increasingly adopted across a wide range of application domains due to its numerous reported benefits on mental workload and task performance [2, 3, 6, 22, 23, 24, 25].

The reduction of mental workload is seen as an important endeavour in the modern workspace as increasing amounts of information, high task complexities and the increasing need to deliver services quickly and accurately burdens human operators with excessive mental load [13, 26, 27, 28]. A worker under high mental and physical duress can negatively impact work performance, which may result in expensive losses for the company or business. The reduction of mental workload can be seen as an important factor in assessing the effectiveness of AR interventions as compared to traditional means, as it serves as a solution to improve effectiveness and efficiency rather than increase mental burden to the user.

Throughout literature, AR has generally been shown to reduce mental workload and improve various metrics of task performance such as task completion time, error rate and task accuracy. However, some studies reported that the use of AR can introduce new challenges to the user which may otherwise compromise its reported benefits. Ergonomic and usability issues such as discomfort over long-term usage, reduced field of view (FOV), tracking and lagging issues, have all been reported to contribute to negative effects on mental workload and task performance [1, 14, 29, 30]. This paper seeks to get the big picture of how the use of AR systems affect mental workload and task performance. In doing so, we aim to understand the trend in the effectiveness of AR systems as an assistive tool.

2.2. Mental workload and cognitive workload

2.2.1. Mental workload

Mental workload is a complex, multidimensional and versatile construct with no single accepted definition [15, 31]. However, the general

consensus surrounding the concept is that it is very much “a function of the supply and demand of limited human cognitive resources when performing a task” - the cognitive “cost of accomplishing task requirements” [17, 18]. The notion of “what is mental workload” has been broached upon by many researchers over a span of two decades; however there has yet to exist a single encompassing definition of the concept due to its multidimensional and complex nature.

Mental workload has been described as:

- “...the demand imposed by tasks on the human's limited mental resources, whether considered as single or multiple.” Moray [32] as cited in Wickens [33].
- “... an intervening variable, similar to attention, that modulates or indexes the tuning between the demands of the environment and the capabilities of the organism.” Kantowitz [31]
- “....an interaction between the demands of the task and the performance of the operator.” Cain [34] as cited in Byrne [16]
- “...a function of the supply and demand of attentional or processing resources” Tsang and Vidulich [18] as cited in Hertzum and Holmegaard [15]
- “... the interaction between the structure of systems and tasks on the one hand, and the capabilities, motivation, and state of the human operator on the other.” Kramer [35] as cited in Hertzum and Holmegaard [15]
- “...the relation between the function relating the mental resources demanded by a task and those resources available to be supplied by the human operator.” Parasuraman et al. [36]

The concept of mental workload has been a subject of interest since the 1980s [17, 31, 32]. As put forth by Kantowitz [31], a primary motivation to understand and measure mental workload is to be able to predict “how much mental workload is associated with some particular job demand” so employers and designers can know how to “create safe and reasonable levels of mental workload on the job”. As task difficulty (workload) increases: performance usually decreases; response times and errors increase [31]. It is important to maintain safe levels of mental workload in the working environment as the “human cost (e.g., fatigue, stress, illness and accidents) of maintaining performance is unacceptably high” [15, 37].

Workload measurement can also be used to choose between system designs when performance measures are less distinct, allowing designers to optimize system design and performance [37], [38] as cited in [34]. In this regard, workload is considered an important factor in assessing a system's usability and are “now commonly used in industries to identify sources of error and to improve performance” [16].

However, exactly what factors constitute the construct that is mental workload remains an area that requires extensive research. As Cain [34] puts it, “the science is not completely static, particularly in the psychophysiological domain.” “The many definitions that exist in the psychological literature are a testament to the complexity of the construct as are the growing number of causes, consequences and symptoms that have been identified” [37]. Mental workload is made up of combination of interrelating facets (or dimensions) such as feelings, experience, effort, task workload, working environment, individual differences, and other human factors [17, 31].

The concept of mental workload is primarily built upon the assumption of limited human cognitive capabilities, whereby workload is the “cost of accomplishing mission requirements for the human operator” [37]. The concept of mental workload has been attempted to be explained through various theories, models, and frameworks of human cognition, such as the model of attention. Modern measurements of mental workload were developed primarily based on these models and theories. Due to this, different types of techniques were introduced to measure mental workload.

Kantowitz [31] built his theory of mental workload upon the model of attention and the construct of spare capacity. He proposed that hu-

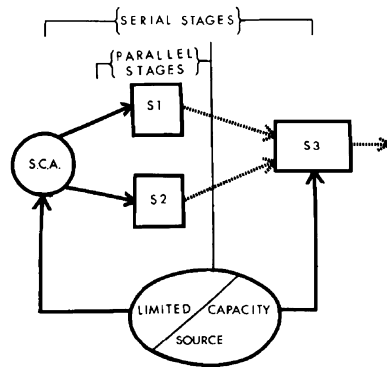


Fig. 1. The Kantowitz-Knight hybrid model of attention and time sharing, whereby a limited capacity source feeds both stage S3 and a static capacity allocator (SCA) which partitions capacity between stages S1 and S2 (from Kantowitz & Knight [40]. Reprinted from *Acta Psychologica*, 40, Barry H. Kantowitz, James L. Knight, *Test tapping timesharing, II: Auditory secondary task*, Pages No. 359, Copyright (1976), with permission from Elsevier.

mans have a single pool of capacity, arguing that it is more suitable in achieving the “practical goal of measuring mental workload” than a multiple-pools model (Fig. 1). He stated that “the construct of spare capacity, derived from models of attention, is the most important component of mental workload” [31].

He proposed the use of secondary tasks to “obtain direct estimates of spare capacity, and hence mental workload”. The secondary-task paradigm requires for “an additional task to be performed in parallel with the primary task of interest [39] as in Kantowitz [31]. “Decrements in secondary-task performance are interpreted as indicating increased mental workload associated with increases in primary task demands” [31].

However, Kantowitz [31] acknowledged that due to the multi-dimensional nature of mental workload, measuring “spare capacity” alone is insufficient. It is recommended to use multiple measures to tap into the different dimensions of mental workload, such as psychophysiological measures and subjective ratings. This sentiment is echoed by modern practitioners, maintaining the advice that the simultaneous application of multiple measures provides “better diagnosticity” [11, 15, 18].

The multiple resource theory is another approach to understand the cause and effect of task demands on limited mental resources, by also implementing a secondary-task (or dual-task) paradigm. Assuming that human’s limited resources can be shared across multiple tasks, the amount of effort expended on a primary task can be predicted by observing the decrements or increments in performance on a secondary-task (with both tasks being performed in parallel) [11, 33].

Wickens [33] distinguished between the concepts of multiple resource theory and mental workload, expressing that mental workload relates more strongly to the component of demand in the multiple resource model architecture, whereby demand is the “demand imposed by tasks on the human’s limited mental resources”. Wickens suggested that the multiple resource model only becomes relevant to mental workload by predicting how much performance will fail when overload is imposed by multiple tasks.

The NASA Task Load Index (NASA-TLX) is a multi-dimensional rating scale that estimates overall subjective or perceived (mental) workload based on the subjective evaluation of six workload-related factors: mental demand, physical demand, temporal demand, performance, effort and frustration level [17, 37, 41]. Hart and Staveland [17] developed the NASA-TLX based on their conceptual framework of variables that influence human performance and workload, assuming that “workload is a hypothetical construct that represents the cost incurred by a human operator to achieve a particular level of performance”. They proposed that mental workload “emerges from the interaction between the

requirements of a task, the circumstances under which it is performed, and the skills, behaviours, and perceptions of the operator” [17].

Apart from the NASA-TLX and the dual-task paradigm, other tools for measuring mental workload include the SWAT workload, “Cooper Harper” workload, “Subjective Workload Dominance” and “Bedford workload” [41]. Grier [41] reported that the NASA-TLX is the “most cited survey-based workload measure”, and has been proven to be more pragmatic, less sensitive to individual differences and more sensitive to workload differences as compared to other measures. The reliability and validity of the NASA-TLX as a measurement of mental workload has been proven over its vast and extensive use in various research areas and application domains over the past two decades [37, 41].

2.2.2. Cognitive workload

Cognitive load (or workload) is the amount of mental effort being utilized in working memory at any given time; it is commonly associated with the Cognitive Load Theory (CLT), which is “concerned with the manner in which cognitive resources are focused and used during learning and problem solving”. Depending on the task’s level of difficulty, the profile of the learner and the instructional design, any task completion or learning process is prone to excessive cognitive load. It proposes that there are three types of cognitive load: intrinsic, extraneous and germane [42]. Intrinsic load is imposed by the task or problem at hand (the intrinsic difficulty of the task); extraneous cognitive load is imposed by how the information is presented (i.e. instructional format) or the effort required to process instruction; and germane cognitive load is the effort to construct schemas or generate knowledge [42, 43].

The CLT emphasizes that humans have a limited working memory capacity: “a high mental load would require the allocation of mental resources to elaborate information” rather than to solve the problem at hand [43]. The addition of intrinsic and extraneous cognitive loads would result in fewer cognitive resources left in working memory to generate schemas or for knowledge construction. Thus, in the design of learning and training environments, it is important to design technologies that reduce cognitive load associated with non-friendly interfaces to free up more cognitive resources devoted to schema construction [44].

In summary, as is the case with the multiple resource theory and mental workload [33], the concepts of mental workload and cognitive workload “overlap but are distinct”. The concept of mental workload is built upon the theory of limited cognitive capabilities whereas cognitive workload is built upon the theory of limited working memory capacity.

