Empirical evidence, evaluation criteria and challenges for the effectiveness of virtual and mixed reality tools for training operators of car service maintenance

Simone Borsci[1], Glyn Lawson[1], Simon Broome[2]

[1]Human Factors Research Group, Faculty of Engineering, The University of Nottingham, Nottingham NG7 2RD, United Kingdom.

[2]Jaguar Land Rover, Abbey Road, Whitley, Coventry, CV3 4LF, United Kingdom

Abstract

The debate on effectiveness of virtual and mixed reality (VR/MR) tools for training professionals and operators is long-running with prominent contributions arguing that there are several shortfalls of experimental approaches and assessment criteria reported within the literature. In the automotive context, although car-makers were pioneers in the use of VR/MR tools for supporting designers, researchers started only recently to explore the effectiveness of VR/MR systems as mean for driving external operators of service centres to acquire the procedural skills necessary for car maintenance processes. In fact, from 463 journal articles on VR/MR tools for training published in the last thirty years, we identified only eight articles in which researchers experimentally tested the effectiveness of VR/MR tools for training service operators' skills. To survey the current findings and the deficiencies of these eight studies, we use two main drivers: i) a well-known framework of organizational training programmes, and ii) a list of eleven evaluation criteria widely applied by researchers of different fields for assessing the effectiveness of training carried out with VR/MR systems. The analysis that we present allows us to: i) identify a trend among automotive researchers of focusing their analysis only on car service operators' performance in terms of time and errors, by leaving unexplored important pre- and post-training aspects that could affect the effectiveness of VR/MR tools to deliver training contents – e.g., people skills, previous experience, cibersickness, presence and engagement, usability and satisfaction and ii) outline the future challenges for designing and assessing VR/MR tools for training car service operators.

1. Introduction

Virtual reality (VR) systems are defined as human-computer environments in which users are immersed in, and able to perceive, act and interact with a three-dimensional world (Bowman, Gabbard, & Hix, 2002; Milgram & Kishino, 1994). However, as Milgram and Kishino (1994) underlined, immersive systems represent only one, and an extreme, point in the continuum from virtual to real world. In fact, along this continuum researchers and manufacturers, by mixing virtual and real worlds, have developed different hybrid technologies to serve different goals. These Mixed Reality (MR) systems are designed, for instance, to augment the user experience of the real environment with virtual information (augmented reality) or to augment the virtual systems through real inputs (augmented virtuality). The debate on the application of VR/MR tools among researchers started in the late Sixties (see: Sutherland, 1968), but due to the high costs of these simulation tools only a restricted group of

experts had access to this debate. It was only in the Nineties that a larger community of experts started to explore the potential of VR/MR systems thanks to the price drops of hardware and the concurrent increase of the technologies' performances (Adam, 1993; Robler, 1998).

The costs of these tools are still today a barrier that excludes many researchers from the access to, and experimental analysis of, these technologies. However, nowadays several different VR/MR systems which vary in terms of software, hardware, functioning and interaction modes (Mantovani, Castelnuovo, Gaggioli, & Riva, 2003) are used daily in different fields – e.g., manufacturing, surgery, education, military – as support for the work of professionals, and as part of programmes to train the skills of employees, specialists and managers (e.g., prototype and assembly, drive, fight, fly, surgery procedures etc. see: Belardinelli, Blümel, Müller, & Schenk, 2008). As Mantovani (2003) underlined, in the context of a training programme, VR/MR tools are used by practitioners to deliver contents and to drive operators to acquire, or increase their previous, know-how – intended as a set of procedural skills to perform in correct order, and in an efficient and effective way all the steps to achieve all the tasks for reaching a goal (Anderson, 1982) – and to engage operators in the exercise of core skills for performing the procedures – such as abstract reasoning, visualization and management of complex information spaces etc. (Salzman, Dede, Loftin, & Chen, 1999).

Among the different fields of VR/MR application, automotive manufacturers have been pioneers in the use of tools for prototyping and assessing a product's design – e.g., computer-aided design – and for verifying the accuracy of assembly and maintenance processes (Gomes de Sá & Zachmann, 1999). However, compared to other industries, auto-makers started only in recent times to look at the application of VR/MR as systems for training operators know-how. This interest in VR/MR tools for training has recently produced several international projects and systems for supporting operators of car service – e.g., the European projects SKILLS (http://www.skills-ip.eu/) and VISTRA (http://www.vistra-project.eu), as well as tools like the Mobile Augmented Reality Technical Assistance created by Metaio and Volkswagen (http://www.metaio.com) and the BMW Augmented Reality glasses (http://www.bmw.com/)

Car manufacturers are interested in both training and providing a support tool for professionals of car service maintenance; intended as a highly specialized multistep process in which operators have to perform in the correct way the sequences: i) disassembly of the car and removal of faulty system components, ii) replacement or repair of these components and, iii) reassembly of the car (Li, Khoo, & Tor, 2003). The training of service operators is a priority for automotive manufacturers because the market of car maintenance and repair, in the US alone, totalled 166.5 billion for 2012 (Mintel Group Ltd, 2013). Moreover, the enhancement of these operators and their accuracy in maintenance procedures is important for auto-makers because service operators are the main interface for the customers (i.e., to solve their cars' issues). Therefore, the performance of these operators could affect the brand image, especially when new models of cars are released on the market and the skills of operators in supporting customers' needs to become an essential factor for the success of these new products (Dombrowski, Engel, & Schulze, 2011; Gaiardelli, Resta, Martinez, Pinto, & Albores, 2014).

VR/MR tools are considered reliable solutions to train operators of service maintenance, at least, for three main reasons underlined in literature (Mantovani, 2003; Mantovani et al., 2003; Mikropoulos, Chalkidis, Katsikis, & Kossivaki, 1997). First, after the initial investment to acquire the systems, and the maintenance costs, VR/MR tools reduce the overall training costs. Second, in line with the learning by doing approach (Bruner, 1966), VR/MR systems allow people to visualize and interact, during a training, with simulated real artefacts. This interactive experience increases the quality of the trainees' acquisition of the skills. Moreover, these systems offer a good adaptability to the people's needs and learning style – i.e., personalization – by generally increasing the trainees' motivation during the training. Third, practitioners and trainers by VR/MR systems can easily

collect a wide set of data about the trainees' performances, to check, assess and calibrate the training process. In line with that, training with VR/MR tools is considered more powerful and effective than a classic training programme. There is, however, a hot debate in the scientific literature on VR/MR tools effectiveness for training. In fact, researchers commonly analyse the efficacy of these tools with small samples and with a limited set of comparable evaluation criteria (Haque & Srinivasan, 2006; Yuviler-Gavish et al., 2013). Therefore, the reliability of the current experimental results is still uncertain.

Some differences could be underlined among the fields of VR/MR tools applications for training. For instance in fields such as surgery or military procedures, comparable evaluation criteria (within each field) can be used by researchers for assessing the VR/MR tool effectiveness, because the tools are applied under similar: i) training programmes aims – i.e., increase the curriculum of specialists, and ii) training conditions and contents– e.g., to train a specific surgery procedures, or to flight or drive a particular vehicle etc. In light of that, by aiming to train one, or more, specific and specialized procedure with a set of standardized rules, researchers could assess the trainees' performance against comparable benchmarks within the specialist fields. Therefore, for these kinds of training applications, although the results of the studies are still under debate, a relatively reliable set of evidence supports the idea that VR/MR tools are effective for training.

Conversely, when researchers aim to assess the training of operators of service maintenance procedures, a set of less stable and comparable evaluation criteria can be applied, because: i) the programmes' aims, the actors involved in the training (i.e., managers, suppliers, trainers, trainees etc. Anastassova, Burkhardt, Mégard, & Ehanno, 2005), and VR/MR tools differ from company to company, and ii) usually operators are trained to perform variable procedures – e.g., assembly of, or maintenance on, different products (Michalos, Makris, Papakostas, Mourtzis, & Chryssolouris, 2010). Therefore, the outcomes of studies on VR/MR tools effectiveness for training service maintenance could vary substantially across the companies, thus producing a high level of uncertainty about the effective application of these systems.

