GUIDING THE DIGITAL EDUCATOR
The Instructional Design, Implementation and Evaluation of Postgraduate Medical E-learning

ROBERT DE LEEUW
Guiding the Digital Educator
Robert de Leeuw

**Cover:** the cover of this thesis shows a symbolic interpretation of the main outcomes of this thesis. Digital Education is symbolised as a tree. The roots are planted in experts opinions, learners needs and scientific evaluations and models. The stem is the core of this thesis; the instructional design and evaluation of e-learning. The main branches are Motivate, Learn and Apply. These themes are then divided into five categories; Motivators, Barriers, Learning enhancers, Learning discouragers and Real-world translators. Each category, then has a specific number of blossoms. Those blossoms are the items discussed in this thesis. Different e-learnings might share the stem and branches, but they can all have their own unique set of blossoms and flowers. Finally, at the bottom there are two triangles. One symbolises the analogue educational strategies, slowly being replaced by the other triangle, the digital education.
Guiding the Digital Educator
The instructional design, implementation and evaluation of postgraduate medical e-learning

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op dinsdag 18 juni 2019 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Robert Adrianus de Leeuw
geboren te Numansdorp
promotor: prof.dr. F. Scheele

copromotor: dr. M. Westerman
“Someday, in the distant future, our grandchildren’s grandchildren will develop a new equivalent of our classrooms. They will spend many hours in front of boxes with fires glowing within. May they have the wisdom to know the difference between light and knowledge.”

Plato
Leescommissie:

prof.dr. M.B.M. Zwekhorts (Voorzitter)
prof.dr. P.W. Teunissen
prof.dr. J. Huirne
prof.dr. S.G. Oei
prof.dr. C. Boer
Content

Chapter 1  General introduction  12

Chapter 2  How do we evaluate postgraduate medical e-learning? A systematic review.
            *JMIR Medical Education 2019*

Chapter 3  Quality specifications in postgraduate medical e-learning: an integrative literature review leading to a postgraduate medical e-learning model
            *BMC Medical Education 2016*

Chapter 4  Quality indicators for learner-centered postgraduate medical e-learning
            *International Journal of Medical Education 2017*

Chapter 5  Consensus on Quality Indicators of Postgraduate Medical E-Learning: Delphi Study
            *JMIR Medical Education 2018*

Chapter 6  A theory- and evidence-based Postgraduate Medical E-learning Development Model: nine steps to an empirical educational experience
            *JMIR Medical Education 2019*
<table>
<thead>
<tr>
<th>Chapter 7</th>
<th>Influencing categories in the implementation of postgraduate medical e-learning: a thematic analysis</th>
<th>164</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>BMC Medical Education under review</em></td>
<td></td>
</tr>
<tr>
<td>Chapter 8</td>
<td>The Development and Validation of an instructional design evaluation survey for Postgraduate Medical E-learning</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td><em>JMIR 2019</em></td>
<td></td>
</tr>
<tr>
<td>Chapter 9</td>
<td>Digital vs face-to-face information provision in patient counseling for prenatal screening: a noninferiority randomized controlled trial</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td><em>Prenatal diagnosis 2019</em></td>
<td></td>
</tr>
<tr>
<td>Chapter 10</td>
<td>General discussion</td>
<td>254</td>
</tr>
<tr>
<td></td>
<td>English Summary</td>
<td>276</td>
</tr>
<tr>
<td></td>
<td>Nederlandse samenvatting</td>
<td>284</td>
</tr>
<tr>
<td></td>
<td>Dankwoord</td>
<td>292</td>
</tr>
<tr>
<td></td>
<td>Publications</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td>Curriculum vitae</td>
<td>304</td>
</tr>
</tbody>
</table>
General Introduction
There is very little more satisfying and inspiring than a beautiful environment in which a passionate teacher is revealing the secrets of your favorite topic. Think of a surgical resident who travels to the most advanced operating theater, then has a one-on-one tutorial with the international laparoscopy guru whom she has read so much about. They start working on cadaver training together. Then, the tutor has the most advanced surgery planned and shows the resident how to perform it. After observing, the resident can do the surgery alongside the tutor and finally, she goes home with not only newly acquired skills, but new knowledge about anatomy and complication risks and a new insight into the organization of care, post-surgical checkups, and team collaboration in an operating theater.

This would, or should, be every resident’s dream, but it will not happen. Time restrictions, patient safety, staff limitations, money, and daily care prevent this educational utopia. Or can it happen? We are far from an ideal digital learning environment, but digital learning might overcome some of the barriers to just such a utopia.

Medicine requires dynamic life-long learning [1] and electronic learning (e-learning) might be the future of this life-long learning [2]. The cost-effectiveness and logistical benefits have already been demonstrated [3] and where 10 years ago it was promising, it is now part of mainstream medical education [4]. However, the overall effectiveness and added value of e-learning over ‘conventional’ educational approaches such as face-to-face learning is debatable and the results in the literature are diverse [5]. Reasons for this may include the lack of evaluation instruments, and the diverse outcomes used to study the added value, making comparisons difficult [1, 6]. Even more contradictory are the results in postgraduate medical e-learning (PGMeL), where the added value of e-learning seems negligible [7]. In 2008 Cook et al performed a meta-analysis on the effectiveness of internet learning with all kinds of health professionals. The 201 studies considered, and the pooled effect in comparison to no intervention, favored internet-based interventions. The authors described positive effects on knowledge outcomes, skills, behavior, and patient effects, but the differences from the effects of non-internet instructional methods were small, suggesting an effectiveness similar to that of traditional methods [8]. Based on that review, in 2010 the authors concluded that the aspects most associated with improved learning outcomes were interactivity, practice exercises, repetition, and feedback [9].

In 2016 Sinclair et al performed a systematic review of randomized controlled trials to assess the effectiveness of e-learning on clinician behavior and patient outcomes. After considering seven trials, they concluded that e-learning was at least as effective as traditional approaches, but no papers were identified that reported patient outcomes [10]. They did show the added value in learning as regards location and motivation. In the same year Tiago et al analyzed 251 articles, seeking answers to the question of
what is sought from computer-based learning intervention in medical education. They described an increase in publications about e-learning, of which only 7% used instructional design principles. There seems a high increase in the development and evaluation of e-learning software in medical education [6].

In 2017 an integrative review by Lawn et al concluded that only a few e-learning approaches use high levels of interactivity, reflection, and practice, which should be the added value of e-learning. They suggest that blended learning, combining face-to-face methods and e-learning, supports learning for health care professionals [11]. Finally, in 2018 a Cochrane review included 16 randomized controlled trials involving 5679 health professionals. Compared to traditional learning, they found that e-learning made little or no difference to patient outcomes or health professionals’ behavior, skills, or knowledge [7].

In conclusion, there is no current consensus about the added value in effectiveness of PGMeL. However, there is added value as regards learning as regards location, “just-in-time” learning, motivation, and possibly cost-effectiveness. PGMeL seems effective, but no more effective than traditional learning. Implementation and evaluation are diverse, and few to no instructional design models are used.

For a better understanding of the added value, apart from effectiveness, of PGMeL over conventional learning methods, this thesis will attempt to provide insight into three aspects that are important for the creation and evaluation of PGMeL:

1. which instruments and outcomes are currently used to evaluate PGMeL;
2. which indicators are known to be able to set best practice norms; and
3. how those indicators be used in the design, implementation, and evaluation of PGMeL.

These insights may enable us to guide the design of PGMeL, and compare different forms of it, in search of the most efficient and effective methods of digital education.

In order to better understand the following chapters, an explanation will be given first of the “e” part of e-learning, and then of the “learning” part. This chapter does not aim to provide a comprehensive insight into learning, and all the psychological learning models, but, rather, an introduction to an e-learning-centered approach to learning models. When discussing the ‘e’ in e-learning, it should be noted that the modernization of e-learning goes further than merely adding ‘electronic’. The development of the internet has greatly influenced e-learning and this chapter will provide a short overview of its evolution (Webs 1.0 through 3.0.) After this introduction, the research questions and contents of this thesis will be presented.
The ‘e’

To understand the current landscape and future of e-learning, a brief history is in order. The most influential development is that of the internet, due to the change it brought in affordances [12]. Affordances are features that provide a potential for action, while constraints are those features that provide the structure of and guidance to such affordances [13]. The internet created new opportunities and affordances for e-learning within new constraints [14]. These affordances are among the indicators that this thesis seeks and, to know how to properly evaluate PGMeL, it is useful to know where the internet came from and where it might go.

Web 1.0 is the web of 1989 which, according to Berners-Lee, its founder, could be considered ‘read-only’. This early web provided limited user interaction and content contribution, and only allowed users to search for and read information. Web 2.0 was defined in 2004 as a ‘read-write web’. It allowed the assembly and management of large global crowds with common interests in social interactions [15]. Users were able to create content, and learners could participate more actively in the learning process. By allowing learners to add content as well, it can be seen that Web 2.0 was becoming more learner-centric. Web 3.0 is still in its formative years, but it is being defined as the ‘semantic web’, which desires to decrease human tasks and decisions and, instead, leave these to machines by providing machine-readable content [15]. This allows e-learning 3.0 to be undertaken anytime, anywhere, while interaction with artificial intelligence (A.I.) provides feedback [16]. While Web 1.0 was mainly read and Web 2.0 was read and write, Web 3.0 adds execution. The next steps are Webs 4.0 and 5.0, which are both in the future. Web 4.0 is a symbolic web, where machines can react to human needs based on mind-controlled interfaces [15]. Web 5.0 belongs to a more distant future and might involve the sensory and emotional aspects of machine learning, providing even bigger challenges in teaching students how to interpret internet content correctly [17]. (See figure 1 for an overview.)

The development of the internet allowed the use of new affordances (see figure 1) and has changed the way educators can use the digital medium for e-learning. Because e-learning increasingly used the internet, many modern e-learning platforms make great use of these affordances and can therefore be of great added value. To understand this value, an understanding of the learning theories on which these affordances are based is required.
Figure 1 – the development of the internet, e-learning, and the affordances that followed

Legend for figure 1: In the first column, the development of the internet is described in general terms. The development of e-learning followed the same steps and educators use increasingly advanced aspects of the internet. The effect of these new aspects is the use of new affordances, based on learning theories.
About (e-)learning

E-learning is a widely (ab)used term to signify any form of digital content. Putting a PowerPoint presentation online or publishing an article in PDF format on a website might be called e-learning by some; however, these certainly do not fulfil the requirements to be considered such. Creating a well-designed, interactive environment where users can explore new knowledge might be an e-learning, but it does not have to be. It is debatable, for example, whether Google Earth could be characterized as an example of e-learning.

There has been much debate around the interpretation of the word e-learning. One international study which evaluated the way educators use the word concluded that and most participants seemed more familiar with the terms ‘e-learning’ and ‘online learning’ than, for example, ‘distance learning’, but the findings showed many differences as to what ‘e-learning’ meant [18]. Because of the diverse interpretation of the word, a Delphi of Sangra et al in 2012 formulated the definition thus:

“E-learning is 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 knowledge, skills and/or behavior/attitude” [19].

