



A review of the use of virtual reality head-mounted displays in education and training

Jensen, Lasse; Konradsen, Flemming

Published in:
Education and Information Technologies

DOI:
[10.1007/s10639-017-9676-0](https://doi.org/10.1007/s10639-017-9676-0)

Publication date:
2018

Document version
Peer reviewed version

Citation for published version (APA):
Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23(4), 1515-1529. <https://doi.org/10.1007/s10639-017-9676-0>

This is a post-peer-review, pre-copyedit version of an article published in *Education and Information Technologies*. The final authenticated version is available online at: <http://dx.doi.org/10.1007/s10639-017-9676-0>

A review of the use of virtual reality head-mounted displays in education and training

Lasse X Jensen¹ & Flemming Konradsen²

¹ Centre for Online and Blended Learning, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

² Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

Abstract

In the light of substantial improvements to the quality and availability of virtual reality (VR) hardware seen since 2013, this review seeks to update our knowledge about the use of head-mounted displays (HMDs) in education and training. Following a comprehensive search 21 documents reporting on experimental studies were identified, quality assessed, and analysed. The quality assessment shows that the study quality was below average according to the *Medical Education Research Study Quality Instrument*, especially for the studies that were designed as user evaluations of educational VR products. The review identified a number of situations where HMDs are useful for skills acquisition. These include cognitive skills related to remembering and understanding spatial and visual information and knowledge; psychomotor skills related to head-movement, such as visual scanning or observational skills; and affective skills related to controlling your emotional response to stressful or difficult situations. Outside of these situations the HMDs had no advantage when compared to less immersive technologies or traditional instruction and in some cases even proved counterproductive because of widespread cybersickness, technological challenges, or because the immersive experience distracted from the learning task.

Keywords

Virtual Reality (VR); Head-Mounted Display (HMD); Education; Training; Educational Technology; Simulation;

1 Introduction

Since the 1960s the term *Virtual Reality* (VR) has been used to describe a wealth of very different technologies, both software and hardware, such as the *Sensorama Simulator* (Heilig, 1962), online virtual worlds (e.g. Second Life), massive multiplayer online role playing games (MMORPGs, such as World of

Warcraft), surgery simulators, rooms where all walls are covered in displays (*Cave Automatic Virtual Environments*, CAVE), as well as a wealth of different Head-Mounted Displays (HMDs).

For decades it has been discussed if VR has the potential to revolutionize education. The argument is that VR can be used for simulation-based education, where students and learners can practice new skills in a simulated environment that enables correction, repetition and non-dangerous failure and at the same time offers access to interaction with expensive or far-away environments. Despite the high hopes, these ideas have been based on speculation more than praxis, and outside of dedicated training simulators for surgeons, pilots, and military personnel the VR technology has not been on a level where it could be applied in education and training at large.

This, however, changed in 2013 when the first developer versions of a HMD from the company Oculus Rift introduced a new generation of consumer-priced VR technology. During the next couple of years a myriad of competitors launched their own HMDs, making this new technology accessible to the wider public and for research and education purposes as well. Hodgson et al. (2015) give an example of comparable VR hardware from 2006 and 2014 that cost USD 45,000 and USD 1,300 respectively. In a 2016 report on technology trends in higher education the *New Media Consortium* predicted that VR technologies will be adopted in the higher education sector within 2-3 years (Johnson, Adams Becker, Estrada, Freeman, & Hall, 2016).

Aside from a much lower price the new generation of HMDs also offered a better quality user experience. An often cited quality difference between pre- and post-2013 HMDs is in the so-called Field of View (FOV). When putting on a HMD, the natural human FOV of 180 degrees is limited both horizontally and vertically and this influences the realism of the VR experience. Before 2013 the typical FOV of HMDs was between 25 and 60 degrees, while most of the new generation of HMDs have FOVs above 100 degrees (Riva, Wiederhold, & Gaggioli, 2016).

1.1 Previous work

There is a substantial body of older research which is looking at VR and education. This research, however, does not limit the scope to HMDs and the vast majority of the work was published before the new generation of low-price high-quality HMDs became available.

Two key concepts of VR theory are *immersion* and *presence*. They are sometimes used interchangeably, but formally immersion describes the experience of using so-called immersive technology. This technology works by exchanging sensory input from reality with digitally generated sensory input, such as images and sounds (Ott & Freina, 2015). If you, subjectively, react to being immersed in a virtual environment in a way where your brain and nervous system behave in a way similar to being in the same situation in the real world, then you are experiencing presence (Slater, 2003). Often researchers make a distinction between *immersive virtual reality*, where the virtual environment surrounds you (as is the case with HMDs or CAVE systems), and *non-immersive virtual reality*, where you look into the virtual environment from the outside, typically accessed through a traditional display of a desktop computer (Ott & Freina, 2015).

In their systematic review, covering 1999-2009, Mikropoulos and Natsis (2011) identified a number of features, so-called *affordances*, of VR that are conducive to learning. These include the first-person experience and sense of presence, which are both related to the unique experience of being in the virtual environment, as well as a number of affordances that are related to the possibility of giving the learner access to phenomena that are otherwise not available to everyday experience. This is in agreement with Ott and Freina (2015) who found that the main motivation to use VR in education is that it makes it possible to experience situations that are either inaccessible (in time or space) or problematic (dangerous or unethical).

Mikropoulos and Natsis (2011) found that the immersive systems only had an advantage vis a vis the desktop systems when the tasks involved “complex, 3D, and dynamic” content. Many of the studies in the same review reported that their study participants felt a sense of presence while being in the virtual environment, and in three of the reviewed studies this was correlated with better learning outcomes. The few studies examining the use of haptic devices for interacting with the virtual environments did not show an increase in learning, leading Mikropoulos and Natsis (2011) to conclude that “carefully designed learning activities are more important than an exotic interface that contributes to intuitive interaction”. In fact, even though they found that both educators and learners share a positive attitude towards using VR for education, the reviewers did not find evidence to conclude anything regarding retention of knowledge gained this way.

Being motivated by the emergence of a new technology this review followed a hardware centred approach that focussed narrowly on VR simulations that are accessed through HMDs which completely block out visual access to the surroundings, and are offering diagonal FOV above 70 degrees. The review included all relevant published research on education and training with these HMDs. The objective was to systematically assess the quality of the studies and synthesise and discuss their findings with a particular focus on how the VR learner experience affects the learning outcomes.

2 Methods

This review includes all peer-reviewed research documenting experimental or quasi-experimental studies that are relevant to the objective.

