Systematic Review of Augmented Reality Training Systems

Isidro M. Butaslac, III, Yuichiro Fujimoto, *Member, IEEE,* Taishi Sawabe, *Member, IEEE,*
Masayuki Kanbara, *Member, IEEE,* and Hirokazu Kato, *Member, IEEE*

Abstract—Recent augmented reality (AR) advancements have enabled the development of effective training systems, especially in the medical, rehabilitation, and industrial fields. However, it is unclear from the literature what the intrinsic value of AR to training is and how it differs across multiple application fields. In this work, we gathered and reviewed the prototypes and applications geared towards training the intended user's knowledge, skills, and abilities. Specifically, from IEEE Xplore plus other digital libraries, we collected 64 research papers present in high-impact publications about augmented reality training systems (ARTS). All 64 papers were then categorized according to the training method used, and each paper's evaluations were identified by validity. The summary of the results shows trends in the training methods and evaluations that incorporate ARTS in each field. The narrative synthesis illustrates the different implementations of AR for each of the training methods. In addition, examples of the different evaluation types of the current ARTS are described for each of the aforementioned training methods. We also investigated the different training strategies used by the prevailing ARTS. The insights gleaned from this review can suggest standards for designing ARTS regarding training strategy, and recommendations are provided for the implementation and evaluation of future ARTS.

Index Terms—Augmented reality training systems, training method, evaluation, narrative synthesis, systematic review

✦

1 INTRODUCTION

HUMAN skills are acquired through training. Martin et al. defined training as, "the development and delivery of information that people will use after attending it" [1]. As defined, the acquisition of a skill often involves a personal or social necessity. This includes the activity to acquire surgical and diagnostic skills by medical students, the activity to learn how to assemble new products by factory workers, and rehabilitation by hemiplegic patients to use their upper limbs. The mechanisms of skill acquisition through training are closely related to cognitive processes and motor learning theory [2], [3]. Throughout the vast history of research, numerous "training methods" based on these theories have been proposed to effectively acquire skills.

Computer-mediated training uses a computer to aid human training from multiple perspectives, thereby increasing efficiency and lowering costs, such as minimizing training time and providing a more individualized training format [4], [5]. This paper focuses on computer-mediated training that utilizes augmented reality (AR) technology, which can superimpose information on the real world. We refer to systems that provide this type of training as augmented reality training systems (ARTS). Task support, which helps people to efficiently perform specific tasks, is one of the most common applications of AR. However, in the last couple of decades, it has also been gradually used for training. Although the use of AR for training has the potential to bring benefits, it is not straightforward. Many studies have reported its effectiveness in task support systems in helping

people understand and perform certain tasks efficiently by adding the aid content using AR. The user can "rely" on the system in that case. AR task support systems are typically not designed for the user to stop using them in the future. By contrast, in many training contexts, the training system should be designed such that the user can perform the task without the system in the future.

An ideal ARTS should include special techniques and approaches (i.e., training methods) to achieve this goal. We adopted Martin et al.'s classification of training methods and analyzed how each of the existing training methods is extended by the proposed ARTS [1].

Another important aspect of training is the evaluation. Measuring the effectiveness of ARTS training can be more complex than one of the AR task support systems. The reason for this is that it is not possible to directly assess the training effect from the training itself, and it is necessary to check the acquisition of the target skill by the trainee after training. Another reason is that most training requires continuity. The requirement for continuity not only increases the cost of the experiment, but it also creates the need to separate the skill improvement of the user by the continuous training from the skill improvement that is brought by the ARTS. We refer to the evaluation types of Barsom et al. that are widely used, especially for medical systems and applications [6]. As described in Section 2.4.2, this is a five-point framework for the validity of a system to be guaranteed in order for it to reach social implementation.

In this paper, we categorize ARTS technologies that are proposed in various fields (medical, rehabilitation, industrial, etc.) in terms of their training methods [1], their use in AR are summarized, and their trends and characteristics are discussed. In addition, five evaluation types [6] are used to identify trends in the evaluations adopted in the

• *I. Butaslac, Y. Fujimoto, T. Sawabe, M. Kanbara, and H. Kato are with the Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara, 630-0192, Japan. E-mail: {isidro.butaslac.hw2,yfujimoto,t.sawabe,kanbara,kato}@is.naist.jp*

Manuscript received {*place here manuscript details.}

existing ARTS papers for each application field and training method. To the best of our knowledge, there is no paper that discusses ARTS in multiple fields from the viewpoint of the training method and evaluation type. The training methods and evaluation types emphasized in each field potentially differ, but these have not been sufficiently clarified. Furthermore, the intrinsic values of AR to training are difficult to ascertain by focusing on just one specific field; they only become apparent by looking at the similarities and differences across multiple fields. Our main objective is to discuss the intrinsic value of AR in training by enumerating these across disciplines. Our research questions are as follows:

R1   What are the training methods for implementing ARTS and how are these methods used in the different fields?

R2   What types of evaluations are used to assess the existing ARTS and how do they contribute to the validity of each training method?

R3   How is AR utilized for each training strategy?

2 BACKGROUND

2.1 Definition of Training

The definition of training has evolved since the dawn of the industrial revolution. A literature review conducted by Somasundaram et al. indicated that one of the earliest definitions of training was given by Black in 1961: "imparting job knowledge to employees so that they can carry out orders smoothly, efficiently, and cooperatively" [7]. As we approach the 21st century, training according to Goldstein and Ford is defined as the "systematic acquisition of skills, rules, concepts, or attitudes that results in improved performance in another environment" [8].

It can be observed that such a common word used in everyday life has a plethora of interpretations. From this, Somasundaram et al. resolved this ambiguity by analyzing and synthesizing the available definitions of training and development, and listed the dependent variables, area of focus, and core elements of these definitions [9]. Based on the results from their review paper, Somasundaram et al. proposed three major categories for elucidating the purpose of training: develop/gain knowledge, develop/gain skills, and improve performance.

2.2 Definition of Augmented Reality

Around the 1990s, the term "augmented reality" was coined by Thomas Caudell and David Mizell as a technology that is "used to augment the visual field of the user with information necessary in the performance of the task" [10]. Thirty years later, their definition of AR resurfaced as the prevailing phrase used to characterize the modern computing paradigm, and it has changed how people and the world interact.

AR can be considered a variant of virtual environments (VE), otherwise known as virtual reality (VR). VE/VR systems work by completely immersing the user in a digital world. In comparison, AR helps the user to see the real world and it allows virtual objects to be superimposed on or combined with the real world. In this sense, AR expands the reality instead of replacing it altogether. As a result, the user will assume that virtual and real objects can coexist in the same workspace. With this consideration, Azuma established three defining characteristics of AR systems [11]: (1) they combine real and virtual objects; (2) they are interactive in real time; and (3) they are registered in three dimensions.

The definition above by Azuma allows for other systems other than head-mounted displays (HMDs) to be considered as AR. The usage of AR technology techniques, such as computer vision and object recognition, can help the user interact and digitally manipulate the data surrounding the real world. With the help of AR, information about the environment can be overlaid and augmented in the user's real-world view. AR has the capability of improving the perception and interaction between a user and the real world. This is because virtual objects show details that the user cannot perceive directly with their senses. The extra information the virtual objects convey can help the users perform real-world tasks [11].

Owing to AR's multidisciplinary aspect, researchers from various backgrounds have come to define it a bit differently. As Santos et al. pointed out, what is considered to be AR is debatable depending on the quality of the implementation. They cite the example of "imitating the effect of AR... is simulated only by flashing relevant information on a screen. It does not employ any kind of tracking," which is still considered by some scholars as AR [12]. We have decided to include them because they also provide useful insights about the training methods and strategies, even if they are, in the strict definition, not AR. In this loosened definition of AR, we have found in our collected articles that authors tend to use the term Mixed Reality (MR), as long as it satisfies the elements of physical venue, virtual medium, and the user's interactive imagination [13]. Examples of training setups that fall into this category are works that use special input equipment such as fetoscopes [14] or phantom haptic interfaces [15], [16], displayed in a stationary device such as a large monitor.

2.3 Current State of AR Training

2.3.1 AR Training in Practice

As a growing trend in technology, AR is already being used in a variety of fields. In the survey that we conducted, we identified three major fields in which AR is used: medicine, rehabilitation/exercise, and industry.

It is difficult to gain hands-on expertise executing treatments without danger in medicine. AR helps to overcome this barrier by letting medical students understand anatomy and perform procedures. Understanding the interior anatomy of the human being can be a challenging task as it is outside the everyday experience and requires more creative imagination. With the help of AR, the gap between these ideas and operating processes is simplified further [17], [18], [19].

Rehabilitation and exercise in AR can enable more individuals to obtain tailored training at home [20], and perhaps increase adherence and facilitate more regulated execution of physical training sessions [21]. AR in different forms has been used effectively in skill training for doctors/healthcare staff, and it is considered to be appropriate for instruction

that demands strong motor skills and meticulous spatial movement [22].

AR technology significantly benefits industry by allowing the transfer of information from expert to novice to go from just conceptual knowledge to usable hands-on experience in a swift manner [23]. For example, HMDs are trending in industrial AR as they enable the delivery of real-time, step-by-step guidance and feedback from trainers during practice. This technology is ideal for industrial workers since it allows trainees to learn faster and practice more frequently [24].

We may have identified three major categories that are the most common applications for ARTS; however, its usefulness is not limited to only these aforementioned fields. For example, we identified research on ARTS that focuses on general work activities, autonomous driving, or some form of course education such as business ethics or musical instruction. Despite the fact that these are included, they are labeled as "Other" for scoping reasons.

2.3.2 Distinction over Existing Surveys

In the last couple of years, AR has proliferated and a number of survey papers have emerged that describe its trends. One example is a survey that was conducted by Santos et al., in which they performed a meta-analysis of AR learning experiences (ARLE) [12]. They concluded in their study that "there is a need for valid and reliable questionnaires to measure constructs related to ARLEs to iteratively improve ARLE design" [12].

A few years later, in 2020, Bianchi et al. reported that the results of their Systematic Review verified that there is no standard protocol for evaluation, at least in the scope of the medical teaching–learning process content of AR [25]. This sentiment however can be generalized to the broader sense of AR training (i.e., across multiple fields) by doing a preliminary search in the IEEE Xplore search engine using the search string [augment* AND reality AND train* AND review] on all metadata.

There are Systematic Literature Reviews that consider the evaluation methods of AR applications in the educational scenario, however saying that these methods approach are quantitative, qualitative, mixed, or not specified may be too general of a classification to draw concrete conclusions about the training [26]. There are also Literature Reviews about AR usage performance that extracted learnability factors from Kolb's Experiential Learning Theory, and validated these factors through surveying students and academicians [27]. This study by Hafsa et al. is a good example of linking the effectiveness of AR with relation to the user's learning; however, this method of validation is quite costly timewise and moneywise; thus, a "survey" type of validation would prove rather difficult if it is applied for a generalized study scope.

Barsom et al. [6] also checked the validity of AR systems and proposed five validity stages through which ARTS should be assessed to complete a full validation process. However, their work limits the search to AR applications for the purpose of training or educating medical professionals. They focused more on how AR is feasible to the medical field at different stages of validation.

Barsom et al.'s work is different from ours as our motivation is to gain a general understanding of the trends of ARTS research through the categorization by training methods and evaluations. We then use this understanding of the trends to extract the utilization of AR that is useful for the design of training. Hopefully, this will serve as a reference to people who want to develop ARTS but do not know what strategies they should adopt and how they should evaluate it.

In summary, the following are the distinct characteristics of this work: We consider studies that have a high impact on the AR research community. We identify the common application fields in which ARTS is currently used. We categorized ARTS by its training methods and evaluation types.

It is also important to note that this is not the first survey paper to use the study impact as a consideration for the inclusion criteria. For example, Dey et al. [28] used Google Scholar to find the total number of citations of each paper to calculate the average citation count (ACC) per year since it was published. For this paper, we have utilized Google Scholar's h-5 index to indicate the impact of a publication in the last five years. The decision to adopt this strategy rather than the ACC is to address the issue of missing out on recently published research. The trade-off between this and Dey et al.'s method is further elucidated in Sections 3.3 and 3.4.

Another thing to take note is that the screening process for the papers does not include a limiting range for the date of publication. The search process includes the related studies until the day the search was performed. This is because this paper is the first to review ARTS that is not restricted to any application field and it summarizes them by their training methods and evaluation types.

2.4 Training Categorizations

2.4.1 Categorization by Training Method

In a study by Martin et al., they developed an exhaustive list of possible methods that can describe and encapsulate the different types of training systems. This was accomplished by documenting the strategies, techniques, and procedures that are associated with the core process of the sample training systems. From their study, they have determined 13 core training methods that are able to represent any kind of training. These include: a case study, games-based training, internship, job rotation, job shadowing, lecture, mentoring and apprenticeship, programmed instruction, role-modeling, role play, simulation, stimulus-based training, and team-training [1].

However, these results show the core methods of training in a general sense. Since one of the goals of this paper is to define the key characteristics of training systems that is specific only to AR, the authors have narrowed down these 13 core methods. Based on the results of the method that is described in Section 3, each of the 64 studies included for synthesis utilizes one of these training methods: simulation, programmed instruction, games-based training, job shadowing, and mentoring.

Table 1 provides a summary of how Martin et al. defined these five training methods. With regards to job shadowing and mentoring, one could possibly classify these two

TABLE 1
Definition of training methods relevant to AR
(directly taken from Martin et al. [1])

Method	Definition
Simulation	Involves the use of a simulator where specific skills are developed through repeated practice with a multisensory experience of imitated conditions.
Programmed instruction	Involves the delivery of training through instruction that is delivered by a program via some electronic device without the presence of an instructor.
Games-based training	Trainees compete in a series of decision-making tasks which allows them to explore a variety of strategic alternatives and experience the consequences which affect training the other players, but with without risk to the individuals or the organization.
Job shadowing	Involves a trainee closely observing someone perform a specific job in the natural job environment for the purpose of witnessing first-hand the details of the job.
Mentoring	Involves a one-on-one partnership between a novice employee with a senior employee. Mentorship aims to provide support and guidance to less experienced employees.