Ellaway et al stated that e-learning should contain a pedagogical approach, and typically aspires to be flexible, engaging, and learner-centric. We agree with Ellaway et al and believe that e-learning consists of more than just an electronic element. Making an educational tool requires the knowledge and use of known insights into how people learn, and how to educate. This means that affordances in PGMeL should at least be supported by learning theories. Because Sangra et al combined ‘electronic’ and ‘learning’, their definition has been used throughout this thesis.

Learning theory is a very diverse topic, with many interpretations and models. At a theoretical level, there seems to be increasing agreement about the psychological fundamentals [20], but an increasing number of working models. One way of organizing the mass of terminology is shown by the pyramid in figure 2. At the basis are the fundamental learning theories. Then, depending on the target audience (in our case medical postgraduates) and learning objectives, specific learning theories can be followed. These theories lead to instructional design models for pedagogical frameworks. From these models, concrete instructional principles are formulated which can be used to actually create teaching and learning activities. These five levels are examined individually below.
Figure 2 – the learning theory building blocks of e-learning

Legend for figure 2: after the development of certain fundamental learning theories, researchers started to develop learning theories for specific target audiences and learning objectives. Based on those theories, instructional design models were built. Instructional design models in their turn provided educators with principles they can use in daily practice when teaching and during learning activities.
Fundamental learning theories

As with all theories, there are many routes to follow. A rather pragmatic way of looking at these fundamental theories is to create three clusters of broad perspectives, as per Greeno et al [21]:

1. Learning as an activity (behaviorism);
2. Learning as achieving understanding (cognitivism); and
3. Learning as a social practice (social-culturalism).

Box 1 at the end of this chapter, presents a brief summary of each perspective to enable further background reading. This thesis, however, will focus on PGMeL and the use of electronic and multimedia learning media, and look at an adult target audience in medicine. According to Khalil et al, the relevant specific learning theories for PGMeL are:

1. cognitive load theory;
2. multimedia learning; and
3. adult learning theory [22].

A short introduction to these three learning theories is provided below, with an indication of their relevance to this thesis.

Cognitive load theory (CLT)

Lifelong learning requires the desire to learn, and the ability to learn [23]. One way to create a working model for the ability to learn is the cognitive load theory (CLT) [24, 25]. CLT provides a model of a cognitive architecture that can be used to align instructional principles [22]. This model consists of a sensory memory, working memory, and long-term memory. As can be seen from figure 3, a sensory input is processed in the brain in order to remember. The sensory memory is vast, but very short-lived. The working memory is limited, but is the bridge to long-term memory. Things pass from working memory to long-term memory by cognitive schemes that can be helped by repetition and the chunking of information. Learning is described as the construction of such schemas. These are assumptions based on the fundamental cognitivism and constructivism theories. The working memory can handle a maximum load. According to CLT, there are three kinds of cognitive loads: intrinsic (load from the complexity of the performed task and the expertise of the learner); extraneous
(load generated by suboptimal faulty presentation of instructional materials); and
germane (load caused by learning processes from interactions with the appropriate
instructional design that contributes to learning) [26]. To balance these loads,
extraordinary load should be minimal, intrinsic load titrated to the development stage of
the learning, and all unused working memory dedicated to the germane load [27].

CLT may not be relevant to teaching simple tasks, but has been developed to help
complex learning that imposes high load on the learner’s cognitive system and is
therefore very useful in medical education [26, 27]. This theory forms the foundation
for a series of instructional principles in medical education, which in their turn can be
used in daily practice [26]. CLT is therefore very well suited to support instructional
principles for medical education, and because the model leans on sensory memory and
describes the input of information as visual and auditory, it is also very well suited for
e-learning. Principles found and used in this thesis are largely based on CLT. It forms
the foundation of a model frequently used in e-learning, the multimedia theory, which
is discussed in the next section.

*Figure 3—the cognitive load theory*
Multimedia learning is defined as learning from words and pictures, preferably verbal and pictorial learning. This theory is based on three principles derived from CLT:

1. the dual channel principle, which proposes that learners have separate channels for processing verbal and pictorial material;

2. the limited capacity principle, which proposes that learners can process only a few elements in each channel at any one time; and

3. the active processing principle, which proposes that meaningful learning occurs when learners engage in appropriate cognitive processing during learning, including attending to relevant material, mentally organizing it into a coherent cognitive representation, and integrating it with prior knowledge activated from long-term memory [28].

The goal of multimedia theory and CLT is to use evidenced-based principles for instructional design that reduce extraneous cognitive processing and help to manage the amount of information presented to learners to avoid overloading their cognitive capacity [22]. Multimedia theory is frequently used to create e-learning. Since e-learning is in essence a form of multimedia (usually containing a combination of visual and auditory stimuli), the above described principles are frequently used and supported in this thesis.

Because PGMeL targets adult learners, the above-mentioned theories must be placed in the perspective of adult learning.
Adult learning theory

In 1968 Malcolm Knowles proposed to distinguish adult learning from preadult learning. The concept of andragogy (helping adults learn) was contrasted with pedagogy (helping children learn) [29]. There are five assumptions underlying andragogy that differentiate how adults learn compared to children [29]:

1. Self-concept: an adult learner is independent and self-directing in his/her learning;
2. Experience: an adult learner greatly depends on a broad amount of experience which is a resource for learning;
3. Readiness to learn: an adult learns to develop the tasks belonging to his/her social roles;
4. Orientation to learning: adults’ learning is problem-centered and needs immediate application; and
5. Motivation: a person’s motivation to learn is internal rather than external.

Based on this theory, instructional principles follow which can be used when designing PGMeL. Moreover, while CLT and multimedia learning focus more on the use of instruments, andragogy focusses on the target audience [30]. This theory was greatly critiqued in the 70s and 80s; indeed, the suggestion was made that it is not so much a theory but a description of “what the adult learner should be” [31]. A second area of criticism is the question of whether these assumptions are for adult learners only, or whether children have the same needs and characteristics [31]. The debate around these principles has continued, and adult learning in medicine has paid increased attention to self-regulated learning, or lifelong learning [32, 33]. Lifelong learning refers to the systematic acquisition, renewal, updating, and completion of knowledge throughout one’s career [24]. According to Hojat et al, it is a “set of self-initiated activities and information-seeking skills with a sustained motivation to learn and the ability to recognize one’s own learning needs” [34]. Despite the long debate, variations of the adult learning theory are still used and evaluated [35].

The choice of adult learning theory may be justified by the focus of this thesis on postgraduates. Despite the critique, it can be said that graduates (learning fulltime and motivated by tests and tuition fees) have different learning needs and motivators from postgraduates (learning while working and motivated more by intrinsic needs and peer pressure). It therefore makes sense to focus the questions of this thesis on a specific group, instead of mixing the needs and motivations of different groups. The above paragraphs describe the most influential learning models for PGMeL (according to Kahlik et al [22]). A basic knowledge of these models helps to understand the indicators, affordances, and evaluation methods discussed in this thesis.
E-learning and learning theories

Along with the changes in the web and new insights from learning theories, e-learning itself changes as well. E-learning 1.0 evolved along with many Learning Management Systems (LMS) that used various aspects of traditional learning and allowed the incorporation of the learning theories of behaviorism and cognitivism [16]. E-learning 2.0 allowed the use of other learning theories, such as constructivism and social constructionism [16]. E-learning 3.0 had the capabilities of e-learning 1.0 and 2.0, and a rich 3D virtual learning environment with links to big data, cloud computing, and the semantic web [16, 17].

Having introduced learning, and the evolution of the internet and e-learning, the current section links theory with practice. In 1974 Snellbecker et al introduced the term ‘Instructional Design’ (ID) as “a link between the science of how people learn and daily practice as a process for designing instruction based on empirical principles” [36]. Another way to describe ID is a way to manage the development of a learning instrument effectively so that it will increase competent performance by a student [37]. ID is also sometimes called Curriculum Design (CD) and the difference depends greatly on one’s definition of curriculum. Kemp et al defined a curriculum as ‘All learning which is planned and guided’ [38], which suggests a great overlap between ID and CD, although the latter can be interpreted as a broader design than a series of instructions. Merrill et al compared different ID theories and models for their principle of e-learning (as described above) and concluded that they all share a series of first principles, although no one theory or model includes all principles. Differences can be based on different theoretical insights or in the details following the first principles, depending on, for example, the target audience [39].

Therefore, when evaluating PGMeL, there are two aspects that should be evaluated:

1. the learning aim (knowledge, skills, or attitude/behavior); and

2. the Instructional Design
Because the ID will greatly influence the effectiveness of the learning instrument, it is necessary to know how good the design was, in order to gauge whether there is a more efficient way to reach the learning aim. It can be debated that the first is the result of the second; therefore, you can only measure the learning aim. If the learning aim is not met, this is probably due to an incorrect or inappropriate design. However, there is a flaw in that reasoning. Let us suppose that face-to-face learning is compared to e-learning with knowledge as the primary learning aim. If the face-to-face group scores significantly higher in the outcome test, the conclusion may be that face-to-face learning was a better medium for this topic. However, it must be asked what caused this: whether the digital medium of e-learning was the correct choice, or whether the poor ID of the e-learning was at fault. Perhaps with a better ID, the e-learning could have been more effective and better than, or as good as, the face-to-face learning.
Central research questions and overview

This chapter demonstrates the need for a learning theory, making use of ID principles and showing that the current literature rarely does this. The conclusion is that e-learning is used more and more frequently, but poorly evaluated in light of the learning theories discussed above. E-learning offers valuable new opportunities for PGMeL [2]. E-learning also offers participants an advantage in that they can choose a comfortable and accessible place and time to study, which is important in postgraduate and continuous medical education [3]. However, one of the problems is that the results of studies directly comparing technology-assisted education with traditional teaching often conflict, and frequently do not demonstrate or propose best practices [4, 5]. It is known that e-learning should be targeted to the needs of a specific audience (in this case postgraduates) [40, 41]. The success of e-learning programs has also been linked to the use of a theoretical framework or a learning theory [42]. However, no evaluation instruments or design models exist for PGMeL. There is also no way to properly evaluate the ID of PGMeL.

The central research questions within this thesis are:

1. What instruments or outcomes are currently used to evaluate PGMeL?
2. Which indicators are known in the current literature that can set best practice norms for PGMeL?
3. Which of these indicators are acknowledged and which are missing according to the main stakeholders, namely users, educational experts, and content creators?
4. How can these indicators be used to evaluate and design PGMeL?
5. When this e-learning is created, which factors influence the implementation and how can the team prepare for successful implementation?
To answer those questions, eight research projects have been undertaken. Chapter two presents a narrative review to evaluate which e-learning affordances (functions and characteristics such as feedback, quiz questions, etc.) are described as quality indicators for PGMeL, which will be used to answer question two. Chapter three consists of a systematic review, aiming at a list of primary outcomes and previously used models. This will allow the determination of which models are already available and how they compare to the indicators from the first review, and answer question two. Chapter four presents a focus group discussion to further evaluate the indicators with experienced users, commercial e-learning creators, and educational experts. Chapter five consists of an international Delphi with a group of users and widely published e-learning experts re-evaluating the list of items. Chapters two, four and five will together answer question three.