To minimize this uncertainty about the effective application of VR/MR tools, automotive researchers have typically limited the effectiveness evaluation of the training carried out by these systems to the estimation of the improvement of operators' performance in terms of time and errors (Tang, Owen, Biocca, & Mou, 2003). However, as several authors underlined, VR/MR systems are tools applied under the conditions of a training programme – intended as a set of objectives, designed rules, actions and tools for reaching the learning needs in tune with a company requirements– and the effectiveness of these systems cannot be assessed outside of the training programme framework (De Freitas, Rebolledo-Mendez, Liarokapis, Magoulas, & Poulovassilis, 2010; Mantovani, 2003; Mantovani et al., 2003; Moskaliuk, Bertram, & Cress, 2012). Therefore, to fully assess the effectiveness of VR/MR tools it is necessary to extend the attention from the measurement of the training outcomes – i.e., trainees' performance – to both the training process experienced by trainees, and the match between training programme objectives and outcomes (Mantovani, 2003) .

The main components of organizational trainings and the key factors for evaluating these programmes are well-established in literature (Arthur Jr, Bennett Jr, Edens, & Bell, 2003; Campbell, 1971; Mantovani, 2003; Mantovani et al., 2003; Salas, Tannenbaum, Kraiger, & Smith-Jentsch, 2012; Taylor, Russ-Eft, & Chan, 2005; Tennant, Boonkrong, & Roberts, 2002), nevertheless, researchers have rarely referred to this background in the definition of the evaluation criteria for testing the effectiveness of VR/MR tools.

This paper, taking into account the importance of the training programme under which a VR/MR tool is applied, aims to survey current approaches of assessment of VR/MR effectiveness for training the operators' procedural skills in service maintenance of a product, with a particular interest on effectiveness data related to the

automotive field. We used a well-known framework of training programme design (Arthur Jr et al., 2003; Salas et al., 2012; Taylor et al., 2005) and a set of evaluation criteria applied in experimental studies of different fields for training people , as drivers to: i) identify the current results and the limits of these studies in terms of evaluation criteria applied by researchers, and iii) outline the future challenges in training car service maintenance with VR/MR tools.

2. Current studies of virtual and mixed reality tools effectiveness for training operators of service maintenance

VR/MR tools are considered by professionals as expensive one-off creations tied to company needs (Mantovani et al., 2003). Thus after the investment for designing their own tools, manufacturers are often reluctant to add an extra budget to test the efficacy of these systems by removing several specialized operators from their work to take part in an evaluation process (Haque & Srinivasan, 2006). This economic barrier usually prevents researchers to experimentally assess the effectiveness of VR/MR tools for training. This lack of experimental evidence is particularly clear in the manufacturing and automotive sectors in which researchers often try to justify, instead of analyse, the efficacy of VR/MR tools for training on the basis of experimental outcomes coming from other fields (Tang et al., 2003). Moreover, where VR/MR tools effectiveness has been experimentally assessed, researchers only evaluated the ability of the tools to optimize operators' time in performing accurate maintenance procedures, independently from the company training programme aims under which the VR/MR tools were applied (Anastassova & Burkhardt, 2009; Mantovani, 2003; Tang et al., 2003). This lack of consideration of organizational needs and environment – i.e., actors and their relationships – has often led researchers to obtain inconsistent and non-comparable results about the effectiveness of VR/MR tools for training service maintenance (Haque & Srinivasan, 2006; Mantovani, 2003; Mantovani et al., 2003; Yuviler-Gavish et al., 2013).

The use of VR/MR systems, however, for training professionals and operators in different fields is a growing research area. As Figure 1 shows, the number of published articles on training with VR/MR tools constantly increased in the last thirty years. We identified a total sample of 463 peer reviewed papers (48 published from 1990 to 1999; 225 from 2000 to 2009; and 190 from 2010 to 2013), with titles that contain the keywords 'Virtual Reality' or 'Augmented Reality', 'Mixed reality' and 'Training' in Scopus, Web of Knowledge and IEEExplore from 1990 to 2013. 60% (278) of these articles aim to analyse the functioning of one or more models of VR/MR training system for specific procedures and to discuss the design approach and principles (e.g., Belardinelli et al., 2008; Xu & Gao, 2011). 30.5% (141) of the articles aimed to discuss issues related to the interaction and the trends of use of VR/MR systems in different fields – such as technology acceptance, usability, benefit of use etc. (e.g., Fong et al., 2010). Finally, in tune with the researchers' concerns about the lack of effectiveness assessment, only 9.5% (44) of the papers are systematic review and experimental articles in which trainees' performances are assessed – in terms of time, learning curve, psychomotor skills improvement etc.(e.g., Grantcharov, Bardram, Funch-Jensen, & Rosenberg, 2003; Seymour et al., 2002).

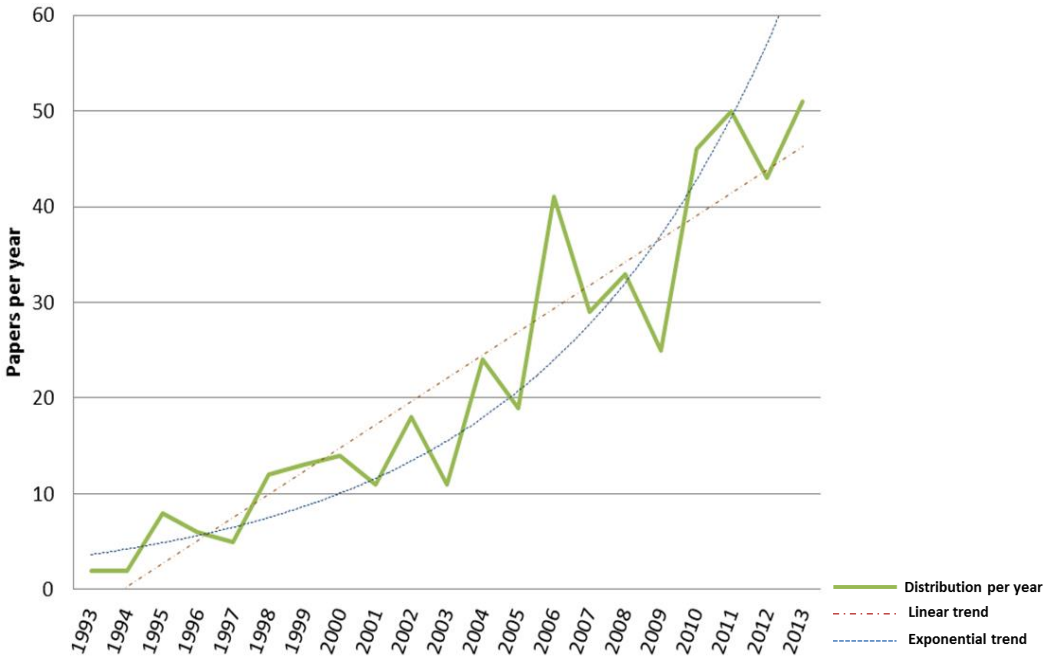


Fig. 1. Number of articles per year with titles that contains the keywords 'Virtual Reality' or 'Augmented Reality', 'Mixed reality' and 'Training' in Scopus, Web of Knowledge and IEEExplore from 1990 to 2013. The linear trend shows the constant growth among the years of publishing articles on VR/MR tools for training, while the exponential trend estimated the expected growth over the years.