2.1 Search strategy

To ensure a comprehensive search, eight research databases were identified. Of these three are interdisciplinary (SCOPUS, Web of Science, EBSCOhost), and the other five cover the fields of biomedicine and health (PubMed), computer science and engineering (IEEE Xplore), education (ERIC), psychology (PsycINFO), and social sciences (International Bibliography of the Social Sciences IBSS).

Each database was searched in March 2017 with keywords based on this Boolean search string:

(virtual reality OR head-mounted display) AND (education OR training OR learning)

The search was limited to only include peer-reviewed publications. Since the focus is on the current generation of HMDs, it was also limited to studies that were published since 1 January 2013. To validate the search string a number of test searches with alternative keywords (*virtual world, cyberspace, HMD*) were undertaken. They did not generate any relevant references that were not already located when searching for *virtual reality* and *head-mounted display*.

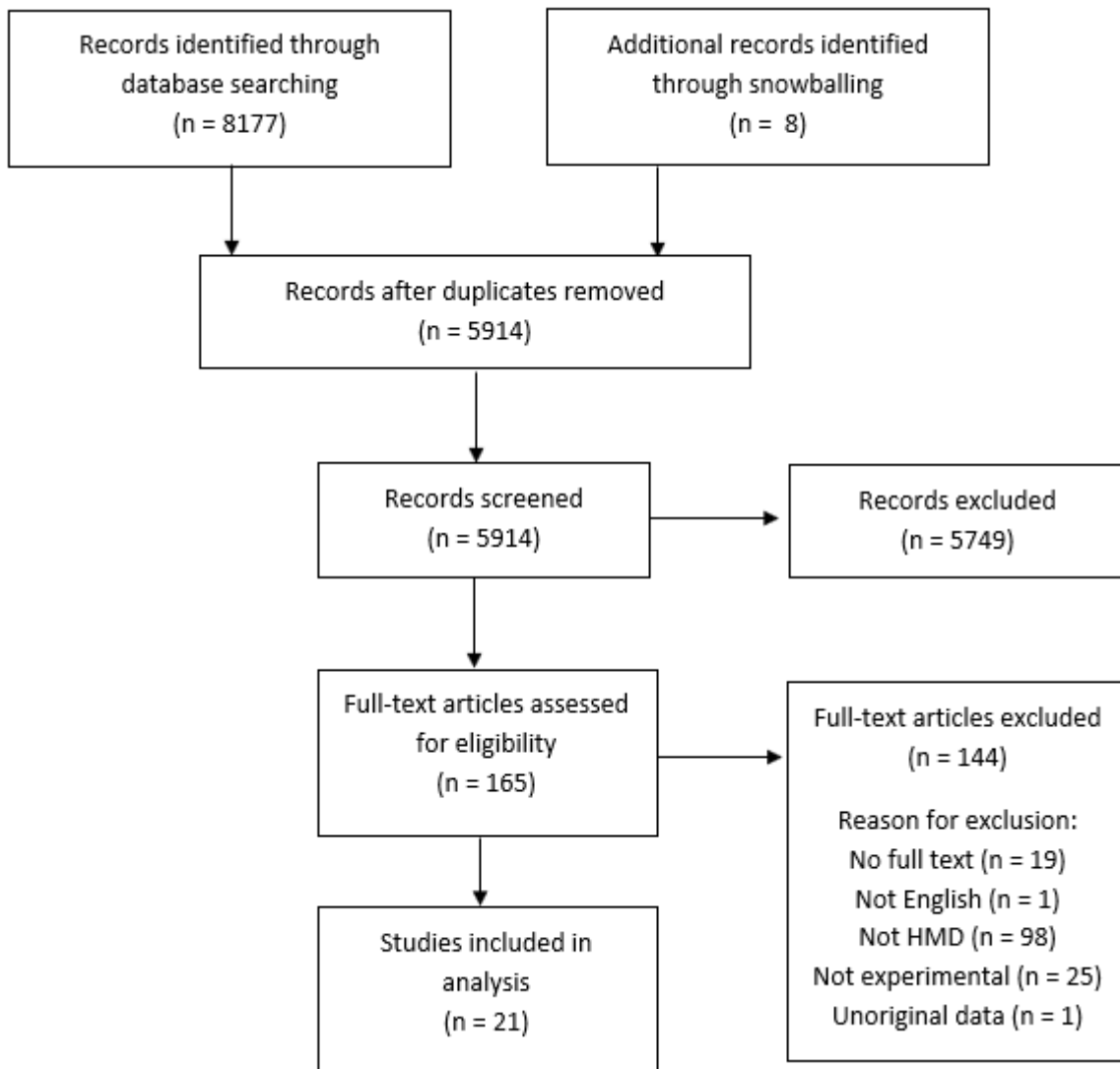


Fig. 1: Selection process flowchart

The search generated a total of 8177 references to journal articles, conference papers, and book chapters. A small number of further documents (n=8) were identified in reference lists or as ‘similar articles’ suggestions on journal and database websites, and they were included in the selection process. As a consequence of the very wide search strategy this raw list of references included many duplicates (n=2271). After they were removed, 5914 unique documents for primary sorting based on title and abstract remained.

2.2 Selection process

The high number of references in the gross list reflects the fact that some keywords (*virtual reality*, *learning*, *education*, and *training*) are very general terms, and can be found in a great number of publications that are not dealing with HMDs in an educational context. However, the interdisciplinary nature of the field means that there is no agreed-upon vocabulary which could have limited the search without risking missing relevant studies. Documents focussing upon studies of non-HMD technologies, such as virtual worlds, surgery simulators, and learning management systems (often called *virtual learning environments*) were excluded. Also studies of VR use in rehabilitation and health care, as well as non-experimental technical descriptions of hardware or software were excluded.

This primary sorting, based on titles and abstracts, left us with just 165 documents. Next step was a systematic sorting, based on the full text versions of the remaining documents. This sorting was based on a five step approach, where a document had to pass every step to be included in the analysis. The criteria were:

1. Full text version accessible and available
2. Full text version is in English
3. Describes the use of a HMD with high FOV
4. Describes an experimental or quasi-experimental study of the educational use of HMDs
5. Reports original data that is not analysed more thoroughly by the same authors in another of the included documents

Of the 165 documents, 21 passed all five steps, and were thus deemed appropriate for inclusion in the analysis.

