Cochrane Database of Systematic Reviews | Protocol - Intervention

Serious Gaming and Gamification interventions for health professional education

Sarah Gentry, Beatrice L'Estrade Ehrstrom, Andrea Gauthier, Julian Alvarez, David Wortley, Jurriaan van Rijswijk, Josip Car, Anneliese Lilienthal, Lorainne Tudor Car, Charoula K Nikolaou, Nabil Zary
Authors' declarations of interest

Version published: 03 June 2016 Version history
https://doi.org/10.1002/14651858.CD012209

Abstract

This is a protocol for a Cochrane Review (Intervention). The objectives are as follows:

To evaluate the effectiveness of Serious Gaming and Gamification interventions for delivering pre- and post-registration health professional education compared with traditional learning, other types of eLearning, or other Serious Gaming and Gamification interventions. We will primarily assess the impact of these interventions on students' knowledge, skills, professional attitudes and satisfaction.

Background

Description of the condition

Many healthcare systems worldwide are suffering from a critical shortage of trained health workers. In 2013 the World Health Organisation (WHO) estimated that this shortage numbered 7.2 million (WHO 2013), with low- and middle-income countries most affected.

A dramatic demographic and epidemiological change is happening worldwide, with ageing populations and increasing prevalence of non-communicable diseases, resulting in increasing demand for healthcare workers in all settings (Lopez 2006). Some of the poorest countries are facing a "triple burden" of disease, as non-communicable diseases add to existing communicable and socio-behavioural illnesses (Sen 2000; WHO 2014). Many low- and middle-income countries, whose resources are already severely limited, face the further depletion of staff as many students and trained health workers migrate to wealthier countries, known as the “brain drain” phenomenon (Kuehn 2007). This has the overall effect of subsidising the healthcare systems of wealthier countries, as health workers, whose training has been paid for or subsidised by their home country, then migrate.

A further challenge has been posed for post-registration health professional education by the working hour restrictions in place in many countries, such as the European Working Time Directive in the European Union. Such restrictions are important for ensuring patient safety, but require the development of innovative methods of training healthcare professionals in increasingly complex techniques in a shorter time, whilst ensuring cost-effectiveness. New methods of learning that allow training to occur in a safe environment outside of the workplace are necessary.

The shortage of health workers is aggravated by the inadequacy of many training programmes (Chen 2010) and the lack of available healthcare teachers and lecturers (Dorman 2009). The content, organisation and delivery of training programmes often fail to equip health workers with the skills, competencies, experience and expectations needed to satisfy the changing health needs of the populations they are to serve (Frenk 2010).

It is essential to enable, develop and promote innovative educational programmes which increase the number of trained health providers, whilst ensuring that the quality and relevance of training meets these new challenges (WHO 2011). Of particular importance is the ability of health professionals to mobilise and search for available knowledge, along with a capacity and willingness to engage in critical reasoning and collaborative practice (Frenk 2010). The increased use of information and communication technologies is recognised as one of the key strategic platforms on which to build strong education and training systems (Crisp 2008). Innovative ways of teaching and learning are required to respond to the need for health professional education, tackle the shortage of trained healthcare workers and ultimately improve patient care. eLearning may be one such innovation.

This review is one of a series of Cochrane reviews assessing the scope for, and potential impact of, a range of eLearning resources for different levels of healthcare education and training. These arose in response to a report commissioned by the WHO Department of Knowledge Management and Sharing, which involved a systematic review evaluating the efficacy of eLearning interventions for undergraduate health professional education (George 2014; Rasmussen 2014; WHO 2015). eLearning may encompass a variety of interventions characterised by their tools, contents, learning objectives, pedagogical approaches and setting of delivery. eLearning can include, but is not limited to, offline and online computer-based eLearning, Massive Open Online Courses (MOOCs), virtual reality environments, simulation, mLearning, Serious Gaming and Gamification. This review will focus on the use of Serious Gaming and Gamification interventions for pre- and post-registration health professional education.

Description of the intervention

eLearning can be defined as "an approach to teaching and learning, representing all or part of the educational model applied, that is based on the use of electronic media and devices as tools for improving access to training, communication and interaction and that facilitates the adoption of new ways of understanding and developing learning" (Sangrà 2012). The field of eLearning is growing due to advances in modern technology and current applications of its use, accentuated by the increased volume of, and access to, information (Frenk 2010). When eLearning is combined with traditional methods of education, such as face-to-face teaching, it is described as ‘blended learning’.