The primary motivation behind the concept of mental workload and measuring it is to “quantify the mental cost of performing tasks in order to predict operator and system performance” [34]. The final goal is to set “standards for the appropriate levels of mental workload in industry”, so that employers and designers can know how to design and create a working environment in which the mental workload imposed does not exceed their capacity and lead to unacceptable levels of performance [15, 18, 31, 34].

The concept of cognitive workload is built upon the theory of limited working memory capacity and is more focused towards the design of instructional formats. It suggests that intrinsic and extraneous cognitive loads are additive, and that inappropriately designed learning formats or instructions can impose a high mental load, reducing cognitive resources dedicated to learning or performing the task [43]. The end goal is to design training and learning environments or instructions that does not impose unnecessary amounts of extraneous cognitive load and free up more cognitive resources for schema construction [44].

This paper is interested to understand the effects of the use of AR on mental workload, or more specifically, the human operator’s limited mental resources. Henceforth, this paper adopts Tsang and Vidulich’s [18] general, simplified definition of the concept: “a function of the supply and demand of attentional or processing resources” when meeting task demands. This umbrella definition can encompass the various forms of mental resources: attentional and processing resources, inclu-

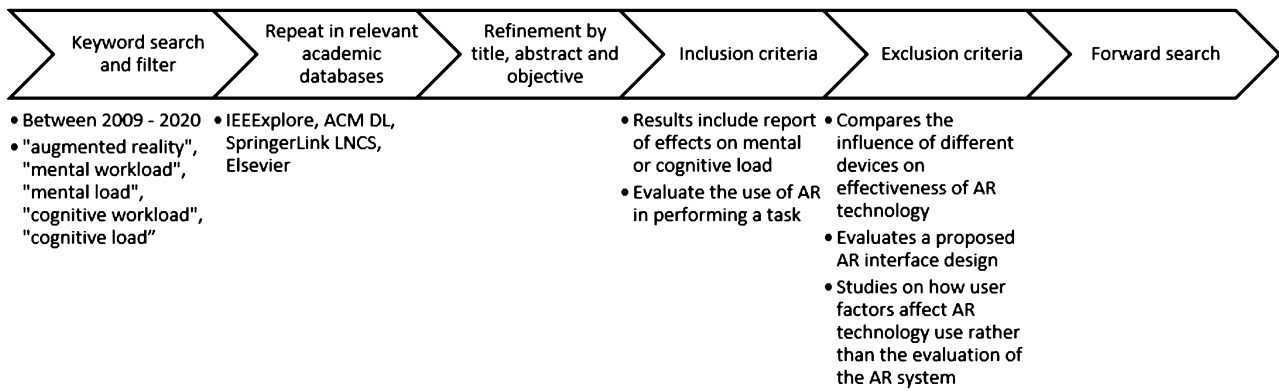


Fig. 2. Summary of literature review process.

Table 1. Search strings.

#	Search String
1	"augmented reality" AND "mental workload"
2	"augmented reality" AND "mental load"
3	"augmented reality" AND "cognitive load"
4	"augmented reality" AND "cognitive workload"

sive of working memory, which is the facet of mental workload that the CLT focuses on. Thus, "mental workload" in this paper refers to both the concepts of mental workload and cognitive load.

3. Methodology

The review begins by searching for relevant papers that investigate the effects of the use of AR technology. There are few to no studies that specifically and exclusively investigate the effects of the use of AR technology on mental workload. Rather, the effects of AR on mental workload are more often reported as a part of the results of a study that evaluates the use of an AR application in a specified use case.

The search strings comprise of combining "augmented reality" with the following phrases: "mental workload", "mental load", "cognitive workload", and "cognitive load". The rationale for using these variants of the term "mental workload" is because these phrases are used interchangeably in literature. Table 1 lists the combination of phrases and resulting search terms used. The phrases are enclosed in quotation marks to ensure that the algorithms return results with the specified phrase.

The search is repeated in the following academic databases: IEEEExplore, ACM Digital Library, SpringerLink's Lecture Notes in Computer Science, and Elsevier. The IEEEExplore, ACM DL and SpringerLink's (LNCS) were selected as they have special focus on computer science [24]. The Elsevier database was included in the search as it has journals focused on computer science research as well to help broaden the search results. To ensure the timeliness and relevance of the results, papers published between 2009 - 2020 were selected to ensure relevance and range.

The initial search results are then refined by the title, abstract and the objective of the study. As a general inclusion criteria, research articles evaluating the use of an AR application in performing a task (in any domain) are included, provided that the paper also reports an analysis of user's mental workload when using the AR application. Papers detailing frameworks, descriptions of architectures, methods, effects of human factors on AR technology use, requirements and guidelines, proposed applications or systems (which are unevaluated) are excluded. Finally, a forward search of the selected articles was conducted. Fig. 2 summarizes the literature review process.

Papers whose abstract mention an analysis of mental workload as part of the results are included for further review. Comparison studies between different AR displays or technologies on a specified use case are also included, as long as the paper also reports an analysis on user's

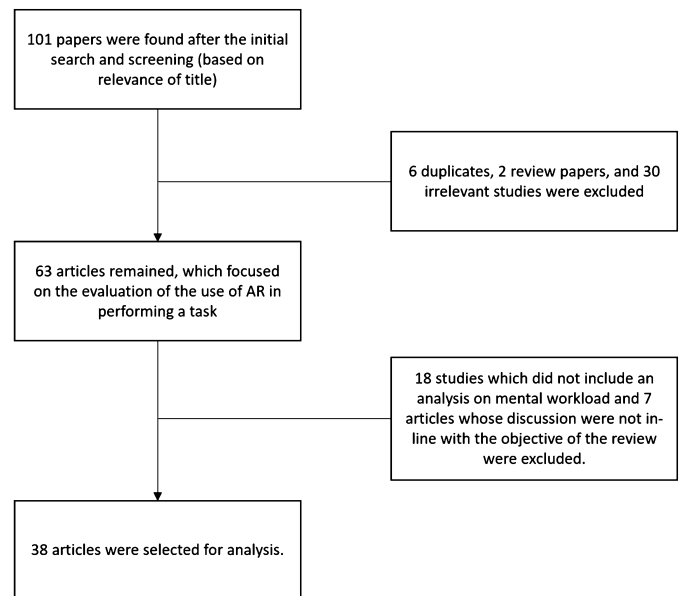


Fig. 3. Literature selection flowchart.

mental workload. The rationale for this is that AR factors such as displays and information presentation could be considered as influencers of how well an AR system affects mental workload. If a paper whose objective is in line with the objective of the review but does not mention any reports on mental workload in the title or abstract, it is put on hold for further analysis in the second phase of review, as some papers do include measurement and analysis of mental workload but is not specifically mentioned in the abstract.

Fig. 3 shows the literature selection process in a graphical format.

A total of 38 research articles are selected for review. Table 2 summarizes the number of articles from different sources and categorized by application domain. To answer the research questions, the following data are gathered: type of comparison study conducted; subjective and objective measurements used; the effects of AR on mental workload; the effects of AR on measures of task performance.

4. Results and discussion

The 38 articles selected conducted different variations of comparison studies and were divided into four categories: "Comparison between conventional/traditional methods and AR intervention", "Comparison between AR system elements (interfaces, displays, content)", "Comparison between AR system elements and conventional/traditional methods", and "Comparison between conventional method, AR system in-

Table 2. Number of articles based on article source and categorized by application domain.

Domain	Source			Num of articles
	Journal	Conference	Symposium	
Education	1	0	1	2
Manual Assembly/Manufacturing	6	10	0	16
Automotive Interface/ In-Vehicle Map Navigation	2	2	0	4
Medicine Surgery	3	0	0	3
Robot Programming	0	0	1	1
Maintenance	1	2	0	3
Online Retail	1	0	0	1
Not specified	1	0	1	2
Rehabilitation/Handicapped Aids	1	1	0	2
Construction Layout	2	0	0	2
Logistics	1	0	0	1
Cybersecurity		1	0	1
Total	18	13	3	38

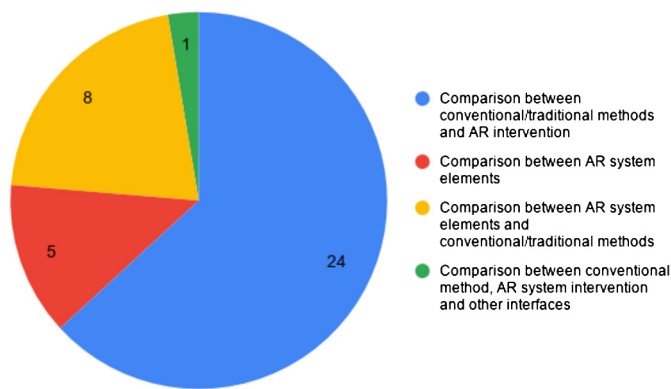


Fig. 4. Number of articles grouped based on types of comparison studies conducted in selected articles.

intervention and other interfaces”. Fig. 4 shows the frequency of articles based on type of comparison study.

During data collection, it was found that some articles conducted more than one study or provided more than one set of results based on user characteristics or demographics. Whereas some studies provided “overall” results, other studies only provided separate sets of results. In the case of an article conducting more than one studies, the results of the other study/studies were included if relevant. In the case of multiple results, the “overall” results are selected if provided. However if not presented, then all results are included for review. This is because the different sets of results may differ in statistical significance and we have no method nor justification of selecting a particular set of results over the other. Based on this, it should be clarified that though 38 articles were selected for the review, 40 sets of results were analyzed.