With the final resulting items, an ID model for PGMeL is created and presented in chapter six. Chapter seven creates a survey and validates the usefulness, usability, and added value of an evaluation tool, thus answering question four, together with the previous chapter. Chapter eight will address question five, namely how to implement an e-learning based on the model. Experienced authors are interviewed about the factors that influence the implementation of e-learning, and their answers are compared to Rogers’ theory on the diffusion of innovation. Given that it was necessary to evaluate each item on the list of items, design model, and evaluation tool separately, chapter nine presents the findings of a randomized controlled trial comparing patient education with and without interactive elements. The question asked is whether digital information provision, with or without interactivity, can be just as satisfying, effective, and time-efficient as a face-to-face meeting.

Finally, chapter ten will address all the main findings and overall conclusions drawn from the research projects. Strengths and weaknesses are described, and future research discussed. The thesis concludes with a personal vision of the future of e-learning.
Box 1 – fundamental learning theories

Behaviorism (learning as behaviour)

Behaviorism assumes that learning is essentially passive and a learner is responding to environmental stimuli as positive and negative reinforcement [43]. Positive and negative reinforcement will increase the probability that the antecedent behavior will happen again, while punishment (both positive and negative) decreases the likelihood. Behaviorism is concerned with what students do (responses) as a result of environmental stimuli and environmental consequences. A true behaviorist has no interest in what or how students think or feel [44]. Therefore, it is teacher-centered: the teacher controls the environment and stimuli, and reinforces the wanted behavior. A lot of early work was done with animal experiments by people such as Thorndike, Pavlov, Skinner, and Watson.

Cognitivism (learning as achieving understanding)

While behaviorism focusses on observable output, cognitive theories lean on how knowledge is acquired, constructed, and represented in the mind and subsequently remembered [44]. In cognitivism, learners are active participants and each learner has a different level of knowledge, skill, and motivation that influence the outcome. In response to behaviorism, it posits that people do not merely respond to stimuli, but require active participation to learn and that actions are a consequence of thinking [45]. One framework from this line of thinking is the dividing of the memory into a sensory registrar, a short-term memory, and a long-term memory. This framework helps to guide students from hearing something to gaining long-term knowledge. A variant is the social cognitive theory, mainly proposed by Albert Bandura. Bandura argued that observations and experiences can be symbolically represented in the mind and subsequently accessed to inform future behavior [44]. He placed the focus on attention, retention, production, and motivation. Examples of important contributors are Merrill, Gagne, and Schank.
Socio-constructivism (learning as social practise)

Theorists such as Piaget and Vygotsky viewed learning as a social process wherein new knowledge is implemented based on already constructed schemes on two levels: the interpersonal (external level) and the intrapersonal (internal level) [44]. Learning is linked to prior knowledge; thus, mental representations are subjective. The learner is, therefore, not a blank slate (tabula rasa), but brings past experiences and cultural factors to the learning process. According to Vygotsky, learners cannot learn internally, without external exposure: “semiotic mechanisms mediate social and individual functioning and connect the external and the internal, the social and the individual.” Examples of those semiotic mechanisms are any number of cultural tools, language, symbol systems, calendars, processes, art, maps, writing, writing utensils, technology, and machinery [22].

From these three fundamental theories, many other learning models have arisen. Readers are recommended to visit www.learning-theories.com, where over 100 learning theories offer greater insight into this area.
Literature


How do we evaluate postgraduate medical e-learning?

A systematic review

De Leeuw RA
de Soet A
van der Horst S
Walsh K
Westerman M
Scheele F

JMIIR Medical Education
February 2019
Abstract

Background: Electronic learning (e-learning) in postgraduate medical education has seen a rapid evolution; however, we tend to evaluate it only on its primary outcome or learning aim, whereas its effectiveness also depends on its instructional design. We believe it is important to have an overview of all the methods currently used to evaluate e-learning design so that the preferred method may be identified and the next steps needed to continue to evaluate postgraduate medical e-learning may be outlined.

Objective: This study aimed to identify and compare the outcomes and methods used to evaluate postgraduate medical e-learning.

Methods: We performed a systematic literature review using the Web of Science, PubMed, Education Resources Information Center, and Cumulative Index of Nursing and Allied Health Literature databases. Studies that used postgraduates as participants and evaluated any form of e-learning were included. Studies without any evaluation outcome (e.g., just a description of e-learning) were excluded.

Results: The initial search identified 5973 articles, of which we used 418 for our analysis. The types of studies were trials, prospective cohorts, case reports, and reviews. The primary outcomes of the included studies were knowledge, skills, and attitude. A total of 12 instruments were used to evaluate a specific primary outcome, such as laparoscopic skills or stress related to training. The secondary outcomes mainly evaluated satisfaction, motivation, efficiency, and usefulness. We found 13 e-learning design methods across 19 studies (4% 19/418). The methods evaluated usability, motivational characteristics, and the use of learning styles or were based on instructional design theories, such as Gagne’s instructional design, the Heidelberg inventory, Kern’s curriculum development steps, and a scale based on the cognitive load theory. Finally, 2 instruments attempted to evaluate several aspects of a design, based on the experience of creating e-learning.

Conclusions: Evaluating the effect of e-learning design is complicated. Given the diversity of e-learning methods, there are many ways to carry out such an evaluation, and probably, many ways to do so correctly. However, the current literature shows us that we have yet to reach any form of consensus about which indicators to evaluate. There is a great need for an evaluation tool that is properly constructed, validated, and tested. This could be a more homogeneous way to compare the effects of e-learning and for the authors of e-learning to continue to improve their product.
Introduction

Background

Electronic learning (e-learning) in postgraduate medical education has seen a rapid evolution [1,2]. Moreover, e-learning has become a central part of education, whether stand-alone, part of hybrid learning, or an essential element in the successful flipped classroom concept [3-5].

Although postgraduate medical e-learning (PGMeL) is becoming part of mainstream education, its effectiveness has been subject to debate. A Cochrane review from 2018 concludes that comparing e-learning with traditional learning seems to reveal little to no difference [6]. Yet, other studies show great benefits with regard to primary outcomes [7,8] or secondary aspects such as environmental impact [9].

A possible reason for this discrepancy can be the heterogeneity in instructional design and other elements of e-learning that are poorly evaluated [10]. PGMeL is frequently evaluated by means of a pre- and posttest of the primary learning aim (eg, new knowledge) [11]. However, every educational instrument has functionalities and elements that are used to optimize its effect. The elements required for a specific e-learning model are defined in the so-called instructional design. These elements are also called affordances and have the purpose of maximizing the effect, effectiveness, and usefulness of an educational instrument [12]. Therefore, the affordance of an instrument is an action made possible by the availability of that tool (eg, interactive videos) [13]. Although several reviews of the effects of e-learning have been carried out, little has been written about the ways in which an e-learning’s instructional design may be evaluated [6,14]. A valuable introduction to the design of e-learning was given by Ellaway and Masters, who provide certain guidelines but no method of evaluation [15].

We believe it is of great importance to have a better insight into the current PGMeL evaluation methods and which outcomes (primary or secondary) are used. The aim of this study was to provide an overview of the outcomes used to evaluate PGMeL and the evaluation methods of the models used. To do so, we first need to provide a working definition of e-learning for this review.
Electronic Learning Definitions

The definition of e-learning changed with the evolution of the internet, and most definitions fail to describe the subtleties and certain important aspects of e-learning. It does not simply consist of placing documents in an electronic format via the internet. It should encourage interaction, collaboration, and communication, often asynchronously [15]. For this literature review, we have chosen the following, slightly adapted, definition from the study by Sangra et al [16]:

E-learning is 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 knowledge, skills and/or behaviour/attitude.
Methods

Study Design

A systematic review was carried out to determine how PGMeL can be evaluated and which outcomes are used. Some studies compared e-learning with other learning methods in trials or cohorts, whereas others were conducted from case reports by authors who evaluated a newly used e-learning method alone. We followed all the steps laid out in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines because the risk of bias is not relevant in answering our question [17]; given that we are not looking at the results of the outcomes but, rather, at the content of the outcomes themselves, we did not evaluate the risk of bias.

Types of Studies and Participants

The types of studies included are trials, reviews, and other descriptive evaluation studies as well as all the studies that evaluated any form of e-learning, as defined above, that have postgraduate medical professionals as a target audience.

Study Eligibility

The inclusion criteria were as follows:

1. Any e-learning evaluation study (studies without any evaluation outcome were excluded)
2. Postgraduate target audience for the e-learning
3. Published in English
4. Published after the introduction of web 1.0 (1994)

Type of Intervention and Outcomes

The type of intervention was any form of e-learning, as discussed in the introduction. Given that the purpose of this review was to overview the kinds of outcomes used, all outcomes were included. We differentiated between primary and secondary outcomes. A primary outcome was defined if the study described the outcome as a primary outcome, if a sample size was calculated based on that outcome, or when the authors defined the outcome in the research question. If it was not clear what the primary outcome was, all outcomes were used as primary outcomes.
Study Identification

The literature search was performed in November 2017, searching PubMed, Education Resources Information Center, Cumulative Index of Nursing and Allied Health Literature and Web of Science databases separately. The search string was quite extensive and used a combination of Medical Subject Headings terms and a combination of title and abstract keywords. The complete string may be found in Multimedia Appendix 1.

Study Selection

Working independently and in duplicate, reviewers (RDL, ADS, and SVH) screened all article titles and abstracts. Potentially eligible abstracts and abstracts with disagreement or insufficient information were screened in full text. Disagreements were handled by discussing the full text and the majority counts. The dataset supporting the conclusions of this study is available in the Postgraduate ME Model repository [18].
Results

Search Results

The initial search identified 5973 articles, of which 4691 were left after removing all duplicates. The titles and abstracts were read to determine the relevance, outcomes, and target audience. After handsearching and snowballing, 824 possible studies remained for review. After reading the full texts of these articles, we rejected 406 as not being targeted at the right audience or not evaluating the e-learning but only describing it. We used 418 final articles for our analysis, as shown in the flow diagram in Figure 1, which all evaluated an educational intervention that satisfied our definition of e-learning. For a list of all 418 studies, please refer to Multimedia Appendix 2.
Figure 1. Search and article selection process. e-learning: electronic learning; CINAHL: Cumulative Index of Nursing and Allied Health Literature; ERIC: Education Resources Information Center.
General Characteristics

The types of studies were trials (n=201), prospective cohorts (n=110), case reports (n=98), and reviews (n=9). We found a variation of e-learning methods and combined these into 4 categories: serious gaming (n=8), virtual reality (n=90), simulation (n=79), and theoretical knowledge–aimed e-learning (n=241). We added augmented reality into the virtual reality group (Figure 2). Most of the e-learning was created for general medicine (n=86), followed by surgery (n=84), internal medicine (n=59), pediatrics (n=32), gynecology (n=28), and family medicine (n=23; Figure 3). Studies were grouped under general medicine when they were multidisciplinary. A group of 16 studies had no specified target audience. Family medicine was grouped together with primary care.