An educational game can be defined as "an instructional method requiring the learner to participate in a competitive activity with pre-set rules" (Fitzgerald 1997) and generally refers to gaming interventions delivered via any medium. A number of systematic reviews have assessed the efficacy of educational games for health professional learning. A Cochrane review by Akl 2013 included two studies of non-digital games for health professional education and was unable to "confirm or refute the utility of games as a teaching strategy". Another Cochrane review of games for mental health professionals included one small study of a non-digital game and found that participants in the intervention group had better test scores a few hours post intervention (Bhoopathi 2006). A systematic review of games for medical and dental student education included five studies, three of which suggested that gaming interventions had a positive effect on student knowledge. However all of the included studies were deemed to be of low methodological quality and only one study involved an intervention delivered via a digital device (Akl 2010).

The idea of ‘Serious Games’ was first outlined by Abt in 1970, who considered them to be games that "have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement" (Abt 1970). This was before the widespread use of digital technologies, and so referred to games played on any medium, but the term is now more commonly used to refer specifically to digital games (Sawyer 2003). The concept of games with ‘serious intent’ is being applied within a broad set of industries, including health, defence, education, politics, training and ecology, with a very diverse range of approaches (Alvarez 2012). As a result, a variety of other terms have been used to describe some of these interventions, including Digital Game-Based Learning, Alternative Purpose Games, Games for Good (Sawyer 2008) and Edugaming (Angarita 2005), and no specific definitions or domain boundaries have been agreed upon amongst industry professionals or in the academic literature (Djaouti 2011).

There is some debate as to whether the term ‘Serious Game’ can be used more broadly than as defined by Abt 1970, to include games used for education without being designed for the purpose, such as using commercial off-the-shelf games (COTS) for education. Additional challenges in definition are posed by the perhaps contradictory idea of whether something can be both 'fun' and 'serious', and if it can, what factors separate a serious game from one that is not serious.

For the purposes of this review we will use terminology as defined by Alvarez 2015 and Alvarez 2012. The term ‘Serious Game’ will be used to refer only to games designed specifically for the ‘serious’ purpose of providing health professional education, delivered via a digital device. The term ‘Serious Diverting’ will be used to refer to the use of games originally designed primarily for entertainment used without modification, as part of an intervention intended to be used for health professional education, delivered via a digital device. The term ‘Serious Modding’ will be used for games originally designed for entertainment, which have been modified in order to be used ‘seriously’ for health professional education. ‘Serious Gaming’ will be used to refer to any use of digital games for health professional education, thereby encompassing ‘Serious Games’, ‘Serious Diverting’ and ‘Serious Modding’.

A related but separate concept is that of ‘Gamification’, which for the purposes of this review, can be defined as "the application of the characteristics and benefits of games to real world processes or problems" (Wortley 2013). Gamification differs from Serious Games in terms of the design intention, with Gamification interventions involving the application of game elements to something with a utilitarian purpose, and Serious Games designed as full-fledged games for a purpose other than just entertainment (Deterding 2011a). Both may be experienced by the user as a complete game. Gamification has the potential to allow for greater involvement of the user in setting their own objectives or outcomes, personalisation of the intervention and cost effectiveness (Wortley 2013). Most, but not all, uses of the term refer to interventions involving the use of enabling digital technologies.

Serious Gaming and Gamification interventions may take place on a number of platforms, including personal computers and mobile phones. Approximately three quarters of the world's population has access to a mobile phone, and growth in mobile communication is particularly notable in low- and middle-income countries, where many new mobile

applications are designed (World Bank 2012), and the number of people with access to personal computers and the internet is growing.

Interventions that: aim to directly replicate real experiences (such as a cardiac arrest simulation); involve a psychomotor skills trainer (such as a laparoscopic surgery simulator); or involve virtual reality environments, will only be included in this review if there is a clear game-based component and the intervention is delivered via a digital device.