Of the 21 studies 14 examined the learner experience (Andreoli et al., 2016; Bharathi & Tucker, 2015; Fernandes et al., 2016; Gutiérrez-Maldonado, Ferrer-García, Plasajuanelo, Andrés-Pueyo, & Talarn-Caparrós, 2015; Janssen, Tummel, Richert, & Isenhardt, 2016; Kleven et al., 2014; Loup, Serna, Iksal, & George, 2016; Madrigal, Prajapati, & Hernandez-Prera, 2016; Moesgaard et al., 2015; Pan et al., 2016; Polcar & Horejsi, 2015; Ray & Deb, 2016; Reiners, Wood, & Gregory, 2014; Stavroulia et al., 2016). The main aspects of learner experience being examined in the included studies can be divided into three groups, namely: (1) Perceived presence, immersion and realism; (2) Physical discomfort when using HMDs; and (3) Learner attitudes towards VR in education.

Of the 21 studies 11 measured learning outcomes for study participants using HMDs (Alhalabi, 2016; Gutiérrez-Maldonado et al., 2015; Huang, Churches, & Reilly, 2015; Kahlert, van de Camp, & Stiefelhagen, 2015; Moesgaard et al., 2015; Nomoto et al., 2016; Polcar & Horejsi, 2015; Ragan et al., 2015; Rasheed, Onkar, & Narula, 2015; Ray & Deb, 2016; Sportillo, Avveduto, Tecchia, & Carrozzino, 2015). Learning outcomes are often divided into three domains, namely the cognitive, affective and psychomotor domains (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). Of the 21 studies in the review one study looked at affective skills acquisition, six studies examined cognitive skills acquisition, and five studies examined psychomotor skills acquisition.

Online Resource 1 includes a detailed description of the 21 studies, including study purpose, findings, educational topic, study participants, and type of HMD used.

2.3 *Quality assessment tool*

Of the 21 included studies 19 use exclusively or primarily quantitative methods. The remaining two are primarily qualitative (Fernandes et al., 2016; Reiners et al., 2014). To assess the quality of the 19 quantitative studies the *Medical Education Research Study Quality Instrument* (MERSQI) was employed (Reed et al., 2007). MERSQI itself has been rigorously assessed, to evaluate its correlation with other quality assessment tools and different proxies for research quality (David A Cook & Reed, 2015). Although the instrument was developed for medical education it is in praxis discipline neutral and can also be used to assess the quality of non-medical education research.

MERSQI consists of ten items, covering six domains, namely (1) study design, (2) sampling, (3) type of data, (4) validity evidence for evaluation instrument scores, (5) data analysis, and (6) outcome. A total score is calculated by simply adding the scores of the individual items. Each domain has a maximum score of 3, making the maximum total score for a study 18.

3 Findings

3.1 *Quality of included studies*

Out of the 21 identified studies 19 were quantitative. Of these, the most used study design was comparison of two non-randomized groups (n=13). None of the studies lived up to the criteria for randomized controlled trials. The studies all had a high (above 75%) response rate, and in the majority of the studies the study participants came from just one institution (n=15). A strength in the studies is that most of them (n=15) measured objective data (as opposed to self-reporting).

The greatest weakness across the pool of studies relates to the item that deals with validity of the evaluation instrument. Here the plurality of studies (n=9) scored zero points, and only one got the maximum score of 3. The validity of the evaluation instrument is especially important when measuring learning outcomes as such data is completely dependent on the appropriateness and difficulty of the test. Most (n=16) of the studies were rated as having an appropriate analysis, but just six of them were given points for going beyond descriptive analysis. The last item, *outcome*, shows that the majority of the studies deal with acquisition of skills and knowledge (n=13), while 6 have a primary focus on attitudes and perceptions of the learners. No studies examined outcomes related to behaviour or organisational change. Online Resource 2 shows the detailed MERSQI scores for each of the 19 quantitative studies.

The mean MERSQI score of the quantitative studies was 10.9, with a range of 6.0–14.5. In a study of 26 reviews that all used the MERSQI to assess the quality of educational research, researchers found that across the reviews the average of their overall total scores was 11.3, with a range of 8.9–15.1 (David A Cook & Reed, 2015). This lower score, and lower research quality, in the present review is influenced by the fact that ten of the studies took the form of *user evaluations*, and had less focus on scientific rigorosity. The MERSQI mean of these ten studies was 10.3, with a range of 6–13.8, compared to a mean of the other studies of 11.7, with a range of 9.0–14.5. However, since the search and selection process only identified 21 studies for inclusion, even the ones that scored very low on the MERSQI were retained in the pool.

There were two primarily qualitative studies, which could not be included in the quantitative MERSQI analysis. In Fernandes et al. (2016) the researchers described six experiments each of which included two sources of data: observations of learners playing a VR based educational game and a questionnaire that the learners filled out after the experience. Their methodology follows an iterative design where elements, such as the questionnaire or the game itself, are adjusted and improved between the experiments. In Reiners et al. (2014) the researchers conducted a series of seven short experiments to examine learner attitudes towards using VR technologies in the classroom. Their study included just 13 study participants recruited from both inside and outside the university. After each VR experience each study participant is interviewed. The interviews were based on a guide, which included both open ended and Likert scale questions. Furthermore the study included observations of study participants during the VR experiences. Both of the qualitative studies focussed on learner experience and did not examine learning outcomes.

3.2 *Factors influencing immersion and presence*

Creating a sense of presence through immersion is a main motivation for using immersive VR both for education and in other domains. In the included studies a number of different factors influencing immersion and presence were identified. Pan et al. (2016) found that shortcomings in the visual presentation of virtual patients, such as lagging graphics, was limiting the sense of presence felt by medical professionals during the VR experience. When evaluating an educational video game, Fernandes et al. (2016) found that the awareness of people watching you while you were wearing the HMD was limiting the sense of presence.

Another factor that was identified was that standing up, as opposed to sitting down, led to increased sense of presence (Reiners et al., 2014). In a study looking at correlations between personality traits and learner experience Janssen et al. (2016) found that people with more anxious or reserved personalities not only had a less positive experience in VR, but also felt less immersed. This led the researcher to conclude that learners with certain individual traits and characteristics will benefit less from learning in VR.

3.3 *The influence of immersion and presence on learning*

All studies in the review were based on the idea, that more immersion has a positive influence on learning outcomes. Examples of this are Loup et al. (2016) who found that learners with HMD were more engaged, or Reiners et al. (2014) who observed that their study participants took the more immersive VR simulations more seriously. This meant avoiding bumping into things in the virtual environment and approaching dangers with greater care. Furthermore, when comparing three VR systems, Alhalabi (2016) found that study participants in the most immersive system voluntarily spent more time on the learning task. Findings like these all point towards the affordances of immersive technologies such as HMDs. This however, is somewhat contrasted by Fernandes et al. (2016), who found that increasing immersion by adding 3D sounds and a graphical rendering of user's own hands to the virtual environment confused some study participants and distracted from the learning task.