Figure 2. Types of electronic learning (%). Knowledge refers to any acquaintance with facts, truths, or principles. Simulation refers to any form of digital imitation of enactment that is not virtual reality. Virtual reality refers to a simulation of a 3-dimensional environment, experienced or controlled by movement of the body. Serious gaming refers to a learning environment with gamification elements aimed at learning rather than entertainment. e-learning: electronic learning.
Study Outcomes

The learning aims of the included studies were knowledge (n=286), skills (n=130), and attitude (n=2), which reflected the primary outcomes. Knowledge was tested by pre- and postcourse tests, and 12 instruments were used to evaluate an e-learning-specific primary outcome (see Table 1), such as laparoscopic skills or stress related to training.

The secondary outcomes of the studies were both more diverse and more focused on the design (see Table 2). The most prevalent evaluated outcomes were satisfaction (n=99), self-efficacy (n=60), adherence in practice (n=33), and time spent (n=32). Overall, 28 studies had some sort of qualitative evaluation, such as focus discussions or personal interviews. To prevent too diverse a series of outcomes, we grouped comparable outcomes together. Therefore, satisfaction can be measured by using a Likert scale but also by asking if someone would recommend the e-learning to other residents. Adherence in practice can be self-reported practice change or objective changes in practice, for example, subscription practice. We used the term self-efficacy for each form of self-assessed confidence, understanding, or comfort in clinical or theoretical settings.
A total of 5 studies used Kirkpatrick’s levels of evaluation. These levels were more used as secondary outcomes of the learning aim than as a design evaluation method [29,33-36]. Kirkpatrick described a 4-level framework of evaluation for measuring the effectiveness of training or human performance technology programs originally aimed at corporate human resources [37]. The levels are reaction, learning, behavior, and results. Aitken et al evaluated their radiology e-learning material based on the first 2 levels, using the framework to build an evaluation questionnaire [34]. Sim et al focused on learning, behavior change, and impact on workplace by quantitative pre-, mid- and postmodule surveys; qualitative Web-based discussions; and short facilitator meetings [33]. In 2016, Bowe et al evaluated their e-learning program by means of the Kirkpatrick framework, but a narrative review provided them with the 3 other evaluation tools discussed below as well [29]. Finally, Patel et al undertook a review to establish the effectiveness of simulation in interventional radiology and evaluated which level of Kirkpatrick’s hierarchy the studies reached, with only 1 reaching level 4. No proper validation of PGMeL has been carried out, and there are many concerns about the overgeneralization and misunderstandings that compromise its evaluation [38]. One study by Sears et al [39] used Robert and McDonald’s revision of Kirkpatrick’s levels, where the third and fourth levels fall into an overall practice domain and a new level, value, is added to better suit current technologies and continuing education approaches.
Table 1. Discipline of skill-specific outcome measurement tools.

<table>
<thead>
<tr>
<th>Name</th>
<th>Evaluation topic</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vandenberg and Kuse Mental Rotations Test</td>
<td>Laparoscopic skills</td>
<td>Ahlborg [19]</td>
</tr>
<tr>
<td>Arthroscopic Surgery Skill Evaluation Tool</td>
<td>Arthroscopic skills</td>
<td>Waterman [20]</td>
</tr>
<tr>
<td>Stanford Microsurgery and Resident Training Scale</td>
<td>Microsurgery skills</td>
<td>Satterwhite [21]</td>
</tr>
<tr>
<td>Global Operative assessment of laparoscopic Skills</td>
<td>Laparoscopic skills</td>
<td>Rinewalt [22]</td>
</tr>
<tr>
<td>McGill inanimate system for training and evaluation of laparoscopic skills</td>
<td>Laparoscopic skills</td>
<td>Martinez [23]</td>
</tr>
<tr>
<td>Objective Structured Assessment of Technical Skills (OSAT)</td>
<td>Laparoscopic skills</td>
<td>Tomaz [24]</td>
</tr>
<tr>
<td>Evaluating the Attitude towards research test</td>
<td>Attitude</td>
<td>Pelayo [25]</td>
</tr>
<tr>
<td>Survey of Attitudes towards Achieving Competency in Practise-Based Learning and Improvement and System-Based Practise (SAAC)</td>
<td>Managed care competencies and attitude</td>
<td>Yedidia 2000 [26]</td>
</tr>
<tr>
<td>Attitude, belief and behaviour survey regarding domestic violence</td>
<td>Domestic violence attitude</td>
<td>Harris [27]</td>
</tr>
<tr>
<td>State Trait Anxiety Inventory</td>
<td>Stress</td>
<td>Samakar [28]</td>
</tr>
<tr>
<td>Mini Mental State Exam</td>
<td>Stress</td>
<td>Tomaz [29]</td>
</tr>
<tr>
<td>Attitude towards Health Care Teams Scale</td>
<td>Teamwork</td>
<td>Bowe [29], Leipzig [30]</td>
</tr>
<tr>
<td>Assessment care for the vulnerable elderly</td>
<td>Elderly care</td>
<td>Halmboe [31]</td>
</tr>
<tr>
<td>Cumulative sum analysis (CUSUM) for colorectal histology</td>
<td>Histology</td>
<td>Patel [32]</td>
</tr>
<tr>
<td>Outcome</td>
<td>Statistics, n (%)</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>satisfaction</td>
<td>88 (19.9)</td>
<td></td>
</tr>
<tr>
<td>self-efficacy</td>
<td>60 (13.6)</td>
<td></td>
</tr>
<tr>
<td>adherence in practice</td>
<td>31 (7)</td>
<td></td>
</tr>
<tr>
<td>long term knowledge</td>
<td>28 (6.3)</td>
<td></td>
</tr>
<tr>
<td>qualitative evaluation</td>
<td>28 (6.3)</td>
<td></td>
</tr>
<tr>
<td>time spent</td>
<td>27 (6.1)</td>
<td></td>
</tr>
<tr>
<td>skills</td>
<td>25 (5.7)</td>
<td></td>
</tr>
<tr>
<td>attitude</td>
<td>20 (4.5)</td>
<td></td>
</tr>
<tr>
<td>usefulness</td>
<td>16 (3.6)</td>
<td></td>
</tr>
<tr>
<td>efficiency</td>
<td>8 (1.8)</td>
<td></td>
</tr>
<tr>
<td>confidence</td>
<td>8 (1.8)</td>
<td></td>
</tr>
<tr>
<td>usability</td>
<td>8 (1.8)</td>
<td></td>
</tr>
<tr>
<td>acceptability</td>
<td>6 (1.4)</td>
<td></td>
</tr>
<tr>
<td>preference</td>
<td>6 (1.4)</td>
<td></td>
</tr>
<tr>
<td>costs</td>
<td>5 (1.1)</td>
<td></td>
</tr>
<tr>
<td>presentation quality</td>
<td>5 (1.1)</td>
<td></td>
</tr>
<tr>
<td>knowledge</td>
<td>4 (0.9)</td>
<td></td>
</tr>
<tr>
<td>motivation</td>
<td>4 (0.9)</td>
<td></td>
</tr>
<tr>
<td>stress</td>
<td>3 (0.7)</td>
<td></td>
</tr>
<tr>
<td>patient satisfaction</td>
<td>2 (0.5)</td>
<td></td>
</tr>
<tr>
<td>agreement</td>
<td>1 (0.2)</td>
<td></td>
</tr>
<tr>
<td>discomfort</td>
<td>1 (0.2)</td>
<td></td>
</tr>
<tr>
<td>overall reaction</td>
<td>1 (0.2)</td>
<td></td>
</tr>
<tr>
<td>participation</td>
<td>1 (0.2)</td>
<td></td>
</tr>
<tr>
<td>readiness to change</td>
<td>1 (0.2)</td>
<td></td>
</tr>
<tr>
<td>screening percentage</td>
<td>1 (0.2)</td>
<td></td>
</tr>
<tr>
<td>cognitive load</td>
<td>1 (0.2)</td>
<td></td>
</tr>
</tbody>
</table>
Electronic Learning Design Evaluation Methods and Theories

Overall, 19 studies (4%) used some form of tool to evaluate the e-learning design, and 13 tools were described in these studies. These 19 studies alone provided us with the methods and theories at which our initial research question was aimed.

Two instruments focused on usability, namely, the System Usability Scale (SUS) and the Software Usability Measurement Inventory (SUMI).

The System Usability Scale (n=5)

This is a 10-item questionnaire developed by Brooke that measures the usability of computer systems in 3 domains: effectiveness, efficiency, and satisfaction. It has been freely available since 1986 and has been cited in more than 1200 publications [40]. Davids et al used the SUS first to evaluate an e-learning resource for electrolyte and acid-base disorders [41] and again in 2014 to evaluate the effect of improving usability [42]. The SUS was also used by Gorrindo et al [43], Diehl et al [44], and Gillespie in 2017 [45].

The Software Usability Measurement Inventory (n=1)

According to Deraniyagala et al, there are multiple approaches to measuring usability, but the gold standard is the SUMI because of its extensive validations and long track record of success in evaluation [46]. It consists of a 50-item questionnaire devised in accordance with psychometric practice and was inspired by the 1993 ISO 9241 definition by Kiralowski and Corbett [47].

A total of 3 instruments attempted to evaluate the motivational characteristics of the design.

The Motivated Strategies for Learning Questionnaire (n=1)

Ahlborg et al used a few items from the Motivated Strategies for Learning Questionnaire to evaluate self-efficacy [19] and Cook et al validated the entire questionnaire [48]. It consists of a self-reported, Likert scale instrument developed by Pintrich et al in 1993, which aims to assess the motivation and use of learning strategies by college students [49]. Cook et al concluded that the scores are reliable and offer meaningful outcomes for residents in a Web-based course.
Keller’s Instructional Attention, Relevance, Confidence, and Satisfaction Motivation Model (n=2)

This proposes to assess the motivational characteristics of instructional materials or courses using an Attention, Relevance, Confidence, and Satisfaction (ARCS) model of motivation and was validated by Cook et al with 124 internal medicine residents [50]. Although the data were limited, they support the validity of the survey. Kawamura et al used the system as well to determine factors of motivation in serious gaming [51].

Instructional Materials Motivation Survey (n=1)

Cooke et al validated the Instructional Materials Motivation Survey (IMMS) to assess the motivational characteristics of a course [50]. The IMMS is an instrument developed by Keller using his ARCS model. The aim of the tool is to improve a course design generally or to adapt a course to an individual’s needs.

The 2 scales focused on the use of learning styles as described in the following sections.

The Learning Style Index (n=2)

The Learning Style Index [52,53], developed in 1988 by Richard Felder and Linda Silverman, is designed to capture the most important learning style of engineering students, differentiated by 4 dimensions (active-reflective, visual-verbal, sensing-intuitive, and sequential-global) [54]. Cook et al evaluated whether the preferred learning style had any effect on a Web-based course and questions. Cognitive and learning styles had no apparent influence on learning outcomes [53].

Riding’s Cognitive Style Analysis (n=1)

Riding’s Cognitive Style Analysis (RCSA) determines whether an individual has a particular cognitive style or a preferred way of processing information [53]. The RCSA test measures the cognitive style on a verbal-imagery dimension and a holistic-analytic dimension [55].