How the intervention might work

eLearning interventions have the potential to provide learners with greater ease of access and flexibility, portability, an increased number of interactions with tutors and an increased amount of interaction with peers when compared to traditional learning. eLearning interventions may provide monetary savings, be scalable, free up lecturer time for tutor-led workshops, ease development and updating of materials and allow for practice of skills prior to practice with patients (or even allow for practice when patients are not available). However, it may be more time consuming, inhibit in-depth discussions and remove opportunities for students to clarify points with their tutors or gain experience with patients.

Educational games may be part of an active learning environment which allows learners to practice solving problems and making decisions in risk-free surroundings (Akl 2013). The motivational properties and intensiveness of gaming have the potential to be harnessed for educational purposes (Garris 2002). Serious Gaming and Gamification interventions have the potential to combine the advantages of eLearning with the advantages of game-based learning. The role of simulation in health professional education is growing rapidly, giving trainees opportunities to practice skills in a safe environment, improving patient safety and allowing for greater experience. Serious Gaming interventions may have the potential to allow for greater experience in a safe environment in a similar way (Allery 2004).

Whether cost is an advantage of Serious Gaming and/or Gamification interventions, or a barrier to its use, will likely depend on the type of intervention. The reusable nature of these interventions may reduce costs, in a similar way as proposed for digital learning objects (Ruiz 2006). Alternatively costs of development and maintenance may be a barrier, as has been suggested for other forms of eLearning (Childs 2005). Serious Gaming and Gamification interventions could be developed as standalone interventions focused on a particular area of health professional education, or integrated into education programmes, settings and environments (Breuer 2010). These interventions could allow for greater learner engagement, particularly for those who have grown up in an environment in which they are immersed in information technology and digital media (Breuer 2010; Prensky 2003).

A variety of technical and non-technical skills could be targeted, including, but not limited to, analytical skills, strategic thinking, knowledge, multitasking, decision making, communication and psychomotor skills (Susi 2007), with multiplayer functions providing opportunities for collaborative learning (Prensky 2003). Serious Gaming and Gamification interventions have the potential to be used to teach decision-making skills, as in a study by Boreham 1989, which assessed the effect of a computer game teaching medical students about phenytoin dosage decision-making skills compared with no intervention, or the Serious Game being developed by Petit dit Dariel 2013 for the development of nurses' clinical reasoning in community settings. These interventions may also have the potential to be used for practical skills training. A small study of medical students by Enochsson 2004 found that those with greater computer games experience performed significantly better in a surgical simulation task.

Serious Gaming and Gamification interventions could also be used to create a learner-centred environment, moving away from the role of the teacher as information provider, and creating variety within a training programme (Allery 2004). Serious Gaming and Gamification interventions for learning should be developed with learning objectives in mind, and aligned with educational theory.

The digital and reusable nature of these interventions may allow adult learners to choose the time and place in which they learn, to receive feedback on their performance on which they can base future learning, to set and achieve learning objectives and to develop their skills in a setting with relevance to real life (Akl 2013; Allery 2004). They also offer the potential for experiential learning to occur in an environment that is safe and comfortable for the learner and the patient, who does not need to participate in early phases of students' learning process (i.e. the phase when mistakes are more likely to be made).

Serious Gaming and Gamification interventions are consistent with Knowles' theory of andragogy, which suggests that adult learners are: 1. self-directed; 2. build on their previous experiences when learning; 3. are goal-orientated; and 4. their learning is problem centred (Knowles 1970). These interventions may allow adult learners to choose the time and place in which they learn, to receive feedback on their performance on which they can base future learning, to set and achieve learning objectives and to develop their skills in a setting with relevance to real life (Akl 2013). Kolb's experiential learning theory suggests that a learner has an experience, observes and reflects on it, analyses it and forms abstract concepts, and actively experiments, applying what they have learned to their surroundings to see what the results are (Kolb 1984). Each of these stages could be targeted by these interventions, allowing learning to occur from the gaming/gamification experience.

Ritterfeld 2006 suggest that there are three main approaches that may be taken by a Serious Gaming or Gamification resource to combine education and gaming:

1. a reinforcement paradigm, where entertainment is offered as a reward for learning;
2. a motivation paradigm, where entertainment is used to gain the learner's interest and attention to prepare them for learning; or
3. a blended paradigm, where the learning procedure itself aims to be entertaining.