Furthermore, the results of studies were explained and discussed differently between articles. Whereas some articles clearly show when positive or negative results is statistically significant or insignificant, others are more vague with a simple description of effects (e.g. “faster task completion time”, “reduced error rates”) or a declaration of “no statistical difference”. Statistically significant or insignificant results, where specified, are classified accordingly under “Significant positive effect”, “(Insignificant) positive effect”, “(Insignificant) negative effect”, and “Significant” negative effect. When results are mentioned to be “positive” or “negative” without mentioning of statistical significance/insignificance, it is assumed to be statistically insignificant e.g. “AR improved task completion time”, it is classified under “(Insignificant) positive effect” or “(Insignificant) negative effect”. When the results are deemed too vague (even after referring to provided data) or declared to have no statistical difference, it is classified under “No differences found or mentioned”.

The reason for classifying the effects based on statistical significance and to differentiate them is for identifying the relationship between

effects on mental workload and effects on task performance. Studies which are in the “Comparison between AR system elements and conventional/traditional methods” category typically provide different sets of results. In this case, the most significant result is selected.

Results of studies which compare AR systems against other interfaces and studies which compare between AR system elements or factors (such as FOV, display devices) are shown and discussed differently. This is due to the diversity of types of displays, information presentation and other variables used. Table 3 shows a summary of the different types of comparisons that were studied. Section 4.4.5 summarizes the results of these studies.

4.1. How are mental workload and task performance measured?

Due to the diversity of application domains and tasks, there are many different measures of task performance. The most frequently used objective task measurements are task completion time (TCT) and error rate(ER). This is also because 14/34 studies are in the “Manual assembly/Manufacturing” domain, and the most commonly used measurements for manual assembly tasks are TCT and ER. Some articles in the “Manual assembly/Manufacturing” domain analyse the TCT and ER of manual assembly subtasks rather than the manual assembly procedure as a whole. Whilst this may give a deeper insight into how effectively AR systems assist operators in the different stages of manual assembly, it made it challenging to classify the results. Thus it is suggested for future studies to provide a general overview of results alongside with the more specific results for ease of analysis, classification, and comparison with other studies.

Other domains with a lower article frequency, such as Education, Logistics, Online Retail, Medicine & Surgery either have domain-specific measurements such as “needle navigation precision” [4], “driving performance” [10, 11, 45], “learning gain” [54], occurrences of “puncturing the phantom dura” [4]; or measurements that are not present or used in other studies (i.e. occurred only once) such as decision time or success rate. This wide range of what constitutes task performance in each respective domain is grouped together under “Other measures of task performance” for ease of analysis. Fig. 5 shows the occurrences of objective measures across studies.

Similarly, various subjective measures were employed to measure the effect of AR on mental workload, the most commonly used measure being the NASA-TLX [41], followed by the NASA-RTLX [37], which is another version of the former. Some studies employed domain-specific versions of the NASA-TLX, such as the Surgical Task Load Index (SURG-TLX) [27] and Driving Activity Load Index (DALI) [10], whereas some studies adapted the NASA-TLX according to their user group as is done by Funk et al. [48]. The NASA-TLX is typically employed as a subjective measure alongside objective measurements of task performance. Kim and Dey [11] proposed supplementing NASA-TLX, a subjective post-hoc mental workload measurement, with objective real-time measurements of mental workload as measured by psychophysiological

Table 3. Summary of comparisons made in studies comparing different AR system elements.

Reference	Type of AR system element compared	Comparison made
Bolton et al. [45]	Information presentation	Different types visualizations to aid in in-vehicle road navigation: arrows vs. dynamic arrows vs. boxes
Fan et al. [7]		Different combinations of high and low levels of simulated physical control (SPC) and environmental embedding (EE)
Fischer et al. [27]		X-ray augmentation on 2d video vs. 3D surface reconstruction augmented by digital reconstructed radiographs and live tool visualization
Kishishita et al. [46]		In-view labelling vs. In-situ labelling
Lampen et al. [47]		3D in-situ projections and instructions vs. Human simulation data
Kim et al. [30]		Information mode: text-based UI vs. graphic-based UI Information availability: always on vs. on demand
Funk et al. [48]		Different types of projected AR instructions: pictorial vs. video vs. contour-based
Wang et al. [49]		Paper manual vs. augmented video instructions on Google Glass vs. augmented annotations using arrows, symbols and icons on Google Glass
Kim et al. [50]		Different configurations of: - icon position, shape and size and - text position for an in-vehicle AR Heads up Display
Baumeister et al. [29]	Display device used	Gear VR vs. Hololens vs. Projector
Alves et al. [51]		Handheld mobile AR vs. Mobile AR using a tripod vs. Spatial AR using Kinect and Projector
Kim et al. [30]		Head Worn Display (HMD) type: Binocular (Epson Moverio BT-200) vs. Monocular (Vuzix M100)
Funk et al. [52]		Head Mounted Display (HMD) vs. Tablet vs. In-situ (using a projector)
Kishishita et al. [46]	Field of view (FOV)	Different sets of horizontal and vertical FOV angle parameters
Park et al. [53]	Tracking method	AR marker-based method vs. Proposed method using deep-learning based object detection

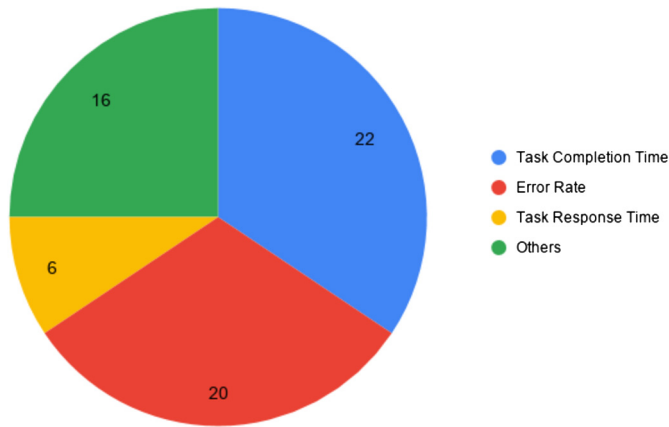


Fig. 5. Frequency of objective measures employed across studies.

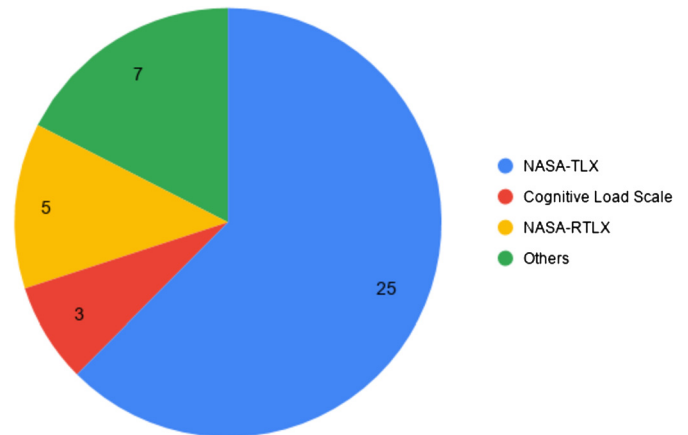


Fig. 6. Frequency of subjective measures employed across articles.

measures such as electroencephalogram (EEG) readings, electrocardiogram (ECG) readings, galvanic skin response (GSR) and eye-tracking measures. This is so that subjective measures of mental workload can be correlated with objective real-time measures to understand what is really happening in real-time.

Measures of cognitive load consist of the Paas mental effort scale [29, 55] and custom questionnaires [7, 56]. Several studies employed a dual-task methodology to measure cognitive load, whereby a user performs a secondary task in tandem with the primary task [29, 46]. “According to the dual-task theory, performance on the secondary task will deteriorate in response to increasing extraneous load inherent in the primary task” [29]. Subjective measures which do not fall into the categories of “NASA-TLX”, “NASA-RTLX”, or “Cognitive Load Scale” are classified into “Others”. Fig. 6 shows the occurrences of subjective measures across studies. For ease of analysis and discussion, effects on mental workload and cognitive load are both referred to as “effect on mental workload”.

4.2. What are the effects of AR use on mental workload and task performance?

A majority of studies reported positive effects (both significant and insignificant) of AR use on mental workload and task performance when

compared against conventional methods (Fig. 7). In most instances of AR system for manual assembly studies, AR is shown to decrease task completion and time and decrease error rate. In in-vehicle map navigation situations, AR is shown to reduce cognitive load in both young and elder drivers while improving driving and navigation performance. The works of Ameri et al. [4] and Fischer et al. [27] show that AR can help surgeons and clinicians perform safer, more accurate and efficient procedures by improving visualization, navigation, positioning and alignment of equipment as well as reduce mental workload. Fan et al. [7] found that a high level of environmental embedding and simulated physical control resulted in lower cognitive loads and improved cognitive fluency, which then improved users attitude towards online products.

In the instances where AR use is reported to increase mental workload and negatively impact task performance, it is usually attributed to the AR device’s usability issues. Deshpande and Kim [1] found that AR use increased task completion time for simple tasks and increased mental workload and task completion time in more complex tasks. They reported that users encountered usability issues regarding the AR device’s interaction, the spatial alignment of the overlaid graphics, and the effectiveness and efficiency of voice commands.