By considering the uncertainty about the results of VR/MR tools effectiveness for training, and to extend the reliability of experimental analyses, researchers in the surgical field have often performed meta-analytic reviews to aggregate data of studies about the efficacy of tools for training specialist skills (Haque & Srinivasan, 2006). Although meta-analysis is a reliable solution to clarify the effectiveness of VR/MR tools for training, currently too few experimental and homogenous studies – i.e., terms of the training contents – are available in literature for a meaningful meta-analysis on the effectiveness of these systems for training automotive and manufacturing operators of service maintenance. In fact, among the 44 papers which analysed the trainees performances after a training with VR/MR tools, we identified only ten peer-reviewed articles in which researchers declared in the title, or in the text of the paper, to focus their attention on the VR/MR tools effectiveness for training the procedural skills for performing the entire (or a part of a) service maintenance process.

Among these ten papers, only two are experimental studies in the automotive field (see, Table 1: Anastassova & Burkhardt, 2009; Anastassova et al., 2005). However, further four articles were included in the list as although these were not explicitly referred to automotive field the authors discussed and tested core skills for service maintenance (e.g., visual spatial abilities, perceptual feedbacks etc. see: Abate, Guida, Leoncini, Nappi, & Ricciardi, 2009; Yuviler-Gavish, Yechiam, & Kallai, 2011) or they proposed a comparative assessment of different tools in training industrial maintenance processes (Webel et al., 2013; Yuviler-Gavish et al., 2013). Four journal articles were excluded from the list because authors discussed the features and the use of VR/MR tools for training, without presenting any experimental analysis of effectiveness of these systems (Li et al., 2003; Neumann & Majoros, 1998; Stork et al., 2012; Xu & Gao, 2011).

Finally, we manually added to the six articles selected three conference papers in which authors presented the results of experiments on VR/MR tools for training operators in one or more essential step for service or maintenance, such as object recognition for disassembly and reassembly of a product (Dawei, Bhatti, & Nahavandi, 2009; Yuviler-Gavish, Gutierrez Seco, et al., 2011).

	Authors	Publication Year	Kind of article	Kind of training tool discussed	Main Focus of the article
1.	Anastassova	2005	Journal paper, peer review	MR	Match between training aims and user performance
2.	Anastassova and Burkhardt	2009	Journal paper, peer review	MR	Match between training aims and user performance
3.	Dawei, Bhatti, and Nahavandi	2009	Conference paper, peer review	VR	Improvement of user performance
4.	Abate et al.	2009	Journal paper, peer review	VR	Perception of performance inprovment
5.	Yuviler-Gavish, Yechiam, and Kallai	2011	Journal paper, peer review	MR	Improvement of user performance
6.	Yuviler-Gavish et al.	2011	Conference paper, peer review	VR and MR	Improvement of user performance
7.	Webel et al.	2013	Journal paper, peer review	MR	Improvement of user performance
8.	Yuviler-Gavish et al.	2013	Journal paper, peer review	VR and MR	Improvement of user performance and comparative assessment of tools

Table 1. List of current experimental studies on virtual and mixed reality tools for training operators of car service maintenance

Table 1 shows that the majority of the studies (6 out of 8) investigated MR training tools, with a particular focus on augmented reality systems. Among these, two articles also compared VR and MR systems effectiveness (Yuviler-Gavish, Gutierrez Seco, et al., 2011; Yuviler-Gavish et al., 2013), while two articles assess the training with only a VR tool (Abate et al., 2009; Dawei et al., 2009). The trend among researchers is to consider MR tools as more effective than VR systems for training maintenance procedures because the former can be used both during the training and in the daily activities of operators – e.g., as real time instruction instruments – while the latter can be used only for delivering the training contents. There are, however, two problems related to this assumption. First, although the advantages to received real-time instruction could simplify the operators' work, outside of experimental scenarios training programmes are usually centralized by car manufacturers (Anastassova & Burkhardt, 2009; Anastassova et al., 2005) and operators may use MR tools only at training centres, because the availability of these systems in real world (i.e., car service centres) is still far from daily practice. Second, currently only one study compared the effectiveness of different VR and MR tools for training maintenance processes, and in tune with the authors, it is not yet clear whether the MR tools are more effective than VR tools for training service procedures (Yuviler-Gavish et al., 2013). In light of that, there is actually not enough evidence for considering MR tools better than VR for training service maintenance, but only that MR tools could be reliably used in the future for driving the operators' during their work activities when these technologies will become more affordable and commonplace.

Moreover table 1 clearly shows that a large number of researchers select evaluation criteria only for estimating how much VR/MR tools increase the performances of trainees. Nevertheless, the real aim of a training effectiveness analysis is to show the pertinence and usefulness of a VR/MR tool as part of a training programme

with specific aims (Arthur Jr et al., 2003; Mantovani, 2003; Salas & Cannon-Bowers, 2001; Salas et al., 2012). In line with that, the effectiveness assessment is the process by which researchers could identify the shortfalls and the possible customizations of training tool for satisfying the needs of an organization (Anastassova & Burkhardt, 2009; Anastassova et al., 2005; Stone, Watts, Zhong, & Wei, 2011).

In light of that, to fully compare the results and to analyse the limits of current studies on VR/MR tools for training operators of car service maintenance it is necessary to define the main factors of a training programme that researchers have to consider for defining the criteria of effectiveness evaluation. In order to model these factors we adopted, for our aims, a well-known framework proposed by Arthur et al. (2003) of training programme design.

3. Training programme as the rationale of effectiveness assessment of systems

A training programme can be defined as a process in which novices acquire skills, through a series of well formalized actions, becoming experts (Fitts & Posner, 1967). Practitioners design training processes to allow trainees to (Ellis, 1965): i) acquire a set of specific skills, procedures, and actions; ii) generalize these skills – i.e., ability to respond/use skills in the same way to similar stimuli and, iii) use these skills among similar application fields i.e., skills transfer, intended as the use of previous/trained skills to learn new skills or to perform under similar conditions a set of skills.

Design and assessment of training programmes for an organization is therefore a complex process in which experts have to balance different factors (budget, tools, people needs etc.) with the company needs. In tune with Arthur et al. (2003) the trade-off among all the instances and needs drives experts to define the main components of a programme:

a) Objectives – e.g., train one or more new procedure, or train procedure on a new product model etc.;
b) Contents of the training – i.e., skills and procedures that people have to acquire in tune with the objectives of the training;
c) Training method – i.e., how training contents are delivered and experience by trainees, such as classroom, on-job training, VR/MR tools;
d) Set of evaluation criteria and expected outcomes – i.e., measure how much the applied methods match the training objectives.

These components can be manipulated by practitioners in order to optimize the transfer of training contents – i.e., skills and procedures – and to select the evaluation criteria for matching the expected outcomes – i.e., requested by company – with the training programme objectives.

Figure 2 exemplifies the experts' decision making (i.e., manipulation of a,b,c and d as proposed by:Arthur Jr et al., 2003) that could affect the outcomes of a training programme.