TABLE 2
Validity types for augmented reality applications (ARA)
(directly taken from Barsom et al. [6])

Validity	Definition
Face	The degree of resemblance between an ARA and the educational construct as assessed by medical experts (referents) and novices (trainees).
Content	The degree to which the ARA content adequately covers the dimensions of the medical content it aims to educate (or is associated with) ("the truth, whole truth, and nothing but the truth").
Construct	Inherent difference in outcome between experts and novices on outcome parameters relevant to the educational construct.
Concurrent	Concordance of subject outcome parameters using tie ARA compared to outcome parameters on an established instrument or method, believed to measure the same educational construct (preferably the golden standard training method).
Predictive	The degree of concordance of ARA outcome parameters and subjects' performance on the educational construct it aims to resemble in reality.

together as both deal with the handing-down of expert performance and skill to the novice. However, it is still more advantageous to distinguish the two as job shadowing shines in the presentation of the desired result, whereas mentoring fosters the mentor–mentee relationship to stimulate the skills and knowledge transfer. Limbu et al. defined usage of an approach similar to that of Martin's, where "demonstration of the task" corresponds to job shadowing, and the "modelling the task with task analysis" corresponds to the mentoring style of training [29].

2.4.2 Categorization by Evaluation Type

In the most general sense, an evaluation is characterized as the process of judging the worth or value of something. In the nomenclature of research design, an evaluation is achievable by using the concepts of measuring constructs. According to Nelson, there are two important dimensions when considering evaluation measurement methods [30].

The first dimension is reliability, which simply refers to the consistency of a measurement. Although reliability is one important aspect in research design, this is not discussed as it is safe to assume that studies that are published in high impact conferences/journals assure the reliability of their results.

The second dimension, validity, is defined as the "extent to which the scores from a measure represent the variable they are intended to," as stated by Chiang et al. [31]. Barsom pointed out that a full validation process is needed for a training system to be ready for implementation to the real-life environment [6]. This full process compromises the face, content, construct, concurrent, and predictive validity. Barsom et al. uses the terminologies of the validity types to judge whether the evaluations used for augmented reality applications are sufficient for training and education. All of these are defined by Barsom et al. and are summarized in

Table 2. Finally, there is also a need to check whether these evaluations are in relation to the user, as the primary goal of a training system is to increase the knowledge, skills, and abilities rather than the promotion of a tool [32].

2.4.3 Categorization by Training Strategy

According to Salas et al., training can be divided into four stages: information, demonstration, practice, and feedback. These stages are applied in terms of the content that is to be learned (i.e., training strategies) [33]. As different contents are mainly learned and trained in each strategy, the manners in which AR is utilized in each strategy is expected to differ.

The first strategy is to convey information to the trainees (i.e., the concepts, facts, and information they need to learn). To obtain complex skills such as surgical skills, it is necessary to learn sufficient information that is relevant to each step in the task prior to practice.

The second training strategy is to demonstrate the desired behavior, cognition, and attitudes to trainees. According to Schmidt's schema theory, learned movements are not stored by individual concrete motor programs, but by abstracted schemas [3].

The third strategy is to create opportunities to practice the knowledge, skills, and abilities that need to be learned. As an example of surgical technique training, the trainee learns the necessary skills by performing procedures on the physical body models of the patient.

The fourth strategy is to give feedback to the trainee on how they are doing with respect to learning; consequently, it allows for remediation. Here, we define feedback as information provided to the user that is generated adaptively according to the user's actions or their results.

Along with the categorization, we investigate how AR works and what benefits it can bring to each stage.

3 METHODS

3.1 Search for Prototypes

A systematic literature search was conducted using the search string [(augment* OR mix*) AND reality AND train* AND system]. This was performed on January 11, 2021. In this study, the main search engine used was from the IEEE Xplore digital library, complemented with additional papers from PubMed and SciFinder. The search resulted in 985 articles from IEEE Xplore and 121 articles from PubMed and SciFinder.

3.2 Inclusion Criteria

To identify the trends of the training systems in the field of AR, we defined the following criteria that must be met for an article to be considered as part of the database used in this analysis.

1) The paper was submitted in a top AR/VR conference, an IEEE Transaction, or a conference/journal with a h-5 index higher than 20.
2) The number of pages is more than four.
3) The full research paper is publicly accessible.
4) The work can be classified to AR/MR.
5) The work claims to be classified as a training system.
6) The system is evaluated at least once.

The purpose of the first three criteria is to identify the papers that have a big impact and influence over the latest trends in ARTS research. The purpose of the bottom three criteria is to determine whether the article really did work on ARTS that is aligned with our definition as described in the previous section.

3.3 Study Selection

Figure 1 presents the flow diagram of the articles included in the analysis, which follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [34]. In the process of identification, we gathered 985 articles from IEEE Xplore and 121 articles from other digital libraries. Afterwards, we removed duplicates that were present across each database. In the screening process, we filtered the 1,102 articles to follow inclusion criteria 1–3 and we were left with a total of 281 articles. For criterion 1, we used the conference/journal's h-5 index as displayed in Google Scholar metrics, not the individual articles themselves. The reasoning why the score of 20 was chosen is based on the original paper by Hirsch et al., where this score characterizes a successful scientific activity [35]. The caveats of using this metric are also explained in that paper; however, this number can give a rough estimate to the quality of the research.

In the eligibility process, the articles were read carefully and the authors confirmed whether they satisfied inclusion criteria 4–6. For criterion 4, we refer to the loosened definition of AR/MR described in Section 2.2. This includes training systems such as some haptic interface plus monitor setup, like the works of [14], [36], [37], [38], [39]. What is excluded are works that describe purely VR systems; however, comparative studies between AR and VR such as the work of Qin et al. [40] are included.
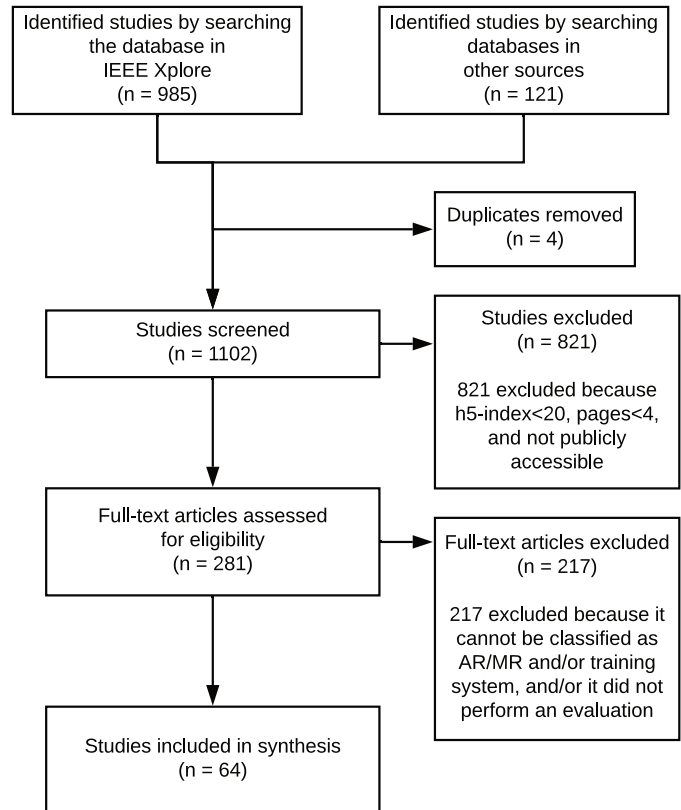

Fig. 1. Flow diagram following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [34].

As for criterion 5, we examined whether the goal of the developed system is in fact for the improvement of a person's knowledge, skills, or abilities, aligned with the earlier definition of training in Section 2.1. Examples of false cases that passed the initial screening are machine learning-related studies, as they also provide hits for the keyword "training," although not for people but for neural networks.

Finally for criterion 6, we checked whether the study contains evaluation conducted on the users. As we want to check for the evaluation as a training system, this metric should be in relation to the user, not the system performance such as tracking speed or rendering time of virtual objects. After removing the articles that did not meet all of the inclusion criteria, 64 articles remained for the analysis and synthesis.

3.4 Limitations

One limitation is the selection of the database used. Our aim is to identify the current AR trends that encompass a wide variety of application fields while also having a huge impact on the AR community. The best candidate we identified that has a huge influence in the AR community is the IEEE Xplore database. To complement the IEEE Xplore results in order to have a wide variety of application samples, we used PubMed and SciFinder only.

Another important limitation of note is the use of the h-5 index. The use of this tool is important to determine the impact; however, this impact that we have measured is not the impact of the paper itself but from the conference/journal
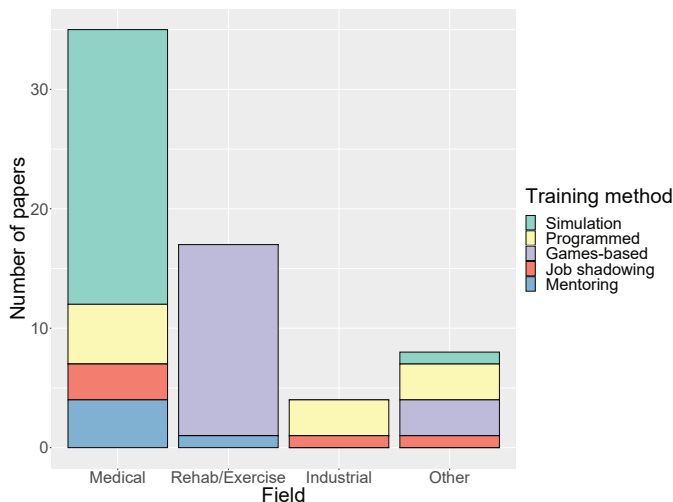
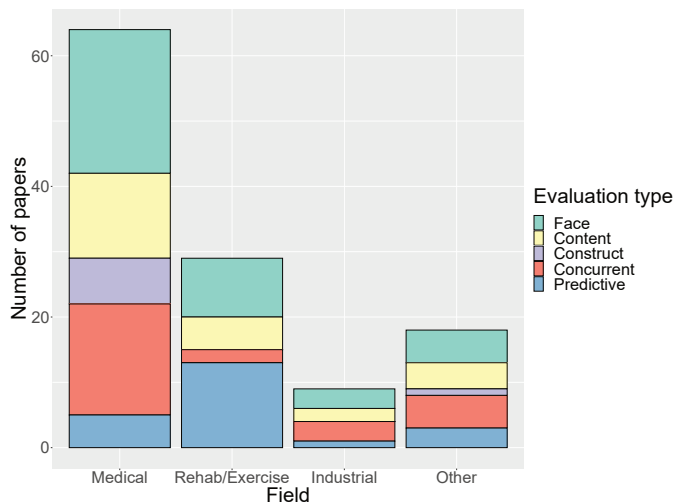Fig. 2. Number of papers in each training method and application field.

Fig. 3. Number of papers in each evaluation type and application field.

venue it was published. This opens the weakness of being vulnerable to missing out on important articles that were published in venues that were not as impactful. However, our decision to filter these by using the h-5 index rather than the actual paper citation itself is to overcome the problem of missing out on important papers that have recently been published; thus, having a low citation count. This trade-off has been considered while thinking about the inclusion criteria. There are also several impactful fields contributing to AR/VR that do not have journals with h-5 index greater than 20. This is currently an important limitation of the study. Hopefully, future authors of review/survey papers can propose better suggestions on how to quantify paper impact other than the two aforementioned methods.

Finally, we set the condition that the work should be considered as AR or MR; however, different authors have different terminologies and usage. In this paper, we discuss the definition of AR according to Azuma; however, we also includ works that are a little different from that definition but consider themselves as AR/MR systems.

4 SUMMARY OF RESULTS

4.1 Overview

In this section, the results of the qualitative analysis are described. A list of all 64 papers that meet all the criteria in Section 3.2 is shown in Table 3. The table includes the reference number, application field, purpose, training method, evaluation type, AR device used, and user that was employed for the evaluation. As for the application field, the training for a medical skill (e.g., surgery, diagnosis) was accounted for in 35 papers. This was followed by rehabilitation/exercise (17 papers), industry (four papers), and other fields (eight papers). It is noteworthy that there is very little training in the industrial field in contrast to the extremely large amount of support systems that use AR.

4.2 Categorization by Training Method

Figure 2 shows the number of papers for each training method in each field. In descending incidence order:

simulation (24 papers), games-based training (19 papers), programmed instruction (13 papers), job shadowing (four papers), and mentoring (four papers). The majority of the simulation studies (23 papers) were used in the medical field. Training dolls differ in many ways from the actual human body, which makes the learning transfer difficult. AR enables the appearance of a training doll to be similar to that of the actual human body by superimposing the patient's body textures and organ models on it. This is most likely the main reason why simulations are widely used in this field. On the other hand, the other fields, where imitation of the real environment is not comparatively necessary, employ different strategies.

Games-based training (19 papers) was the second most adopted, and 74% (14 papers) of these papers were in the field of rehabilitation/exercise. The majority of users in this field are people who have lost or are losing certain functions of their bodies owing to disease or injury. Rehabilitation often requires long-term continuity, although existing rehabilitation programs are often monotonous repetitions of tasks [97], [98]. One of the biggest problems is the difficulty in maintaining patient motivation [99]. Games-based training is an effective means of incorporating a mechanism for continuity into these monotonous structures through gamification. It is easy to understand that users in the rehabilitation field would find it helpful if the training system adopted the games-based method for the tedious tasks they typically have to endure.

Programmed instruction (13 papers) has been equally adopted in other fields, except rehabilitation. This may be because the normal rehabilitation process does not include general knowledge acquisition through classroom lectures.