Deshpande and Kim [1] added that the difficulty of seeing objects in space through the small FOV may have increased frustration thus NASA-

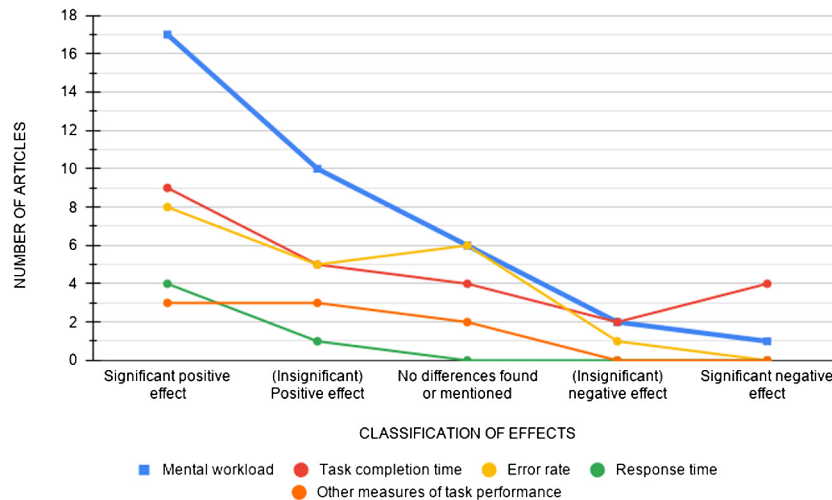


Fig. 7. Frequency of effects of AR use on mental workload and measures of task performance as classified by significance.

TLX ratings. Stadler et al. [14] attributed the longer task completion time with AR to a combination of lag and no spoken time limit. Alves et al. [51] found that some users encountered perceptual issues when viewing AR information due to the augmentation's transparency. The work of Baumeister et al. [29] shows that one or multiple restrictions of HMDs can manifest as both a detrimental effect on performance and additional extraneous cognitive load.

4.3. Is there a relationship between effects on mental workload and task performance?

The use of AR has been shown to positively impact mental workload and task performance across application domains. However, it is unclear whether an effect on mental workload would cause the effect on task performance. Fig. 7 only shows the number of occurrences of positive, negative or absence of effects of AR use on mental workload and task performance. It is insufficient to derive from this the relationship between mental workload and task performance, if any. We are interested to find whether a positive effect on mental workload would result in positive effects on task performance. To achieve this, we attempted to map the effects on task performance against effects on mental workload. To do this, significant positive effect, (insignificant) positive effect, no difference found or mentioned, (insignificant) negative effect and significant negative effect were each assigned a numerical value of 2, 1, 0, -1, -2 respectively. With mental workload being the independent variable (x-axis), and task completion time, error rate, response time and other measures of task performance being the dependent variables (y-axis).

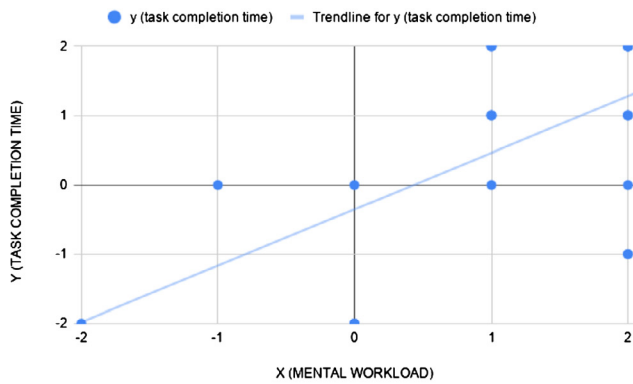
Fig. 8 shows the relationship between effects on mental workload and task completion time, error rate, response time and other measures of task performance in individual scatter graphs. Fig. 9 presents a wholistic view of the relationship between effect on mental workload and task completion time, error rate, response time and other measures of task performance. The trendlines for task completion time, error rate, response time and other measures of task performance show a positive correlation. Most of the points are scattered in the upper right side of the graphs, indicating that most studies which reported positive effects on mental workload also reported positive effects on task performance. Based on this, it can be derived that there is a positive linear relationship between effects on mental workload and effects on task performance: if the effects on mental workload are positive, the effects on task performance are likely to be positive as well. A drawback of presenting the data in this method is that it fails to show the frequency of same effects. For example, 6 studies found significant positive effects on both mental workload and task performance [9, 12, 13, 27, 30, 57]. However, this is not reflected in the graph.

Additionally, it is important to consider the outliers in the data. As shown in Fig. 8a and Fig. 8b, though few, a positive effect or no effect on mental workload can have negative effects on task performance measures. Stadler et al. [14] observed that though the use of AR decreased cognitive load, users sacrificed task completion time for accuracy. It is also important to note that though some studies found no statistically significant difference in overall NASA-TLX ratings, significant differences were found in individual subscales. As in the case of Kretschmer et al. [26], though total mental workload score did not reveal any significant differences, analyses of TLX subdimensions indicated significant mean difference between palletising with a paper list and AR glasses. Palletising with paper list was found to be significantly more mentally and temporally demanding than with AR glasses. The notion that an analyses of TLX subdimension scales can reveal significant differences is also supported by the works of Kim et al. [30], Dixon et al. [58], and Kyong-Ho and Kwang-Yun [10], whereby the use of AR was shown to significantly affect individual subscales. More often than not, AR is shown to influence mental and temporal demand more than other subscales.

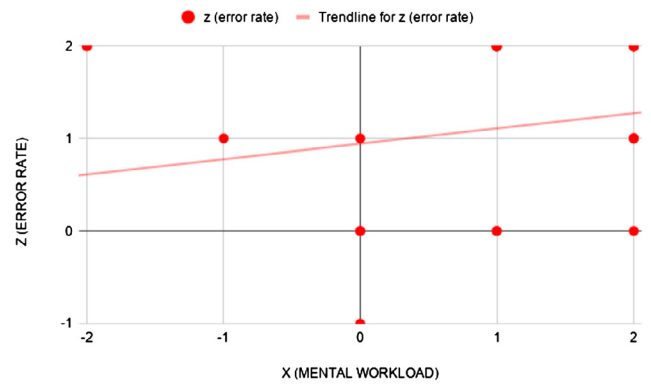
To summarise, there is a positive correlation between effects on mental workload and effects on task performance: when there are positive effects on mental workload, effects on task performance are most likely to be positive as well. However, it would be assumptions to say that effects on task performance are the resultant of effects on mental workload. Rather, the benefits on task performance may be the resultant of the effects of AR use on cognitive resources, as in, the freeing up of mental resources such as working memory [7, 14, 59]. This may not necessarily translate into an overall lower workload rating, but as a lower rating of the mental demand subscale or lower extraneous cognitive load. By freeing up more resources, operators can direct them into other efforts that is, performing the task, which then translates into improved performance [8]. As is suggested by the CLT: when extraneous load is decreased, more cognitive resources can be focused on generation of schemas, problem-solving, or in this case, performing the task. This notion is further discussed in Section 4.4.

4.4. How does AR affect mental workload?

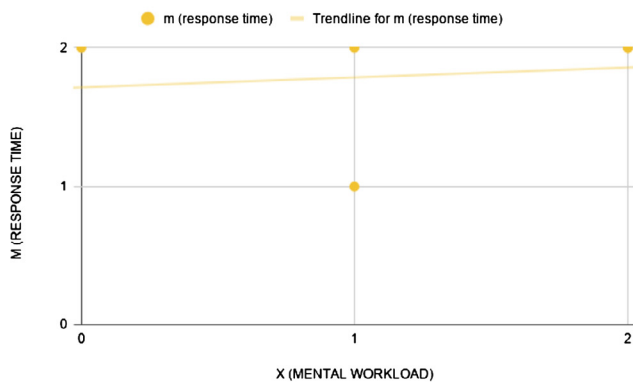
AR is shown to have a positive effect on mental workload by reducing it, whether the difference is significant or otherwise. Based on findings from the review of the selected studies, generally, AR can reduce mental workload by facilitating the perception of instruction. AR's primary capability to augment relevant virtual information upon the real world enhances perception and understanding of information in instances where the operator would have to refer to secondary sources



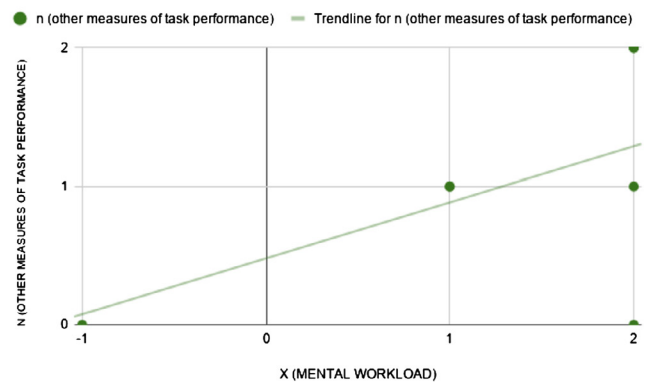
(a) Effects on task completion time based on effects on mental workload



(b) Effects on error rate based on effects on mental workload



(c) Effects on response time based on effects on mental workload



(d) Effects on other measures of task performance based on effects on mental workload

Fig. 8. The relationship between effects on mental workload and effects on measures of task performance shown in individual graphs.