Breuer 2010 suggest that the third approach is likely to be most effective and has the potential to be able to harness the inherent enjoyment of learning, in addition to enjoyment provided by the game or game elements. Koster 2004 suggests that games and learning are fundamentally linked, with the ‘fun’ from games occurring as a result of the learning. Deterding 2011b suggests that the role of gaming in making learning fun is by providing optimal conditions in which the learning can take place, and that using games to provide extrinsic motivation can actually be harmful, as it reduces intrinsic motivation.

Why it is important to do this review

Previous reviews of the efficacy of eLearning interventions for health professional education and barriers to its use have underlined its potential, but also stressed the need for further research and reviews on the topic (Childs 2005; Cook 2010; Feng 2013; George 2014; Lahti 2014; Rasmussen 2014; Rowe 2012; WHO 2015). This is largely due to the limited scope of existing evaluations, in terms of: outcomes (use, enjoyment and satisfaction, as opposed to assessment of students' knowledge and skills or patient outcomes); duration (short term rather than long term); professional field (nurses, medical education); educational context (mostly high-income countries) and technology used (online and offline computer-based, virtual reality). Other common limitations include a lack of appropriate study design, lack of tools available for unobtrusive data gathering and a lack of validated scales for assessing outcomes (Mayer 2012). Although past reviews have explored the impact of game-based learning on a number of skills-based and cognitive outcomes (Garris 2002), it is still unclear which modes of design, context, content and delivery of such interventions are most effective for health professional education.

Whilst Serious Gaming and Gamification interventions appear to have much potential, rigorous evaluation is required to assess whether they can lead to effective learning. There is the potential for the game or game elements to become a distraction rather than a facilitator of learning, with the method "more memorable than the message" (Allery 2004), and so quality of learning must be the focus, as opposed to the capabilities of the technology used (Vogel 2002).

Our review aims to help address the existing gaps by:

- updating the rapidly growing body of evidence on Serious Gaming and Gamification, especially at a time when new technologies, including mobile technologies, online games, virtual worlds and alternate reality games (ARGs) have expanded the ways in which games have traditionally been played;
- focusing on Serious Gaming and Gamification interventions across various professional fields of health sciences education at pre- and post-registration levels;
- evaluating the impact of interventions on learners' knowledge, skills, attitudes and satisfaction;
- evaluating the impact of interventions for post-registration health professionals on patient outcomes;
- capturing risks/side effects of Serious Gaming and Gamifcation interventions;
- including evidence from low-, middle- and high-income countries;
- being integrated in a series of reviews and final overview which will provide a systematic approach to the multiple uses and applications of eLearning in terms of channels (including online and offline interventions, simulated environments and blended learning, as well as Serious Gaming and Gamification) and training stages (pre- or post-registration health professional education);

We hope that our evaluation will allow us to make recommendations for improvements to the design of future randomised and cluster randomised controlled trials in this area, including types of interventions that require further evaluation, recommended comparisons, outcomes, methods of outcome assessment and duration of follow up.

Objectives

To evaluate the effectiveness of Serious Gaming and Gamification interventions for delivering pre- and post-registration health professional education compared with traditional learning, other types of eLearning, or other Serious Gaming and Gamification interventions. We will primarily assess the impact of these interventions on students' knowledge, skills, professional attitudes and satisfaction.

Methods

Criteria for considering studies for this review

Types of studies

Randomised controlled trials (RCTs) and cluster randomised controlled trials (cRCTs).

We will include RCTs with unclear or high risk of bias for sequence generation. If meta-analysis of included studies is feasible and appropriate, we will include all RCTs regardless of their sequence generation bias rating. However, we will also conduct sensitivity analyses excluding those at unclear or high risk of bias, to examine the robustness of the meta-analysis results to methodological limitations of the included studies. We will exclude cross-over trials due to the high likelihood of carry-over effect.

Types of participants

We will include studies with participants who are enrolled either in:

- a pre-registration, undergraduate, health-related university degree or basic, health-related vocational training programme, defined for the purpose of this review as any type of study leading to a qualification that: (i) is recognised by the relevant governmental or professional bodies, and (ii) entitles the qualification holder to apply for entry level positions in the healthcare workforce. For this reason, graduate medical education courses, such as those common in the United States (USA), will be included; or
- a post-registration health professional education programme, defined as any type of study after a qualification which is recognised by the relevant governmental or professional bodies that enables the qualification holder entry into or continuation of work in the healthcare workforce in a more independent or senior role, including participation in continuing professional development (CPD) or similar activities.