3.4 *Cognitive skills acquisition*

Six studies looked specifically on the acquisition of cognitive skills. Of these, five compared the learning outcomes when using HMD to that of less immersive technologies or traditional classroom instruction. Just one study found clear evidence of better learning outcomes for HMD than for CAVE and desktop based systems (Alhalabi, 2016). Rasheed et al. (2015) found that HMDs were better for spatial awareness but classroom teaching better for remembering facts, and Ray and Deb (2016) found that only after two weeks of biweekly use did students in a group where HMDs were integrated in the classroom instruction do slightly better than students who received traditional classroom instruction. When Polcar and Horejsi (2015) compared knowledge acquisition across HMD, CAVE and desktop systems the HMD actually led to the lowest acquisition of knowledge. As the only study Moesgaard et al. (2015) compared two variations of a HMD-based learning experience: One version had the information delivered by voice-over and one had it presented in dialogue. Although their sample was too small for statistical inference their findings indicated that voice-over was better for remembering factual knowledge, such as dates, whereas the dialogue was better for understanding connections between phenomena.

3.5 *Psychomotor skills acquisition*

Psychomotor skills are often trained with a training simulator in which the learner repeatedly goes through the actions being trained until a level of proficiency has been reached. For earlier non-HMD-based training simulators it has long been accepted that improved realism leads to better learning outcomes. This is sometimes referred to as *simulator fidelity* (Hays & Singer, 1989). Some, however, have argued that realism should only be applied to select elements in training simulations because realistic representation of very complex environments can confuse the learner (National Research Council, 1994, pp. 52-55).

For HMDs the quality of simulator fidelity depends on the peripheral devices that enable kinaesthetic input and haptic output in relation to bodily movements. Of the five studies that dealt with psychomotor skills acquisition, three used a form of hand tracking (Kahlert et al., 2015; Nomoto et al., 2016; Sportillo et al., 2015), one used a pointing device (Ragan et al., 2015), and one used a traditional gaming joystick (Huang et al., 2015). Of these the hand tracking is the most immersive, i.e. has the highest simulator fidelity.

In Kahlert et al. (2015) the researchers tested learning transfer for a HMD-based simulator designed to teach people to juggle three balls. After an average of 27 mins using the very simplified simulator three of nine study participants were able to juggle with real balls. This was a simple pre-post-test, with no control group, but nonetheless clearly a case where psychomotor skills learned in virtual reality can transfer into real life skills. Ragan et al. (2015) examined a system, designed for training a *visual scanning technique* where the task was to identify targets (armed persons) in a virtual environment resembling an urban setting. Subsequent testing in the most complex scenario showed that those who had trained in more realistic scenarios with more visual complexity were better at adhering to the proscribed technique. Because they did not study skills transfer to a real world visual scanning task, but rather measured performance in the most realistic virtual simulation, it cannot be ruled out that the training simply made the study participants better at ‘playing’ the simulation. Something similar was seen in Sportillo et al. (2015), where the researchers found that study participants wearing HMDs indeed got better at a virtual assembly task, but that this improvement did not translate into better performance on equivalent real world assembly task.

3.6 *Affective skills acquisition*

Although five of the studies described VR simulations where the learner interact with a virtual agent – systems that can be used for training interpersonal, communicative and other affective skills – there was only one study that actually tried to measure the learning outcome of such a system (Gutiérrez-Maldonado et al., 2015). This study compared two different formats for teaching diagnostic interview skills, one with HMD and one on a desktop computer with stereoscopic/3D glasses, but found no statistically significant difference in skills acquisition between the formats.

3.7 *Physical discomfort*

Eight of the 21 studies examined the problem of physical discomfort and cybersickness. The frequencies of cybersickness symptoms reported vary from very rare (Fernandes et al., 2016; Madrigal et al., 2016) to almost every participant (Kleven et al., 2014). In some instances the cybersickness symptoms made study participants drop out of the experiments (Reiners et al., 2014). When present, cybersickness influenced the learner attitude towards the technology negatively, and was correlated with lower learning outcomes (Polcar & Horejsi, 2015). Some studies found that participants with extensive 3D gaming experience reported less symptoms of cybersickness (Andreoli et al., 2016; Reiners et al., 2014), and older study participants reported more cybersickness symptoms than younger (Andreoli et al., 2016).

3.8 *Learner attitudes towards HMD technology*

The 10 studies that examined the learner attitudes towards HMDs were in most cases based on the self-reported opinion of the study participants. Generally, these studies examined if the experience is perceived to be useful for learning and if the experience is perceived to be exciting/interesting. Across all of the studies the researchers found that study participants were very positive towards both of these aspects. When comparing HMDs to desktops they found a moderate preference for HMDs (Bharathi & Tucker, 2015; Kleven et al., 2014). Aside from the issues of physical discomfort described above just one study identified a number of less positive attitudes, namely a feeling of unsafety, because the HMD blocks out access to your actual surroundings, and a feeling of boredom and emptiness, because the user is alone in the VR simulation (Reiners et al., 2014). One study went beyond the self-reported attitudes and opinions and included observational methods to determine learner attitudes. This study found that the HMDs triggered emotions such as “joy, satisfaction, delight, and enthusiasm” (Fernandes et al., 2016). None of the studies measured if these attitudes change over time.

4 Discussion

As is clear from the findings HMDs do not automatically cause learning to occur, but they can be used as a medium to access simulations in which learning *may* take place. Because of this, the question is not so much whether HMDs as such are useful for education and training, but rather if a certain simulation is useful. The discussion below will examine what it is that makes some training simulations and virtual experiences useful for learning and use this knowledge to answer questions of relevance to instructors, trainers and VR professionals.

4.1 Using HMDs for cognitive skills acquisition

Often the immersive experience of being in a simulated reality overshadows the cognitive skills acquisition. This is seen in for instance Moesgaard et al. (2015) where the study participants reported they were “too enthralled by the [virtual] environment to notice the information that was presented to them”. However, when used specifically to help the learner remember and understand visual and spatial aspects of a place, the low-interaction virtual experience seems to have an advantage over non-HMD instruction. This advantage, however, is limited by a number of factors, most importantly the learners’ unfamiliarity with the technology and their tendency to cybersickness. All the cognitive skills being evaluated in the reviewed studies related to what Bloom’s taxonomy calls *lower level* cognitive skills, characterized by remembering or understanding facts (Bloom et al., 1956). No research has examined the use of HMDs to teach higher level cognitive skills.