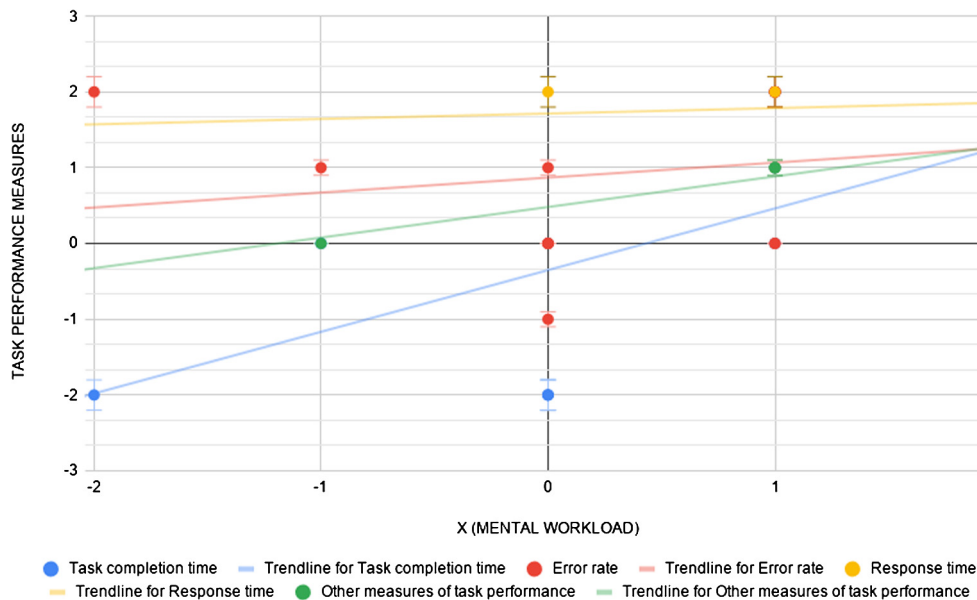


Fig. 9. The relationship between effects on mental workload and effects on measures of task performance.

of information, such as a secondary display or paper-based instructions [8, 12]. The extent of the effectiveness of AR systems in reducing mental workload and enhancing task performance is also affected by: type of AR device used [29], relevance of content, AR presentation [30, 45], user characteristics, task characteristics, and environmental characteristics.

4.4.1. AR's inherent characteristics facilitates the perception of instruction

AR's effect on reducing mental workload is largely attributable to its core feature of augmenting relevant information upon the physical world in real-time. When compared to conventional methods that detracts a user's or an operator's attention from the primary task, such as the use of a secondary display or paper-based instructions [59], AR

proves to be more effective because the augmenting of information upon the real world negates the need to switch between performing the task and searching for information to perform the task [14]. The AR device shows relevant task information directly onto the workspace, mitigating mental activities concerning interpretation and mapping of instructions to the AR device [59]. Simply put, AR facilitates the perception and understanding of contextual information by literally placing the information in the context.

With the support of AR, mental activities related to the searching and interpretation of information needed to perform a task is mitigated, and the only effort required is related to the actual execution of the task [59]. Users or operators would not need to perform typical actions or steps which would otherwise incur more physical and/or mental effort, such as taking measurements, marking positions [12], contextual switching [11], searching for information [8], remembering information [14], confirming instructions [60] and such. This offloading of mental effort is suggested to relieve time demanding mental activities and result in lower mental and temporal demands, contributing to an overall lower NASA-TLX rating [59].

This is in line with the Cognitive Load Theory, whereby presenting information in context is suggested to reduce extraneous load, reducing the effort required to understand the information in context and facilitate the generation of schemas or the execution of the task at hand. Divided attention has been identified to induce more mental load, as users or operators need to spend more cognitive resources to switch between contexts [11]. This is supported by the results of the work of Thees et al. [54], whereby AR was found to reduce extraneous cognitive load in a learning setting.

Furthermore, AR applications which present 3D instructions in the form of 3D models and animations facilitate spatial cognition and mental representation. Indeed, the work of Dan and Reiner [44] shows that the cognitive load of processing 3D information is lower than that of 2D. In the contexts of online shopping [7], mental rotation training [56] and manual assembly [1], presenting instructions in 3D format makes it easier to understand 3D information as it reduces the amount of mental effort that would typically be required to translate a 2D image into a 3D one [56]. By removing the need to imagine or visualize how an object is supposed to look like in 3D, AR reduces the consumption of user's cognitive resources by off-loading the mental task of visualization to the AR application [7].

Comparison studies between AR system elements (or factors) show that certain aspects of AR influence the degree of its effectiveness more so than other factors. For instance, Kim et al. [30] found that job performance, workload and usability were more affected by UI design than HWD type. Several other factors were shown to influence how effectively AR affects mental workload and task performance, namely: type of AR display device used, relevance and timeliness of content, information presentation, user characteristics and task characteristics.

4.4.2. Type of AR display device used

The type of AR display device used influences, to a degree, the usability of AR systems. Some types of AR devices are more effective than others. For instance, studies investigating and comparing Spatial AR (SAR) against conventional methods and other types of AR (such as Head Mounted Displays (HMD)/Head Worn Displays (HWD) or mobile AR) have consistently shown that SAR leads to better mental workload ratings and task performance [13, 29, 51, 52, 61, 62]. Not only does SAR free up both hands and allow for a greater degree of freedom and movement, it is also not restricted to a limited Field of View (FOV) [29, 51]. Though the use of AR HMD or HWD are popular in AR studies, they are commonly reported to have usability issues stemming from limited FOV and discomfort over prolonged use, all of which may negatively influence workload [1, 12, 47]. However, Kishishita et al. [46] found that a wider FOV does not necessarily lead to better performance (in search tasks) and that performance is more impacted by how information is presented rather than FOV. Furthermore, mobile AR devices and

HMDs/HWDs are prone to lagging and technical issues such as loss of tracking, all of which can negatively affect both mental workload and task performance.

To summarize:

- Spatial AR outperforms other types of AR displays (Mobile AR, HMD/HWD) in terms of mental workload and task performance
- The effects of FOV on mental workload, task performance and usability needs to be further investigated

4.4.3. Task characteristics

The effectiveness of AR presentations is influenced by the type of task. Deshpande and Kim [1] found that using AR in a low-complexity assembly task is slower compared to paper instructions. On a similar note, when the task complexity is too low, using an AR system may produce insignificant differences when compared to conventional methods or other approaches [60, 61]. Participants may make few errors and thus produce insignificant results or insufficient results for any statistical analysis to be made [26, 59]. Alves et al. [51] suggested that "AR methods could be more beneficial when applied to more complex tasks".

To summarize, in low task complexity situations, using AR may increase mental workload rather than decrease it. A low task complexity may also make it more challenging to obtain statistically significant results. The design of experiments must be designed so much so that the level of complexity is able to produce a statistically significant or measurable number of errors.

4.4.4. Relevance and timeliness of AR content

The effectiveness of an AR system is also influenced by the relevance of its content. In comparing between a multi-view system and a multi-view system with virtual information in a tele-operated crane task, Chi et al. [60] found that there was no significant difference in performance, which was attributed to low task complexity and lack of critical task-related information. Critical task-related information is deemed to be more beneficial to the user to solve problems than global information which could be deciphered from the environment [60]. Though an information may be relevant, it may not be critical. As such system designers should carefully aim at understanding what type of information the user needs to solve the problem to design an effective system.

4.4.5. AR presentation

The degree of effectiveness of AR systems on mental workload is largely influenced by AR presentation. AR presentation denotes how the virtual information is presented in the real world and comprises of several aspects which were identified from the studies such as: the type of annotation used [29, 45, 46, 47], modalities used [30], information availability [30] and information representation [48].

For an in-vehicle map navigation system, Bolton et al. [45] compared between four types of road navigation information presentations (Fig. 10): conventional fixed arrows display and three variations of AR presentations: dynamically changing arrows augmented on the road, landmarks highlighted with arrows and landmarks highlighted in boxes. The conventional approach was observed to be the most mentally demanding out of all four presentation alternatives. However, the AR presentations were not that different from each other in terms of NASA-TLX scores. Though insignificant, the NASA-TLX scores for landmark-based presentations were relatively low, suggesting that this style of presentation is likely to impose less workload and allow drivers to maintain attention on the primary driving task.

Fan et al. [7] experimented with high and low levels of simulated physical control (SPC) and environmental embedding (EE) and found that a high level of SPC and EE resulted in significant positive effects on cognitive load and cognitive fluency, which in turn increases a customer's attitude towards a product (Fig. 11). Fischer et al. [27] found that 3D surface reconstruction augmented by digitally reconstructed radiographs and live tool visualization resulted in a significantly lower



Fig. 10. Different types visualizations for an in-vehicle AR-HUD as investigated by Bolton et al. [45]. (Republished with permission of ACM (Association for Computing Machinery), from *An investigation of augmented reality presentations of landmark-based navigation using a head-up display*, Large, David R; Burnett, Gary; Bolton, Adam, 2015; permission conveyed through Copyright Clearance Center, Inc.)

task duration compared to conventional x-ray imaging in a simulated k-wire placement task. It also had better SURG-TLX ratings than x-ray augmentation on 2D video, suggesting that placement of the k-wire is best supported with a multi-view 3D visualization.

Kishishita et al. [46] investigated the effects of FOV on mental workload and search performance tasks. In their study, they employed two types of annotations: in-view labelling (Fig. 12a) and in-situ labelling (Fig. 12b). Their results did not show any effects of FOV on mental workload and search performance; a wider FOV did not necessarily lead to better search performance. Rather, search performance was affected by annotation type. With in-view labelling, discovery rate dropped as FOV increased; with in-situ labelling, discovery rate rose as FOV increased. They suggested that “it is likely more important to consider method of annotation than FOV for search related tasks”. This is supported by the work of Kim et al. [30], where they concluded that usability is more affected by information presentation rather than HMD type: “always-on” graphic-based UI was shown to significantly reduce overall NASA-TLX scores, task completion time and number of errors than conventional paper pick list method as compared to text-based UI or UI that is shown “on-demand”.