4.3 Categorization by Evaluation Type

As described in Section 2.4.2, we classified the evaluation into five types, while referring to Barsom et al. [6]. Figure 3 shows the number of papers of each evaluation type in each field. Note that some papers used more than one type of evaluation. The number of papers that adopted face evaluation is the largest (39 papers), followed by concurrent

TABLE 3
Summary of ARTS works by publication venue, field of application, purpose, training method, evaluation types, test users, and AR device used.
* A=Face, B=Content, C=Construct, D=Concurrent, E=Predictive

Ref.	Venue	Field	Purpose	Training Method	A	B	C	D	E	User	Device
[41]	JMIR	Medical	Digital Rectal Examination	simulation	✓	✓	-	-	-	actual	HMD
[42]	J. Healthc. Eng.	Medical	Orthopaedic Open Surgery	simulation	✓	-	-	-	-	mixed	HMD
[14]	IJCARS	Medical	Fetal Minimally Invasive Surgery	simulation	✓	✓	-	-	-	actual	Stationary
[43]	Heliyon	Medical	Cardiopulmonary Resuscitation Training	simulation	✓	-	-	-	-	actual	HMD
[40]	J. Healthc. Eng.	Medical	Peg Transfer Training	simulation	✓	✓	✓	-	-	actual	HMD
[44]	JMIR	Medical	Basic Life Support Training	simulation	✓	-	-	-	-	mixed	HMD
[36]	Anesth. Analg.	Medical	Anesthesiology Training	simulation	-	✓	-	-	✓	mixed	Stationary
[45]	Clin. Simul. Nurs.	Medical	Anatomy Nursing Skills	simulation	✓	✓	-	✓	-	actual	Handheld
[46]	Healthc. Technol. Lett.	Medical	Endoscopic Third Ventriculostomy Surgery	simulation	-	-	✓	✓	-	actual	HMD
[37]	Neurosurgery	Medical	Neurological Surgery	simulation	-	-	-	✓	-	alternative	Stationary
[38]	BMC Bioinform.	Medical	Artificial Cervical Disc Replacement Surgery	simulation	✓	-	-	-	-	actual	Stationary
[47]	Turk. Neurosurg.	Medical	Spinal Surgery	simulation	-	-	-	✓	-	actual	Handheld
[48]	MBEC	Medical	Ultrasound Guided Needle Insertion	simulation	✓	✓	✓	-	-	alternative	Stationary
[39]	World J. Surg.	Medical	Laparoscopic Surgery	simulation	✓	✓	-	-	-	mixed	Stationary
[49]	BJUI	Medical	Urethrovesical Anastomosis Surgery	simulation	✓	-	-	✓	-	actual	Stationary
[50]	IJMRCAS	Medical	Soft Tissue Surgery	simulation	✓	-	-	-	-	actual	Stationary
[51]	IEEE TBME	Medical	Surgical Interventions Planning	simulation	-	-	✓	-	✓	mixed	HMD
[52]	IEEE TVCG	Medical	Micro-CT Analysis	simulation	✓	-	-	✓	-	alternative	HMD, Projected
[15]	IEEE TOH	Medical	Tumor Probing	simulation	-	-	-	✓	-	alternative	Stationary
[16]	IEEE TBME	Medical	Spinal Anesthesia Procedures	simulation	✓	✓	✓	-	-	mixed	Stationary
[53]	IEEE TBME	Medical	Needle Placement for Facet Joint Injections	simulation	-	-	-	-	✓	actual	Stationary
[54]	IEEE THMS	Medical	Anatomy Training	simulation	✓	✓	-	✓	-	actual	HMD
[55]	CCECE	Medical	Neurological Surgery	simulation	-	-	-	✓	-	alternative	HMD, Stationary
[56]	BMC Med. Educ.	Medical	Health Science Education	programmed	✓	-	-	-	-	actual	HMD
[57]	BMC Med. Educ.	Medical	Bladder Catheter Placement	programmed	✓	-	-	✓	✓	actual	HMD
[58]	Biomed. Eng. Online	Medical	Electrocardiogram device operation	programmed	-	-	-	✓	-	actual	HMD
[59]	ISMAR	Medical	Neonatal Endotracheal Intubation	programmed	-	-	-	✓	-	actual	HMD
[60]	IEEE VR	Medical	Sepsis Prevention Education	programmed	✓	-	-	-	-	actual	HMD
[61]	World J. Urol.	Medical	Surgical Training	job shadowing	✓	✓	-	-	-	actual	Stationary
[62]	ISMAR	Medical	Anesthesia Education	job shadowing	✓	✓	-	-	✓	actual	HMD
[63]	ISMAR	Medical	Surgical Interventions Planning	job shadowing	-	-	-	✓	-	actual	HMD
[64]	Surgery	Medical	Surgical Instruction	mentoring	-	-	-	✓	-	actual	Handheld
[65]	NPJ Digit. Med.	Medical	Surgical Instruction	mentoring	-	-	✓	✓	-	actual	HMD
[66]	Sensors	Medical	Point of Care Ultrasound Training	mentoring	✓	-	-	✓	-	actual	HMD
[67]	IEEE VR	Medical	Number Matching Task/Austere Surgery	mentoring	✓	✓	-	✓	-	actual	HMD
[68]	JMU	Rehab	Gait and Balance Rehabilitation	games-based	✓	-	-	✓	-	actual	HMD
[69]	IJERPH	Rehab	Exergames	games-based	✓	-	-	-	-	actual	Stationary
[70]	Behav. Res. Methods	Rehab	Mirror and Imagery Training	games-based	-	-	-	-	✓	alternative	HMD
[71]	JNER	Rehab	Upper Limb Stroke Rehabilitation	games-based	-	-	-	-	✓	actual	Stationary
[72]	PLoS One	Rehab	Gait Rehabilitation	games-based	✓	-	-	-	-	alternative	HMD
[73]	JESF	Rehab	Tai-Chi Training	games-based	-	-	-	-	✓	actual	Stationary
[74]	ARM	Rehab	Balance and Mobility Rehabilitation	games-based	-	-	-	-	✓	actual	Stationary
[75]	JRRD	Rehab	Parkinson disease Gait Rehabilitation	games-based	-	-	-	-	✓	actual	HMD
[76]	Exp. Brain Res.	Rehab	Trunk-arm Rowing	games-based	✓	-	-	-	✓	alternative	Projected
[77]	Cogn. Behav. Neurol.	Rehab	Mild Cognitive Impairment Rehabilitation	games-based	-	-	-	-	✓	actual	HMD
[78]	IEEE VR	Rehab	Tai Chi Chuan Learning	games-based	✓	✓	-	-	-	actual	HMD
[79]	IEEE VR	Rehab	Eye-Hand Coordination Training	games-based	✓	✓	-	✓	✓	actual	HMD, Stationary
[80]	IEEE Access	Rehab	ADHD Children Treatment	games-based	✓	✓	-	-	✓	actual	HMD
[81]	IEEE TVCG	Rehab	Exercise for Reducing Obesity	games-based	✓	✓	-	-	✓	actual	Stationary
[82]	IEEE TNSRE	Rehab	Hemiparesis Stroke Rehabilitation	games-based	-	-	-	-	✓	actual	Stationary
[83]	IEEE/RSJ IROS	Rehab	Wheelchair Assistance	games-based	✓	-	-	-	✓	alternative	HMD
[84]	Biomed Res. Int.	Rehab	Chopsticks Telerehabilitation	mentoring	-	✓	-	-	✓	alternative	HMD
[85]	ISMAR	Industrial	Scenario-Based Training Authoring	programmed	✓	✓	-	✓	-	alternative	HMD, Stationary
[86]	IEEE Access	Industrial	Assembly Instruction	programmed	-	-	-	✓	-	actual	Stationary
[87]	IEEE RO-MAN	Industrial	Transfer Task for Exploration Training	programmed	-	-	-	-	✓	mixed	Stationary
[88]	ISMAR	Industrial	Origami and Building Blocks	job shadowing	✓	✓	-	✓	-	alternative	Stationary
[89]	ACM/IEEE HRI	Other	Autonomous Driving	simulation	-	-	-	✓	✓	actual	HMD
[90]	EAIT	Other	Business Ethics	programmed	-	-	-	-	✓	actual	Handheld
[91]	Front. Psychol.	Other	Musical Instruction	programmed	-	-	-	-	✓	actual	Projected
[92]	IEEE ToE	Other	Remote Laboratory Education	programmed	✓	✓	✓	-	-	actual	Stationary
[93]	IEEE VR	Other	Neurofeedback Training	games-based	✓	-	-	✓	-	actual	Stationary
[94]	IEEE Access	Other	Hands-on Experiential Learning	games-based	✓	✓	-	✓	-	actual	HMD, Stationary
[95]	IEEE Access	Other	Spatial Memory Learning	games-based	✓	✓	-	✓	-	actual	Handheld
[96]	ISMAR	Other	General Work Activities	job shadowing	✓	✓	-	✓	-	actual	HMD

(27 papers), content (24 papers), and predictive (21 papers), whereas construct (eight papers) is the smallest by far. Face evaluation is performed to assess the degree of resemblance between training with the system and the educational construct with a questionnaire or a small interview. Owing to the implementation ease, the number of adoptions is large. The concurrent evaluation is a comparison with the existing training methods. The wide use of this evaluation is also understandable because it is already common practice for studies to do comparisons of their proposed methods against that of the golden standards. Note that 21 studies conducted the predictive evaluation, which is a significant amount. Predictive evaluation is the most important type

because it examines direct training effects (i.e., actual skill acquisition). However, it is also expensive because it requires training with the system and the evaluation to be carried out in separate steps. Thus, we initially presumed that the papers that included the predictive evaluation would be quite limited, although the results are contrasting. The construct evaluation examines the differences in the outcomes between two types of subjects with different skill levels when using the system, i.e., ascertain if the system is reflected with some skill that is possessed by the expert. The lack of its adoption can be interpreted as self-evident, or it is due to the high cost of the evaluation (two different groups of subjects—experts and novices—are required).
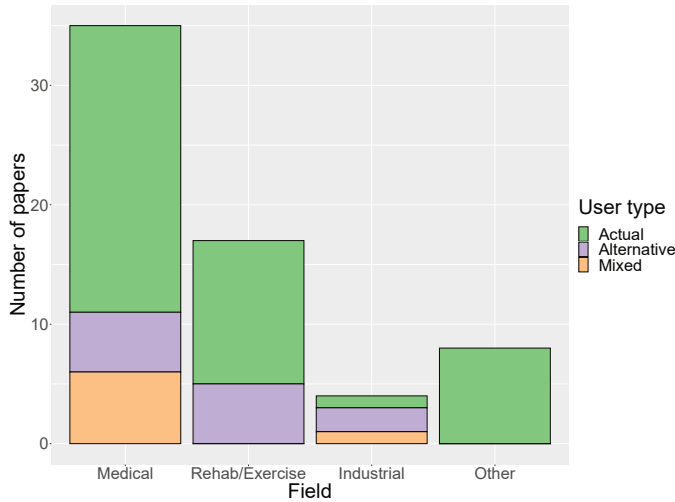
Fig. 4. Number of papers for each user type for the evaluation.

Looking at the differences among the fields, the adoption rate of the predictive evaluation is relatively high in rehabilitation. As mentioned earlier, games-based training is mainly used in the rehabilitation field. This method tends to transform the skills that need to be acquired by incorporating game elements. Therefore, it is understandable that many papers confirmed the impact of the gamified training content relative to the acquisition of the intended skills to be learned.

In addition, we analyzed the number of different types of assessments made in each paper. Although 43% (28/64 papers) had only one type of evaluation, more than half of the papers had a combination of two (18 papers), three (17 papers), and four (one paper) types of evaluations. There were no papers with five types of evaluations. Barsom et al. stated that the new training system can be considered for social implementation only when all these five types of validity are guaranteed [6].

4.4 Categorization by Users in the Evaluation

In addition, we analyzed the users that were used for the evaluation in each field. The users were divided into "actual" users (i.e., people who actually need to be trained, e.g., students in medical school in the case of training for surgical skills) and "alternative" users (i.e., people who do not need to train the target skill). In addition, papers that use both users are classified as "mixed". Figure 4 shows the number of papers in each of the three categories. In some fields, it is very expensive to collect actual users of ARTS as test subjects, so it was assumed that many alternative users were reluctantly used. However, contrary to this, more than two-thirds of the papers conducted evaluations with actual users.

4.5 Categorization by Type of AR Device Used

We also looked at the type of AR device used for each study. For the classification of these devices, we adopted the categorization of AR displays according to distance from eye to display from Schmalstieg et al.'s book, "Augmented

Reality: Principles and Practice" (i.e., HMD, Handheld, Stationary, Projected) [100]. However, this does not consider the quality of the AR device; for example, HoloLens and Google Cardboard (in which a smartphone is used as a VST device) are both considered HMD.

In descending order, the distribution for the AR device used are as follows: HMD (34 studies), Stationary (27 studies), Handheld (five studies), and Projected (three studies). HMD and Stationary devices are the popular choices for implementing ARTS. When looking at the purpose for which these ARTS are used, they usually involve some type of activity that require the user's hands/body. Taking this point into consideration, it is logical that handheld displays are not optimal for training systems in general. Projected displays such as CAVE or other spatial AR displays are also rarely used as there are no additional training benefits they can offer compared with the World space alternative (i.e., Stationary displays) which are easier to implement. For the distribution of the type of AR device with regards to the training method used or the application field, no trends can be drawn as it is distributed evenly.

5 NARRATIVE SYNTHESIS

In Section 4, we summarized the results in terms of the training method, evaluation type, users, and device used in the evaluation. However, from these results, we determined that the current implementation of ARTS has heterogeneity in regard to the study design; thus, statistical pooling cannot be performed. Because a meta-analysis is still difficult to accomplish in the current stage of ARTS, the most logical approach is to conduct a narrative synthesis. This section describes each of the training methods that are effective for ARTS by presenting the representative studies that constitute the essence of each method. We also look at how each method is evaluated by narrating some exemplars for each evaluation type.

5.1 Simulation

Simulation training simply consists of training under a simulated environment and the goal is to develop specific skills. The most obvious advantage of simulation training is that it can provide a risk-free environment that can be considered to be very risky if it is performed in a real-life environment. Another benefit with simulation training is that it offers the trainee the opportunity to do rehearsals and practice the process repeatedly [1]. The use of AR/VR simulation to supplement traditional teaching in surgery skills for example, is a better alternative ethically compared to cadaver dissections [18], [19]. In terms of the realism and didactic value, Botden's study [39] ascertained that the ProMIS AR was the better simulator for practicing laparoscopic skills in comparison with the LapSim VR simulator.