We will include candidates for, and holders of, the qualifications listed in the Health Field of Education and Training (091) of the International Standard Classification of Education (ISCED-F, IUS 2013), except the students of traditional, alternative and/or complementary medicine. We have chosen not to include this group due to differences in regulation of practice, use of evidence to demonstrate efficacy, and registration of professionals compared with other healthcare groups, and the lack of standardised curricula and training outcomes and integration of these services with mainstream health care in many countries. We will include students from the following categories: dental studies, medicine, nursing and midwifery, medical diagnostic and treatment technology, therapy and rehabilitation, and pharmacy. Participants will not be excluded on the basis of age, gender or any other socio-demographic variable.

Types of interventions

We will include studies investigating any type of Serious Gaming or Gamification intervention delivered to pre- and/or post-registration healthcare professionals (as defined above) in which education is the primary purpose of intervention delivery.

The following interventions will be included, if the primary purpose of the intervention is pre- and/or post-registration health professional education, and the intervention is delivered via a digital device:

1. Serious Games – gaming interventions designed specifically for health professional education;
2. Serious Diverting interventions – interventions using games originally designed for entertainment, which have been used for health professional education unmodified;
3. Serious Modding interventions – interventions using games originally designed for entertainment, which have been modified for use in health professional education;

4. Gamification interventions – interventions for health professional education for which game elements have been added, or any other intervention delivered via a digital device for which the primary purpose is the delivery of health professional education, and which uses game mechanics and design techniques to engage and motivate participants to achieve their goals.

Digital devices include, but are not limited to: personal computers, laptop computers, notebooks, tablets, games consoles and mobile phones.

We will include studies where Serious Gaming and/or Gamification methods are the sole intervention. We will also include studies where Serious Gaming and/or Gamification methods are delivered as part of a complex, multi-component intervention (i.e. blended learning), but only where the Serious Gaming and/or Gamification component is evaluated separately. For example, if two complex interventions were compared, where the only difference was the Serious Gaming and/or Gamification method, this would be included as the Serious Gaming and/or Gamification element has been evaluated separately.

We will include studies assessing the effectiveness of commercial off-the-shelf (COTS) games if delivered as part of an intervention in which delivery of educational content is the primary purpose. We will exclude interventions aimed primarily at entertaining the user that may also result in improvements in knowledge and skills. For example, we would include an intervention involving a COTS game with the aim of improving the communication skills of healthcare students, but would exclude assessment of whether playing a game aimed primarily at entertaining the user also happened to improve psychomotor skills.

Gamification interventions will be included if they are primarily delivered via a digital device. For example, an intervention in which learners compete to achieve a higher score on a digital leader board by answering electronic questions on a digital device would be included, whereas an intervention in which learners participate in game-based activities in a classroom but have a leader board on a digital device would be excluded.

Interventions which aim to directly replicate real experiences (such as a cardiac arrest simulation), which involve a psychomotor skills trainer (such as a laparoscopic surgery simulator), or which involve virtual reality environments will be included only if there is a clear game-based component and the intervention is delivered via a digital device. Classical simulations without a gaming component will be evaluated in another Cochrane review in this eLearning series.

We will include studies that make the following intervention comparisons:

- Serious Gaming or Gamification intervention versus no intervention;
- Serious Gaming or Gamification intervention versus traditional learning;
- Serious Gaming or Gamification intervention versus other types of eLearning intervention;
- Serious Gaming or Gamification intervention versus another type of Serious Gaming or Gamification intervention.

Types of outcome measures

We will include studies which report on at least one of the following primary or secondary outcomes:

Primary outcomes

- Patient-related outcomes, for example, time to blood pressure control (only for interventions delivered to post-registration students).