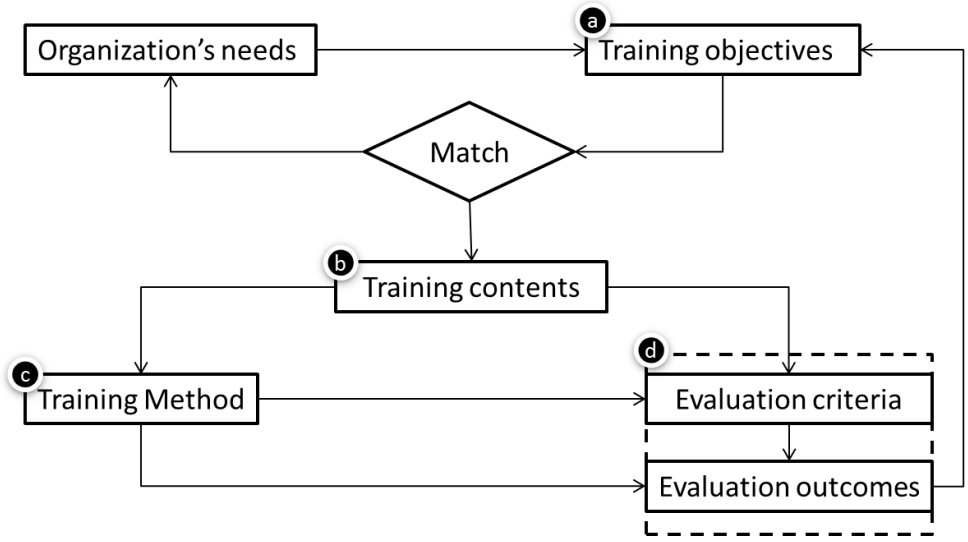


Figure 2. Flowchart of the practitioners' decision making in design a training programme (originally created on the basis of: Arthur Jr et al., 2003)

The effectiveness of a VR/MR tool – i.e., training method – in accordance with Mantovani (2003) varies in line with the concepts that trainees have to learn. For instance, when the training scope (point a, in figure 2) is to deliver a specialized content (point b, in figure 2), such as a specific and standardized surgery procedure, they could select as training method (point c, in figure 2) a VR/MR system with a high level of physical fidelity (Seymour et al., 2002; Yuviler-Gavish, Gutierrez, Webel, Rodriguez, & Tecchia, 2011). In these cases, surgeons by simulating procedures through a VR/MR tool which reproduces (with different degrees of accuracy) the real world environment, tasks and actions are driven to learn a specific set of skills and actions that they could directly transfer and use for performing in the real world. On the other hand, when the training contents are more variable in terms of operations, as in the case of service maintenance processes applied on different models of cars or for training programme that aims to teach basic and advanced versions of the same procedure (Gallagher et al., 2005), VR/MR tools which drive trainees to perform a specific (i.e., replicated) procedure could not exhaustively help operators to face the variability of real world. In these scenarios, practitioners could prefer to use for training cognitive fidelity VR/MR systems, instead of physical fidelity tools. In fact these kinds of systems train operators to perform psychomotor and cognitive activities that are similar to the ones involved in the possible variations of the real-world tasks, by aiming to increase trainees abilities to: i) recognize in real world similarities and relationships among the stimuli –i.e., the different procedures –, and ii) use the learned skills to perform in a correct way the actions for achieving the goal – i.e., response to the stimuli (Hochmitz & Yuviler-Gavish, 2011). In tune with that, cognitive fidelity systems drive trainees to learn skills and actions that they could generalize in real world to perform procedures that are similar to the ones experienced during the training (Gallagher et al., 2005; Yuviler-Gavish, Gutierrez, et al., 2011).

Therefore, the aims and contents of training, as well as the selection of one VR/MR tool over another, significantly affect the researchers' choice of evaluation criteria (point d, in figure 2). For instance, by looking to train a surgery procedure in which specific contemporary, and accurate, hands movements is one of the core factors to achieve the overall goal – i.e., perform a safe and accurate procedure – researchers could use the analysis of the economy of movements to estimate how much surgeons increased their proficiency and accuracy after different iterations of training (Grantcharov et al., 2003). Concurrently, among the criteria for assessing the effectiveness of training operators of car service procedures researchers, for instance, could decide to measure

the economy of movement. However, in contrast to the surgery field, the accuracy of car service maintenance procedures is not directly affected by the number of unnecessary actions and movements of an operator. Therefore, in the context of automotive, and for the specific aims and contents of car service maintenance training, the criteria of economy of movement can be used mostly as a measure of operators' proficiency (i.e., efficiency), but it is not directly related to the efficacy of the performance..

Finally, as Mantovani (2003) suggested, training programme aims and organization needs, composed the rationale of the VR/MR tools effectiveness assessment by defining the contents and the kind of VR/MR tools applied for the training. Therefore, to measure the efficacy of the transfer of training contents mediated through a VR/MR tool, researchers have to: i) select a set of criteria in line with the training aims and contents, and ii) observe the relationship among the selected variables before, during and after the training in order to analyse the distance between the training outcomes and the overall training programme aims.

3.1. Effectiveness criteria in training of service maintenance

In the literature, researchers have proposed different qualitative and quantitative evaluation criteria to estimate the effectiveness of training approaches and tools (Arthur Jr et al., 2003; Kirkpatrick, 1994; Mantovani, 2003; Roussos et al., 1999; Salas & Cannon-Bowers, 2001; Salas et al., 2012; Taylor et al., 2005). A basic set of evaluation criteria, that researchers usually observe in their experimental analyses of VR/MR effectiveness, is composed by the following two components: i) the demographic information of the participants – e.g., sex, age etc. – and, ii) trainees observable reactions to the training contents manipulation such as: time of performance for each task of the procedure, number of unsolved and recovered errors, time for error recovery, number of tasks without errors etc. (see, for instance: Grantcharov et al., 2003; Haque & Srinivasan, 2006; Seymour et al., 2002; Webel et al., 2013; Xu & Gao, 2011). However, associated to this basic set of criteria, several variables that could affect the interaction of, as well as the training with VR/MR tools, were discussed in literature. To identify a list of effectiveness evaluation criteria, we extrapolated from our database of 44 experimental and review studies (see, above in section 2) a list of nine criteria that researchers have tested, or referenced as important for the effectiveness assessment. Table 2 provides an overview of these eleven criteria, by also associating to them a list of highly cited studies – at least 100 citations as shown by Google Scholar – in which these criteria were proposed or discussed.

	Evaluation Criteria	**References** (Authors, Years – *Number of citations*)
1.	Cognitive skills - e.g. visuospatial abilities etc.	Seymour et al., 2002 – *1384*
2.	Levels of trust/acceptance of VR/MR tools	Gallagher et al. 2005 – *425*
3.	Motivation in use	Grantcharov, et al. 2003 – *236*
4.	Participants attitude	Kneebone et al. 2004 – *199*
5.	Previous experience	Stefanidis. Et al. 2005 – *119*
6.	Cybersickness	Kennedy et al., 1993 – *1050* Stanney, Mourant and Kennedy, 1998 – *371*
7.	Physiological Reactions – e.g., attention shift, cognitive load, stress	Witmer and Singer, 1998 – *1927*
8.	Level of presence and engagement	Sanchez-Vives and Slater, 2005 – *387*
9.	Technical aspects and tools features – e.g. Effect of designed features, expected and experienced system functioning	Bowman, Gabbard, and Hix, 2002 – *229*

Table 2. List of evaluation criteria reported in literature as important for testing the effectiveness of VR/MR tools interaction and training

To discuss whether and how researchers applied these criteria in the eight studies on VR/MR training for training service maintenance processes, we divide the criteria in three main groups: pre-training, performance and post-training.

Moreover, together with these evaluation criteria, we also checked whether researchers analysed two other aspects that are considered important in literature for the effectiveness assessment. The first is the analysis of the match between training programme aims and VR/MR tool training outcomes. In fact, as we said above although it is not a very cited assessment aspect in VR/MR studies, the analysis of match is considered by researches of classic training assessment a core factor for estimating the effectiveness of training programme (Anastassova & Burkhardt, 2009; Anastassova et al., 2005; Arthur Jr et al., 2003; Kirkpatrick, 1994; Salas & Cannon-Bowers, 2001; Salas et al., 2012). The second is the variability of experimental conditions tested by researchers for assessing the VR/MR effectiveness. In fact, researchers agree that simple demonstrations of tools functioning carried out under proof-of-concept approaches are not enough for exhaustively explaining the potential and the effectiveness of VR/MR tools as medium of training content delivery (Tang et al., 2003). Therefore, for practitioners it is more important to increase reproducibility and reliability of VR/MR tools analysis by designing experiments that aim to: a) test trainees experience with different levels and variations of tasks complexity during the training procedure and, b) compare the outcomes of training, delivered to operators through different VR/MR systems – i.e., features, functioning etc.