Lampen et al. [47] found that human simulation data with AR induced the least amount of cognitive load, followed by 3D in-situ instructions and pictorial instructions with significant differences between the three approaches. However, the reduction in task completion time and number of errors were insignificant when using the human simulation data with AR. Funk et al. [48] compared between three types of projected AR instructions: pictorial instructions (Fig. 13a), video instructions and contour instructions (Fig. 13b). Contour instructions were found to significantly reduce mental demand compared to no visualizations and had the best NASA-TLX ratings compared to other

visualizations. Though the difference between the other visualization approaches were insignificant, contour visualizations were also perceived as the easiest. Similarly, Wang et al. [49] compared the use of AR annotations and video instructions using Google Glass and found that the latter induced the highest amount of mental load, number of errors as well as the longest task completion time.

In another study, Funk et al. [52] found that in-situ projections induced the least amount of cognitive load in an assembly task compared to mobile AR (using a tablet) and HMD-based AR. They found a significant difference between in-situ projections and HMD, whereby HMD induced a higher cognitive load. Additionally, holding a tablet during assembly tasks was found to interfere with user’s movements when assembling with both hands.

To summarize:

- Information presentation has a larger degree of influence than HMD (and thus FOV)
- Different information representations induce different amounts of cognitive load
- 3D presentations outperform 2D presentations
- Graphic-based UI is better than text-based UI
- Video instructions may not be suitable as a type of AR visualization
- Information that is always available is more effective than that appears on demand

4.4.6. User characteristics

The effects of AR on mental workload and task performance also depend on user characteristics. The effects of AR on mental workload differ between individuals depending on characteristics such as age demographics [11], level of expertise in the domain [8, 12, 14, 60, 63] and level of familiarity with AR technology [14].

Funk et al. [63] studied the effects of in-situ instructions between expert and untrained workers in an assembly workplace over three working days. They found that expert workers were significantly slower using AR. Untrained workers made more errors and took a longer amount of time using the in-situ instructions during the learning phase. However, they were observed to “assemble the product significantly faster and without making any error” after three days of using it.

Loizeau et al. [8] analysed the effects on mental workload and task performance between three categories of domain expertise: beginner, intermediate and expert. They found that the impact of AR on intermediate user profiles is greater than beginners. They suggested that this could be because beginners need more time to process information whereas intermediate users are already familiar with the maintenance process and can take full advantage of AR information. Though beginners gave a higher rating for the AR intervention compared to the conventional method, it was lower than those given by intermediate and expert users. Loizeau et al. [8] explained that this could be because beginners are less at ease with the maintenance process.

In comparing the results of AR support for robot programming between novice and expert robot programmers, Stadler et al. [14] discovered that AR support has a beneficial impact on the workload ratings of expert robot programmers only. They suggested that this could be due to the expertise-reversal effect, whereby “instructional techniques that are highly effective with inexperienced learners can lose their effectiveness and even have negative consequences when used with more experienced learners”. They proposed that the presentation of the task-based information were more effective for robot programming experts. They also found that users who have had first-hand experience with AR (used AR at least once) had a longer task completion time using AR than without, even though they experienced significantly lower mental workload. They assumed that this could be because “users with first-hand AR experience were motivated to explore the capabilities” of their setup.

In a tele-operated crane operation task, Chi et al. [60] found that expert users using AR UI completed faster than novice users as their op-

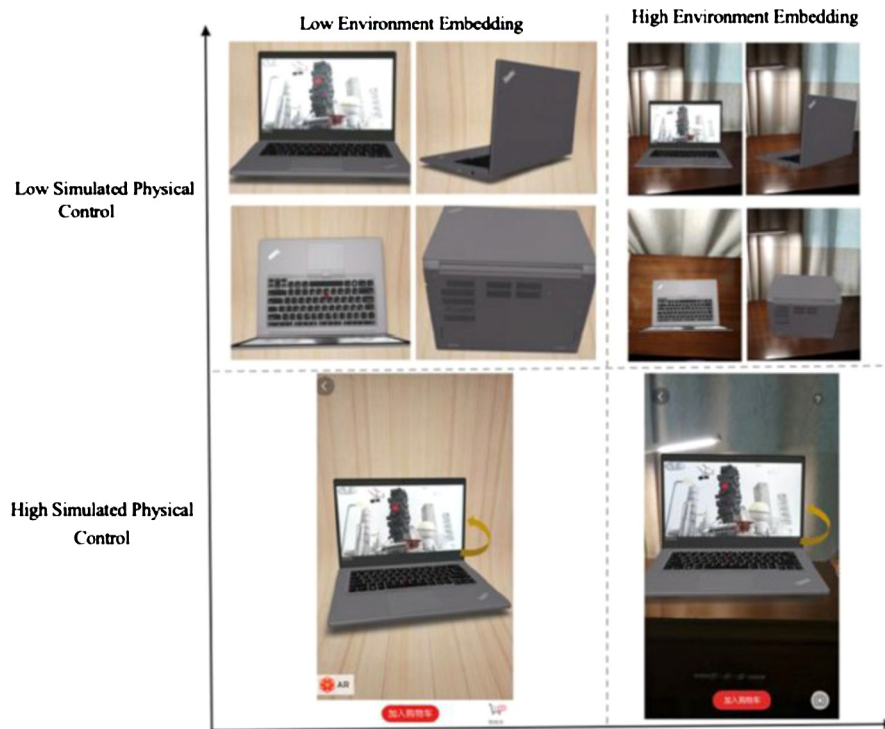


Fig. 11. Different visualizations as a result of the combination of high/low SPC and high/low EE as investigated by Fan et al. [7]. Reprinted from Journal of Retailing and Consumer Services, 53, Xiaojun Fan, Zeli Chai, Nianqi Deng, Xuebing Dong, Adoption of augmented reality in online retailing and consumers' product attitude: A cognitive perspective, Pages No. 8, Copyright (2020), with permission from Elsevier.

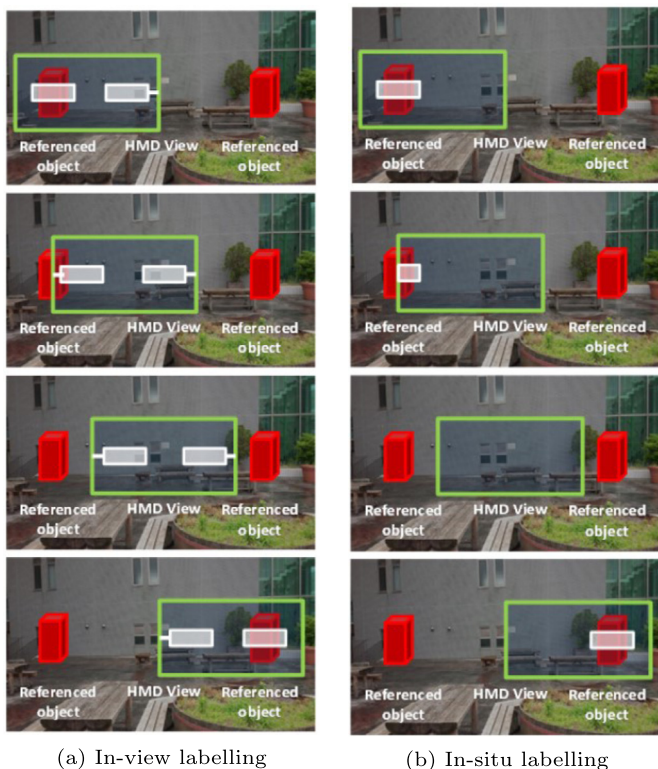


Fig. 12. “Schematic views of labelling techniques” as investigated by Kishishita et al. [46]. © 2014 IEEE. Reprinted, with permission, from IEEE Proceedings, Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks, Naohiro Kishishita, 2014.

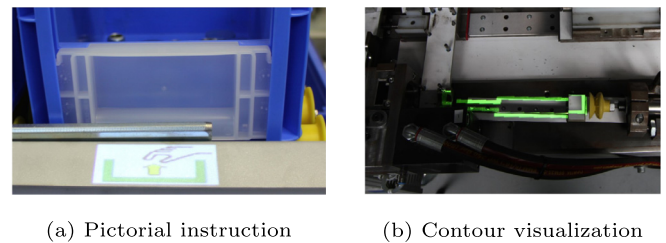


Fig. 13. Examples of pictorial and contour visualizations as investigated by Funk et al. [48]. Markus Funk, Andreas Bächler, Liane Bächler, Oliver Korn, Christoph Krieger, Thomas Heidenreich, and Albrecht Schmidt. 2015. Comparing projected in-situ feedback at the manual assembly workplace with impaired workers. In Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '15). Association for Computing Machinery, New York, NY, USA, Article 1, 1–8. Fig.3a and 3f. DOI:<https://doi.org/10.1145/2769493.2769496>.

erating style was quicker and more efficient. Chalhoub and Ayer [12] discovered that though AR led to significantly better performance across all six subcategories for expert (more than 10 years experience), intermediate (6 to 10 years experience), novice (1 to 5 years experience) and beginner (less than 1 year experience) users, experts performed significantly slower than other user categories. Chalhoub and Ayer [12] suggested that this is because practitioners are more familiar with paper instructions and have a harder time using the new technology. Even so, it is interesting to note that regardless of level of expertise, users were significantly faster when using AR.