- Students' knowledge, using any validated or non-validated instrument to measure difference in pre- and post-test scores, or post-test scores only if no pre-test has been reported. If several post-test results are available, data as to when those tests were conducted will be recorded and the difference between the pre-test and the first post-test will be used. When applicable the difference between the pre-test and the last test available will be used for sensitivity analysis.
- Students' skills, measured using any validated or non-validated instrument (e.g. pre- and post-test scores, time to perform a procedure, number of errors made whilst performing a procedure).
- Students' professional attitudes towards patients (e.g. awareness of moral and ethical responsibilities involved in patient contact) and/or towards new clinical knowledge or skills measured using any validated or non-validated instrument.
- Students' satisfaction with the learning intervention measured using any validated or non-validated instrument.

Secondary outcomes

- Education economics outcomes (e.g. cost, cost-effectiveness).
- Adverse and/or unintended effects of the intervention (e.g. the game elements are a distraction rather than a facilitator of learning, poorer quality learning, focus on the capabilities of the technology as opposed to the learning itself).

Search methods for identification of studies

Electronic searches

The following databases will be searched:

- MEDLINE (via Ovid SP)
- EMBASE (via embase.com)
- Web of Knowledge (WoK)
- Educational Resources Information Centre (ERIC)
- Cochrane Central Register of Controlled Trials (CENTRAL, *The Cochrane Library*)
- PsycINFO (via Ovid SP)
- Cumulative Index to Nursing and Allied Health Literature (CINAHL) (via EBSCOhost)

A common search strategy will be defined and used for all of the reviews in this eLearning series. First a MEDLINE search strategy will be developed, and this will then be adapted to search the other databases. The keywords presented in Appendix 1 will be used. Allocation of studies to a review will be done through the screening of titles and abstracts. We will search from 1990 to the present.

Searching other resources

We will screen the reference lists of all included studies and of systematic reviews identified by our electronic searches. We will also search the International Clinical Trials Registry Platform Search Portal and Current Controlled Trials metaRegister of Controlled Trials. We will contact study authors for further information when necessary and to ask

whether they are aware of any other studies that might meet our inclusion criteria.

Data collection and analysis

Selection of studies

We will merge search results across databases using Endnote, and remove duplicates. Titles and abstracts will then be assessed for eligibility by at least two independent review authors according to our pre-specified inclusion criteria. Screening will be calibrated between the two review authors using the first 500 citations. The full texts of studies which potentially meet our inclusion criteria will be retrieved and reviewed independently by two authors. We will contact study authors for further information if information is missing or unclear. Any disagreements will be resolved by discussion between the two authors with a third adjudicating if required. Studies that appear to fulfil the inclusion criteria but are later excluded from the review at the full text screening stage will be detailed in the 'Characteristics of excluded studies' table, along with their reasons for exclusion. Two review authors will verify the final list of included studies.

Data extraction and management

Data will be extracted independently by at least two review authors using a standardised data extraction sheet derived from the Cochrane EPOC Group data extraction template (EPOC 2002). The data extraction form will be piloted and adapted in response to feedback. Disagreements will be resolved by discussion and a third review author will adjudicate if required. Study authors will be contacted if information is unavailable or unclear. Extracted data will be entered into Review Manager and entries checked for accuracy against the original data by another author.

Data to be extracted will include:

- study citation;
- study design: study design; aims and objectives; country in which study was conducted; study duration; method of participant recruitment;
- participants: inclusion/exclusion criteria; number of participants recruited/excluded/declined; demographic characteristics of participants;
- intervention and comparison: baseline differences; type of qualification participants are pursuing and year of study; description of intervention and control conditions; exposure; type of technology/devices used to deliver intervention; educational theory used; assessment method; "Game/Purpose/Scope (G/P/S)" classification (Alvarez 2012);
- outcomes: outcomes measured; instrument used; data collection method (e.g. videocapture, survey, metrics);
- results: quantitative data for all relevant outcomes;
- key conclusions of each study;
- 'Risk of bias' data, in order to perform 'Risk of bias' assessment, as outlined below.

Assessment of risk of bias in included studies

Risk of bias in RCTs and cRCTs will be assessed as described by the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011), by at least two independent authors. Review authors will pilot the 'Risk of bias' assessment prior to completing it. Disagreements will be resolved by consensus, with discussion with another review author if

necessary. Study authors will be contacted for additional information or clarification of study methods if required. Results of the 'Risk of bias' assessment will be presented in 'Risk of bias' tables, graphs and a narrative summary.