4.2 Using HMDs for psychomotor skills acquisition

The studies examining psychomotor skills acquisition all found that repeatedly using a training simulator made the study participants better at doing well *in the simulator*. Only two studies examined training transfer (Kahlert et al., 2015; Sportillo et al., 2015), and although one of these found no transfer, it was shown that transfer is *possible* and that successful psychomotor skills transfer depends less on the HMD and more on the quality and realism of a peripheral haptic/tactile device. In training simulators that do not include HMDs, such as many surgery simulators, research has shown significant levels of skills transfer between the simulator and the real world task (Dawe et al., 2014). These simulators have the ability to increase complexity slowly and have achieved a level of simulator fidelity where the most difficult simulation is very close to the real world experience, in part due to the fact that the reality being simulated is also mediated by technology. The usefulness of HMDs in psychomotor skills acquisition is therefore very limited. In cases where the psychomotor skill is related to the movement of the head, such as visual scanning or observational skills, the current technology offers high simulator fidelity. For the plethora of other psychomotor skills that require physical interaction with specific artefacts and your surroundings in general efficient psychomotor skills acquisition with HMDs will not be possible until there are significantly improved peripheral technologies for including the user’s body movements into the simulation.

4.3 Using HMDs for affective skills acquisition

Affective skills, like psychomotor skills, require repetition in order to achieve mastery, and a successful VR training simulator for affective skills must be highly interactive. Many affective skills are related to interpersonal skills, and here the ability of the technology to create a believable simulation of a virtual human or social situation is crucial. Although the topic is almost unexamined in current research, affective skills acquisition seems to be a good place to use HMDs. Training simulators for affective skills are less dependent on immersive peripheral devices that include your bodily movements in the simulation, and more dependent on the ability of the simulation to evoke an emotional response in the learner. If the emotional response can be created by exposure to sound and image in a HMD, then this can be used in training. VR technology is already being used to treat irrational fear and phobias (Anderson et al., 2013) and for practicing

stress management strategies (Pallavicini, Argenton, Toniuzzi, Aceti, & Mantovani, 2016). As artificial intelligence improves there will be more and more affective skills which can be trained virtually.

4.4 Current barriers to the use of HMDs in education and training

The optimistic predictions of recent years about the imminent introduction of HMDs into the classroom were based on the fact that the hardware was now much better and much cheaper. However, two fundamental barriers were overlooked.

The first barrier relates to lack of content. Production of VR simulations is prohibitively expensive, and for instructors this means that they have to use the content that is provided by VR content producers. However, most educational VR simulations on the market are directed at self-learners. They were not designed as a tool to be used at different educational levels and with different pedagogical approaches, but as stand-alone learning experiences. This makes them less suitable for classroom use. For HMDs to become a relevant tool for instructors they must have the ability to produce and edit their own content. This is starting to happen with content based on 360 degree video footage, and currently the most promising use of HMDs in education may not be to use educational VR simulations, but to use the HMD as a viewer of 360 degree video content which can form the basis of subsequent educational activities such as classroom discussions, written analysis, group work, or assessments.

The second barrier relates to the hardware. Current HMDs are primarily entertainment devices. They were not designed for classroom use, and require a level of technical skills that is a challenge to many instructors. Furthermore, they require frequent software updates and issues with streaming or preloading materials, and user profiles make it hard for instructors to manage more than just a couple of HMDs. The practical alternative would be the bring-your-own-device (BYOD) philosophy, but that requires all learners to have high quality VR ready smartphones that are compatible with a Google Cardboard headset. This would mean a less immersive experience, and raise questions of equity.

4.5 Issues of equity and equal access to education

As described in the findings the learners reported very different experiences when using HMDs. This ranged from mild feelings of unsafety to severe cases of cybersickness that prompted some to leave the experiment. Furthermore Janssen et al. (2016) showed that personality traits had a strong influence on the way the VR simulations are experienced, and this in turn will influence any learning outcomes. A VR simulation which is designed to engage the learner emotionally can be a very intense experience. In light of this, it is important to not only consider the diversity of the users, but also to make sure that the context of use is perceived to be safe and that the simulations are designed to avoid or minimize physical discomfort.

4.6 Limitations of this review

The studies included in this review came from a wide spectrum of training and educational contexts, and although some had very positive findings it was not possible to make general statements about the benefits of HMDs in education. This limitation was caused partly by the low number of studies and partly by the low quality of the research. In regards to research quality an important weakness in many studies was the measurement of learning outcomes without giving arguments for the validity of the evaluation instrument. Furthermore, the review showed that much of the research was in fact user evaluations of VR products that did not seek to uncover more general knowledge about learning in VR, and at the same time had an unavoidable bias towards a positive evaluation. Another weakness in the evidence base was the high number of media-comparative studies where HMD-based learning is compared to either classroom or online learning. Some researchers have argued that this type of research is “logically impossible because there are

no valid comparison groups” (David A. Cook, 2005). According to this view, new insights can be reached by comparing different variations of learning within the same media. This type of research will not help answer which media or technology is *best* for instruction, but rather will inform our practice when using a certain media or technology. These limitations are general for much education research, but maybe especially pronounced for research in the nexus of learning and technology.

4.7 Future research

To gain more knowledge about the use of HMDs in education and training the interesting question is not *if* HMDs should be used, but rather *how* and *for what* should HMDs be used. Even though this review could identify particular types of skills where the HMDs are useful there is still a great need for more research to inform best-practice.

With the exception of a single study all the reviewed research focussed on very short term use in an experimental setting. Future research should focus on prolonged and repeated usage and examine, with both quantitative and qualitative methods, how the findings relating to motivation, enthusiasm and time-on-task change when the learners become familiar with the technology.

To better understand the barriers and uncover ways of mitigating them, the future research should move away from laboratory style experiments, and examine the use of educational VR in an authentic setting, as part of a real educational or training programme. As presented in the findings above the actual context of use has a great influence on the learner experience and so far no best practice can be formulated.

5 Conclusion

The motivation for using HMDs in education is that it can expose learners to challenging or educational situations and allow them to repeatedly practice new skills in an environment that enables correction and non-dangerous failure. At first sight these affordances seem ideal for teaching almost any skill, and the increased immersion offered by new VR technology seems well suited for successful educational approaches and theories such as constructivism, active learning, or simulation-based learning. However, this review paints a more complex picture, with a much more limited usefulness. While the studies found that learners are generally very positive about using HMDs there are still substantial barriers to the use, especially in regards to cybersickness symptoms, lack of appropriate software, and technical limitations of peripheral devices.