A total of 4 tools were based on previous instructional design theories: Gagne’s instructional design, the Heidelberg inventory, Kern’s curriculum development steps, and a scale based on cognitive load theory.
Gagne’s Events Instructions (n=1)

The instructional design by Gagne et al has been a classic in learning since 1974 and is a general, instructional design theory [56]. It has 9 parts, mirroring Gagne’s idea of the cognitive stages associated with adult learning [57]. The model is used as a framework for designing any adult education instrument.

Heidelberg Inventory for the Evaluation of Teaching (n=1)

The Heidelberg Inventory for the Evaluation of Teaching [58] is a standardized, psychometric questionnaire for the didactic quality assessment of the whole program. It consists of 13 domains and 42 items/questions and was developed to evaluate teaching methods for German undergraduate students [59].

Kern’s 6-Step Curriculum Development for Medical Education (n=1)

This approach [60], described by Kern et al in 2009, aimed to create a planned educational experience with a logical, systematic approach [61].

Learner’s Scale (n=1)

This series of scales [62] is composed of learner satisfaction, self-efficacy, mental effort, and time on task. The questions used for these scales are based on cognitive load principles and multimedia learning, which are based on the work by Clark and Mayer [63] and van Merrienboer [64].

Finally, 2 instruments attempted to evaluate several aspects of a design, based on the experience of creating e-learning.

The 10 Golden Rules for Software Design (n=2)

Created to help in designing software in medical education, this [36,65] starts with a 51-item questionnaire based on the Context, Input, Process, and Product model by Stufflebeam [66]; the Convenience, Relevance, Individualization, Self-assessment, Interest, Speculation, and Systematic criteria [67]; and Kirkpatrick’s 4 levels of evaluation. The questionnaire was then piloted and used to evaluate an interactive distance education course in obstetrics and gynecology [36]. From the qualitative data, 10 common items were identified and represented in the form of *10 golden rules*. 
Quality Improvement Knowledge Application Tool-Revised (n=1)

A revision of the original Quality Improvement Knowledge Application Tool, validated to assess practice-based learning and the system-based practice of residents, the Quality Improvement Knowledge Application Tool-Revised (QIKAT-R) [29,68] consists of 3 subjects—aim, measure, and change—and participants are asked to score the presented scenarios on these subjects.

Apart from these evaluation methods, we found 4 studies that did not evaluate e-learning but did use evaluation methods to create their e-learning. These used instruments to create e-learning with a focus on outcomes, motivation, and technology acceptance:

The Formative Process and Outcome Evaluation Tool by Dunet

Dunet et al [69] described the evaluation process by which they created a course—formative evaluation (content and design), process evaluation (knowledge gain, motivation, and usefulness), and outcome evaluation.

The Website Motivational Analysis Checklist

The authors reviewed an education database and did not find any validated tools. Therefore, they used the Website Motivational Analysis Checklist [70], which was originally created to assess service-based commercial websites in 2000 [71].

Davis’s Technology Acceptance Model and Laurillard’s Model of Interactive Dialogue

A realistic review by Wong et al [72] identified these 2 main theories as having a significant focus on perceived advantage, ease of use, interactivity, and feedback.

Finally, Rosen et al describe a statistical tool to apply to the study of teleoperation, human manipulation actions, and manufacturing applications (Hidden Markov Model), which they suggest might also be useful for other evaluation methods [73].

The above mentioned evaluation models all evaluate certain domains, a summary of which is presented in Tables 3 and 4 as an overview. In the final column, we have added the domains evaluated by de Leeuw et al in previous studies [74].
Table 3. Domains and methods for evaluating postgraduate medical electronic learning design (part 1).

<table>
<thead>
<tr>
<th>Factor</th>
<th>A*</th>
<th>B*</th>
<th>C*</th>
<th>D*</th>
<th>E*</th>
<th>F*</th>
<th>G*</th>
<th>H*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning aims/objectives</td>
<td>-a</td>
<td>xb</td>
<td>x</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Measurement of performance</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Aim for change/transfer to the job</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Usability and control</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Integration or recall of prior learning</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Confidence</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Suitability/usefulness/relevance/helpfulness</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Attention</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sensing or intuitive learning</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Visual or verbal learning</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Active or reflective learning</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sequential or global learning</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Content accountability</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Multimedia use</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
</tr>
<tr>
<td>Problem-based setting</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Impetus for use/motivation</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Costs</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Feedback and interactivity</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Challenge</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Commitment and maintenance</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Implementation</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Rehearsal</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Time management</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Tasks</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Efficiency</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>User expectation</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
</tr>
<tr>
<td>Factor</td>
<td>I*</td>
<td>J*</td>
<td>K*</td>
<td>L*</td>
<td>M*</td>
<td>N*</td>
<td>O*</td>
<td>P*</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Learning aims/objectives</td>
<td>xa</td>
<td>-b</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Measurement of performance</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Aim for change/transfer to the job</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Usability and control</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Integration or recall of prior learning</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
</tr>
<tr>
<td>Confidence</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Suitability/usefulness/relevance/helpfulness</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Attention</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
</tr>
<tr>
<td>Sensing or intuitive learning</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Visual or verbal learning</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Active or reflective learning</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sequential or global learning</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Content accountability</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>—</td>
</tr>
<tr>
<td>Multimedia use</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Problem-based setting</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
</tr>
<tr>
<td>Impetus for use/motivation</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Costs</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
<td>—</td>
</tr>
<tr>
<td>Feedback and interactivity</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Challenge</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Commitment and maintenance</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Implementation</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Rehearsal</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Time management</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Tasks</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Efficiency</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>User expectation</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Legend for table 3 and 4. \(^a\) Factor present in the model, \(^b\) Factor not present in the model. Explanation of models:
A = Quality Improvement Knowledge Application Tool-Revised (model of improvement) B = System Usability Scale C = Instructional Materials Motivation Survey D = Attention, Relevance, Confidence, and Satisfaction motivation model E = Index of learning styles F = 10 golden rules G = Gagne’s events instructions H = Heidelberg inventory for the evaluation of teaching I = Riding’s Cognitive Style Analysis J = Kern’s six steps K = Motivated strategies for learning questionnaire L = Software Usability Measurement Inventory M = Dunet model N = Website Motivational Analysis Checklist O = Davis’s model P = de Leeuw quality indicators
Discussion

There are many ways to evaluate PGMeL, and evaluation is clearly focused on the outcomes of the intervention. We found 14 e-learning-specific and 3 general primary outcomes, 27 secondary outcomes, and 13 evaluation tools. More than half of the studies (60%) had knowledge gain as their primary aim, which is almost the same finding as that in the 2016 review by Taveira-Gomes et al [2], who looked at all kinds of education. We are looking at PGMeL only and found that 38% were simulation and virtual reality studies. This kind of e-learning was not mentioned specifically in the study by Taveira-Gomes et al but might be comparable with the skills outcome (14.6%). The difference could be the result of postgraduates’ need to undertake more task- and real-life-related e-learning, as described in our focus groups [74]. The experts from that study emphasized real-world translation as an important factor for PGMeL. Looking at the outcomes of the studies, Seagull identified similar domains in surgical simulation studies [75]. Self-efficacy, satisfaction, relevance/adherence in practice, and attitude are frequently used as outcomes of e-learning in both our study and that by Seagull et al. Table 1 shows a list of methods used to evaluate an outcome, which focus on the defined outcome rather than the method used to achieve it. Many other instruments are available (such as the critical thinking index [76]), but they are either not yet used in a PGMeL e-learning evaluation setting or were not revealed by our search.

Our research question asked which evaluation methods are used. As mentioned above, only 4% used a method, and of those methods, we can differentiate between theories and instruments. Of the theories, Kirkpatrick’s hierarchy is the most used to evaluate or create e-learning. A 2017 review by Patel et al evaluated the effectiveness of simulation in interventional radiology training [35]. It also found different studies using the levels of Kirkpatrick’s hierarchy to establish or evaluate the success of the e-learning. Of the educational instructional theories, 2 are leading in e-learning in general and were also found in our studies: Gagne’s principles of instructional design and Mayers and Clark’s e-learning and the science of instruction, also referred to as Mayer’s multimedia learning. Mayers and Clark base their instructions on the cognitive load theory, which provides design guidelines based on a model of human cognitive architecture. Cook et al validated a cognitive load index in 2017 [77]. The last theory from our search is from Kern’s curriculum development for medical education: a 6-step approach. All these theories are either based on education in general (eg, the work of Gagne and Mayer) or medical education (eg, the work of Kirkpatrick and Kern), but none of the theories are aimed at PGMeL. They are used to evaluate PGMeL but not specifically aimed at this audience. The Heidelberg inventory for the evaluation of teaching is even aimed at undergraduate students and only used because of the lack of a better alternative [59].
Apart from these theories, some instruments focused on 1 aspect of the design. Although these instruments have a specific focus, Table 1 shows that they cover a wider range of domains. Instruments that aim to evaluate the course as a whole are QIKATR, 10 golden rules, and Dunet’s formative process and outcome evaluation tool. The QIKATR is an answer to the Accreditation Council for Graduate Medical Education, which required practice-based learning and improvement. It is a description of 3 scenarios depicting quality problems. Although the domains are not very specific (describe the aim, measure the effect, and require change), they are aimed at postgraduates and provide a good basis for any education. Conversely, they are not aimed at e-learning education [68]. In 2002, Jha et al created an e-learning model for gynecology called the Distance Interactive Learning in Obstetrics and Gynaecology. They then evaluated the e-learning, and the lessons learned were described as 10 golden rules [65]. These golden rules are aimed at postgraduates and are specific to e-learning. The most significant downside of these rules is that they are based on 1 e-learning experience only; therefore, they may be incomplete or biased by the single case that created the fundament. Finally, Dunet’s formative process and outcome evaluation tool is the result of an evaluation plan based on the experience of creating a hemochromatosis training course for continuing education credits and continuing nurse education. The course has been intensively evaluated by several experts, and the key findings can be summarized in 2 domains: instructional design and content, and usability and navigation. Although aimed at postgraduate education and specific to e-learning, it is based on 1 course only and might, therefore, lack important domains and items that were not available in that course.

As demonstrated in Tables 3 and 4, Gagne’s science of instruction covers most of the domains. Our search did not identify any e-learning evaluation methods that are not expert opinion–based or the result of a single evaluation and aimed at PGMeL. A previous study by our group identified all these domains in literature [78], then evaluated their relevance with the focus groups [74] and an international Delphi [79]. The domains are added in the last column of Tables 3 and 4, which illustrates that all domains, except learning styles, are identified as important in these studies. The learning styles were identified in the review, but the effect of learning style–specific education is disappointing [53]. The conclusion was that it was better not to evaluate the learning style but to offer a diversity in each e-learning [74].
Strengths and Limitations

We believe that the biggest limitation is our search. Had we included papers not aimed at postgraduate education, we would have found many more papers and evaluation models. We could also have included papers that did not actually evaluate a course but only described a theoretical model. However, our research question asked not what is available but what is actually used. We also believed in differentiating between graduate and postgraduate education, hence the choice in our search. However, we also believe that making this distinction is a strength. This paper provides an insight into the diversity of evaluating e-learning and how little is known of and targeted at the right audience. Almost all quality models signify the importance of knowing your target audience but our evaluation tools do not.