3.1.1. Pre-training criteria

This group includes criteria commonly used by researchers for clustering the participants' performances during training with a VR/MR tool (see, for instance: Haque & Srinivasan, 2006; Hart & Karthigasu, 2007; Seymour et al., 2002), such as: i) users' demographic information ii) their previous experience iii) participants' attitude – e.g., participants' learning styles – motivation in use, and trust/acceptance of technology and, iv) cognitive skills – e.g., participants' visuospatial and perceptual abilities. Moreover, among these variables researchers must also control participants' previous experiences of cybersickness (Kennedy, Lane, Berbaum, & Lilienthal, 1993). In fact, especially virtual and immersive tools induce cybersickness symptoms, reducing the effectiveness of the training experience (Sharples, Cobb, Moody, & Wilson, 2008). Therefore, researchers may exclude participants, with questionnaires before training, by checking the cybersickness tendency of trainees, and have to control, after the training, the occurrence of cybersickness among the participants to estimate the negative effect of training tools (Stanney, Mourant, & Kennedy, 1998).

Other criteria, proposed in literature, pertain to analysis of training tools technical aspects (e.g., system functioning, usability, user experience and satisfaction in use. see: Mantovani, 2003; Mantovani et al., 2003) measured by questionnaires before – e.g., pre-training interaction with the technology – and after the training. In fact, different levels of tool functioning – e.g., a low level of tool usability – may significantly affect the transfer of training contents by compromising the user experience (ISO, 2010) during the interaction with the tool. Concurrently, features such as different levels of training gamification can improve or decrease trainees' motivation, presence and engagement (Kapp, 2012; Yamabe & Nakajima, 2013). In light of that, a pre and post-training analysis of these criteria helps researchers to fully manage and observe the effect of the tools for delivering contents on trainees' performances.

As Table 2 shows, several differences can be underlined in terms of pre-assessment analysis among the current studies of VR/MR tools for training service operators. Although all the authors control the age and the sex of the trainees, in only six studies is trainees' expertise considered an important variable for clustering trainees' performance during and after the training. However, researchers collect data on operators' expertise in different ways. In some cases the trainees declared their expertise as part of demographic data, in other cases users' expertise is estimated through a pre-training exercise, or by a set of closed-ended questions administered to the participants before and after the training.

Finally, current experimental studies in service maintenance training do not include among the evaluation criteria: the observation of trainees' attitude and cognitive skills, their motivation and trust/acceptance, as such as the analysis of the system functioning perceived by the participants. However, in at least three studies, the perception of effectiveness in performing the task by the tool (Dawei et al., 2009) and the ease of use (Yuviler-Gavish et al., 2013; Yuviler-Gavish, Yechiam, et al., 2011) are assessed after the training (see Table 4) as an overall measure of trainees experience.

Authors, Years	**Pre-training criteria**		
	Demographic information	Expertise declared	Expertise estimated
Anastassova et al., 2005	**X**	**X**	
Anastassova and Burkhardt, 2009	**X**	**X**	
Dawei, Bhatti, and Nahavandi, 2009	**X**		
Abate et al. 2009	**X**		
Yuviler-Gavish, Yechiam, and Kallai, 2011	**X**		**X**
Yuviler-Gavish et al., 2011	**X**	**X**	
Webel et al., 2013	**X**	**X**	**X**
Yuviler-Gavish et al., 2013	**X**		**X**

Table 2. Pre- training criteria observed in the current studies on service maintenance.

3.1.2. Performance criteria

Researchers commonly collect a wide range of data about trainees' performances during the training, with a primary focus on variables such as: i) time of performances for each task of the procedure, ii) number of unsolved and recovered errors, iii) time for error recovery, and iv) number of tasks without errors (see, for instance: Grantcharov et al., 2003; Haque & Srinivasan, 2006; Seymour et al., 2002; Webel et al., 2013; Xu & Gao, 2011). These criteria aim to measure people' reactions to the delivered contents (Arthur Jr et al., 2003; Kirkpatrick, 1994) in order to define the trainees' learning curve after each training iteration (Grantcharov et al., 2003), and to compare the improvement of the users' ability under different training conditions, such as: different levels of task complexity or with different VR/MR tools (Seymour et al., 2002).

Although the primary interest of researchers is focused on trainees' observable reactions to contents, measured in terms of time and errors, several authors also proposed to measure physiological reactions of trainees during the performances by gathering data, for instance, through eye-tracking tools – to measure the participants' attentional focus during the procedure (Law, Atkins, Kirkpatrick, & Lomax, 2004; Yeh & Wickens, 2000) – or through systems for heart rate analysis – to estimate the cognitive load in performing tasks (Sanchez-Vives & Slater, 2005; Zimmons & Panter, 2003). Moreover, at least two other additional criteria for measuring the improved abilities of the trainees were proposed in literature of VR/MR studies for training. The first one is the measurement of trainees' ability to recall the acquired skills and actions for a correct performance, such as, the economy of movements, and the accuracy of posture and actions measured in terms of time and accuracy within the training sessions (Ebert, Deller, Steffen, & Heintz, 2009; Grantcharov et al., 2003; Kneebone, Scott, Darzi, &

Horrocks, 2004; Stone et al., 2011). The second criterion is the skill-decay (Arthur Jr, Bennett Jr, Stanush, & McNelly, 1998; Arthur, Day, & Bennett, 1997; Hall, Stiles, & Horwitz, 1998; Stefanidis et al., 2005). This variable can be measured as the latency of time in recall of the trained skill after a period of non-use. Usually this criterion is applied as a post-training measure of effectiveness, when operators perform in real world the learned procedure after a long period from the training experience. However the trainees' levels of skill decay can be also measured after different iterations of training to assess the efficacy of VR/MR tool of minimizing the loss of trained skills (Hall et al., 1998).

As table 3 shows in our list all the researchers analysed the training of service operators by using the criteria of reactions to contents, while in only one case physiological measures were gathered for additional information (Yuviler-Gavish, Yechiam, et al., 2011). Moreover, a minimal set of different experimental conditions are considered by researchers. In fact, in just one study trainees are exposed to tasks with different levels of complexity (Dawei et al., 2009), and only in three experimental analyses researchers vary the conditions under which the VR/MR tools are used by trainees, or the VR/MR tools used by trainees for the training (Webel et al., 2013; Yuviler-Gavish, Gutierrez Seco, et al., 2011; Yuviler-Gavish et al., 2013). Finally, none of the authors considered among the evaluation criteria the skills recall and decay.

Authors, Years	**Performance criteria**		**Experimental conditions**	
	Reactions in terms of time and errors	Physiological reactions	Task complexity	Comparison among tools/ training conditions
Anastassova et al., 2005	**X**			
Anastassova and Burkhardt, 2009	**X**			
Dawei, Bhatti, and Nahavandi, 2009	**X**		**X**	
Abate et al. 2009	**X**			
Yuviler-Gavish, Yechiam, and Kallai, 2011	**X**			**X**
Yuviler-Gavish et al., 2011	**X**	**X**		**X**
Webel et al., 2013	**X**			**X**
Yuviler-Gavish et al., 2013	**X**			**X**

Table 3. Performance criteria and the experimental conditions observed in the current studies on service maintenance.

3.1.3. Post-test criteria

Post-test criteria are used by practitioners to match the outcomes with the training aims, thus showing the return for the company in the use of a training tool and estimating future developments of tools (Arthur Jr et al., 2003; Kirkpatrick, 1994; Mantovani, 2003; Stedmon & Stone, 2001; Stone et al., 2011). Researchers include among these criteria the perceived effectiveness of the training, the trainees' embodiment, engagement and presence, usually assessed by questionnaires. Moreover, as mentioned above, data about the ease of tool use, cybersickness, skills recall and decay could also be assessed also after the training.