The review identified a number of situations where HMDs are useful for skills acquisition. These include cognitive skills related to remembering and understanding spatial and visual information and knowledge; psychomotor skills related to head-movement, such as visual scanning or observational skills; and affective skills related to controlling your emotional response to stressful or difficult situations. Obviously HMDs can be used as a medium for training any skill, but it will have no added value vis-à-vis cheaper and less immersive formats, and in some cases be counterproductive, because the immersive experience actually distracts from the learning task.

The low average quality and limited number of studies in this review point to a need for further and more rigorous research that examines the most promising uses of HMDs in an authentic educational or training context.

6 Compliance with Ethical Standards

Data from this review will be made available by contacting the first author directly. No human participants were used in this study. The authors do not have any conflicts of interest in relation to the present work.

7 References

- Alhalabi, W. S. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behaviour & Information Technology*, 35(11), 919-925. doi:10.1080/0144929x.2016.1212931
- Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., & Calamaras, M. R. (2013). Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial. *Journal of consulting and clinical psychology*, 81(5), 751.
- Andreoli, R., Corolla, A., Faggiano, A., Malandrino, D., Pirozzi, D., Ranaldi, M., . . . Scarano, V. (2016). Immersivity and playability evaluation of a game experience in cultural heritage. *Vol. 10058 LNCS. 6th International Euro-Mediterranean Conference on Digital Heritage, EuroMed 2016* (pp. 814-824): Springer Verlag.
- Bharathi, A. K. B. G., & Tucker, C. S. (2015). *Investigating the impact of interactive immersive virtual reality environments in enhancing task performance in online engineering design activities*. Paper presented at the ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, IDETC/CIE 2015.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives, handbook I: The cognitive domain: New York: David McKay Co Inc.
- Cook, D. A. (2005). The Research We Still Are Not Doing: An Agenda for the Study of Computer-Based Learning. *Academic Medicine*, 80(6), 541-548.
- Cook, D. A., & Reed, D. A. (2015). Appraising the Quality of Medical Education Research Methods: The Medical Education Research Study Quality Instrument and the Newcastle–Ottawa Scale-Education. *Academic Medicine*, 90(8), 1067-1076.
- Dawe, S. R., Windsor, J. A., Broeders, J. A. J. L., Cregan, P. C., Hewett, P. J., & Maddern, G. J. (2014). A Systematic Review of Surgical Skills Transfer After Simulation-Based Training: Laparoscopic Cholecystectomy and Endoscopy. *Annals of Surgery*, 259(2), 236-248. doi:10.1097/sla.0000000000000245
- Fernandes, L. M. A., Matos, G. C., Azevedo, D., Nunes, R. R., Paredes, H., Morgado, L., . . . Cardoso, B. (2016). Exploring educational immersive videogames: an empirical study with a 3D multimodal interaction prototype. *Behaviour & Information Technology*, 35(11), 907-918. doi:10.1080/0144929x.2016.1232754
- Gutiérrez-Maldonado, J., Ferrer-García, M., Plasanjuanelo, J., Andrés-Pueyo, A., & Talarn-Caparrós, A. (2015). Virtual Reality to Train Diagnostic Skills in Eating Disorders. Comparison of two Low Cost Systems. *Vol. 219. Stud Health Technol Inform* (pp. 75-81): IOS Press.
- Hays, R. T., & Singer, M. J. (1989). Simulation fidelity as an organizing concept *Simulation Fidelity in Training System Design* (pp. 47-75): Springer.
- Heilig, M. L. (1962). Sensorama simulator: US Patent 3,050,870.
- Hodgson, E., Bachmann, E. R., Vincent, D., Zmuda, M., Waller, D., & Calusdian, J. (2015). WeaVR: a self-contained and wearable immersive virtual environment simulation system. *Behavior Research Methods*, 47(1), 296-307. doi:10.3758/s13428-014-0463-1
- Huang, Y., Churches, L., & Reilly, B. (2015). *A Case Study on Virtual Reality American Football Training*. Paper presented at the 2015 Virtual Reality International Conference, VRIC 2015.
- Janssen, D., Tummel, C., Richert, A., & Isenhardt, I. (2016) Towards measuring user experience, activation and task performance in immersive virtual learning environments for students. & I. f. I. S. Graz University of Technology & M. Computer (Vol. Ed.): *Vol. 621. 2nd International Conference on Immersive Learning Research Network, iLRN 2016* (pp. 45-58): Springer Verlag.
- Johnson, L., Adams Becker, S., Estrada, V., Freeman, A., & Hall, C. (2016). *The NMC Horizon Report: 2016 Higher Education Edition*.