Kim and Dey [11] found that the benefits of lower cognitive load manifested in different positive effects on in-vehicle navigation and driving task performance in younger and elder adults. For elder adults, AR improved the awareness of information related to in-situ decision making (situational awareness). Elder drivers were able to take advantage of the benefit of the route guidance aids while encountering driving and navigation difficulties when using AR. Whereas for younger adults,

reduced workloads in attention improved awareness of upcoming road networks (global awareness). However, Kim and Dey [11] observed that using the AR system reduced younger adult's perceptual capability in responding to in-situ events such as overhead traffic signals.

To summarize:

- Reduction in mental workload can manifest in different types of task performance benefits depending on user's age and level of expertise
- When using AR, the expertise-reversal effect can occur whereby the information presentation employed may benefit a particular expert-level group over the other
- The benefits of AR on mental workload and task performance may not be apparent in beginner- or novice- level users as they still need to familiarise themselves with the task
- Users with higher level of expertise may benefit more with AR support as they are already familiar with the task, know what to expect and know what information they need
- When designing AR systems, it may be useful to consider user characteristics and adapt information presentation accordingly

4.5. Limitations of current studies

It is important to acknowledge limitations in current studies as the ecological validity of AR is a prime concern when concerning widespread industrial adoption [1, 61]. Based on the review, the following limitations have been identified:

- Technical difficulties such as lagging and tracking issues
- Small sample size
- Controlled environment of experiment does not reflect real-world scenarios
- Low task complexity leading to insignificant results or lack of statistical analysis
- Short duration of experiment
- Discomfort over prolonged use of AR device

One of the most commonly reported issues are technical difficulties stemming from the hardware used. AR devices are prone to lag and tracking issues, most of which are out of the control of the experiment [1, 14]. Users reported on the increasing discomfort of HMDs/HWDs over long periods of time which may also influence the design of experiments in terms of duration [1, 12]. Furthermore, time constraints due to participant's availability may also influence experiment duration [13, 63]. Due to this, it is challenging to replicate real-world working hours and scenarios for usability studies, whereby external factors such as "congestion, noise, restricted FOV, connectivity, charging and other challenges could theoretically reduce the expected performance benefits reported" [12].

Some studies failed to obtain statistically significant results due to a small sample size [4, 58]. This problem is especially common in the Medicine & Surgery domain, whereby it may be difficult to get a hold of targeted users and their time. A total of 13/34 studies had a sample size of less than 20, with 3 of them having less than 10 participants.

The impact of prolonged use of AR systems in real-world scenarios need further investigation [12, 28]. Current reported benefits of AR are limited to experiments conducted in controlled environments which may not necessarily reflect real-world environments. This sentiment is echoed by Chalhoub and Ayer [12], whose concern was the applicability of their results in real-world environments as it presents a new set of safety and operability challenges that were not addressed in the study. Studies investigating AR systems for manual assembly typically employ a general assembly task model consisting of pick and place activities, which does not fully represent real-life manufacturing settings. Additionally, there may also be variances in workplace and shop floor designs [13, 52].

Another limitation that can influence the statistical significance of results is the experimental design, in particular, the task complexity [61]. Tasks with low complexity may induce too few errors for any statistical analysis to be performed [26, 59]. The design of experiments must be designed so much so that tasks with differing levels of complexity is able to produce statistically significant results.

5. Conclusion

This paper sought to understand the effects of AR systems on mental workload and task performance and identify the relationship between them. A positive correlation was found when the effects on task performance were mapped against effects on mental workload: if the effects on mental workload are positive, then the effects on task performance are more likely to be positive as well, and vice versa. The effects of AR on mental workload and task performance is attributable to its inherent characteristic to overlay relevant virtual information upon the real-world in real-time. By mitigating the mental tasks of visualizing [7, 56], searching [59], remembering [14], procuring [12], understanding and processing information to the AR system, AR reduces cognitive burden and free up mental resources, allowing for operators to focus more on performing the task [59]. The effectiveness of AR systems is influenced by the type of AR display device used, relevance and timeliness of content, information presentation, user characteristics and task characteristics. In developing AR systems to assist operators, it may be useful to consider these factors when designing user interfaces and interaction techniques. It may also be interesting to consider applying adaptation techniques to AR systems that can adapt to unique users and changing task scenarios and environments. Limitations of current studies include technical difficulties and applicability of results to real-world scenarios [1, 14, 61].

Declarations

Author contribution statement

All authors listed have significantly contributed to the development and the writing of this article.

Funding statement

This work was supported by Yayasan Universiti Teknologi PETRONAS (Grant No: 015LCO-001).

Data availability statement

Data included in article/supplementary material/referenced in article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

- [1] A. Deshpande, I. Kim, The effects of augmented reality on improving spatial problem solving for object assembly, *Adv. Eng. Inform.* 38 (2018) 760–775.
- [2] D. Chytas, E.O. Johnson, M. Piagkou, A. Mazarakis, G.C. Babis, E. Chronopoulos, V.S. Nikolaou, N. Lazaridis, K. Natsis, The role of augmented reality in anatomical education: an overview, *Ann. Anat. Anat. Anz.* 229 (2020) 151463.
- [3] M.-B. Ibáñez, C. Delgado-Kloos, Augmented reality for stem learning: a systematic review, *Comput. Educ.* 123 (2018) 109–123.