For RCTs we will evaluate the following individual elements: random sequence generation; allocation sequence concealment; blinding (participants, personnel); blinding (outcome assessment); completeness of outcome data; selective outcome reporting; and other sources of bias (e.g. baseline imbalance, inappropriate administration of an intervention and contamination). We will consider blinding separately for different outcomes where appropriate (for example, blinding may have the potential to differently affect subjective versus objective outcome measures). We will judge each item as having a high, low or unclear risk of bias according to the criteria provided by Higgins 2011, and include a relevant quote from the study and a justification for our judgement for each item in the 'Risk of bias' table.

We will judge a study as being at the highest risk of bias if it is scored as at high or unclear risk of bias for either the sequence generation or allocation concealment domains, based on growing empirical evidence that these factors are particularly important potential sources of bias (Higgins 2011).

For cRCTs we will also assess and report the risk of bias associated with an additional domain: selective recruitment of cluster participants.

Measures of treatment effect

For continuous outcomes we will calculate the mean difference (MD) and 95% confidence intervals (CI). For dichotomous outcomes we will calculate the risk ratio (RR) and 95% CI.

For cRCTs we will account for the effects of clustering by adjusting each trial to its 'effective sample size' using intra-class coefficients (ICCs), where available, or using external estimates from similar studies (Deeks 2008).

If more than one study measures the same outcome using different tools, the mean differences for each study will be recalculated into standardised mean difference (SMD) by dividing the study mean between groups by the standard deviation of outcome among participants.

Unit of analysis issues

For cluster RCTs (cRCTs) we will consider whether or not the study accounts for unit of analysis error (i.e. if the authors incorrectly analysed participants as independent individuals rather than the unit in which they were randomised) (Higgins 2011). If this has already been accounted for we will simply extract and report effect estimates and use these in any meta-analysis.

If unit of analysis error is not accounted for we will try to re-analyse cRCTs if we can obtain information on the size and number of clusters, and on the intracluster correlation coefficient value from the study report or by contacting the authors. If the intracluster correlation coefficient is not available we will use external estimates from similar studies (Deeks 2008). We will then meta-analyse using a generic inverse-variance method in Review Manager, which accounts for clustering of data.

If we are unable to obtain these data we will report summary effect measurements extracted from each cluster. The number of clusters will be considered the sample size and meta-analysis performed as if the trial was individually analysed. Note that this will reduce the statistical power of the analysis.

Dealing with missing data

We will contact study authors to obtain missing data or clarify areas of uncertainty. If we are unable to obtain sufficient data we will use data available from the studies and assess the risk of bias through the criterion 'incomplete outcome data'. We will not impute any missing data. Where possible we will conduct analyses on an intention-to-treat basis. Implications of missing data will be considered in the 'Discussion' section of the review.

Assessment of heterogeneity

We will assess heterogeneity qualitatively to decide whether the included studies are similar enough, in terms of populations and interventions, for pooling of data to give meaningful conclusions. We will combine in meta-analyses those deemed sufficiently homogenous and for which sufficient data are available.

Where meta-analyses are performed statistical heterogeneity will be examined by visual inspection of the scatter of effect estimates in the forest plots and using the I^2 statistic (Higgins 2011), after using the inverse variance method. This statistic gives the percentage of the variability in effect estimates that can be attributed to heterogeneity rather than chance (Deeks 2008). A value of greater than 50% will be considered to be substantial heterogeneity

If substantial clinical, methodological or statistical heterogeneity is found we will not perform meta-analyses and will instead present a narrative synthesis.

Assessment of reporting biases

Reporting bias will be assessed qualitatively based on the characteristics of included studies (e.g. if we identify mostly small studies with positive findings we will have a high degree of suspicion) and through contacting the study authors (e.g. if their responses suggest relevant unpublished studies we will have a high degree of suspicion). If at least 10 RCTs are found we will also assess reporting bias using a funnel plot regression weighed by the inverse of the pooled variance.

Data synthesis

Data will be reported using Review Manager. Extracted data will be entered into tables grouped by study design and type of intervention to create a descriptive synthesis. The results of individual RCTs and cRCTs will be reported as mean differences for continuous variables and odds ratios for dichotomous variables with 95% confidence intervals.