- Kahlert, T., van de Camp, F., & Stiefelhagen, R. (2015) Learning to juggle in an interactive virtual reality environment. *Vol. 528. 17th International Conference on Human Computer Interaction, HCI 2015* (pp. 196-201): Springer Verlag.
- Kleven, N. F., Prasolova-Førland, E., Fominykh, M., Hansen, A., Rasmussen, G., Sagberg, L. M., & Lindseth, F. (2014). *Training nurses and educating the public using a virtual operating room with Oculus Rift*. Paper presented at the 20th International Conference on Virtual Systems and Multimedia, VSMM 2014.
- Loup, G., Serna, A., Iksal, S., & George, S. (2016) Immersion and persistence: improving learners' engagement in authentic learning situations. *Vol. 9891 LNCS. 11th European Conference on Technology Enhanced Learning, EC-TEL 2016* (pp. 410-415): Springer Verlag.
- Madrigal, E., Prajapati, S., & Hernandez-Prera, J. C. (2016). Introducing a Virtual Reality Experience in Anatomic Pathology Education. *Am J Clin Pathol*, 146(4), 462-468. doi:10.1093/ajcp/aqw133
- Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56(3), 769-780.
- Moesgaard, T., Witt, M., Fiss, J., Warming, C., Klubien, J., & Schoenau-Fog, H. (2015). *Implicit and explicit information mediation in a virtual reality museum installation and its effects on retention and learning outcomes*. Paper presented at the Proceedings of the European Conference on Games-based Learning.
- National Research Council. (1994). *Learning, remembering, believing: Enhancing human performance* (D. Druckman & R. A. Bjork Eds.): National Academies Press.
- Nomoto, A., Ban, Y., Narumi, T., Tanikawa, T., Hirose, M., & Acm. (2016). Supporting Precise Manual-handling Task using Visuo-haptic Interaction. *Proceedings of the 7th Augmented Human International Conference (Augmented Human 2016)*, 8. doi:10.1145/2875194.2875216
- Ott, M., & Freina, L. (2015). *A literature review on immersive virtual reality in education: state of the art and perspectives*. Paper presented at the Conference proceedings of » eLearning and Software for Education «(eLSE).
- Pallavicini, F., Argenton, L., Toniuzzi, N., Aceti, L., & Mantovani, F. (2016). Virtual Reality Applications for Stress Management Training in the Military. *Aerosp Med Hum Perform*, 87(12), 1021-1030. doi:10.3357/amhp.4596.2016
- Pan, X., Slater, M., Beacco, A., Navarro, X., Rivas, A. I. B., Swapp, D., . . . Delacroix, S. (2016). The Responses of Medical General Practitioners to Unreasonable Patient Demand for Antibiotics - A Study of Medical Ethics Using Immersive Virtual Reality. *PLoS One*, 11(2), 15. doi:10.1371/journal.pone.0146837
- Polcar, J., & Horejsi, P. (2015). Knowledge acquisition and cyber sickness: a comparison of VR devices in virtual tours. *MM Science Journal*(June 2015), 613-616. doi:10.17973/MMSJ.2015_06_201516
- Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., & McMahan, R. P. (2015). Effects of Field of View and Visual Complexity on Virtual Reality Training Effectiveness for a Visual Scanning Task. *IEEE Trans Vis Comput Graph*, 21(7), 794-807. doi:10.1109/tvcg.2015.2403312
- Rasheed, F., Onkar, P., & Narula, M. (2015). *Immersive virtual reality to enhance the spatial awareness of students*. Paper presented at the 7th International conference on Human Computer Interaction, IndiaHCI 2015.
- Ray, A. B., & Deb, S. (2016). *Smartphone Based Virtual Reality Systems in Classroom Teaching - A Study on the Effects of Learning Outcome*. Paper presented at the 8th IEEE International Conference on Technology for Education, T4E 2016.
- Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *JAMA - Journal of the American Medical Association*, 298(9), 1002-1009. doi:10.1001/jama.298.9.1002
- Reiners, T., Wood, L. C., & Gregory, S. (2014). *Experimental study on consumer-technology supported authentic immersion in virtual environments for education and vocational training*. Paper presented at the 31st Annual Conference of the Australian Society for Computers in Tertiary Education, ASCILITE 2014.
- Riva, G., Wiederhold, B. K., & Gaggioli, A. (2016). Being different. The transformative potential of virtual reality. *Annu Rev Cybertherapy Telemed*, 14, 1-4.

- Slater, M. (2003). A note on presence terminology. *Presence connect*, 3(3), 1-5.
- Sportillo, D., Avveduto, G., Tecchia, F., & Carrozzino, M. (2015) Training in VR: A preliminary study on learning assembly/disassembly sequences. *Vol. 9254. 2nd International Conference on Augmented and Virtual Reality, AVR 2015* (pp. 332-343): Springer Verlag.
- Stavroulia, K. E., Ruiz-Harisiou, A., Manouchou, E., Georgiou, K., Sella, F., & Lanitis, A. (2016). *A 3D virtual environment for training teachers to identify bullying*. Paper presented at the 18th Mediterranean Electrotechnical Conference, MELECON 2016.

Online Ressource 1: Description of the 21 studies included in the review

Publication	Learning outcome focus	Learner experience focus	Educational topic	Study participants	Type of HMD	Resource format	Purpose	Findings
Alhalabi et al. (2016)	Cognitive skills		Engineering	48 (students)	Oculus Rift	Unspecified	Compare learning outcomes for different VR systems and traditional teaching	Stand-alone VR (Oculus Rift) was more efficient than VR, CAVE and traditional teaching
Andreoli et al. (2016)		Physical discomfort; Immersion and realism; Learner attitudes	History	72 (mostly students)	Oculus Rift DK2	Serious game	Evaluate learner perception of immersivity and playability of a serious game on cultural heritage	Study participants reported high motivation and ease-of-use. Most experienced some form of discomfort (e.g. nausea or dizziness)
Bharathi & Tucker (2015)		Physical discomfort; Immersion and realism; Learner attitudes	Engineering	54 (students)	Oculus Rift	Virtual lab	Compare task performance and attitudes of learners exposed to VR through either HMDs or a desktop computer	HMD group performed task faster. Both groups self-reported high appreciation of VR learning, especially the group using HMD
Fernandes et al. (2016)		Physical discomfort; Immersion and realism; Learner attitudes	History	437 (various)	Oculus Rift DK2	Virtual agent	Evaluate the user experience of an educational game that included both a VR (HMD) and an AR (Google Glass) player	VR-player was much more enthusiastic than AR-player. Many hardware limitations influenced the experience. Sense of immersion was reduced by the presence of real world audience
Gutiérrez-Maldonado et al. (2015)	Cognitive and affective skills	Immersion and realism;	Health care	52 (students)	Oculus Rift DK1	Virtual agent	Compare efficacy and usability of two VR systems, one using HMD and one using a stereoscopic computer screen	No statistical difference in learning outcome between groups. HMD group reported better usability
Huang et al. (2015)	Psychomotor skills		American football	17 (football players)	Oculus Rift DK1	Virtual agent	Evaluate the efficacy of VR training for quarterbacks	VR training can improve the ability of quarterbacks to pass at the right moment
Janssen et al. (2016)		Immersion and realism;	Problem solving	10 (students)	Oculus Rift DK2	Serious game	Examine the role of 'individual-related variables' (character traits) on learning experience in VR	Nervous and anxious people are less immersed in VR environment
Kahlert et al. (2015)	Psychomotor skills		Juggling	9 (unspecified)	Oculus Rift	Virtual artefact	Examine if psychomotor skills acquired in VR training can be transferred to skills in reality	Three of the nine study participants learned to juggle after only training in VR
Kleven et al. (2014)		Physical discomfort; Learner attitudes	Health care	12 (students)	Oculus Rift	Virtual agent	Evaluate the attitudes of nursing students towards using VR to train communication with patients	The nursing students saw role playing games in a virtual world (via HMDs) as a useful way to teach communication and social skills
Loup et al. (2016)		Learner attitudes	Science	57 (high school students)	Oculus Rift	Serious game	Compare learner engagement and motivation in a serious game accessed through either a HMDs or a desktop computer	More engaged behaviour from learners using HMD as compared to learners using desktop interface. No difference in motivation between the two groups
Madrigal et al. (2016)		Learner attitudes	Medicine	9 (medical professionals)	Phone based HMD	Visualization	Compare learner attitudes towards 2D (desktop) and 3D (HMD) video in anatomical pathology teaching	Two thirds of study participants preferred 3D over 2D, and learners did not encounter significant simulator sickness
Moesgaard et al. (2015)	Cognitive skills	Immersion and realism;	History	40 (unspecified)	Oculus Rift	Virtual field trip	Compare differences in learning and retention for in-universe dialogue and narrative information in a VR experience	No significant difference in retention and learning outcomes, but a tendency for dialogue to foster dynamic understanding and for narration to foster factual knowledge