As Table 4 shows, researchers of service maintenance training rarely explored the match between the VR/MR tools effectiveness outcomes and training programme aims and objectives under which the systems were used for delivering the know-how. In fact, only in two articles authors attempted to establish this connection and to explore the lack and the possible developments of the training tools in line with the company needs (Anastassova & Burkhardt, 2009; Anastassova et al., 2005). Moreover, researchers largely applied qualitative scales to check

the users' perceived effectiveness of training, and the perceived VR/MR tools easiness of use, but none of the articles analysed the occurrence of cybersickness, and the participants' levels of skill recall or decay.

Authors, Years	**Post-training criteria**			
	Declared match with training aims	Perceived effectiveness of training	Presence, engagement	Experience of system functioning and features
Anastassova et al., 2005	**X**			
Anastassova and Burkhardt, 2009	**X**			
Dawei, Bhatti, and Nahavandi, 2009		**X**	**X**	**X**
Abate et al. 2009		**X**		**X**
Yuviler-Gavish, Yechiam, and Kallai, 2011		**X**		
Yuviler-Gavish et al., 2011		**X**		**X**
Webel et al., 2013		**X**		**X**
Yuviler-Gavish et al., 2013		**X**		**X**

Table 4. Post-training criteria observed in the current studies on service maintenance

4. Current results and limits of training operators of service maintenance

The current results of experimental studies on the effectiveness of VR/MR tools for training service maintenance operators can be summarized as follows:

- Researchers have considered MR systems more useful tools for training assembly and disassembly processes compared to VR systems, though a limited set of evidence supports this idea (Dawei et al., 2009; Webel et al., 2013; Yuviler-Gavish, Gutierrez Seco, et al., 2011; Yuviler-Gavish et al., 2013; Yuviler-Gavish, Yechiam, et al., 2011).
- Training on service maintenance with VR/MR tools results in a low number of unsolved errors and training time compared to traditional training (Webel et al., 2013; Yuviler-Gavish, Gutierrez Seco, et al., 2011).
- Trainees usually perceived the training experience with VR/MR tools as more positive than classic training because these tools help them to increase the understanding of tasks and procedures (Dawei et al., 2009).
- VR/MR tools can be used with different levels of effectiveness on the basis of the operators' expertise. For instance, visual information could accelerate the process of skills acquisition of experts' operators. However, too much stimulation could impair less expert trainees' performance when used during on-job-the trainings, or in real world procedures(Yuviler-Gavish, Yechiam, et al., 2011).

Our analysis reveals two main limitations of the current studies on service maintenance training with VR/MR tools. First, researchers had a very low interest in the match between the training programmes' aims and the training outcomes. Therefore, current results are only focused on the improvement of operators' performances, without any considerations of the companies' training programmes, aims and contexts, and with limited usefulness for informing the adaptation of the training tools for the specific organizations' needs (Anastassova & Burkhardt, 2009; Anastassova et al., 2005; Arthur Jr et al., 2003; Kirkpatrick, 1994). Second, researchers applied a reliable, but very limited, set of evaluation criteria for testing the effectiveness of training tools compared to the experimental approaches usually applied in other fields (e.g., see:Grantcharov et al., 2003; Haque &

Srinivasan, 2006; Hart & Karthigasu, 2007; Seymour et al., 2002). Associated with these limits, a lack of uniformity among the evaluation criteria and experimental approaches emerged from our analysis of the literature. In fact, researchers are so concentrated in the assessment of time and errors of trainees' performances by aiming to show the effectiveness of their experimental platforms of training, to leave broadly unconsidered several factors, such as: i) cybersickness, ii) skill recall and decay, iii) trainees' levels of motivation, acceptance and trust of technologies, iv) the users' previous attitude and cognitive skills. Moreover, some evaluation criteria are minimally considered, though only as post-training factors (Table 4), such as: presence, engagement, perceived training effectiveness and tools ease of use. However, researchers often measured these criteria by using invalidated qualitative measures, instead of standardized instruments – e.g., Presence questionnaire (Witmer & Singer, 1998) or System Usability Scale (Borsci, Federici, & Lauriola, 2009).

Finally, although researchers recently started to propose comparative analyses of different kinds of VR/MR training tools (see Table 3), currently the authors did not included in their experimental paradigms a set of evaluation criteria to go beyond the simple assessment of performances based on time and errors estimation. Usually these comparative analyses presented only a minimal variation of both the experimental conditions and the features of the tools for training.

5. Future challenges on virtual and mixed reality tool for training car service operators

The effectiveness of VR/MR tools for training manufacturing and post-manufacturing activities is an ongoing discussion in which experts of different disciplines – e.g., engineers, training and learning experts, psychologists and human factors practitioners – will be involved, in the next years, to define a common framework for designing and assessing these kind of training systems. In fact, the lack of a common set of rules, as we claimed above, lead each researcher to marginally consider the relationship among VR/MR tools, training aims and needs of car-makers, and to explore the training carried out with these tools mostly in terms of time and errors. Therefore, the main challenge for the community of VR/MR researchers will be the increase of experimental reliability, and the definition of even more sophisticated and accurate criteria for the effectiveness assessment.

Researchers have already started to discuss the definition of a set of design and assessment guidelines of VR/MR training systems (Yuviler-Gavish, Gutierrez, et al., 2011). Recently, researchers, seeking for a common framework, started to increase the comparative analysis of the tools (Webel et al., 2013; Yuviler-Gavish, Gutierrez Seco, et al., 2011; Yuviler-Gavish et al., 2013). Concurrently, practitioners attempted to explore the characteristics of MR/VR tools in order to define the most important, current and future, functionalities that could positively affect the training effectiveness of these systems, such as: portability, wearability, usability, and adaptability of system to the trainees' characteristics (Anastassova & Burkhardt, 2009; Anastassova et al., 2005; Dawei et al., 2009).

Despite these efforts, as our analysis of the current experiments showed, in the field of automotive training of service operators there are still several unexplored aspects about the training experience with VR/MR tools, such as, for instance: i) the assessment of pre and post training operators understanding, recognition and retention (skill recall and loss) of product components and actions; ii) the effect of operators' cognitive, spatial and visual abilities during the acquisition of skills; iii) the role of personal, technical and safety aspects that could prevent or improve the ability of trainees to acquire skills during the VR/MR training – e.g., motivation, attention, presence, usability, satisfaction, cibersickeness etc.

These uncharted aspects leave open several questions about the VR/MR tools effectiveness for training: Which are the short and long term effects of these tools on trainees' performances, skills and abilities? Are these effects comparable to the ones of classic training approach and tools? Which kind of tool (VR or MR) is more effective to train specific tasks and procedures? When is it more convenient to use a VR or a MR tool, compared to other, more classic, training approaches?

Although the increase and the extension of experimental analysis of effectiveness is a common aim for the scientific community to remove any space of uncertainty about the training effect of VR/MR tools, for the specific field of training car service operators, researchers shall face at least other two important challenges in the next decade.

The first challenge, in line with Anastassova and colleagues (2005), will be the widening of experimental focus from the training of practical and sequential operations to reach the goal of a service procedure – i.e., disassembly, repair and reassembly – to the application of VR/MR tools for training and supporting operators in the decision-making process before the service: i) identification of car problems, and ii) definition of a correct strategy of service maintenance – i.e., the diagnostic process. In fact, there is a growing request from manufacturers to develop and control the effect of VR/MR tools to help operators to acquire diagnostic skills for avoiding incorrect fault diagnosis, as well as a non-identification of product issues or a miscomprehension of problems (Liang, 2010). Currently there are no experimental studies that have explored the effectiveness of VR/MR tools for training diagnostic reasoning of car service operators and their strategies of maintenance.