5.1.1 Implementation - Simulation

When considering how training in simulators is implemented and how the skills in this process are accumulated, it is important to discuss the concepts of skills generalization and skills transfer. Gallagher states that "skills generalization refers to the training situation where the trainee learns
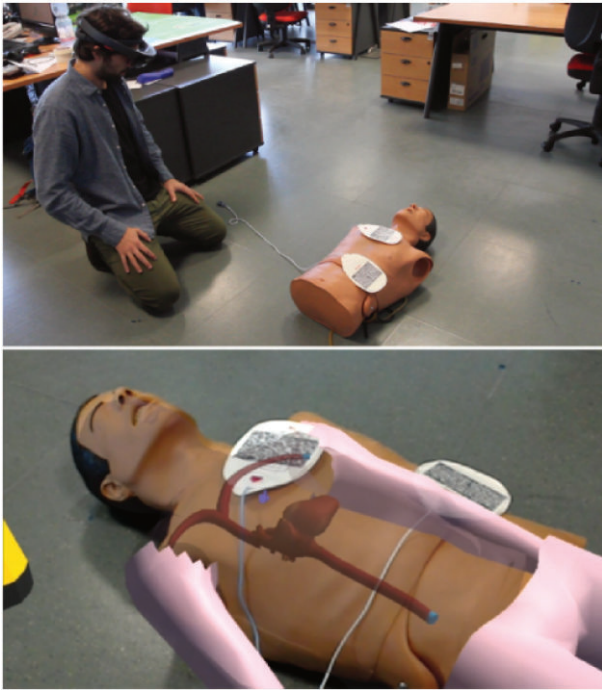
Fig. 5. Environment for defibrillation training [44].

Fig. 6. Peg transfer simulator on box, AR, and VR platforms [40].

fundamental skills that are crucial to completion of the actual operative task or procedure. Skills transfer refers to a training modality that directly emulates the task to be performed in vivo or in the testing condition" [101]. Each of these has its own advantages. For example, skills generalization can greatly boost the rate of learning while not having to invest substantially in the cost of the technology (in this case, the realism of the simulator). Meanwhile, skills transfer provides a more realistic approach and it generates a venue for rehearsal and practice.

Ingrassia et al. [44] presented a nice example of skills transfer training. They developed a system that is able to realistically reproduce the scenario of defibrillation training. The goal was to provide inexperienced trainees with an environment for self-instruction training to perform the CPR procedure, as shown in Figure 5. In this environment the trainees are able to do natural gestures and body movements, as if they were in the actual environment.

Another example of skills transfer training is provided by Muangpoon et al. [41]. The goal of the system was to provide clinicians and medical students an environment to learn, teach, and practice digital rectal examination (DRE). In this system, they are able to visualize a virtual hand that is overlaid on their real hand and the internal organs are overlaid on a benchtop model. Because the technique and manner of operation is vital for DRE, they have tracked the movement of the hand together with the amount of applied pressure.

For the skills generalization training, Qin et al. [40] offer a good example. They utilized a multiple platform simulator system (AR, VR, etc.) for the peg transfer task, as illustrated in Figure 6. Instead of recreating the whole surgical process itself, this type of training focuses more on reproducing the specific surgical maneuvers to gain the target skill.

Aside from looking at the perspective of the task itself, there are also training systems that deal with the planning of the operation. Abhari et al. [51] proposed a system that is able to facilitate training for the planning of a neurosurgical procedure. They overlaid patient-specific data onto a mannequin head to aid the planning of the operation. Their results show that the performance index of the non-clinicians significantly improved when they used their system. Furthermore, the performance time of the clinicians was significantly faster in comparison to using conventional planning environments. The reason for these improvements is that the participants were able to develop spatial reasoning ability, which cannot be gained in traditional methods.

5.1.2 Evaluation - Simulation

• Face and content evaluation example:

Javaux et al. [14] evaluated the face and content by requesting surgeons that performed fetal minimally invasive surgery to complete certain objectives using a developed simulator. The skills targeted included basic fetoscopic and procedural skills for laser surgery. A five-point Likert scale was used to assess the face evaluation, which contained aspects such as the realism of the trainer, the body wall phantom, and the rendering of the environment (e.g., image quality, light propagation, and depth perception). Similarly, the content evaluation was accomplished by measuring the training capacities (e.g., scope handling and lasering) and the usefulness of each task.

• Construct evaluation example:

Qin et al. [40] conducted a Null-Hypothesis Significance Testing (NHST) between novice thoracic surgeons with four years of post-graduate experience and experts with 12 years of experience. They also considered the experience of the participant in regard to the box trainers, VR games, and HMDs. Meanwhile, Botden et al. [39] acquired 30 novices, 30 intermediates, and 30 experts of laparoscopy and the demographics included interns, surgical residents, and surgeons.

The laparoscopic experience of each group was decided based on experience, such as the number of times they performed suturing in a clinical setting, or the number of times they participated in the surgical procedure, either as a spectator, a camera handler, or an assistant.

• Concurrent evaluation example:

An example of a concurrent evaluation is presented by Aebersold et al. [45]. The main goal of their study was to investigate whether the students had a better understanding and learning of nursing skills such as placing nasogastric tubes. Students were randomly appointed to either the control group (usual training method) or the AR group, which received the training module of anatomy simulation by using an iPad. There were 34 and 35 participants in each group, respectively. The nursing skills were assessed through a 17-item checklist scoring system and an NHST was carried out between these two groups. In another study, Chowriappa et al. [49] explored the effects of using an AR-based training module for robot-assisted urethrovesical anastomosis. In the user study, 52 participants were randomized to be either in the hands-on surgical training (HoST) group or the control group. They applied five global evaluative assessment of robotic skills (GEARS) to establish a comparison between the two groups and they determined the results that were based on the computed p-value.

• Predictive evaluation example:

In an investigation by Sappenfield et al. [36], a MR simulator was used to train anesthesiology residents to achieve the goal of improving supraclavicular access to the subclavian vein. When considering the evaluation metrics, they used an automated scoring system to objectively score the time, success, and errors/complications of the participant's performance. When considering the design of their study, they randomized the residents into two groups. The first group was exposed to real-time 3D visualizations in the first trial but not in the second trial. The second group was not exposed to 3D visualizations in the first and second trials, but they were later asked to experience the playback of the second trial's 3D visualizations before carrying out the third trial without the visualization guide. The comparison of scores between the subsequent trials and the differences between the groups that experienced real-time visualization, no visualization, and delayed visualization point toward the evidence of this study's predictive evaluation.

5.1.3 Discussion - Simulation

In this training method, it can be argued that the simulator's quality of realism affects the enhancement of the learning transfer. For this reason, 15 out of the 24 total simulation training papers that evaluated their system used the face evaluation type. Studies also usually perform content with the face because these evaluations used the same kinds of questionnaires such as five-point Likert scales. In the skills transfer training example, Muangpoon et al. [41] assessed the DRE clinicians and medical students with face and content evaluation. However, no evaluation metrics were performed for the other evaluation types. For the skills transfer training, the most important part to consider during the evaluation is the realism and quality of the simulation; thus, studies tend to evaluate the face and content.

In the investigation by Qin et al. [40], novices and experts were recruited for the peg transfer skill, and they evaluated the face, content, and construct. The lack of scenario information for skills generalization (i.e., patient/mannequin operation) puts more emphasis on the manner of doing the task. Because the difference in the expert's technique and novice's technique is emphasized, the value of performing a construct evaluation increases. Furthermore, Abhari et al. [51] performed the construct evaluation, which makes sense since the construct is arguably the most important when considering the planning phase for surgical procedures. In the simulation training method, there are 10 studies that performed a concurrent evaluation whereas only four had a predictive evaluation. It is notable that a lot of studies have adapted the practice of comparing their results to the golden standard (concurrent evaluation). However, we suggest that more studies should also investigate the effects of their simulators not only in the AR environment, but also how well these acquired skills from the simulators can translate toward real implementation (predictive evaluation).

5.2 Programmed Instruction

The use of programmed instruction with AR provides a lot of benefits in comparison to using more traditional approaches such as manuals and instruction books. For instance, programmed instruction can offer more flexibility when instructions and training procedures are updated to more recent techniques. Similar to the simulation training method, programmed instruction also enables the user to repeatedly practice and rehearse the specific skill [102], [103]. The unique point of programmed instruction over the other training methods is its potential to take advantage of multisensory features such as sound, text, and animations [1].

5.2.1 Implementation - Programmed Instruction

There are many ways of presenting instruction, but when deliberating about learning theory and instructional design, one of the more crucial aspects that need to be considered is the instruction sequencing [104]. Many studies have suggested that the sequence and arrangement of these learning activities impact how the information and knowledge is being handled [105], [106], [107]. According to theories of instructional design, one approach is to follow a simple-to-complex sequence [104]. There are also other design approaches such as simply conforming to the traditional methods (e.g., how kids learn the alphabet from A to Z). We refer to this delivery as the sequential instruction approach. The presentation of sequential instruction can also come in two forms: pre-recorded or real-time.

An example of a pre-recorded sequential instruction is presented in Sankaran et al. [60]. They used a 360-degree video recording session to enable the viewpoint of what it is like to experience the real clinical practices for sepsis prevention. To accelerate the training of medical students, the students were exposed to the recorded simulated environment along with the augmented information such as the patient graphical data or the pop-up information, as shown in Figure 7. At certain checkpoints in the program, the students were asked to answer a pop-quiz to assess their learning progress.

Fig. 7. Pop-up information about sepsis condition [60].

Furthermore, an example of a real-time sequential instruction is provided by Zhao et al. [59]. They developed an effective ARTS that is able to give a comprehensive understanding of the endotracheal intubation procedure by providing real-time instruction and evaluation. This was accomplished by overlaying 3D see-through visualizations of data such as the depth, penetration, and time to the manikin. In addition, the visualization of the instructions for each phase of the procedure was displayed to the user that was wearing the HMD.

By contrast, the other approach of the instruction delivery is the non-sequential instruction. In line with the component display theory of instructional design that was presented by Merrill, learner control is considered to be a very important component [108]. Learner control is "the idea that learners can select their own instructional strategies in terms of content and presentation components" [104]. In regard to this approach, the most essential point is the fact that students are able to have more control so they can learn at their own pace.

The study by Marquez et al. [92] presents the concept of non-sequential instruction. This work proposed the idea of the implementation of an augmented remote laboratory (ARL). Inside the ARL, students were free to experience and explore the environment and do laboratory activities that are analogous to traditional laboratory classes.

5.2.2 Evaluation - Programmed Instruction

• Face and content evaluation example:

Sankaran et al. [60] carried out face and content evaluation by using the system usability scale. Sankaran et al. asked 28 novice students for a score that had a scale ranging from one (strongly disagree) to five (strongly agree). The face evaluation questions contained statements such as "I needed to learn a lot of things before I could get going with this system." Content evaluation statements included, "I thought the system was easy to use," or, "I found the various functions in the system were well integrated."

• Construct evaluation example:

In addition, Marquez et al. [92] performed construct evaluation using the results of the questionnaire and compared the scores between the evaluations from the teachers and the evaluations from the students.

• Concurrent evaluation example:

Schöb et al. [57] developed an MR system teaching tool that can provide students with instructions while learning

a new practical task. Schöb et al. recruited 164 medical students to experience the bladder catheter placement. In this study, 107 students were assigned to the control group, and received instructions only from the instructor. The 57 remaining students received instructions from the MR guidance system while using HoloLens. Both groups were asked to take a standardized, non-timed objective structured clinical examination and were assessed by their learning outcomes. The control and study group were compared by performing an NHST.

• Predictive evaluation example:

Sari et al. [90] used AR techniques to teach ethics and moral imagination. They recruited 142 students that were taking a business ethics course. The study utilized the 3x2 experiment method, which consists of three training modes and two time periods. The three training modes comprised AR-based, paper-based, and no training. As evidence of the predictive evaluation, the authors performed an analysis of variance (ANOVA) to calculate the F-value and p-value for between-subjects and within-subjects. They also conducted a post hoc analysis test of the training methods to determine if the groups differed in terms of the moral imagination.

5.2.3 Discussion - Programmed Instruction

This section introduces two programmed instruction approaches inspired from the concepts of instructional design, sequential instruction, and non-sequential instruction. As most studies used the sequential approach, it is difficult to determine the differences between their evaluations. However, the low adoption number does not discredit the effectiveness of the non-sequential design of instruction. The non-sequential design shines when thinking about individually matching the pace of each student's personal growth.

As discussed in Section 4, when looking at the relationship between the programmed instruction and application field, it can be observed that programmed instruction is not utilized in the rehabilitation and exercise area. This is because programmed instruction is focused more on the development of knowledge rather than the development of skills. For the purposes of rehabilitation and exercise, the training needed require systems that allow for the forming of an ability and capability to do something, whether physical or psychological. On the other hand, programmed instruction is very effective when it is used for the endowment of knowledge. For this reason, most of the studies categorized to programmed instruction (except one) have their target user as the students that will actually use that specific knowledge (e.g., medical students [57], [60], [92], and business ethics students [90]).

5.3 Games-based training

The culture of games has certainly risen in influence, especially for today's younger generation. It is no surprise that the use of game-like mechanics and gamification techniques can motivate users to engage in activities that might otherwise be less interesting [109], [110], [111]. The clear advantage of games-based training over other training methods is in its competitive nature, because it instills motivation that leads toward learning. The downside of this method though is that too many elements and components are integrated
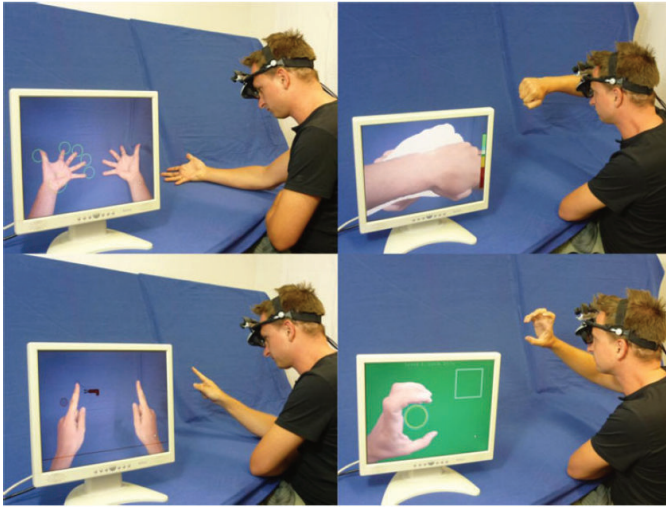
Fig. 8. Example of activities for hand-mirroring ARTS (monitor view for demonstration purposes only. In the actual implementation, the users experienced these activities through an HMD) [70].

into the game. This may be difficult to clearly determine and guarantee which parts contributed to learning [1]. When comparing other games-based methods, AR games-based training stands out because it provides a nice balance between immersion in the game and being rooted in the real world. The latter is especially important when considering if the users have physical or psychological disorders.