If included studies are sufficiently homogenous in terms of population, inclusion criteria, interventions and outcomes, we will consider meta-analysis of data relating to students' knowledge and skills. The decision as to whether meta-analysis is performed will be made by consensus of all review authors. The choice of model would depend on the heterogeneity (assessed as described in Assessment of heterogeneity) of the studies included in the meta-analysis. We will conduct the analysis according to the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). We expect that if meta-analysis is feasible, we will use a random-effects model, which provides a more conservative estimate of effect and can be used where there is moderate heterogeneity. If studies are deemed too heterogeneous. we will present a narrative synthesis of findings, with effect sizes calculated for outcomes where there are sufficient data. When possible an assessment will be made of the quality, size of effect observed and statistical significance of studies.

Using Miller's classification of clinical competence (Miller 1990) the different types of tests for students' knowledge and skills will be grouped and analysed together. For example, multiple choice questions assessing knowledge (i.e. knows) will be analysed together, and essay questions assessing competence (i.e. knows how) will be analysed together. The focus will therefore be on the testing method rather than the delivery method (i.e. if skills were assessed by a knowledge test it would be categorised as knowledge).

Where studies report more than one measure for each outcome, the primary measure as defined by the primary study authors will be used in the analysis. Where no primary measure has been reported, a mean value of all the measures for the outcome will be calculated and used in the analysis.

For students' professional attitudes, the different types of assessment will be grouped and analysed as cognitive attitudes, behavioural attitudes or affective attitudes as described by Martin 2002. Students' satisfaction will include the satisfaction and attitudes towards the learning intervention to which they were exposed. Students' professional attitudes and satisfaction will only be assessed narratively, as preliminary work conducted by the Global eHealth Unit (George 2014; Rasmussen 2014; WHO 2015) suggests that there is a high level of heterogeneity in the operational definition of these outcomes across different studies.

Blended and non-blended learning interventions will be analysed separately. We will use intention to treat data in all meta-analyses.

Subgroup analysis and investigation of heterogeneity

We anticipate that the following subgroup analyses are likely to be appropriate; those which:

- compare low-, middle- and high-income countries;
- compare pre- and post-registration interventions;
- compare the different qualifications listed in the Health Field of Education and Training (091) of the International Standard Classification of Education (ISCED-F) included in this review, i.e. dental studies, medicine, nursing and midwifery, medical diagnostic and treatment technology, therapy and rehabilitation, and pharmacy (IUS 2013);
- compare one-off interventions to repeated interventions;
- compare interventions provided on difference devices (e.g. personal computer, tablet, mobile phone);
- compare quartiles of adherence/time spent on the intervention. We will recalculate and present the measure of adherence/time spent on the intervention as a percentage to account for the difference in intervention duration between studies.

We acknowledge that there are many other subgroup analyses that could be performed, for example, comparing interventions according to learning objectives and interactivity of interventions. People conducting future reviews after completion of our series of initial reviews are in the best position to do so because such comparisons would be most meaningful from the perspective of an educator if multiple methods of eLearning were to be compared.

Sensitivity analysis

Sensitivity analyses will be considered to explore the impact of the 'Risk of bias' dimensions on the review outcomes. Studies deemed to be at high risk of bias after examination will be removed from meta-analyses. We will evaluate the effect of removal on pooled effect size, based on the following factors:

- high risk of bias studies (as specified above);
- small studies;
- sources of funding, divided into the following categories: industry sponsorship (solely industry funded); mixed sponsorship (public and industry funded, including free provision of study material only); non-industry sponsorship (solely public funded and no free provision of material); or not described/unclear;

- time lapse between end of intervention and first post-test (quartiles).

'Summary of findings' table

We intend to prepare a 'Summary of findings' table to present the meta-analysis results, based on the methods described in chapter 11 of the *Cochrane Handbook for Systematic Reviews of interventions* (Schünemann 2011). We will present the results of meta-analyses for the major comparisons of the review, for each of the major primary outcomes as well as potential adverse effects, as defined in the Types of outcome measures section. We will provide a source and rationale for each assumed risk cited in the table(s). Two authors will use the GRADE criteria to rank the quality of the evidence using the GRADEprofiler (GRADEpro) software (Schünemann 2011). If meta-analysis is not feasible, we will present results in a narrative 'Summary of findings' table format, such as that used by Chan 2011 (CCCRG 2014).