Conclusions

It may be asked what comes next. We have reached the point at which we should stop evaluating only the outcomes of e-learning as an educational intervention and start evaluating the e-learning design that goes with it. However, to do so, we need a validated instrument to help us assess the nuances of all the different electronic education instruments. We believe that our previous studies have provided us with validated content for such a tool [74,79] and that this paper emphasizes the need for such a system.

PGMeL is evaluated in very diverse ways, almost exclusively based on its outcomes or learning aims. Although there is a need to evaluate the e-learning itself as well, we lack the instruments to do so. This paper provides an overview of available instruments; however, they are not aimed at postgraduate medical education, not expert opinion–based, or the result of lessons learned from a single case study. With the increasing ease of creating and distributing e-learning, the need for a content-validated evaluation tool is of ever greater importance.
Authors' Contributions

All authors participated in the study design and manuscript revisions. RDL performed the search with help from Hans Ket (see Acknowledgments). RDL, ADS, and SVH reviewed the search results. RDL, KW, and FS drafted the first version of the manuscript, added background data, and participated in the interpretation of the results. MW, ADS, and SVH revised the manuscript accordingly. All authors read and approved the final manuscript.

Abbreviations

**ARCS:** Attention, Relevance, Confidence, and Satisfaction

**e-learning:** electronic learning

**IMMS:** Instructional Materials Motivation Survey

**PGMeL:** postgraduate medical e-learning

**QIKATR:** Quality Improvement Knowledge Application Tool-Revised

**SUMI:** Software Usability Measurement Inventory

**SUS:** System Usability Scale
References


18. Research Survey. URL: http://researchsurvey.nl/reviews/PGMEevaluation [accessed 2019-02-21] [WebCite Cache ID 76JWb0CB8]


67. Stufflebeam DL. The Relevance of the CIPP Model for Educational Accountability. Presented at: Annual Meeting of the American Association of School Administrators; February 24, 1971; Atlantic City, NJ.


Multimedia Appendix 1 - search string

Databases used

1. PubMed
2. CINAHL
3. ERIC
4. Web of Science

postgrad med

e-learning tools

RCT’s, SR’s


OR

Multimedia Appendix 2 - Search results (sorted by year)

Due to its big size, this appendix can be found online at: www.MotivateLearnApply.com
Quality specifications in postgraduate medical e-learning: an integrative literature review leading to a postgraduate medical e-learning model

De Leeuw RA
Westerman M
Nelson E
Ket JCF
Scheele F

BMC Medical Education
June 2016
Abstract

Background: E-learning is driving major shifts in medical education. Prioritising learning theories and quality models improves the success of e-learning programs. Although many e-learning quality standards are available, few are focused on postgraduate medical education.

Methods: We conducted an integrative review of the current postgraduate medical e-learning literature to identify quality specifications. The literature was thematically organised into a working model.

Results: Unique quality specifications (n = 72) were consolidated and reorganised into a six-domain model that we called the Postgraduate Medical E-learning Model (Postgraduate ME Model). This model was partially based on the ISO-19796 standard, and drew on cognitive load multimedia principles. The domains of the model are preparation, software design and system specifications, communication, content, assessment, and maintenance.

Conclusion: This review clarified the current state of postgraduate medical e-learning standards and specifications. It also synthesised these specifications into a single working model. To validate our findings, the next-steps include testing the Postgraduate ME Model in controlled e-learning settings.
Background

E-learning plays a prominent role in conventional education, adult education, and medical training because of its flexibility, broad resource-sharing capacity, and cost-effective scalability [1]. E-learning has become central to medical education, and web technologies offer valuable new opportunities for both under- and postgraduate medical education [2]. E-learning also offers participants an advantage in that they can choose a comfortable and accessible place and time to study, which is important in postgraduate and continuous medical education [3].

There are many studies comparing e-learning methods. However, one of the problems is that the results of studies directly comparing technology-assisted education with traditional teaching often conflict and often do not demonstrate or propose best practices [4, 5]. Critical evaluation of the quality and efficiency of e-learning is warranted [6]. Therefore, there is a need to develop a consensus-based quality assurance standard for postgraduate medical e-learning [7, 8].

It is known that e-learning should be targeted to the needs of the specific audience (in this case postgraduates) [9, 10]. The success of e-learning programs has also been linked to the use of a theoretical framework or a learning theory [11]. Standards for e-learning exist and have been evaluated [12].

Although learning theories are broad and diverse, there is progressive agreement about the psychological fundamentals [13]. Previous research suggests that the constructivist approach (founded by Jean Piaget, among others) is compatible and appropriately designed for e-learning. According to this theory, humans are active learners and construct new knowledge based on prior experiences and interactions [14]. An example is problem-based learning, which has been shown to be effective in medical education [15, 16]. Another theory based on the constructivist approach is the cognitive load theory (CLT) developed by Sweller [17]. The constructivist approach also forms the foundation of the cognitive theory of multimedia learning, a well-evaluated learning theory developed by Mayer [18]. This theory has been specifically adjusted for e-learning, and is believed to provide a good basis for an e-learning standard [18]. These theories are not elaborated here. However, it is important to remember that learning theory, not technology, should guide the design and content of e-learning.

Standards for e-learning exist [12], but are often isolated to a specific sector of the industry for which they were developed [19]. There are many industrial standards in education, often published by organisations such as the American Council on Education and the European Association of Distance Teaching Universities [20]. Several organisations have made progress in developing international industrial
standards for universities, including the Open e-Quality Learning Standards and the Leonardo DaVinci program [21, 22].

The most common formal standard in use is the ISO/IEC 19796-1 [23]. This standard was issued by the International Standardisation Organization (ISO) in 2005, and contains the Reference Framework for the Description of Quality Approaches (RFDQ), a framework supported by the European Quality Observatory [24]. Stracke implemented this standard and concluded that it was not only the first quality standard for learning, education, and training, but was a valuable instrument for sustained quality development in e-learning [23]. Little tested two higher education quality standards or rubrics, in professional continuous nursing education in 2009 [25]. The College of Public Health Online Course Standards and the Quality Matters Peer Course Review Rubric were evaluated, and although they look hopeful, little is known about their actual effectiveness. The Quality Matters rubric is not publicly available, making evaluation even less accessible.

Existing literature fails to clarify the methods and strategies used to evaluate the quality of postgraduate medical e-learning. As Clark noted in his well-respected book E-Learning and the Science of Instruction, the target audience should be the first thing considered when designing an e-learning program [9, 10]. Although undergraduates and postgraduates may learn in the same way, adult learning theories suggest their learning strategies and capacities might be dissimilar. Therefore, we focused our review on our target audience: postgraduates and physicians bound to continuous medical education. Ellaway et al. began this process in 2008 by describing a two-part guideline for e-learning in medical education. Although the guideline was not specifically aimed at postgraduate medical education, it served as a foundation for further research [16, 26]. The previously mentioned standards and the Quality Matters Rubric look promising, but are all aimed at different target audiences [25]. The ISO standard (ISO/IEC 19796) covers a lot of domains, but lacks detail regarding their application [27].

It is difficult to establish quality control practices for e-learning materials [28], yet this is an important problem in postgraduate medical education where the quality of training directly influences patient care [29]. Therefore, as technological innovations reshape medical education institutions, the question of quality assurance is at the forefront of university leadership concerns worldwide [30]. In 2010, Cook et al. reiterated the primary importance of defining quality standards in medical education. In their Second Flexner Report, they identified standardisation as a key goal for improving medical education [31]. We believe that it is important to establish a testable quality assurance model to improve the uptake of e-learning and motivate continuous medical learning [5]. This review will add to existing literature by providing an integrative literature review and a working model of quality assurance standards in postgraduate medical e-learning.
Methods

We performed an integrative review to identify and critically evaluate qualitative and quantitative literature associated with current postgraduate medical e-learning quality assurance. We used the updated integrative review methodology developed by Whittemore et al. [32]. This method consists of three steps: 1) a systematic search combined with at least one nonsystematic search method; 2) data evaluation comparing study models and quality scores; and 3) data analysis. During data analysis, we compared individual quality items, clustered them, searched for contrasts and comparisons, discerned patterns, and built an overview of the domains.

Definitions

**E-learning:** An approach to teaching and learning, representing all or part of the educational model applied, 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 [33]. In practice, the definition includes any digital content made to teach and distributed physically or online. Quality. To date, there is no consensus definition of quality. However, a high-quality product is generally defined as one that meets consumer-defined specifications, delights the consumer, consistently meets the standard that the producer has set for itself, and leads to customer satisfaction. Producers should be able to assure this quality [30, 33].

**Postgraduate (and continuous) medical education:** Any form of learning aimed at medical professionals who have graduated from formal training and residency, or used by medical professionals a part of continuous learning to maintain their competency and develop new knowledge [34].

**Standards and specifications:** A standard is a set of specifications that guide an e-learning author in developing an e-learning program. A specification is a specific item that is addressed within the standard [17].
Step 1: Systematic literature search

The primary search terms were distance learning (and all synonyms) [35] and quality (and all synonyms). We conducted the search on May 2, 2015. We searched ISI/Web of Knowledge, PubMed, EBSCO/Cinahl, EBSCO/PsycInfo, and EBSCO/ERIC (Table 1). Google Scholar was also searched, despite debate on its added value [36].

Inclusion criteria

Included articles were peer-reviewed and published in the English language between 1970 and May 2015. The articles had to describe and evaluate specific e-learning characteristics in postgraduate or continuous medical education. The search was kept broad and selection was made on postgraduate medical education manually after reading the titles, abstracts, and the full text (if necessary).

Exclusion criteria

We excluded dissertations, conference abstracts, and articles comparing e-learning with other forms of learning without describing the quality specifications used.

After selecting the titles and abstracts, we sampled 40 random abstracts for independent evaluation by a second (MW) and third author (FS). If there was no consensus, three authors participated (RL, MW, FS) in a discussion until consensus was reached on whether or not to include the article.