5.3.1   Implementation - Games-based training

From the database of papers we collected, games-based training is usually targeted for rehabilitation and exercise as demonstrated in an overwhelming 16 out of 19 papers. In terms of providing solutions for therapeutical treatments, games-based training systems can either address the physical or psychological problems of its target users.

One case of using games to treat physical problems is the work that is presented by Johnsen et al. [81]. In this study, Johnsen et al. created a MR system that allows obese kids to interact with a virtual pet. To trigger interactions with these pets, the kids need to input physical activities. Growth of the pet is proportional to the physical activity progress of the kid; thus, the virtual pet becomes a strong motivator for promoting health. Another example is the work by Trojan et al. [70], who proposed an AR hand training system that uses the mirror image approach. They used techniques such as finger flexing, hand posture fitting, and a "snake" video game to train the motor skills by hand-mirroring, as illustrated in Figure 8.

Other studies have also dealt with psychological disabilities. For example, Park et al. [77] examined the effects of their MR training system on patients that have mild cognitive impairments such as Alzheimer's disease. They designed games that recreated day-to-day activities to target cognitive functions such as selective attention, the visual/verbal working memory, and problem solving. One more case of this type is a study conducted by Kim et al. [80]. They developed an MR eye-contact game that is able to treat children with attention deficit hyperactivity disorder (ADHD). The game utilizes face recognition techniques to

confirm the child's attention, trigger the treatment, and develop interpersonal skills.

5.3.2   Evaluation - Games-based training

• Face and content evaluation example:

Batmaz et al. [79] investigated the differences between the effectiveness of AR, VR, and the conventional 2D touchscreens to train professional athletes by using the eye-hand coordination reaction test. This study accomplished the face and content evaluation by collecting subjective opinions of 15 participants from their local university by using a seven-point Likert scale. Statements such as "increased sense of reality" and "having better perception of depth" indicate face and content, respectively.

• Concurrent and predictive evaluation example:

Moreover, Batmaz et al. [79] analyzed the effects of the time, error rate, and throughput while considering the experimental condition (AR, VR, or 2D screen). Aside from this, they also analyzed the effects for the haptic feedback and environments. They compared the effects by computing the F-value and p-value for each of the conditions, which suggests concurrent evaluation. Furthermore, they analyzed the performance improvement of the participants across multiple repetitions by using AR and VR in comparison to a 2D screen, which is indicative of a predictive evaluation.

5.3.3   Discussion - Games-based training

Most cases for games-based training are targeted at rehabilitation and exercise. This suggests that games are a great avenue for developing or recovering physical and psychological skills. The reasoning for this is that tasks that are directed toward these skills are usually tedious, monotonous, and repetitive. Games are excellent at solving this problem because it can stimulate motivation and maintain user involvement.

One thing to note is that for games-based training, no studies have conducted a construct evaluation. This is self-evident as there is no user that has a high skill or a low skill for a physical or psychological disability. However, what matters is the confirmation if there is the transfer of training effects to the real context in terms of whether they acquired or recovered the targeted skill. In this situation, the predictive evaluation is of great importance, which is implemented in many studies (12 out of 19 papers).

5.4   Job shadowing

In companies and enterprises, job shadowing refers to on-the-job training of a trainee that involves observing the model employee that is performing their usual work. When applying AR in training, job shadowing is the instance when a novice diligently observes the performance of an expert by using techniques such as sharing first-person views. The main advantage of job shadowing over the previous training methods is the ability of the novice to experience the perspective of the expert, which gives a broader outlook to the development of skills and techniques of professionals.

5.4.1   Implementation - Job shadowing

Job shadowing in AR in practice can be done directly or indirectly. Direct job shadowing refers to the show-by-example approach where the expert representation (e.g.,
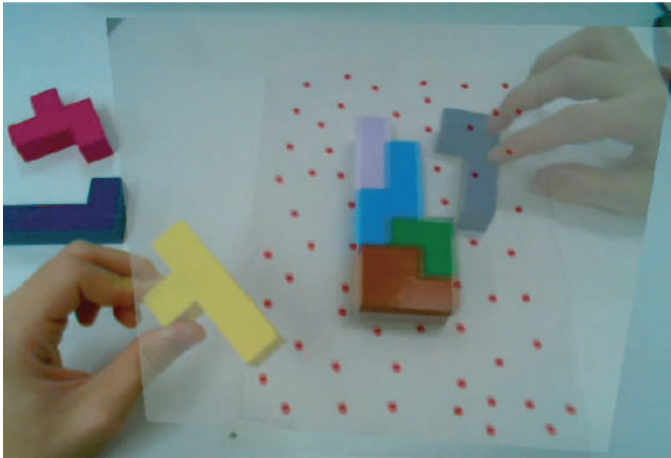
Fig. 9. Instructor's hand overlay seen through an HMD [88].

virtual hand) demonstrates the necessary techniques to the novice. On the other hand, indirect job shadowing refers to the use of tools such as annotations (e.g., circles and pointers) to guide the novice to perform the proper techniques. Programmed instruction and indirect job shadowing may seem very similar at first glance, but they are slightly different in a sense that the former focuses more on developing the knowledge while the latter focuses on forming the skill.

Goto et al. [88] exemplified the concept of direct job shadowing. This system provides visual guidance to the novice by overlaying in the first-person view an example of the correct way of performing the task. This is illustrated by the phantom representation of the instructor's hand as shown in Figure 9. In this study, Goto et al. addressed the problem of visual confusion of the real work environment from the visual guide by using techniques such as changing the transparency and enhancing the contours.

Conversely, Lee et al. [96] demonstrated the concept of indirect job shadowing. They developed a prototype system that is able to capture and share first-person view annotations to share the instructions. These shared instructions are then used by the novice to obtain an understanding of how to do things. Lee et al. used circles and arrows to guide the novice in performing general work activities (e.g., setting up presentation facilities in a seminar room).

5.4.2 Evaluation - Job shadowing

• Face and content evaluation example:

Goto et al. [88] tested the face evaluation by using a five-point Likert scale for the subjective evaluation of an instructional video in terms if it is easy or difficult to follow. Goto et al. tested the content evaluation again by using a five-point Likert scale for the subjective evaluation of the visual effects (e.g., shift in the AR view, alteration of the transparency, etc.) to determine if it is suitable or not suitable for the specific task.

• Concurrent evaluation example:

Going back, Lee et al. [96] performed a user study with 18 participants from university students and staff and asked them to rate the user experience by using a seven-point Likert scale. Lee et al. allowed the user to undergo two conditions: video only and video with spatial cues. Lee et

al. compared the results between the two conditions and did tests (e.g., paired t-test, Wilcoxon Signed Rank test, etc.) for the task completion time, error, and angular difference.

• Predictive evaluation example:

Quarles et al. [62] developed a system where the user can review the past training experience that is overlaid to the current experience of the user. Quarles et al. recruited 19 students and three educators to be a part of the user study. The evaluation metrics that Quarles et al. used included a fault test score and a confidence test score. As evidence of the predictive evaluation, Quarles et al. performed a comparison of the scores before and after using the after action review from the previously mentioned evaluation metrics.

5.4.3 Discussion - Job shadowing

In the survey of papers that were gathered, only five studies contributed to the category of job shadowing. It is difficult to determine the trends with an insufficient sample size; however, this suggests that researchers still have a lot of scope to cover in this area that has not been investigated yet. Because job shadowing has the ability to facilitate the demonstration of an ideal execution and implementation, it can be considered as a great subject that should be focused for future ARTS. Currently, we have determined that there are two approaches to job shadowing: direct and indirect.

5.5 Mentoring

Mentoring is the process where the mentor provides guidance and support to the apprentice. The goal of mentoring is for the apprentice to grow in terms of developing the skill or experience so that they can do the work on their own. The advantage of AR mentoring, just like AR job shadowing, is the ability to facilitate the sharing of each other's views. In terms of the role of the actors, mentoring and job shadowing are similar in that experts share their knowledge and skills to the novice. The main, yet subtle, difference between these two training methods lies in the relationship between the experts and novice. Although job shadowing novices observe the expert performing, mentoring adds the element of the expert being more teacher-esque to the novice.

5.5.1 Implementation - Mentoring

Just like job shadowing, implementation of AR mentoring systems can also be classified as direct and indirect mentoring. Examples of the implementations for direct and indirect mentoring are akin to the strategies that are used in job shadowing (virtual hand overlay or annotations). However, the element of a two-way communication between the novice and expert, along with the condition of it being met in real-time, must be satisfied.

An example of direct mentoring is the work that was done by Wang et al. [66]. They created a telemedicine system that is capable of sharing hand pointing gestures through the use of Leap Motion and HoloLens. In this system, the mentor is able to see the first-person view of the trainee and the trainee is able to see the serialized hand data of the mentor. Above these, they are also able to communicate through a real-time audio stream. Another example is the
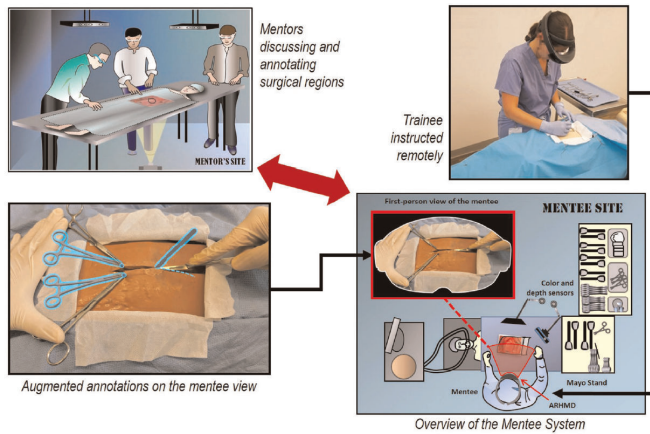
Fig. 10. Telementoring system workflow [65].

work by Chinthammit et al. [84]. They developed Ghost-man, a system in which the therapist and the patient are able to stream each other the ghost image of their hands. This is used to exchange the perspective that can be observed in the augmented environment. They designed the system in a way that the therapist is able to deliver instructions remotely to the patient in terms of how to properly execute the motor skills, and in this case, handling chopsticks.

By contrast, Rojas et al. [65] demonstrated the idea of indirect mentoring. They developed a telementoring system for cricothyroidotomies training, in which expert surgeons can share annotations and audio guidance to the onsite medical trainees, as summarized in Figure 10. To accomplish this, first, a stabilized first-person view of the working environment is streamed to the experts. Then, in a seamless manner, it provides instructions to the trainee through the use of annotations and 3D model augmentations projected through the AR HMD.

5.5.2 Evaluation - Mentoring

• Face evaluation example:

As described above in Section 5.5.1, Wang et al. [66] conducted a study that is an example of direct mentoring. For the face evaluation, they collected the trainee's and mentor's opinions about the system and asked them to choose answers such as "the technology was easy to setup and use" or "the technology was overly complex" on a five-point Likert scale.

• Content evaluation example:

The Ghostman work by Chinthammit et al. [84] provides good examples of statements that relate to the content evaluation. In this preliminary user study, they recruited 12 participants with no physical deficiencies. A five-point Likert scale of questions regarding the instructions or the overall training program that contributed to the learning of using chopsticks provided evidence for the content evaluation.

• Construct evaluation example:

As mentioned earlier, the work by Rojas et al. [65] is considered to be an indirect mentoring example that performed a construct evaluation. This is evident by comparing the scores between the different experienced subgroups, namely the low first responder experience (participants with

fewer than 10 years of experience) and low cricothyroido-tomy experience (participants with less than seven years of experience performing cricothyroidotomy as part of their training).

• Concurrent evaluation example:

Going back to the work by Wang et al. [66], recall that they assessed the concurrent evaluation by comparing the results of the HoloLens versus the full telemedicine setup (traditional method) by using evaluation metrics. The evaluation metrics include the global rating scale, trainee and mentor opinion, completion time, mental effort, and task difficulty ratings.

• Predictive evaluation example:

In the study conducted by Chinthammit et al. [84], the user experience questionnaire comprised a 2 (group) x 4 (test) mixed design ANOVA for the total skill error and the task completion time. The two groups refer to those who used Ghostman or the face-to-face (traditional instruction) method. The four tests contained statistics on the pretest, posttest, as well as 24 hours and seven days during the study.

5.5.3 Discussion - Mentoring

Similar to job shadowing, mentoring only has five studies allocated to this category. It is also difficult to contribute opinions regarding the trends of this method. Although the two may function in a similar manner, unlike job shadow-ing, which focuses on the demonstration of the ideal result, mentoring focuses more on the teacher–student relationship and how the teacher can guide the student to the ideal result. However, both methods have the potential to flourish the growth of the student/trainee if it is used correctly.

6 UTILIZATION OF AR

Thus far we have described existing ARTS for each training method in each field. In this section, we discuss how the characteristics of AR technology are utilized in training contexts, in a cross-disciplinary manner, based on the discussion up to this point.

According to Salas et al., training can be divided into four stages: information, demonstration, practice, and feedback. These stages are applied in terms of the content that is to be learned (i.e., training strategies) [33].

This section re-summarizes how AR works and what benefits it can bring to each stage.

The main training strategies and the utilization methods of AR adopted in each study are shown in Table 4. Each color indicates the training method employed in each paper. With a few exceptions, it can be seen that AR is utilized in different training strategies, but for the same training method.

6.1 Information

To obtain complex skill acquisition such as surgical skills, it is necessary to learn sufficient information that is relevant to each step in the task prior to practice. This is "learning" and it is generally performed by using textbooks and instructional videos. The effects of ARTS on learning are discussed in detail in Santos et al. [12]. They cite the multimedia

TABLE 4
Training Strategies applied for each study.
Color representation: Simulation , Programmed instruction ,
Games-based training , Job shadowing , and Mentoring .