The second challenge that experts have to face in the next years will be the design and assessment of tools for training car service maintenance process in tune with emerging approaches, such as serious games and gamification (Freitas & Liarokapis, 2011). Currently, these approaches are marginally considered in automotive studies, although several experimental evidences showed how the introduction of gamified elements and scenarios in VR/MR training systems – e.g., for military procedures –increases the engagement of trainees, and force them to learn skills and strategies for a quick and effective reactions to the stimuli (Zyda, 2005). Gamification is still an unexplored element in service maintenance that may play a significant role to depict the tools effectiveness and, much more, to reveal new potentialities of VR/MR tools for training operators.

Acknowledgements

This paper was completed as part of Live Augmented Reality Training Environments (LARTE) – 101509 project. The authors would like to acknowledge the Technology Strategy Board for funding the work.

References

Abate, A. F., Guida, M., Leoncini, P., Nappi, M., & Ricciardi, S. (2009). A haptic-based approach to virtual training for aerospace industry. *Journal of Visual Languages & Computing, 20*(5), 318-325. doi: http://dx.doi.org/10.1016/j.jvlc.2009.07.003

Adam, J. A. (1993). Virtual reality is for real. *Spectrum, IEEE, 30*(10), 22-29. doi: 10.1109/6.237580

Anastassova, M., & Burkhardt, J.-M. (2009). Automotive technicians' training as a community-of-practice: Implications for the design of an augmented reality teaching aid. *Applied Ergonomics, 40*(4), 713-721. doi: http://dx.doi.org/10.1016/j.apergo.2008.06.008

Anastassova, M., Burkhardt, J.-M., Mégard, C., & Ehanno, P. (2005). Results from a user-centred critical incidents study for guiding future implementation of augmented reality in automotive maintenance. *International Journal of Industrial Ergonomics, 35*(1), 67-77. doi: http://dx.doi.org/10.1016/j.ergon.2004.08.005

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369-406. doi: 10.1037/0033-295x.89.4.369

Arthur Jr, W., Bennett Jr, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of Training in Organizations: A Meta-Analysis of Design and Evaluation Features. [Article]. *Journal of Applied Psychology, 88*(2), 234-245. doi: 10.1037/0021-9010.88.2.234

Arthur Jr, W., Bennett Jr, W., Stanush, P. L., & McNelly, T. L. (1998). Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. *Human Performance, 11*(1), 57-101. doi: 10.1207/s15327043hup1101_3

Arthur, J. W., Day, E. A., & Bennett, J. W. (1997). Dyadic Versus Individual Training Protocols: Loss and Reacquisition of a Complex Skill. [Article]. *Journal of Applied Psychology, 82*(5), 783-791.

Belardinelli, C., Blümel, E., Müller, G., & Schenk, M. (2008). Making the virtual more real: research at the Fraunhofer IFF Virtual Development and Training Centre. *Cognitive Processing, 9*(3), 217-224. doi: 10.1007/s10339-008-0216-0

Borsci, S., Federici, S., & Lauriola, M. (2009). On the dimensionality of the System Usability Scale: a test of alternative measurement models. *Cognitive Processing, 10*(3), 193-197. doi: 10.1007/s10339-009-0268-9

Bowman, D. A., Gabbard, J. L., & Hix, D. (2002). A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators and Virtual Environments, 11*(4), 404-424. doi: 10.1162/105474602760204309

Bruner, J. (1966). *Toward a Theory of instruction*. Cambridge, MA: Belknap Press of Harvard University Press.

Campbell, J. P. (1971). Personnel Training and Development. *Annual Review of Psychology, 22*(1), 565-602. doi: doi:10.1146/annurev.ps.22.020171.003025

Dawei, J., Bhatti, A., & Nahavandi, S. (2009, 7-8 Nov. 2009). *Design and evaluation of a haptically enable virtual environment for object assembly training*. Paper presented at the Haptic Audio visual Environments and Games, 2009. HAVE 2009. IEEE International Workshop on.

De Freitas, S., Rebolledo-Mendez, G., Liarokapis, F., Magoulas, G., & Poulovassilis, A. (2010). Learning as immersive experiences: Using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world. *British Journal of Educational Technology, 41*(1), 69-85. doi: 10.1111/j.1467-8535.2009.01024.x

Dombrowski, U., Engel, C., & Schulze, S. (2011, 10-12 July 2011). *Changes and challenges in the after sales service due to the electric mobility.* Paper presented at the Service Operations, Logistics, and Informatics (SOLI), 2011 IEEE International Conference on.

Ebert, A., Deller, M., Steffen, D., & Heintz, M. (2009). *"Where did I put that?" Effectiveness of kinesthetic memory in immersive virtual environments.* Paper presented at the 13th International Conference on Human-Computer Interaction, San Diego, CA.

Ellis, H. C. (1965). *The Transfer of Learning*. New York: The Macmillan Company.

Fitts, P. M., & Posner, M. I. (1967). *Human Performance*. Belmont, CA: Brooks/Cole Publishing.

Fong, K., Chow, K., Chan, B., Lam, K., Lee, J., Li, T., . . . Wong, A. (2010). Usability of a virtual reality environment simulating an automated teller machine for assessing and training persons with acquired brain injury. *Journal of NeuroEngineering and Rehabilitation, 7*(1), 1-9. doi: 10.1186/1743-0003-7-19

Freitas, S., & Liarokapis, F. (2011). Serious Games: A New Paradigm for Education? In M. Ma, A. Oikonomou & L. C. Jain (Eds.), *Serious Games and Edutainment Applications* (pp. 9-23): Springer London.

Gaiardelli, P., Resta, B., Martinez, V., Pinto, R., & Albores, P. (2014). A classification model for product-service offerings. *Journal of Cleaner Production, 66*(0), 507-519. doi: http://dx.doi.org/10.1016/j.jclepro.2013.11.032

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., . . . Satava, R. M. (2005). Virtual Reality Simulation for the Operating Room: Proficiency-Based Training as a Paradigm Shift in Surgical Skills Training. *Annals of Surgery, 241*(2), 364-372.

Gomes de Sá, A., & Zachmann, G. (1999). Virtual reality as a tool for verification of assembly and maintenance processes. *Computers & Graphics, 23*(3), 389-403. doi: http://dx.doi.org/10.1016/S0097-8493(99)00047-3

Grantcharov, T. P., Bardram, L., Funch-Jensen, P., & Rosenberg, J. (2003). Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. *The American Journal of Surgery, 185*(2), 146-149. doi: http://dx.doi.org/10.1016/S0002-9610(02)01213-8

Hall, C. R., Stiles, R. J., & Horwitz, C. D. (1998, 18-18 1998). *Virtual reality for training: evaluating knowledge retention.* Paper presented at the Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998.

Haque, S., & Srinivasan, S. (2006). A meta-analysis of the training effectiveness of virtual reality surgical simulators. *Information Technology in Biomedicine, IEEE Transactions on, 10*(1), 51-58. doi: 10.1109/titb.2005.855529

Hart, R., & Karthigasu, K. (2007). The benefits of virtual reality simulator training for laparoscopic surgery. *Current Opinion in Obstetrics & Gynecology, 19*(4), 297-302. doi: 10.1097/GCO.0b013e328216f5b7

Hochmitz, I., & Yuviler-Gavish, N. (2011). Physical Fidelity Versus Cognitive Fidelity Training in Procedural Skills Acquisition. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53*(5), 489-501. doi: 10.1177/0018720811412777

ISO. (2010). ISO 9241-210:2010 Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems. Brussels, BE: CEN.

Kapp, K. M. (2012). *The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education*. San Francisco, CA: Pfeier.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology, 3*(3), 203-220. doi: 10.1207/s15327108ijap0303_3

Kirkpatrick, D. L. (1994). *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler.

Kneebone, R. L., Scott, W., Darzi, A., & Horrocks, M. (2004). Simulation and clinical practice: strengthening the relationship. *Medical Education, 38*(10), 1095-1102. doi: 10.1111/j.1365-2929.2004.01959.x

Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (2004). *Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment*. Paper presented at the Proceedings of the 2004 symposium on Eye tracking research \& applications, San Antonio, Texas.