- [4] G. Ameri, A. Rankin, J.S. Baxter, J. Moore, S. Ganapathy, T.M. Peters, E.C. Chen, Development and evaluation of an augmented reality ultrasound guidance system for spinal anesthesia: preliminary results, *Ultrasound Med. Biol.* 45 (10) (2019) 2736–2746.
- [5] A. Nee, S. Ong, Virtual and augmented reality applications in manufacturing, in: 7th IFAC Conference on Manufacturing Modelling, Management, and Control, IFAC Proc. Vol. 46 (9) (2013) 15–26.
- [6] F. Lamberti, F. Manuri, A. Sanna, G. Paravati, P. Pezzolla, P. Montuschi, Challenges, opportunities, and future trends of emerging techniques for augmented reality-based maintenance, *IEEE Trans. Emerg. Top. Comput.* 2 (4) (2014) 411–421.
- [7] X. Fan, Z. Chai, N. Deng, X. Dong, Adoption of augmented reality in online retailing and consumers' product attitude: a cognitive perspective, *J. Retail. Consum. Serv.* 53 (2020) 101986.
- [8] Q. Loizeau, F. Danglade, F. Ababsa, F. Merienne, Evaluating added value of augmented reality to assist aeronautical maintenance workers—experimentation on on-field use case, in: P. Bourdot, V. Interrante, L. Nedel, N. Magnenat-Thalmann, G. Zachmann (Eds.), *Virtual Reality and Augmented Reality*, Springer International Publishing, Cham, 2019, pp. 151–169.
- [9] A. Setti, P. Bosetti, M. Ragni, Artool- augmented reality platform for machining setup and maintenance, in: Y. Bi, S. Kapoor, R. Bhatia (Eds.), *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016*, Springer International Publishing, Cham, 2018, pp. 457–475.
- [10] K.-H. Kim, K.-Y. Wahn, Effects on productivity and safety of map and augmented reality navigation paradigms, *IEICE Trans. Inf. Syst.* E94.D (5) (2011) 1051–1061.
- [11] S. Kim, A.K. Dey, Augmenting human senses to improve the user experience in cars: applying augmented reality and haptics approaches to reduce cognitive distances, *Multimed. Tools Appl.* 75 (16) (2016) 9587–9607.
- [12] J. Chalhoub, S.K. Ayer, Exploring the performance of an augmented reality application for construction layout tasks, *Multimed. Tools Appl.* 78 (2019) 35075–35098.
- [13] T. Bosch, R. Könemann, H. de Cock, G. van Rhijn, The effects of projected versus display instructions on productivity, quality and workload in a simulated assembly task, in: *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '17*, Association for Computing Machinery, New York, NY, USA, 2017, pp. 412–415.
- [14] S. Stadler, K. Kain, M. Giuliani, N. Mirnig, G. Stollnberger, M. Tscheligi, Augmented reality for industrial robot programmers: workload analysis for task-based, augmented reality-supported robot control, in: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016, pp. 179–184.
- [15] M. Hertzum, K.D. Holmegaard, Perceived time as a measure of mental workload: effects of time constraints and task success, *Int. J. Hum.-Comput. Interact.* 29 (1) (2013) 26–39.
- [16] A. Byrne, Measurement of mental workload in clinical medicine: a review study, *Anesth. Pain Med.* 1 (2) (2011) 90–94.
- [17] S.G. Hart, L.E. Staveland, Development of nasa-tlx (task load index): results of empirical and theoretical research, in: P.A. Hancock, N. Meshkati (Eds.), *Human Mental Workload*, in: *Advances in Psychology*, vol. 52, North-Holland, 1988, pp. 139–183.
- [18] P.S. Tsang, M.A. Vidulich, *Mental Workload and Situation Awareness*, John Wiley & Sons, Ltd, 2006, pp. 243–268.
- [19] R.T. Azuma, A survey of augmented reality, *Presence, Teleoper. Virtual Environ.* 6 (4) (1997) 355–385.
- [20] P. Milgram, F. Kishino, A taxonomy of mixed reality visual displays, *IEICE Trans. Inf. Syst.* E77-D (12) (1994) 1321–1329.
- [21] T. Narumi, T. Kajinami, S. Nishizaka, T. Tanikawa, M. Hirose, Pseudo-gustatory display system based on cross-modal integration of vision, olfaction and gustation, in: *2011 IEEE Virtual Reality Conference*, 2011, pp. 127–130.
- [22] R. Van Krevelen, R. Poelman, A survey of augmented reality technologies, applications and limitations, *Int. J. Virtual Real.* 9 (2010) 1.
- [23] X. Li, W. Yi, H.-L. Chi, X. Wang, A.P. Chan, A critical review of virtual and augmented reality (vr/ar) applications in construction safety, *Autom. Constr.* 86 (2018) 150–162.
- [24] L.F. de Souza Cardoso, F.C.M.Q. Mariano, E.R. Zorzal, A survey of industrial augmented reality, *Comput. Ind. Eng.* 139 (2020) 106159.
- [25] P. Soltani, A.H. Morice, Augmented reality tools for sports education and training, *Comput. Educ.* 155 (2020) 103923.
- [26] V. Kretschmer, T. Plewan, G. Rinke, B. Maettig, Smart palletisation: cognitive ergonomics in augmented reality based palletising, in: W. Karwowski, T. Ahram (Eds.), *Intelligent Human Systems Integration*, Springer International Publishing, Cham, 2018, pp. 355–360.
- [27] M. Fischer, B. Fuerst, S. Chun Lee, J. Fotouhi, S. Habert, S. Weidert, E. Euler, G. Osgood, N. Navab, Preclinical usability study of multiple augmented reality concepts for k-wire placement, *Int. J. Comput. Assisted Radiol. Surg.* 11 (2016) 1007–1014.
- [28] F. Loch, F. Quint, I. Brishel, Comparing video and augmented reality assistance in manual assembly, in: *2016 12th International Conference on Intelligent Environments (IE)*, 2016, pp. 147–150.
- [29] J. Baumeister, S.Y. Ssin, N.A.M. ElSayed, J. Dorrian, D.P. Webb, J.A. Walsh, T.M. Simon, A. Irlitti, R.T. Smith, M. Kohler, B.H. Thomas, Cognitive cost of using augmented reality displays, *IEEE Trans. Vis. Comput. Graph.* 23 (11) (2017) 2378–2388.
- [30] S. Kim, M.A. Nussbaum, J.L. Gabbard, Influences of augmented reality head-worn display type and user interface design on performance and usability in simulated warehouse order picking, *Appl. Ergon.* 74 (2019) 186–193.
- [31] B.H. Kantowitz, *Mental Workload*, North-Holland Publishing Co., NLD, 1987, pp. 81–121.
- [32] N. Moray (Ed.), *Mental Workload*, vol. 8, 1st edition, Springer US, 1979.
- [33] C.D. Wickens, Multiple resources and mental workload, *Hum. Factors* 50 (3) (2008) 449–455, PMID: 18689052.
- [34] B. Cain, A review of the mental workload literature, Tech. Rep., Defence Research and Development Canada Toronto, 2007.
- [35] A.F. Kramer, Physiological metrics of mental workload: a review of recent progress, *Mult. Task Perform.* (1991) 279–328.
- [36] R. Parasuraman, T.B. Sheridan, C.D. Wickens, Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs, *J. Cogn. Eng. Decis. Mak.* 2 (2) (2008) 140–160.
- [37] S.G. Hart, Nasa-task load index (nasa-tlx); 20 years later, *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 50 (9) (2006) 904–908.
- [38] D.K. Mitchell, Mental workload and arl workload modeling tools, Tech. Rep. ARL-TN-161, Army Research Laboratory, Aberdeen Proving Ground, MD 21005-5425, Apr. 2000.
- [39] J.M. Rolfe, *Measurement of Man at Work*, Taylor and Francis, London, 1971, Ch. The secondary task as a measure of workload.
- [40] B.H. Kantowitz, J.L. Knight, Testing tapping timesharing, ii: auditory secondary task, *Acta Psychol.* 40 (5) (1976) 343–362.
- [41] R.A. Grier, How high is high? A meta-analysis of nasa-tlx global workload scores, *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 59 (1) (2015) 1727–1731.
- [42] P. Chandler, J. Sweller, Cognitive load theory and the format of instruction, *Cogn. Instr.* 8 (4) (1991) 293–332.
- [43] J. Sweller, J.J.G. van Merriënboer, F.G.W.C. Paas, Cognitive architecture and instructional design, *Educ. Psychol. Rev.* 10 (3) (1998) 251–296.
- [44] A. Dan, M. Reiner, Eeg-based cognitive load of processing events in 3d virtual worlds is lower than processing events in 2d displays, in: *Neural Patterns of Learning, Cognitive Enhancement and Affect*, *Int. J. Psychophysiol.* 122 (2017) 75–84.
- [45] A. Bolton, G. Burnett, D.R. Large, An investigation of augmented reality presentations of landmark-based navigation using a head-up display, in: *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '15*, Association for Computing Machinery, New York, NY, USA, 2015, pp. 56–63.
- [46] N. Kishishita, K. Kiyokawa, J. Orlosky, T. Mashita, H. Takemura, E. Kruijff, Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks, in: *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2014, pp. 177–186.
- [47] E. Lampen, J. Teuber, F. Gaisbauer, T. Bär, T. Pfeiffer, S. Wachsmuth, Combining simulation and augmented reality methods for enhanced worker assistance in manual assembly, in: *Ljubljana, Slovenia, June 12-14, 2019.*, *Proc. CIRP* 81 (2019) 588–593.
- [48] M. Funk, A. Bächler, L. Bächler, O. Korn, C. Krieger, T. Heidenreich, A. Schmidt, Comparing projected in-situ feedback at the manual assembly workplace with impaired workers, in: *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '15*, Association for Computing Machinery, New York, NY, USA, 2015.
- [49] C.-H. Wang, N.-H. Tsai, J.-M. Lu, M.-J.J. Wang, Usability evaluation of an instructional application based on Google glass for mobile phone disassembly tasks, *Appl. Ergon.* 77 (2019) 58–69.
- [50] H. Kim, X. Wu, J.L. Gabbard, N.F. Polys, Exploring head-up augmented reality interfaces for crash warning systems, in: *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '13*, Association for Computing Machinery, New York, NY, USA, 2013, pp. 224–227.
- [51] J. Alves, B. Marques, M. Oliveira, T. Araújo, P. Dias, B.S. Santos, Comparing spatial and mobile augmented reality for guiding assembling procedures with task validation, in: *2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 2019, pp. 1–6.
- [52] M. Funk, T. Kosch, A. Schmidt, Interactive worker assistance: comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions, in: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '16*, Association for Computing Machinery, New York, NY, USA, 2016, pp. 934–939.
- [53] K.-B. Park, M. Kim, S.H. Choi, J.Y. Lee, Deep learning-based smart task assistance in wearable augmented reality, *Robot. Comput.-Integr. Manuf.* 63 (2020) 101887.
- [54] M. Thees, S. Kapp, M.P. Strzys, F. Beil, P. Lukowicz, J. Kuhn, Effects of augmented reality on learning and cognitive load in university physics laboratory courses, *Comput. Hum. Behav.* 108 (2020) 106316.
- [55] Z. Yang, J. Shi, W. Jiang, Y. Sui, Y. Wu, S. Ma, C. Kang, H. Li, Influences of augmented reality assistance on performance and cognitive loads in different stages of assembly task, *Front. Psychol.* 10 (2019) 1703.
- [56] I.-J. Lee, C.-H. Chen, K.-P. Chang, Augmented reality technology combined with three-dimensional holography to train the mental rotation ability of older adults, *Comput. Hum. Behav.* 65 (2016) 488–500.
- [57] L. Hou, X. Wang, L. Bernold, Using animated augmented reality to cognitively guide assembly, *J. Comput. Civ. Eng.* 27 (5) (2013).
- [58] B. Dixon, M. Daly, H. Chan, A. Vecsan, I. Witterick, J. Irish, Augmented image guidance improves skull base navigation and reduces task workload in trainees: a preclinical trial, *Laryngoscope* 121 (2011) 2060–2064.

- [59] G.M. Re, J. Oliver, M. Bordegoni, Impact of monitor-based augmented reality for on-site industrial manual operations, *Cogn. Technol. Work* 18 (2016) 379–392.
- [60] H.-L. Chi, Y.-C. Chen, S.-C. Kang, S.-H. Hsieh, Development of user interface for tele-operated cranes, *Adv. Eng. Inform.* 26 (3) (2012) 641–652.
- [61] M. Mengoni, S. Ceccacci, A. Generosi, A. Leopardi, Spatial augmented reality: an application for human work in smart manufacturing environment, in: 28th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2018), June 11–14, 2018, Columbus, OH, USA Global Integration of Intelligent Manufacturing and Smart Industry for Good of Humanity, *Proc. Manuf.* 17 (2018) 476–483.
- [62] E.S. Wilschut, R. Könemann, M.S. Murphy, G.J.W. van Rhijn, T. Bosch, Evaluating learning approaches for product assembly: using chunking of instructions, spatial augmented reality and display based work instructions, in: Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 376–381.
- [63] M. Funk, A. Bächler, L. Bächler, T. Kosch, T. Heidenreich, A. Schmidt, Working with augmented reality? A long-term analysis of in-situ instructions at the assembly workplace, in: Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 222–229.