Table 1 Databases searched and corresponding search strings

<table>
<thead>
<tr>
<th>Database</th>
<th>Search string used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
<td>Query</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>ISI/Web of Knowledge</td>
<td>(assessment OR criteria OR metrics OR characteristics OR measurement OR evaluation OR standards OR quality OR principles) [TI] AND (e-learning OR (electronic learning) OR (distance education) OR (technology-enhanced learning) OR tele-learning OR (web-based learning) OR (web-based education) OR (internet-based learning) OR (computer based learning) OR (computer-assisted instruction) OR (distance learning) OR (online learning)) [AB]</td>
</tr>
<tr>
<td>EBSCO/Cinahl</td>
<td>(assessment OR criteria OR metrics OR characteristics OR measurement OR evaluation OR standards OR quality OR principles) [AB] AND (e-learning OR (electronic learning) OR (distance education) OR (technology-enhanced learning) OR tele-learning OR (web-based learning) OR (web-based education) OR (internet-based learning) OR (computer based learning) OR (computer-assisted instruction) OR (distance learning) OR (online learning)) [AB]</td>
</tr>
<tr>
<td>EBSCO/Psychinfo</td>
<td>(assessment OR criteria OR metrics OR characteristics OR measurement OR evaluation OR standards OR quality OR principles) [AB] AND (e-learning OR (electronic learning) OR (distance education) OR (technology-enhanced learning) OR tele-learning OR (web-based learning) OR (web-based education) OR (internet-based learning) OR (computer based learning) OR (computer-assisted instruction) OR (distance learning) OR (online learning)) [AB]</td>
</tr>
<tr>
<td>EBSCO/Eric</td>
<td>(assessment OR criteria OR metrics OR characteristics OR measurement OR evaluation OR standards OR quality OR principles) [AB] AND (e-learning OR (electronic learning) OR (distance education) OR (technology-enhanced learning) OR tele-learning OR (web-based learning) OR (web-based education) OR (internet-based learning) OR (computer based learning) OR (computer-assisted instruction) OR (distance learning) OR (online learning)) [AB]</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>All in title: (assessment OR criteria OR metrics OR characteristics OR measurement OR evaluation OR standards OR quality OR principles) AND ((e-learning OR (electronic AND learning) OR (distance AND education) OR (technology-enhanced AND learning) OR tele-learning OR (web-based AND learning) OR (web-based AND education) OR (internet-based AND learning) OR (computer AND based AND learning) OR (computer-assisted AND learning) OR (distance AND learning) OR (online AND learning))</td>
</tr>
</tbody>
</table>
Step 2: Data evaluation

We aggregated the selected studies that met the inclusion criteria, flagging those with an unclear method. The aggregate was graded according to criteria adapted from the levels of evidence model of the Oxford Centre for Evidence-Based Medicine (Table 2). Our adaptation consisted of the addition of grade 6, indicating an unclear method. We flagged a study as unclear if it did not fall into any of the five existing categories.

We used the two-tier data evaluation strategy developed by Wittemore et al. to grade the quality of each item. Levels 1, 2, and 3 were identified as high quality (tier 1), and 4 and 5 as low quality (tier 2) [32]. We removed grade 6 studies from the analysis.

Step 3: Data analysis

After aggregating the studies and grading the papers, we analysed each item for common themes and contradictory findings. If items conflicted, we rejected the lower quality study in favor of the higher quality study. We categorised the items according to the ISO/IEC 19796-1 domains, and then generated new themes for items that did not fit the established domains.
Results

Primary selection and analysis

The literature search identified 10,732 articles. Searching Google Scholar gave us access to unanticipated databases such as the Emerald, IEXEE, and Editlib, as well as journals that were not registered with the other databases (International Journal of Information and Educational Technology, Journal of E-Learning and Knowledge Society, Applied Soft Computing, International Educational E-Journal, and Journal of Theoretical and Applied Information Technology). We then manually reviewed the identified studies.

After applying inclusion and exclusion criteria, 884 eligible titles remained. Eligible abstracts were reviewed and final selection criteria were refined as described in the Methods section.

Table 2 Adapted Oxford Centre for Evidence-Based Medicine levels of evidence

<table>
<thead>
<tr>
<th>Level</th>
<th>Design</th>
<th>Tier</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Systematic review of randomized controlled trials or individual randomized controlled trial</td>
<td>high</td>
</tr>
<tr>
<td>2</td>
<td>Systematic review of cohort studies or individual cohort study</td>
<td>high</td>
</tr>
<tr>
<td>3</td>
<td>Systematic review of case-control studies or individual case-control study</td>
<td>high</td>
</tr>
<tr>
<td>4</td>
<td>Case-series</td>
<td>low</td>
</tr>
<tr>
<td>5</td>
<td>Expert opinion</td>
<td>low</td>
</tr>
<tr>
<td>6</td>
<td>Unclear method</td>
<td>low</td>
</tr>
</tbody>
</table>

In total, 36 articles met the final selection criteria (see Fig. 1), representing quality items in postgraduate medical education literature. Of these, 15 were original case reports and expert opinions; 13 were reviews of previous case reports and expert opinions; five were randomised controlled trials; and two were meta-analyses (Table 3). Collectively, these articles represented 16 high tier articles (Oxford Centre for Evidence-Based Medicine levels 1, 2, and 3) and 20 low tier articles (levels 4 and 5). The publication dates ranged from 1995 to 2015.
**Figure 1** Flowchart of the literature search results

- PubMed: n=5,442
- Ebsco/Cinahl: n=831
- Ebsco/Eric: n=796
- Ebsco/Payinfor: n=3,101
- ISI/Web of Science: n=923

Search updated till May 2015

Total Titles for review: n = 11,093

- Google Scholar: n=2,006 (1,482 rejected, 362 duplicated, 162 new titles)
- Duplicates: n=1,036
- Titles rejected: n=9,335

Relevant titles: n = 884

- Learning theory only: n = 76
- General e-learning: n = 27
- Irrelevant topic: n = 417

Relevant full articles: n = 48

- Book reviews: n = 57
- Undergraduate e-learning: n = 64
- Non-medical adult e-learning: n = 175

Rejected after reading full article: 19
- No actual characteristics: 18
- Targeted not at postgraduates: 5
- Total rejected: n = 23

Cross-reference & hand search: n = 11

Final amount of articles relevant for "postgraduate medical e-learning": n = 36

* e-learning not made for a specific audience (specifically not for postgraduate medical learners)
Sub-selection and analysis

After consolidating duplicates for quality specifications (440 in total), there were 72 unique quality specifications remaining. We categorised these specifications according to the seven ISO 19796 domains, combined them, and renamed them if necessary. This produced a model with six final domains, which we called the Postgraduate Medical E-learning Model (Postgraduate ME Model; Fig. 2). In the following paragraphs, we have defined the domains and provided examples from the literature.

**Preparation:** This step should be performed before designing and building the e-learning platform. Twelve articles described the importance of knowing the end users [23, 36-46]. Preparation is a two-step process. First, one must define the platforms used by different types of end users and their expectations of the e-learning platform. If the end users’ level of knowledge is known, the e-learning strategy can be designed appropriately. Special functionalities for individuals with learning disabilities can also be incorporated if applicable. Cook et al. described this step as a part of the needs analysis [10]. If possible, learning should be adapted to the audience’s motivational level, skills, and expectations. Curry et al. [47] and Olson et al. [20] emphasised the importance of selecting e-learning authors based on subject matter expertise. They suggested that academically qualified people must actively participate in development and training to improve e-learning. Four articles advised readers to allocate time to accurately budget and plan for the expected costs [23, 41, 44, 48]. Budgets and policy statements should reflect commitment to the program. Seven authors agreed on the positive effect of planning, describing objectives and expectations, and placing these in a timeline to maximize efficiency [7, 23, 36, 37, 39, 40, 49]. Cook et al. [50] identified potential barriers to implementation, and Olson et al. [20] advised readers to develop a marketing plan for reaching end users. Curry [47], Wong [51], and Sowan [39] advocated efficient e-learning that saves the user time, compared with other forms of learning.

**Design:** This depends on a series of difficult choices that include the types of hardware and software that best fit end-users’ needs. Although, hardware and software are very different, most articles described these in one section. In several articles (n = 5), authors stressed the importance of reliability and emphasized that the combined hardware and software system should be tested and iterated to meet end-user design requirements [7, 23, 39, 42, 49]. Testing should include different browsers and different monitors with varying resolutions, as well as different hand-held devices if a mobile-based platform is developed.

Most (n = 14) articles discussing software design focused on ready access to the e-learning platform. Access should be fast, easy to find, and always available; the platform should have a short loading time, and should provide reliable online access to all external links [7, 36-40, 45, 46, 48, 49, 52, 53, 70]. A secure connection is also
important to support the privacy and legal requirements, copyright rules, and
intellectual property issues [23, 39, 44, 49]. Lewis et al. [52] urged readers to be
mindful of the basics such as grammar and spelling. Cook et al. [6] advised piloting e-
learning websites before publishing them, and Bangert [46] advised the use of a variety
of learning environments. Yavner et al. [53] wanted to give users maximum control
over navigation, an approach that Mayer later challenged [54].

Navigation and layout are important design elements. Navigation should be user
friendly and intuitive with a “less is more” design strategy. The design should be
visually pleasing, adhere to the principles of excellent website design, and use reusable
learning objects for a standard look and feel [37, 45]. Four articles provided a minimal
set of pages/functionalities: glossary of terms, frequently asked questions, concept
map, references cited, abbreviation key, and labelled diagrams [36, 45, 55, 56].

Table 3 Articles used for the postgraduate medical education model identified in the literature search

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Type of study</th>
<th>Domains discussed*</th>
<th>Evidence tier</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khoiny</td>
<td>1995</td>
<td>Expert opinion</td>
<td>1</td>
<td>Low</td>
<td>[57]</td>
</tr>
<tr>
<td>Kim</td>
<td>1999</td>
<td>Systematic review</td>
<td>1, 2, 4, 5, 6</td>
<td>High</td>
<td>[72]</td>
</tr>
<tr>
<td>Sekikawa</td>
<td>2001</td>
<td>Expert opinion</td>
<td>2</td>
<td>Low</td>
<td>[73]</td>
</tr>
<tr>
<td>Doyle</td>
<td>2002</td>
<td>Expert opinion</td>
<td>2, 3</td>
<td>Low</td>
<td>[74]</td>
</tr>
<tr>
<td>Jha</td>
<td>2002</td>
<td>Case study</td>
<td>1, 2, 4</td>
<td>Low</td>
<td>[75]</td>
</tr>
<tr>
<td>Minasian</td>
<td>2002</td>
<td>Expert opinion</td>
<td>1, 2, 5, 6</td>
<td>Low</td>
<td>[55]</td>
</tr>
<tr>
<td>Cook</td>
<td>2004</td>
<td>Review</td>
<td>1, 2, 3, 4, 5, 6</td>
<td>high</td>
<td>[6]</td>
</tr>
<tr>
<td>Olson</td>
<td>2004</td>
<td>Review</td>
<td>1, 2, 3, 4, 5, 6</td>
<td>high</td>
<td>[20]</td>
</tr>
<tr>
<td>Knight</td>
<td>2004</td>
<td>Expert opinion</td>
<td>1, 2, 3, 4, 5, 6</td>
<td>Low</td>
<td>[48]</td>
</tr>
<tr>
<td>Casebeer</td>
<td>2004</td>
<td>Trial</td>
<td>2, 4, 5</td>
<td>high</td>
<td>[44]</td>
</tr>
<tr>
<td>Curry</td>
<td>2005</td>
<td>Expert opinion</td>
<td>1, 2, 4, 5, 6</td>
<td>Low</td>
<td>[47]</td>
</tr>
<tr>
<td>Bangert</td>
<td>2005</td>
<td>Case study</td>
<td>2, 3, 4, 5</td>
<td>Low</td>
<td>[46]</td>
</tr>
<tr>
<td>Garde</td>
<td>2007</td>
<td>Case study</td>
<td>3, 4, 5</td>
<td>low</td>
<td>[43]</td>
</tr>
<tr>
<td>Maor</td>
<td>2007</td>
<td>Review</td>
<td>3, 4</td>
<td>high</td>
<td>[45]</td>
</tr>
<tr>
<td>Posel</td>
<td>2009</td>
<td>Review</td>
<td>1, 2, 4, 5, 6</td>
<td>high</td>
<td>[66]</td>
</tr>
<tr>
<td>Casimiro</td>
<td>2009</td>
<td>Review</td>
<td>2, 4, 5</td>
<td>high</td>
<td>[71]</td>
</tr>
</tbody>
</table>