Nomoto et al. (2016)	Psychomotor skills		Drawing	25 (unspecified)	Oculus Rift DK2	Virtual artefact	Evaluate a visio-haptic HMD based system used to train precise manual handling skills (such as drawing)	The system led to an improvement in drawing precision, but also to a slow-down of working speed
Pan et al. (2016)		Immersion and realism; Learner attitudes	Health care	21 (medical professionals)	Oculus Rift DK2	Virtual agent	Explore attitudes of medical doctors towards the use of virtual agents to train interactions with unreasonable patients	The study participants thought the situation was realistic, and they felt under pressure, as they would in a real world situation
Polcar & Horejsi (2015)	Cognitive skills	Physical discomfort;	No specific topic	45 (students)	Oculus Rift DK2	Virtual environment	Compare learning outcome and physical discomfort for learners using HMD, stereoscopic wall and PC	Higher levels of cybersickness in HMD group led to higher skills acquisition for PC and stereoscopic wall
Ragan et al. (2015)	Psychomotor skills		Military	45 (mostly students)	nVisor SX111	Virtual environment	Explore how field of view (FOV) and visual complexity influence training transfer for visual scanning VR simulation	Neither FOV nor visual complexity was found to influence training transfer (learning outcomes). Study participants exposed to high FOV and visual complexity did better <i>during training</i>
Rasheed et a. (2015)	Cognitive skills		History	20 (pupils, aged 8-10)	Google Cardboard	Virtual environment	Compare the learning outcomes of a VR experience and a traditional lecture	Spatial awareness, including perception of colour, direction, and size were best in the VR experience, while factual data was more accurately remembered after a traditional lecture
Ray & Deb (2016)	Cognitive skills	Learner attitudes	Engineering	40 (students)	Google Cardboard	Visualization	Comparing learning outcomes and learner attitudes for a Google Cardboard based VR system and traditional slides on a projector, both in classroom setting	Study participants in VR group had improved learning outcomes and thought the system was easy and interesting to use
Reiners et al. (2014)		Immersion and realism; Learner attitudes	No specific topic	13 (various)	Oculus Rift DK1	Virtual environment	Compare attitudes to various immersive experiences accessed through a HMD and a desktop computer	Study participants preferred the HMD based experiences, and had a positive attitude towards including HMDs in the class room
Sportillo et al. (2015)	Psychomotor skills		Complex machines	8 (unspecified)	Oculus Rift DK2	Virtual artefact	Evaluate a HMD based system for training of assembly skills, by comparing it to identical real world assembly task	Study participants in HMD group improved their assembly skills inside the simulation, but the evaluation showed poor transfer
Stavroulia et al. (2016)		Immersion and realism; Learner attitudes	Identifying bullying	10 (teachers)	Oculus Rift	Virtual environment	Evaluate learner attitudes towards a HMD based system for training teachers to identify bullying behaviour	The study participants found the simulations realistic and believed the system can be used to train inexperienced teachers

Online Resource 2: Assessing the quality of the 19 quantitative studies with the Medical Education Research Study Quality Instrument (MERSQI)

Study	Design	Sampling		Data	Validity of evaluation instrument			Analysis		Outcome	Total
	Study design	Number of institutions	Response rate	Type of data	Internal Structure	Content	Rel. to other variables	Appropriateness	Sophistication	Outcome	
Alhalabi et al. (2016)	2	0.5	1.5	3	0	0	0	1	1	1.5	10.5
Andreoli et al. (2016)	1	0.5	1.5	3	1	1	1	1	1	1	12.0
Bharathi & Tucker (2015)	2	0.5	1.5	3	0	0	0	1	2	1.5	11.5
Gutiérrez-Maldonado et al. (2015)	2	0.5	1.5	3	0	0	0	1	1	1.5	10.5
Huang et al. (2015)	1.5	0.5	1.5	3	0	0	0	0	1	1.5	9.0
Janssen et al. (2016)	2	0.5	1.5	3	1	1	0	1	2	1.5	13.5
Kahlert et al. (2015)	1	0.5	1.5	3	na	na	na	1	1	1.5	11.4*
Kleven et al. (2014)	2	0.5	1.5	1	0	0	0	1	1	1	8.0
Loup et al. (2016)	2	0.5	1.5	1	0	1	0	1	1	1	9.0
Madrigal et al. (2016)	2	0.5	1.5	1	0	0	0	1	1	1	8.0
Moesgaard et al. (2015)	2	1.5	1.5	3	0	1	0	1	2	1.5	13.5
Nomoto et al. (2016)	2	0.5	1.5	3	na	na	na	1	2	1.5	13.8*
Pan et al. (2016)	1	1.5	1.5	3	0	1	1	1	2	1	13.0
Polcar & Horejsi (2015)	1	0.5	1.5	3	0	0	0	1	2	1.5	10.5
Ragan et al. (2015)	2	1.5	1.5	3	1	0	1	1	2	1.5	14.5
Rasheed et a. (2015)	2	1	1.5	3	0	0	0	1	1	1.5	11.0
Ray & Deb (2016)	2	0.5	1.5	3	0	0	0	1	1	1.5	10.5
Sportillo et al. (2015)	2	0.5	1.5	3	na	na	na	0	1	1.5	11.4*
Stavroulia et al. (2016)	1	0.5	1.5	1	0	0	0	0	1	1	6.0

Possible values for each item

Study design: Single-group cross-sectional or single-group posttest only: 1 • Single-group pretest and posttest: 1.5 • Nonrandomized, two-group: 2 • Randomized controlled trial: 3;

Sampling, institutions: One institution: 0.5 • Two institutions: 1 • Three or more institutions: 1.5;

Sampling, response rate: Not applicable • < 50% or not reported: 0.5 • 50%–74%: 1 • ≥ 75%: 1.5;

Type of data: Assessment by study participant: 1 • Objective: 3;

Validity evidence for evaluation instrument scores: Not applicable • Internal structure: 0/1 • Content: 0/1 • Relationships to other variables: 0/1;

Data analysis, sophistication: Descriptive analysis only: 1 • Beyond descriptive analysis: 2;

Data analysis, appropriate: Data analysis appropriate for study design and type of data: 0/1;

Outcome • Satisfaction, attitudes, perceptions, opinions, general facts: 1 • Knowledge, skills: 1.5 • Behaviours: 2 • Patient/health care outcome: 3;