Li, J. R., Khoo, L. P., & Tor, S. B. (2003). Desktop virtual reality for maintenance training: an object oriented prototype system (V-REALISM). *Computers in Industry, 52*(2), 109-125. doi: http://dx.doi.org/10.1016/S0166-3615(03)00103-9

Liang, J. S. (2010). Design and implement a virtual learning architecture for troubleshooting practice of automotive chassis. *Computer Applications in Engineering Education, 18*(3), 512-525. doi: 10.1002/cae.20219

Mantovani, F. (2003). VR learning: Potential and challenges for the use of 3D environments in education and training. In G. Riva & C. Galimberti (Eds.), *Towards cyberpsychology: Mind, cognitions and society in the internet age* (pp. 207-226). Amsterdam: IOS Press.

Mantovani, F., Castelnuovo, G., Gaggioli, A., & Riva, G. (2003). Virtual Reality Training for Health-Care Professionals. [Article]. *CyberPsychology & Behavior, 6*(4), 389. doi: 10.1089/109493103322278772

Michalos, G., Makris, S., Papakostas, N., Mourtzis, D., & Chryssolouris, G. (2010). Automotive assembly technologies review: challenges and outlook for a flexible and adaptive approach. *CIRP Journal of Manufacturing Science and Technology, 2*(2), 81-91. doi: http://dx.doi.org/10.1016/j.cirpj.2009.12.001

Mikropoulos, T., Chalkidis, A., Katsikis, A., & Kossivaki, P. (1997). Virtual realities in environmental education: the project LAKE. *Education and Information Technologies, 2*(2), 131-142. doi: 10.1023/a:1018648810609

Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems, E77-D*(12), 1321-1329.

Mintel Group Ltd. (2013). Car Service: Maintenance and Repair - US - April 2013 Retrieved 2014, 16 February, from http://store.mintel.com/car-service-maintenance-and-repair-us-april-2013?cookie_test=true)

Moskaliuk, J., Bertram, J., & Cress, U. (2012). Training in virtual environments: putting theory into practice. *Ergonomics, 56*(2), 195-204. doi: 10.1080/00140139.2012.745623

Neumann, U., & Majoros, A. (1998, 18-18 1998). *Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance.* Paper presented at the Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998.

Robler, A. (1998). *Trends In Virtual Reality*. Paper presented at the Virtual Reality 1998, Kuala Lumpu.

Roussos, M., Johnson, A., Moher, T., Leigh, J., Vasilakis, C., & Barnes, C. (1999). Learning and Building Together in an Immersive Virtual World. *Presence: Teleoperators and Virtual Environments, 8*(3), 247-263. doi: 10.1162/105474699566215

Salas, E., & Cannon-Bowers, J. A. (2001). The Science of Training: A Decade of Progress. *Annual Review of Psychology, 52*(1), 471-499. doi: doi:10.1146/annurev.psych.52.1.471

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The Science of Training and Development in Organizations: What Matters in Practice. *Psychological Science in the Public Interest, 13*(2), 74–101. doi: 10.1177/1529100612436661

Salzman, M. C., Dede, C., Loftin, R. B., & Chen, J. (1999). A Model for Understanding How Virtual Reality Aids Complex Conceptual Learning. *Presence: Teleoperators and Virtual Environments, 8*(3), 293-316. doi: 10.1162/105474699566242

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience, 6*(4), 332-339. doi: http://dx.doi.org/10.1038/nrn1651

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual Reality Training Improves Operating Room Performance: Results of a Randomized, Double-Blinded Study. *Annals of Surgery, 236*(4), 458-464.

Sharples, S., Cobb, S., Moody, A., & Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays, 29*(2), 58-69. doi: http://dx.doi.org/10.1016/j.displa.2007.09.005

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human Factors Issues in Virtual Environments: A Review of the Literature. *Presence: Teleoper. Virtual Environ., 7*(4), 327-351. doi: 10.1162/105474698565767

Stedmon, A. W., & Stone, R. J. (2001). Re-viewing reality: human factors of synthetic training environments. *International Journal of Human-Computer Studies, 55*(4), 675-698. doi: http://dx.doi.org/10.1006/ijhc.2001.0498

Stefanidis, D., Korndorffer Jr, J. R., Sierra, R., Touchard, C., Dunne, J. B., & Scott, D. J. (2005). Skill retention following proficiency-based laparoscopic simulator training. *Surgery, 138*(2), 165-170. doi: http://dx.doi.org/10.1016/j.surg.2005.06.002

Stone, R. T., Watts, K. P., Zhong, P., & Wei, C.-S. (2011). Physical and Cognitive Effects of Virtual Reality Integrated Training. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53*(5), 558-572. doi: 10.1177/0018720811413389

Stork, A., Sevilmis, N., Weber, D., Gorecky, D., Stahl, C., Loskyll, M., & Michel, F. (2012, 2-5 Sept. 2012). *Enabling virtual assembly training in and beyond the automotive industry.* Paper presented at the Virtual Systems and Multimedia (VSMM), 2012 18th International Conference on.

Sutherland, I. E. (1968). *A head-mounted three dimensional display*. Paper presented at the Proceedings of the December 9-11, 1968, fall joint computer conference, part I, San Francisco, California.

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). *Comparative effectiveness of augmented reality in object assembly*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA.

Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A Meta-Analytic Review of Behavior Modeling Training. [Article]. *Journal of Applied Psychology, 90*(4), 692-709. doi: 10.1037/0021-9010.90.4.692

Tennant, C., Boonkrong, M., & Roberts, P. A. B. (2002). The design of a training programme measurement model. *Journal of European Industrial Training, 26* (5), 230 - 240. doi: 10.1108/03090590210424902

Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., & Preusche, C. (2013). An augmented reality training platform for assembly and maintenance skills. *Robot. Auton. Syst., 61*(4), 398-403. doi: 10.1016/j.robot.2012.09.013

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225-240. doi: 10.1162/105474698565686

Xu, W., & Gao, F. (2011, 20-22 Aug. 2011). *Research of virtual reality technology in automotive engine assembly teaching.* Paper presented at the Information Technology and Artificial Intelligence Conference (ITAIC), 2011 6th IEEE Joint International.

Yamabe, T., & Nakajima, T. (2013). Playful training with augmented reality games: case studies towards reality-oriented system design. *Multimedia Tools and Applications, 62*(1), 259-286. doi: 10.1007/s11042-011-0979-7

Yeh, M., & Wickens, C. D. (2000). Attention and Trust Biases in the Design of Augmented Reality Displays. University of Illinois, Urbana-Champaign, Savoy, IL: Aviation Research Lab.

Yuviler-Gavish, N., Gutierrez Seco, T., Webel, S., Rodriguez, J., Peveri, M., & Bockholt, U. (2011). Transfer of Skills Evaluation for Assembly and Maintenance Training. *BIO Web of Conferences, 1*, 00028.

Yuviler-Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U., & Tecchia, F. (2013). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 1-21. doi: 10.1080/10494820.2013.815221

Yuviler-Gavish, N., Gutierrez, T., Webel, S., Rodriguez, J., & Tecchia, F. (2011). Design Guidelines for the Development of Virtual Reality and Augmented Reality Training Systems for Maintenance and Assembly Tasks. *BIO Web of Conferences, 1*, 00029.

Yuviler-Gavish, N., Yechiam, E., & Kallai, A. (2011). Learning in multimodal training: Visual guidance can be both appealing and disadvantageous in spatial tasks. *International Journal of Human-Computer Studies, 69*(3), 113-122. doi: http://dx.doi.org/10.1016/j.ijhcs.2010.11.005

Zimmons, P., & Panter, A. (2003, 22-26 March 2003). *The influence of rendering quality on presence and task performance in a virtual environment.* Paper presented at the Virtual Reality, 2003. Proceedings. IEEE.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer, 38*(9), 25-32. doi: 10.1109/mc.2005.297