Strategy	Description and Study	
Information	Display of educational content in relevant locations	
	[56], [60], [86]	
	[45], [54]	[95]
Demonstration	Presentation of ideal behavior	
	[64], [66], [67], [84]	[69], [78]
	[61], [63], [88]	[91]
Practice	Expanding the flexibility of training content by adding information to physical objects	
	[38], [39], [41], [46], [80]	
	[68], [69], [70], [75]	[92]
	Facilitation of the skill transfer through imitation of the actual	
	[14], [42], [43], [44], [45]	
	Task support by presenting relevant information	
	[16], [36], [41], [47], [49], [50], [51], [53]	
	[86], [87]	[83]
Feedback	Feedback on the performance, problems, and ways to improve	
	[39]	[59]
	[65]	
	Real-time feedback for motion compensation	
	[71], [72], [75], [76], [78], [82]	[41]
	Real-time visualization of invisible current conditions	
	[93]	

learning principle [112] and extend it to learning with an AR annotation. They state that "people learn better from annotated virtual words onto physical objects than from separate multimedia (e.g., illustrated manual) and physical objects," in terms of time contiguity. In terms of spatial contiguity, "people learn better when corresponding virtual words and physical objects are presented near rather than far from, each other on the screen." As an example of this, Sankaran et al. promoted the acquisition of necessary knowledge to prevent sepsis by displaying educational content near the relevant location in omnidirectional images [60].

6.2 Demonstration

According to Schmidt's schema theory, learned movements are not stored by individual concrete motor programs, but by abstracted schemas [3]. The construction of this schema requires a body schema (postural schema), which is a cognitive standard for intuition by knowing the current posture of the body, the positional relationships of the body parts, and how much each body part needs to be moved to perform a certain action. A body schema is unconscious, subjective, and it has body-centered spatial coordinates (i.e., first-person perspective) [113]. Therefore, it is expected that motor learning is more efficient to start with the first-person view "demonstration." AR-applied job shadowing (e.g., [61], [63]) is considered to be an effective method for acquiring body schematics because it allows for first-person observation of the (ideal) movements of the expert.

6.3 Practice

The nature of the application of AR to the practice is divided into three categories (which may be applicable to different examples).

6.3.1 Extend the flexibility of training content by adding information to the physical object

Even if a physical body model is elaborately made, the functions and degrees of freedom it possesses are limited. On the other hand, AR can provide versatile training conditions by superimposing the simulated organs model or the surface texture on the physical model. For instance, the medical trainees would need to change the trajectories of cutting the skin according to the shape of the overlaid organs in surgery.

6.3.2 Facilitation of skill transfer by imitating real objects and their superimposed display

When the training environment deviates from the real one, the acquired skill may be difficult to directly apply to the real environment. This is attributed to the enormous cognitive load that is involved in filling in the difference. Therefore, it is generally desirable for the training environment to resemble the real environment. The visual difference between the actual body and physical model can be compensated by imitating the texture of the real environment and superimposing it by using AR. This is expected to facilitate the application of skills that are acquired during training to the real environment.

6.3.3 Assistance in task execution through the presentation of task-related information

AR can reduce the cognitive load that is required for performing a task or it enables the user to perform the task more accurately by displaying cues that are related to the task execution at the relevant location on the real object. Abhari et al. assisted novice surgeons by superimposing a 3D trajectory of the pre-planned instruments on a display [51]. This is equivalent to the AR task support, but care should be taken when using this strategy in a training context. This information is not provided in the real task. If users rely on them, this can potentially lead to failure of the task execution in the real environment. For instance, Hulin et al. revealed that in programming-by-demonstration for an arm robot, the user's performance is impaired if visual effects are applied in the training phase [87]. Hulin et al. concluded that this is because users have come to rely too heavily on visual information during training. It should be used in combination with an appropriate skill transfer, such as training with less auxiliary information after becoming familiar with the task. One of the specific methods is in Section 6.4.2 under feedback.

6.4 Feedback

The effects of AR on this process are broadly classified into the following three categories.

6.4.1 Feedback on the performance, problems, and ways to improve

The simplest feedback is for the system to evaluate the trainee's behavior and its results. The feedback also provides the user the evaluation, the problem, and how to improve it. Along the time axis, this can be classified into real-time feedback and summary feedback. The real-time feedback is presented to the user almost immediately when the system finds a problem with the user's current state or behavior. The summary feedback is a method of presenting a summary of the overall evaluation and areas for improvement at the end of a section of the experiment or during a short break. According to Chollet et al., the former has a motivational maintenance effect, whereas the latter has a substantial effect on skill improvement in the context of public speaking training (note that the system in this paper is not ARTS) [114]. Zhao et al. proposed a CNN-based method for automatically evaluating the user performance in neonatal endotracheal intubation training [59]. In addition, based on this evaluation, the system generates and presents summary feedback, which is color-coded to indicate areas that need more practice.

6.4.2 Real-time feedback for motion compensation

One way to effectively use AR with real-time feedback is to present the difference between the optimal position/posture and the current ones of the user's body or the grasped object. Although this is effective in that the correction content can be intuitively understood, it must be designed so it does not lead to excessive user dependence. Sigrist et al. proposed a feedback method for oar pedaling training that combines sonification (the process of turning information into sounds [115]) of the difference between the current and optimal movements with visual information by using AR [76]. The feedback is designed to disappear when the user reaches the optimal state. They revealed that by iterating the training of constantly adjusting one's own state so that the feedback disappears, the skills improved even in the absence of feedback.

6.4.3 Real-time visualization of invisible current conditions

An example of special real-time feedback is biofeedback, which aims to control one's own internal state (e.g., calming down). This can be interpreted as replacing a skill whose acquisition process is unclear with a different task that is easier to perform: self-regulation of physiological index values (e.g., brain waves and heartbeat) that are presented visually in real-time to meet certain criteria. During this training, it is important to associate the display information with the control target. For example, an attempt has been made to strengthen this connection by using AR to display electroencephalogram (EEG) information on the head of a mirror image of oneself [93]. The authors stated that this can also be used to train users to learn how to coordinate brain activity in some areas of the brain.

7 CONCLUSION

7.1 Overall Discussion

Training plays a crucial role in today's society as a means of developing competent and productive people in the

workforce. With the goal of improving human performance, training serves as the medium for creating opportunities of nurturing and flourishing knowledge, skills, and abilities. There are many forms in which training can be successfully carried out. One such delivery that is proving to be effective is computer-mediated training, with an emphasis on AR training.

To answer R1 (What are the training methods for implementing ARTS and how are these methods used in the different fields?), we adopted Martin et al.'s categorization of training methods, and determined five methods for which AR is effective. Simulation training is particularly valuable when targeting operations that are very dangerous when they are executed in a real-life environment. Programmed instruction proves its usefulness when considering the efficacy of imparting substantial information and knowledge. Games-based training displays its advantage by inspiring motivation from users to induce learning. Job shadowing excels in the demonstration of an ideal performance, whereas mentoring facilitates the mentor-student relationship for effective knowledge and skill handover. We also found trends of these training methods across the different fields, such as high simulation studies for medical, and high games-based and no programmed instruction studies for rehabilitation.

To answer R2 (What types of evaluations are used to assess the existing ARTS and how do they contribute to the validity of each training method?), we described Barsom et al.'s validity types and used them as reference to standardize the evaluation of ARTS. For a training system to be ready for implementation in a real-life environment, Barsom et al. suggested that a "full validation process" needs to be completed. Only when the face, content, construct, concurrent, and predictive aspects of the system are evaluated, this can be considered as the ideal ARTS that can be used and deployed in real practice. Our survey results show that no study has achieved the full validation process; however, we have identified which evaluation types are considered to be more prioritized depending on the training method and strategy that was used.

To answer R3 (How is AR utilized for each training strategy?), we looked at the trends between the training methods and evaluations across the different fields. From these generalizations, we were able to discuss how AR is used based on Salas et al.'s strategy for training and gain an understanding of AR's utilization in relation to the different training methods. The first strategy was to convey the information, which was performed mainly by placing the educational content in the relevant locations. We found that this strategy was used by methods that needed to build up on the basic knowledge and foundation of the learner. The second strategy was to demonstrate the desired behavior, cognition, and attitudes so that learners can imagine the ideal result of their training. This strategy was used by methods that already established the basics, and have learners who want to go to the next step of understanding, seeing from the point-of-view of the expert (i.e., job shadowing and mentoring). The third strategy was to create opportunities to practice the knowledge, skills, and abilities learnt. This strategy was where simulation training methods were most applied as simulation studies tried to create rehearsal stages akin to that of the real environment

or work scenario. The fourth strategy was to give feedback to the relevant things that aid the users in learning. Many of them presented information that intuitively guided the trainee's current body movements to the appropriate ones. This strategy was applied more to games-based training because an appropriate feedback presentation led to the gamification of training.

7.2 Findings and Suggestions

Before reading each full-text article, we presumed that only a limited number of studies would perform a predictive evaluation because it has a high evaluation cost owing to the need for the assessment to be realized in several steps. However, we were pleasantly surprised to find quite a few studies that went out of their way and evaluated their training system in an intricate manner. This is an important point as a predictive evaluation reflects the training effects gained from the system to the actual practice itself, which proves skills acquisition. Similarly, we noticed that more than two-thirds of the studies conducted their evaluations using actual users. Recruiting actual ARTS users as test subjects can sometimes be expensive; hence, the usage of alternative users can be the more practical option. Although not always the case, alternative users sometimes cannot fully reflect assessment results that interpret the performance to actual practice. From these two points, one can argue that the current evaluation design of ARTS is leaning in a positive direction.

The results of our survey indicated that only a handful of studies are focused on job shadowing and mentoring training methods. However, we recommend future ARTS researchers to focus more on this area because there are still numerous concepts and ideas to potentially uncover, which further proves its effectiveness in the application of training. Presently, we have only identified five out of 13 of Martin et al.'s training methods that are effective with AR according to the results of our study. However, this does not mean that the remaining methods are not applicable with AR technology. This merely suggests that no studies have implemented these methods yet. There is still much potential for AR to be implemented by using other methods, maybe even new methods that are not listed in Martin et al.'s core training methods.

We also noticed studies that explicitly performed Barsom et al.'s validity types usually come from the medical field. Barsom et al.'s validity types is becoming (if not already) a standard procedure in the medical field. Although the roots of the test validity concept are from the research design in behavioral and social sciences, adopting this concept in engineering has merits. This is particularly true for studies that handle people as subjects and performance indicators (e.g., ARTS). From this perspective, we can recommend that future ARTS research adopt Barsom et al.'s validity as an option for standardizing evaluations of training systems, regardless of the application field. When the standardization of the procedure for doing evaluations is realized, the next step is to strive for Barsom et al.'s full validation process. It is indisputable that proving the novelty and validity of an idea is paramount in the research community. However, it can also be pointed out that the implementation in actual practice is of great significance. To quote the famous

entrepreneur Scott Belsky, "it's not about ideas, it's about making ideas happen." The true value of a training system becomes clear once it is assessed with a full validation; hence, we recommend that future ARTS not only focus on lab performance (i.e., implementation situated in ideal conditions), but also extend investigations in non-ideal scenarios and check how it will fare in actual implementation.

7.3 Final Thoughts

We have confirmed the use of ARTS and its current implementations, with an emphasis in the medical, rehabilitation/exercise, and industrial fields. In the future, we believe that ARTS will be utilized more in a wider variety of application fields, particularly for training that is deemed to be hazardous or complex when it is performed in conventional practices. As mentioned above, future studies on ARTS should try to evaluate the totality of the system by following the full validation process recommended by Barsom et al. As the current implementations of ARTS have heterogeneity in their study designs, it is difficult to appraise in a quantitative manner and perform statistical pooling of the data. We recommend that future ARTS implementations classify their work based on Martin et al.'s definition of training methods, while considering the utilization of AR in the training context, as described in Section 6. In addition, it would benefit our community to have more investigations performed on the training effects of the less frequently used methods, which include job shadowing and mentoring. When future ARTS studies follow a more structured approach, future review papers can focus more on the statistical effects of ARTS, such as conducting meta-analysis reviews.

REFERENCES

[1] B. O. Martin, K. Kolomitro, and T. C. M. Lam, "Training methods: A review and analysis," *Hum. Resour. Dev. Rev.*, vol. 13, no. 1, pp. 11–35, Oct. 2013.

[2] P. M. Fitts, "Perceptual-motor skill learning," in *Categories of Human Learning*, A. W. Melton, Ed. Academic Press, 1964, pp. 243–285.

[3] R. A. Schmidt, "A schema theory of discrete motor skill learning," *Psychol. Rev.*, vol. 82, no. 4, pp. 225–260, Jul. 1975.

[4] S. Naidu, J. Barrett, and P. Olsen, "Improving instructional effectiveness with computer-mediated communication," *Res. Learn. Technol.*, vol. 3, no. 2, pp. 63–75, 1995.

[5] B. Babbitt, W. Pieper, C. Semple, and D. Swanson, "Effectiveness and efficiency of computer-based training: Recent research findings," in *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 29, no. 2, Oct. 1985, pp. 198–201.

[6] E. Z. Barsom, M. Graafland, and M. P. Schijven, "Systematic review on the effectiveness of augmented reality applications in medical training," *Surg. Endosc.*, vol. 30, no. 10, pp. 4174–4183, Feb. 2016.

[7] J. M. Black, *Developing Competent Subordinates*. New York: American Management Association, 1961.

[8] I. L. Goldstein and J. K. Ford, *Training in Organizations: Needs Assessment, Development, and Evaluation*, 4th ed. Belmont: Wadsworth, 2002.

[9] U. V. Somasundaram and T. M. Egan, "Training and development: An examination of definitions and dependent variables," in *Academy of Human Resource Development Int. Conf. (AHRD)*, Mar. 2004, pp. 850–857.

[10] T. P. Caudell and D. W. Mizell, "Augmented reality: An application of heads-up display technology to manual manufacturing processes," in *Proc. 25th Hawaii Int. Conf. System Sciences*, vol. 2, Jan. 1992, pp. 659–669.

[11] R. T. Azuma, "A survey of augmented reality," *Presence (Camb)*, vol. 6, no. 4, pp. 355–385, Aug. 1997.

[12] M. E. C. Santos, A. Chen, T. Taketomi, G. Yamamoto, J. Miyazaki, and H. Kato, "Augmented reality learning experiences: Survey of prototype design and evaluation," *IEEE Trans. Learn. Technol.*, vol. 7, no. 1, pp. 38–56, Dec. 2013.

[13] C. E. Hughes, C. B. Stapleton, D. E. Hughes, and E. M. Smith, "Mixed reality in education, entertainment, and training," *IEEE Comput. Graph. Appl.*, vol. 25, no. 6, pp. 24–30, Nov. 2005.

[14] A. Javaux *et al.*, "A mixed-reality surgical trainer with comprehensive sensing for fetal laser minimally invasive surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 12, pp. 1949–1957, Jul. 2018.

[15] S. Jeon, S. Choi, and M. Harders, "Rendering virtual tumors in real tissue mock-ups using haptic augmented reality," *IEEE Trans. Haptics*, vol. 5, no. 1, pp. 77–84, Jan. 2012.

[16] C. Sutherland, K. Hashtrudi-Zaad, R. Sellens, P. Abolmaesumi, and P. Mousavi, "An augmented reality haptic training simulator for spinal needle procedures," *IEEE. Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3009–3018, Dec. 2012.

[17] Y. Kim, H. Kim, and Y. O. Kim, "Virtual reality and augmented reality in plastic surgery: A review," *Arch. Plast. Surg.*, vol. 44, no. 3, pp. 179–187, May 2017.

[18] M. R. Desselle, R. A. Brown, A. R. James, M. J. Midwinter, S. K. Powell, and M. A. Woodruff, "Augmented and virtual reality in surgery," *Comput. Sci. Eng.*, vol. 22, no. 3, p. 18–26, Feb. 2020.

[19] J. Roessel, M. Knoell, J. Hofmann, and R. Buettner, "A systematic literature review of practical virtual and augmented reality solutions in surgery," in *2020 IEEE 44th Annu. Computers, Software, and Applications Conf. (COMPSAC)*, Jul. 2020, pp. 489–498.

[20] T. Pisoni *et al.*, "AUSILIA: Assisted unit for simulating independent living activities," in *2016 IEEE Int. Smart Cities Conf. (ISC2)*, Sep. 2016, pp. 1–4.

[21] L. v. Koch, A. W. Wottrich, and L. W. Holmqvist, "Rehabilitation in the home versus the hospital: The importance of context," *Disabil. Rehabil.*, vol. 20, no. 10, pp. 367–372, Oct. 1998.

[22] S. Blomqvist, S. Seipel, and M. Engström, "Using augmented reality technology for balance training in the older adults: A feasibility pilot study," *BMC Geriatr.*, vol. 21, no. 1, Feb. 2021, Art. no. 144.

[23] R. Palmarini, J. A. Erkoyuncu, R. Roy, and H. Torabmostaedi, "A systematic review of augmented reality applications in maintenance," *Robot. Comput. Integr. Manuf.*, vol. 49, pp. 215–228, Feb. 2018.

[24] M. Gattullo, A. Evangelista, A. E. Uva, M. Fiorentino, and J. Gabbard, "What, how, and why are visual assets used in industrial augmented reality? A systematic review and classification in maintenance, assembly, and training (from 1997 to 2019)," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 2, pp. 1443–1456, Aug. 2020.

[25] I. Bianchi, A. L. Zanatta, and R. Rieder, "Augmented reality in medical teaching-learning process content: A systematic review," in *2020 22nd Symp. Virtual and Augmented Reality (SVR)*, Nov. 2020, pp. 129–133.

[26] C. Papakostas, C. Troussas, A. Krouska, and C. Sgouropoulou, "Exploration of augmented reality in spatial abilities training: A systematic literature review for the last decade," *Inform. Educ.*, vol. 20, no. 1, pp. 107–130, Mar. 2021.

[27] S. Hafsa, M. A. Majid, and R. M. Tawafak, "Learnability factors of AR usage performance: Validating through survey," in *2021 Int. Conf. Software Engineering and Computer Systems and 4th Int. Conf. Computational Science and Information Management (ICSECS-ICOCSIM)*, Aug. 2021, pp. 371–376.

[28] A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. Swan, "A systematic review of 10 years of augmented reality usability studies: 2005 to 2014," *Front. Robot. AI*, vol. 5, Apr. 2018, Art. no. 37.

[29] B. H. Limbu, H. Jarodzka, R. Klemke, and M. Specht, "Using sensors and augmented reality to train apprentices using recorded expert performance: A systematic literature review," *Educ. Res. Rev.*, vol. 25, pp. 1–22, Nov. 2018.

[30] A. A. Nelson, "Research design: Measurement, reliability, and validity," *Am. J. Health Syst. Pharm.*, vol. 37, no. 6, pp. 851–857, Jun. 1980.

[31] I. A. Chiang, R. S. Jhangiani, and P. C. Price, *Research Methods in Psychology*, 2nd ed.   BCcampus, Oct. 2015. [Online]. Available: https://opentextbc.ca/researchmethods

[32] Y. Brij and H. Belhadaoui, "Virtual and augmented reality in school context: A literature review," in *2021 3rd Int. Conf. Transportation and Smart Technologies (TST)*, May 2021, pp. 16–23.

[33] E. Salas, S. I. Tannenbaum, K. Kraiger, and K. A. Smith-Jentsch, "The science of training and development in organizations," *Psychol. Sci. Public Interest*, vol. 13, no. 2, pp. 74–101, Jun. 2012.

[34] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *Brit. Med. J.*, vol. 372, Mar. 2021, Art. no. n71.

[35] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 46, pp. 16 569–16 572, Nov. 2005. [Online]. Available: https://www.pnas.org/content/102/46/16569

[36] J. W. Sappenfield *et al.*, "Visualization improves supraclavicular access to the subclavian vein in a mixed reality simulator," *Anesth. Analg.*, vol. 127, no. 1, pp. 83–89, Jul. 2018.

[37] A. Alaraj *et al.*, "Role of cranial and spinal virtual and augmented reality simulation using immersive touch modules in neurosurgical training," *Neurosurgery*, vol. 72, no. S1, pp. 115–123, Jan. 2013.

[38] T. Halic, S. Kockara, C. Bayrak, and R. Rowe, "Mixed reality simulation of rasping procedure in artificial cervical disc replacement (ACDR) surgery," *BMC Bioinform.*, vol. 11, no. S6, Oct. 2010, Art. no. S11.

[39] S. M. Botden, S. N. Buzink, M. P. Schijven, and J. J. Jakimowicz, "Augmented versus virtual reality laparoscopic simulation: What is the difference?" *World J. Surg.*, vol. 31, no. 4, pp. 764–772, Mar. 2007.

[40] Z. Qin *et al.*, "Towards virtual VATS, face, and construct evaluation for peg transfer training of box, VR, AR, and MR trainer," *J. Healthc. Eng.*, vol. 2019, Jan. 2019, Art. no. 6813719.

[41] T. Muangpoon, R. Haghighi Osgouei, D. Escobar-Castillejos, C. Kontovounisios, and F. Bello, "Augmented reality system for digital rectal examination training and assessment: System validation," *J. Med. Internet Res.*, vol. 22, no. 8, Aug. 2020, Art. no. e18637.

[42] S. Condino *et al.*, "How to build a patient-specific hybrid simulator for orthopaedic open surgery: Benefits and limits of mixed-reality using the Microsoft HoloLens," *J. Healthc. Eng.*, vol. 2018, Nov. 2018, Art. no. 5435097.

[43] S. Balian, S. K. McGovern, B. S. Abella, A. L. Blewer, and M. Leary, "Feasibility of an augmented reality cardiopulmonary resuscitation training system for health care providers," *Heliyon*, vol. 5, no. 8, Aug. 2019, Art. no. e02205.

[44] P. L. Ingrassia *et al.*, "Augmented reality learning environment for basic life support and defibrillation training: Usability study," *J. Med. Internet Res.*, vol. 22, no. 5, May 2020, Art. no. e14910.

[45] M. Aebersold *et al.*, "Interactive anatomy-augmented virtual simulation training," *Clin. Simul. Nurs.*, vol. 15, no. 1, pp. 34–41, Feb. 2018.

[46] T. Wright, S. Ribaupierre, and R. Eagleson, "Design and evaluation of an augmented reality simulator using Leap Motion," *Healthc. Technol. Lett.*, vol. 4, no. 5, pp. 210–215, Sep. 2017.

[47] M. G. Boyaci *et al.*, "Augmented reality supported cervical transpedicular fixation on 3D-printed vertebrae model: An experimental education study," *Turk. Neurosurg.*, vol. 30, no. 6, pp. 937–943, Aug. 2020.

[48] D. Magee, Y. Zhu, R. Ratnalingam, P. Gardner, and D. Kessel, "An augmented reality simulator for ultrasound guided needle placement training," *Med. Biol. Eng. Comput.*, vol. 45, no. 10, pp. 957–967, Jul. 2007.

[49] A. Chowriappa *et al.*, "Augmented-reality-based skills training for robot-assisted urethrovesical anastomosis: A multi-institutional randomised controlled trial," *BJU Int.*, vol. 115, no. 2, pp. 336–345, Aug. 2014.

[50] V. Ferrari *et al.*, "Augmented reality visualization of deformable tubular structures for surgical simulation," *Int. J. Med. Robot.*, vol. 12, no. 2, pp. 231–240, Jul. 2015.

[51] K. Abhari *et al.*, "Training for planning tumour resection: Augmented reality and human factors," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 6, pp. 1466–1477, Jun. 2015.

[52] B. Laha, D. A. Bowman, and J. D. Schiffbauer, "Validation of the MR simulation approach for evaluating the effects of immersion on visual analysis of volume data," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 4, pp. 529–538, Apr. 2013.

[53] C. T. Yeo, T. Ungi, P. U-Thainual, A. Lasso, R. C. McGraw, and G. Fichtinger, "The effect of augmented reality training on percutaneous needle placement in spinal facet joint injections," *IEEE. Trans. Biomed. Eng.*, vol. 58, no. 7, pp. 2031–2037, Jul. 2011.

[54] R. Serrano Vergel, P. Morillo Tena, S. Casas Yrurzum, and C. Cruz-Neira, "A comparative evaluation of a virtual reality table and a HoloLens-based augmented reality system for anatomy training," *IEEE Trans. Hum. Mach. Syst.*, vol. 50, no. 4, pp. 337–348, Jun. 2020.

[55] H. Ghandorh, J. Mackenzie, R. Eagleson, and S. de Ribaupierre, "Development of augmented reality training simulator systems for neurosurgery using model-driven software engineering," in *2017 IEEE 30th Canadian Conf. Electrical and Computer Engineering (CCECE)*, Apr. 2017, pp. 1–6.

[56] J. Ferrer-Torregrosa, M. A. Jiménez-Rodríguez, J. Torralba-Estelles, F. Garzón-Fariñós, M. Pérez-Bermejo, and N. Fernández-Ehrling, "Distance learning ECTS and flipped classroom in the anatomy learning: Comparative study of the use of augmented reality, video and notes," *BMC Med. Educ.*, vol. 16, no. 1, Sep. 2016, Art. no. 230.

[57] D. S. Schoeb *et al.*, "Mixed reality for teaching catheter placement to medical students: A randomized single-blinded, prospective trial," *BMC Med. Educ.*, vol. 20, no. 1, Dec. 2020, Art. no. 510.

[58] P. Bifulco, F. Narducci, R. Vertucci, P. Ambruosi, M. Cesarelli, and M. Romano, "Telemedicine supported by augmented reality: An interactive guide for untrained people in performing an ECG test," *Biomed. Eng. Online*, vol. 13, no. 1, Nov. 2014, Art. no. 153. [Online]. Available: http://www.biomedical-engineering-online.com/content/13/1/153

[59] S. Zhao *et al.*, "An intelligent augmented reality training framework for neonatal endotracheal intubation," in *2020 IEEE Int. Symp. Mixed and Augmented Reality (ISMAR)*, Nov. 2020, pp. 672–681.

[60] N. K. Sankaran *et al.*, "Efficacy study on interactive mixed reality (IMR) software with sepsis prevention medical education," in *2019 IEEE Conf. Virtual Reality and 3D User Interfaces (VR)*, Mar. 2019, pp. 664–670.

[61] A. M. Jarc, A. A. Stanley, T. Clifford, I. S. Gill, and A. J. Hung, "Proctors exploit three-dimensional ghost tools during clinical-like training scenarios: A preliminary study," *World J. Urol.*, vol. 35, no. 6, pp. 957–965, Sep. 2016.

[62] J. Quarles, S. Lampotang, I. Fischler, P. Fishwick, and B. Lok, "Collocated AAR: Augmenting after action review with mixed reality," in *2008 7th IEEE/ACM Int. Symp. Mixed and Augmented Reality (ISMAR)*, Sep. 2008, pp. 107–116.

[63] T. Sielhorst, T. Blum, and N. Navab, "Synchronizing 3D movements for quantitative comparison and simultaneous visualization of actions," in *4th IEEE/ACM Int. Symp. Mixed and Augmented Reality (ISMAR)*, Oct. 2005, pp. 38–47.

[64] D. Andersen *et al.*, "Medical telementoring using an augmented reality transparent display," *Surgery*, vol. 159, no. 6, pp. 1646–1653, Jan. 2016.

[65] E. Rojas-Muñoz *et al.*, "Evaluation of an augmented reality platform for austere surgical telementoring: A randomized controlled crossover study in cricothyroidotomies," *NPJ Digit. Med.*, vol. 3, no. 1, May 2020, Art. no. 75.

[66] S. Wang *et al.*, "Augmented reality as a telemedicine platform for remote procedural training," *Sensors*, vol. 17, no. 10, Oct. 2017, Art. no. 2294.

[67] C. Lin *et al.*, "How about the mentor? Effective workspace visualization in AR telementoring," in *2020 IEEE Conf. Virtual Reality and 3D User Interfaces (VR)*, Mar. 2020, pp. 212–220.

[68] J. P. O. Held *et al.*, "Augmented reality–based rehabilitation of gait impairments: Case report," *JMIR mhealth uhealth*, vol. 8, no. 5, May 2020, Art. no. e17804.

[69] M. Chen, Q. Tang, S. Xu, P. Leng, and Z. Pan, "Design and evaluation of an augmented reality-based exergame system to reduce fall risk in the elderly," *Int. J. Environ. Res. Public Health*, vol. 17, no. 19, Oct. 2020, Art. no. 7208.

[70] J. Trojan *et al.*, "An augmented reality home-training system based on the mirror training and imagery approach," *Behav. Res. Methods*, vol. 46, no. 3, pp. 634–640, Dec. 2013.

[71] N. Lehrer, Y. Chen, M. Duff, S. L Wolf, and T. Rikakis, "Exploring the bases for a mixed reality stroke rehabilitation system, part II: Design of interactive feedback for upper limb rehabilitation," *J. Neuroeng. Rehabilitation*, vol. 8, no. 1, Sep. 2011, Art. no. 54.

[72] T. Braga Rodrigues, C. Ó Catháin, N. E. O'Connor, and N. Murray, "A quality of experience assessment of haptic and augmented reality feedback modalities in a gait analysis system," *PLoS One*, vol. 15, no. 3, Mar. 2020, Art. no. e0230570.

[73] P.-J. Chen, I.-W. Penn, S.-H. Wei, L.-R. Chuang, and W.-H. Sung, "Augmented reality-assisted training with selected Tai-Chi movements improves balance control and increases lower limb muscle strength in older adults: A prospective randomized trial," *J. Exerc. Sci. Fit.*, vol. 18, no. 3, pp. 142–147, May 2020.

[74] D. J. Im *et al.*, "Utility of a three-dimensional interactive augmented reality program for balance and mobility rehabilitation in the elderly: A feasibility study," *Ann. Rehabil. Med.*, vol. 39, no. 3, pp. 462–472, Jun. 2015.

[75] A. J. Espay *et al.*, "At-home training with closed-loop augmented-reality cueing device for improving gait in patients with Parkinson disease," *J. Rehabil. Res. Dev.*, vol. 47, no. 6, pp. 573–582, Aug. 2010.

[76] R. Sigrist, G. Rauter, L. Marchal-Crespo, R. Riener, and P. Wolf, "Sonification and haptic feedback in addition to visual feedback enhances complex motor task learning," *Exp. Brain Res.*, vol. 233, no. 3, pp. 909–925, Dec. 2014.

[77] E. Park *et al.*, "Effects of a mixed reality-based cognitive training system compared to a conventional computer-assisted cognitive training system on mild cognitive impairment: A pilot study," *Cogn. Behav. Neurol.*, vol. 32, no. 3, pp. 172–178, Sep. 2019.

[78] P.-Y. Kao, P.-H. Han, Y.-F. Jan, Z.-W. Yang, C.-H. Li, and Y.-P. Hung, "On learning weight distribution of Tai Chi Chuan using pressure sensing insoles and MR-HMD," in *2019 IEEE Conf. Virtual Reality and 3D User Interfaces (VR)*, Mar. 2019, pp. 1457–1464.

[79] A. U. Batmaz, A. K. Mutasim, M. Malekmakan, E. Sadr, and W. Stuerzlinger, "Touch the wall: Comparison of virtual and augmented reality with conventional 2D screen eye-hand coordination training systems," in *2020 IEEE Conf. Virtual Reality and 3D User Interfaces (VR)*, Mar. 2020, pp. 184–193.

[80] S. Kim, J. Ryu, Y. Choi, Y. Kang, H. Li, and K. Kim, "Eye-contact game using mixed reality for the treatment of children with attention deficit hyperactivity disorder," *IEEE Access*, vol. 8, pp. 45 996–46 006, Mar. 2020.

[81] K. Johnsen *et al.*, "Mixed reality virtual pets to reduce childhood obesity," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 4, pp. 523–530, Apr. 2014.

[82] M. Duff *et al.*, "An adaptive mixed reality training system for stroke rehabilitation," *IEEE Trans. Neural Syst. Rehabilitation Eng.*, vol. 18, no. 5, pp. 531–541, Oct. 2010.

[83] M. Zolotas, J. Elsdon, and Y. Demiris, "Head-mounted augmented reality for explainable robotic wheelchair assistance," in *2018 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Oct. 2018, pp. 1823–1829.

[84] W. Chinthammit *et al.*, "Ghostman: Augmented reality application for telerehabilitation and remote instruction of a novel motor skill," *BioMed Res. Int.*, vol. 2014, Apr. 2014, Art. no. 646347.

[85] A. Vargas Gonzalez *et al.*, "A comparison of desktop and augmented reality scenario based training authoring tools," in *2019 IEEE Int. Symp. Mixed and Augmented Reality (ISMAR)*, Oct. 2019, pp. 339–350.

[86] P. Horejsi, K. Novikov, and M. Simon, "A smart factory in a smart city: Virtual and augmented reality in a smart assembly line," *IEEE Access*, vol. 8, pp. 94 330–94 340, May 2020.

[87] T. Hulin, V. Schmirgel, E. Yechiam, U. E. Zimmermann, C. Preusche, and G. Pohler, "Evaluating exemplary training accelerators for programming-by-demonstration," in *19th Int. Symp. Robot and Human Interactive Communication (RO-MAN)*, Sep. 2010, pp. 440–445.

[88] M. Goto, Y. Uematsu, H. Saito, S. Senda, and A. Iketani, "Task support system by displaying instructional video onto AR workspace," in *2010 IEEE Int. Symp. Mixed and Augmented Reality (ISMAR)*, Oct. 2010, pp. 83–90.

[89] D. Sportillo, A. Paljic, and L. Ojeda, "On-road evaluation of autonomous driving training," in *2019 14th ACM/IEEE Int. Conf. Human-Robot Interaction (HRI)*, Mar. 2019, pp. 182–190.

[90] R. C. Sari, M. Sholihin, N. Yuniarti, I. A. Purnama, and H. D. Hermawan, "Does behavior simulation based on augmented reality improve moral imagination?" *Educ. Inf. Technol. (Dordr)*, vol. 26, no. 1, pp. 441–463, Jul. 2020.

[91] J. R. Keebler, T. J. Wiltshire, D. C. Smith, S. M. Fiore, and J. S. Bedwell, "Shifting the paradigm of music instruction: Implica-

tions of embodiment stemming from an augmented reality guitar learning system," *Front. Psychol.*, vol. 5, May 2014, Art. no. 471.

[92] J. M. Andujar, A. Mejias, and M. A. Marquez, "Augmented reality for the improvement of remote laboratories: An augmented remote laboratory," *IEEE Trans. Educ.*, vol. 54, no. 3, pp. 492–500, Aug. 2011.

[93] J. Mercier-Ganady, F. Lotte, E. Loup-Escande, M. Marchal, and A. Lecuyer, "The Mind-Mirror: See your brain in action in your head using EEG and augmented reality," in *2014 IEEE Virtual Reality (VR)*, Mar. 2014, pp. 33–38.

[94] C. Zhou, H. Li, and Y. Bian, "Identifying the optimal 3D display technology for hands-on virtual experiential learning: A comparison study," *IEEE Access*, vol. 8, pp. 73 791–73 803, Apr. 2020.

[95] F. Munoz-Montoya, M. Juan, M. Mendez-Lopez, and C. Fidalgo, "Augmented reality based on SLAM to assess spatial short-term memory," *IEEE Access*, vol. 7, pp. 2453–2466, Dec. 2018.

[96] G. A. Lee, S. Ahn, W. Hoff, and M. Billinghurst, "Enhancing first-person view task instruction videos with augmented reality cues," in *2020 IEEE Int. Symp. Mixed and Augmented Reality (ISMAR)*, Nov. 2020, pp. 498–508.

[97] L. Ada, S. Dorsch, and C. G. Canning, "Strengthening interventions increase strength and improve activity after stroke: A systematic review," *J. Physiother.*, vol. 52, no. 4, pp. 241–248, Feb. 2006.

[98] G. Kielhofner, *Conceptual Foundations of Occupational Therapy Practice*, 4th ed. Philadelphia: F.A. Davis Co., 2009.

[99] N. Friedman *et al.*, "Retraining and assessing hand movement after stroke using the MusicGlove: Comparison with conventional hand therapy and isometric grip training," *J. Neuroeng. Rehabilitation*, vol. 11, no. 1, Apr. 2014, Art. no. 76.

[100] D. Schmalstieg and T. Höllerer, *Augmented Reality: Principles and Practice*. Boston: Addison-Wesley, Jun. 2016.

[101] A. G. Gallagher *et al.*, "Virtual reality simulation for the operating room: Proficiency-based training as a paradigm shift in surgical skills training," *Ann. Surg.*, vol. 241, no. 2, pp. 364–372, Feb. 2005.

[102] B. Sinha, "Programmed instruction as a teaching strategy," *Int. Educ. Sci. Res. J.*, vol. 5, no. 6, pp. 34–38, Jun. 2019.

[103] M. Molenda, "The programmed instruction era: When effectiveness mattered," *TechTrends*, vol. 52, no. 2, pp. 52–58, May 2008.

[104] R. Culatta and G. Kearsley, "Sequencing of instruction," InstructionalDesign.org, accessed 28 June 2021. [Online]. Available: https://www.instructionaldesign.org/concepts/sequence-instruction

[105] S. M. Glynn and F. J. di Vesta, "Outline and hierarchical organization as aids for study and retrieval." *J. Educ. Psychol.*, vol. 69, no. 2, pp. 89–95, 1977.

[106] R. F. Lorch and E. P. Lorch, "Topic structure representation and text recall." *J. Educ. Psychol.*, vol. 77, no. 2, pp. 137–148, Apr. 1985.

[107] J. Van Patten, C. Chao, and C. M. Reigeluth, "A review of strategies for sequencing and synthesizing instruction," *Rev. Educ. Res.*, vol. 56, no. 4, pp. 437–471, Dec. 1986.

[108] M. D. Merrill, "Component display theory," in *Instructional Design Theories and Models*, C. Reigeluth, Ed. Hillsdale, New Jersey: Erlbaum Associates, 1983.

[109] M. Sailer, J. U. Hense, S. K. Mayr, and H. Mandl, "How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction," *Comput. Hum. Behav.*, vol. 69, pp. 371–380, Apr. 2017.

[110] K. Bovermann and T. J. Bastiaens, "Towards a motivational design? Connecting gamification user types and online learning activities," *Res. Pract. Technol. Enhanc. Learn.*, vol. 15, Jan. 2020, Art. no. 1.

[111] R. S. Alsawaier, "The effect of gamification on motivation and engagement," *Int. J. Inf. Learn. Technol.*, vol. 35, no. 1, pp. 56–79, Jan. 2018.

[112] R. E. Mayer, *Multimedia Learning*, 2nd ed. Cambridge: Cambridge University Press, 2009.

[113] J. Paillard, "Vectorial versus configural encoding of body space," in *Body Image and Body Schema*, V. Knockaert and H. D. Preester, Eds. Amsterdam: John Benjamins Publishing Co., 2005, pp. 89–109.

[114] M. Chollet, S. Marsella, and S. Scherer, "Training public speaking with virtual social interactions: Effectiveness of real-time feedback and delayed feedback," *J. Multimodal User Interfaces*, vol. 16, pp. 17–29, Mar. 2021.

[115] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*. Berlin: Logos Verlag, 2011.

Isidro M. Butaslac, III received his B.S. degree in electronics and communications engineering from Ateneo de Manila University (ADMU) in the Philippines in 2016 and his M.S. degree in engineering from Nara Institute of Science and Technology (NAIST), Nara, Japan, in 2020. He is currently pursuing his Ph.D. degree in engineering at NAIST, Nara, Japan. His research interests include augmented reality and human–computer interaction.

Yuichiro Fujimoto (M'14) received his B.S. degree in engineering from Osaka University, Osaka, Japan, in 2010, and his M.S. and Ph.D. degrees in engineering from Nara Institute of Science and Technology (NAIST), Nara, Japan, in 2012 and 2015, respectively. From 2015 to 2018, he was an Assistant Professor in the Engineering Department at Tokyo University of Agriculture and Technology, Japan. He is currently an Assistant Professor in the Division of Information Science at NAIST, Japan. His research interests include analyzing humans from 3D information and spatial augmented reality.

Taishi Sawabe received his B.E. in biomimetics robots from Ritsumeikan University in 2014, and his M.S. and Ph.D. degrees from Nara Institute of Science and Technology (NAIST), Nara, Japan, in 2019 with a focus on autonomous vehicles with passenger comforts. Subsequently, from 2019 to 2020, he was a Postdoctoral Researcher in the IMD-Lab at NAIST. He is currently an Assistant Professor in the Division of Information Science, NAIST, Japan. He is also the CEO of the start-up company Amirobo Tech Inc. His research interests include human–robot interaction, human comforts, and autonomous vehicles.

Masayuki Kanbara received his Ph.D. degree in engineering from Nara Institute of Science and Technology (NAIST) in 2002. He was an Assistant Professor in the Information Science Department at NAIST from 2002 to 2010. He was a Visiting Researcher at the University of California, Santa Barbara in 2008–2009. He has been an Associate Professor at NAIST since 2010. His fields of research are focused on augmented reality, computer vision, and human-robot interaction.

Hirokazu Kato is a Professor at Nara Institute of Science and Technology (NAIST), Japan. He received his B.E., M.E., and Dr. Eng. degrees from Osaka University in Japan in 1986, 1988, and 1996, respectively. He joined the Faculty of Engineering Science at Osaka University from 1989 to 1999. In 1999, he joined the Faculty of Information Sciences at Hiroshima City University in Japan. In 2003, he joined the Graduate School of Engineering Science at Osaka University. Since 2007, he has been with the Graduate School of Information Science at NAIST. He studied pattern recognition and computer vision. When he joined the Human Interface Technology Laboratory (HIT Lab) at the University of Washington as a Visiting Scholar in 1998, he started conducting research on augmented reality, which is currently his main research topic. He received the Virtual Reality Technical Achievement Award from IEEE VGTC in 2009. He also received the Lasting Impact Award at the 11th IEEE International Symposium on Mixed and Augmented Reality in 2012.