Olson et al. [20] and Lau [42] suggested choosing appropriate, intuitive, and user-friendly hardware devices that advise the learner of the skills and technology required for success. Desired learning outcomes should drive the choice of technology. Khoiny [57], Garde et al. [43], and Mhouti et al. [58] described the importance of the environment where the e-learning is used. The physical setting should be a stimulating and motivational place to learn.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Study Type</th>
<th>Domains (numbered)</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cook</td>
<td>2010</td>
<td>RCT</td>
<td>3, 4, 5</td>
<td>high</td>
</tr>
<tr>
<td>Wong</td>
<td>2010</td>
<td>Review</td>
<td>1, 2, 3, 4, 5</td>
<td>high</td>
</tr>
<tr>
<td>Mayer</td>
<td>2010</td>
<td>Review</td>
<td>4</td>
<td>high</td>
</tr>
<tr>
<td>Short</td>
<td>2010</td>
<td>Review</td>
<td>2, 3, 4, 5</td>
<td>high</td>
</tr>
<tr>
<td>Alexander</td>
<td>2010</td>
<td>Expert opinion</td>
<td>1, 2, 3, 5, 6</td>
<td>low</td>
</tr>
<tr>
<td>Friedlander</td>
<td>2011</td>
<td>Review</td>
<td>4</td>
<td>high</td>
</tr>
<tr>
<td>Chang</td>
<td>2011</td>
<td>Case study</td>
<td>1, 2, 3, 4</td>
<td>low</td>
</tr>
<tr>
<td>Issa</td>
<td>2011</td>
<td>Cohort study</td>
<td>4</td>
<td>low</td>
</tr>
<tr>
<td>Mounsey</td>
<td>2012</td>
<td>RCT</td>
<td>4</td>
<td>high</td>
</tr>
<tr>
<td>Raymond</td>
<td>2012</td>
<td>Expert opinion</td>
<td>1, 3, 4, 5</td>
<td>low</td>
</tr>
<tr>
<td>Sowan</td>
<td>2013</td>
<td>Expert opinion</td>
<td>1, 3, 4, 5</td>
<td>low</td>
</tr>
<tr>
<td>Mhouti</td>
<td>2013</td>
<td>Review</td>
<td>1, 3, 4, 5</td>
<td>high</td>
</tr>
<tr>
<td>Bluestone</td>
<td>2013</td>
<td>Meta-analysis</td>
<td>4, 5</td>
<td>high</td>
</tr>
<tr>
<td>Gordon</td>
<td>2013</td>
<td>RCT</td>
<td>4</td>
<td>high</td>
</tr>
<tr>
<td>Shaw</td>
<td>2014</td>
<td>Case study</td>
<td>2, 4, 5</td>
<td>low</td>
</tr>
<tr>
<td>Lewis</td>
<td>2014</td>
<td>Review</td>
<td>1, 2, 4, 5</td>
<td>high</td>
</tr>
<tr>
<td>Yavner</td>
<td>2014</td>
<td>Expert opinion</td>
<td>1, 2, 4, 5, 6</td>
<td>low</td>
</tr>
<tr>
<td>Lau</td>
<td>2014</td>
<td>Review</td>
<td>2, 4, 5</td>
<td>high</td>
</tr>
<tr>
<td>Davids</td>
<td>2014</td>
<td>RCT</td>
<td>2, 4, 5</td>
<td>high</td>
</tr>
<tr>
<td>Cook</td>
<td>2015</td>
<td>Review</td>
<td>1, 2, 3, 4, 5, 6</td>
<td>high</td>
</tr>
</tbody>
</table>

Communication: This includes all forms of internal user-oriented communication and external expert-oriented communication. Articles included in our review provided descriptions of several forms of communication, including: (i) communication about the program that informs the user about learning objectives, costs, and support options; (ii) communication that allows users to contact the faculty/course authors; and (iii) communication between users as they collaborate on coursework. Several articles (n = 8) emphasised the second form of communication, recommending opportunities for live interaction with experts/authors, possibly in a group context [38, 41, 43, 45, 55, 59]. The credibility of the authors should be well established, and should include the authors’ credentials and disclosure of sponsors and conflicts of interest.

Content: This is the aggregate material used to build the lessons (e.g., words, images, videos). Content was a central theme in the articles we reviewed. Problem-based learning was favoured by 13 articles as the best way to incorporate motivation and better understanding [7, 15, 40, 43, 46–48, 52, 56–60]. Introducing interactivity to encourage higher-order thinking was also important (n = 16 articles) [23, 36, 39–41, 45–47, 51, 52, 56–59, 61]. The user should be provided with time and impetus to learn with as little stress as possible and made responsible for the learning process to create a feeling of belonging, and the platform should provide learning exercises [7, 41, 46, 48, 61]. Learning modules should end with summaries, consist of short paragraphs, state a timeline, and use milestones [23, 36, 39, 41, 43, 44, 46, 47, 49, 50, 59]. Lewis et al. [52] emphasised the use of educational standards, and most other articles suggested using cognitive load principles. Cognitive load principles are described in the Discussion. Bluestone et al. [61] advocated for reminders in e-learning systems, and Friedlander et al. [62] and Cook [59] suggested incorporating rewards and reinforcements to maintain motivation.

Assessment: This refers to all of the possible ways to test end users and formalize their knowledge gain. Almost all of the articles prioritized assessment and feedback on performance. Seventeen articles described assessment as most effective when used in a direct, continuous, and personalized way [7, 15, 23, 36–38, 40, 41, 45–47, 50–52, 56, 58, 63]. Self-assessment was also an important part of learning [7, 23, 36, 39, 40, 43–45, 47, 54, 57, 59]. Additionally, Lau [42], Short et al. [64], and Wong et al. [51] stressed the importance of continuing to provide the e-learning software and associated tools after the end of the course. Cook et al. focused on the importance of assessing user experience and satisfaction [50]. Assessment is also a way of evaluating learning outcomes [10].
**Maintenance:** This includes the steps taken to avoid the loss of knowledge after a user finishes the e-learning process. Maintenance also includes reliable long-term access to the platform to allow an end user to return to the platform. Maintenance-related articles \((n = 10)\) described the importance of evaluating a platform's user experience, effectiveness, usability, and cost \([7, 23, 36–38, 40, 44–46, 59]\). Technical maintenance included protecting and verifying hyperlinks. Several articles \((n = 6)\) emphasized that modifications to e-learning programs should be enabled and updated, and those proven to be unsuccessful removed \([7, 23, 37, 38, 44, 45]\). E-learning developers should also estimate the reusability and sustainability of new platforms, as these factors are important in the platform’s maintainability \([10]\).
Figure 2 Postgraduate medical e-learning specifications model

Legend figure 1. Postgraduate Medical e-learning Specification Model the ME Model. “The details from “4A. Cognitive Load principles” are described in the discussion session.”
**Discussion**

We identified 72 features as important in postgraduate medical e-learning, and grouped these into six domains. The domains also provided a model framework for educators involved in drafting e-learning strategies or evaluating e-learning initiatives. The content domain is the most widely described and discussed domain of the model, and we discuss this in detail below followed by the limitations of our review.

**The content domain**

Perhaps the most important part of e-learning is the content, which was emphasised in all articles reviewed. Content is the heart of e-learning, and the design merely delivers content. A common pitfall is developing e-learning simply for the sake of using a new technology. Instead of making e-learning technology-centred, developers should subscribe to a learning theory to ensure the design is guided by pedagogical principles [14]. Mayer described the cognitive theory of multimedia learning, based on the CLT [17]. CLT aims to develop instructional design guidelines based on a model of human cognitive architecture [65]. CLT states that working memory is limited in its capacity to selectively attend to and process incoming sensory data. This theory is concerned with the way in which a learner focuses and uses cognitive resources during learning and problem solving. It suggests that for an instruction to be effective, it must be designed in a way that does not overload the mind's capacity for processing information [40]. Based on this theory, Mayer defined a set of principles, which describe the effect of different design techniques on learning. These principles form guidelines for using multimedia in a learning environment. In the content domain, we found 20 of 36 articles tested or recommended one or more of Mayer’s principles. Because the current literature pays so much attention to these principles, we considered them to be the foundation of the content domain [18].

Not all authors agreed with all of these principles. For example, Yavner et al. [53] proposed giving the learner maximum control, in contradiction to Mayer’s assertion that giving control to learners yields no benefit because learners may have too many options [17]. Curry et al. [47], Mhouti et al. [58], and Posel et al. [66] also highlighted the importance of learner-centred e-learning and supported individualised, rather than standards-based e-learning. Therefore, consensus on the principles of the content of e-learning is lacking.
Limitations

The major limitations of this review pertain to the methods. We performed an integrative review instead of a systematic review or meta-analysis, because the majority of published studies related to postgraduate medical education did not meet the parameters required for a systematic review. The major limitation of integrative reviews is the potential for bias from the inclusion of non-peer-reviewed information or lower-quality studies. Although Cook et al. [50] conducted a meta-analysis, it was limited to a few domains in e-learning. However, the literature from which the authors drew their conclusions was largely consensus-based.

The second limitation was the fast-changing technology that threatens to render our results obsolete. E-learning is rapidly changing the landscape of medical education and is developing faster than research can evaluate it [67]. This pace of change is a limitation of our review because research is always one step behind technology [67]. Examples of these fast changes are two disruptive innovations in medical education: Massive Open Online Courses and social media on mobile devices [68]. Both impact on e-learning and might dramatically change the education landscape [69]. In this landscape, it is almost impossible to evaluate an innovation properly before it is already outdated. In addition, social media is expected to become important to the collaboration domain of e-learning [70].

Despite these limitations, we believe that our six-domain Postgraduate ME Model will generate discussion and increase the quality of new e-learning courses. Our e-learning model could be interpreted as a general framework rather than postgraduate-specific, although we have not provided evidence to extend it to other settings, due to the limitation in our search strategy. We have limited the search, to target our audience as good the literature allowed us. Clark et al. clearly states that the target audience should be as specific as possible [9, 10]. The articles represented in this review were selected with focus on postgraduate learners; further analysis will be required to determine the applicability of our Postgraduate ME Model to other audiences. Even if e-learning developers reject our model, we feel that it is better to have reasons for not using a model than to have no model at all [71].
Conclusion

In summary, our Postgraduate ME Model aimed to provide a practical framework that can be used to build postgraduate medical e-learning programs that are learner centred, interactive, well planned and designed, based on cognitive load theory, and easy to maintain. E-learning should be about the learner, not the technology. Our proposed model may guide e-learning designers who are developing quality e-learning targeted to postgraduates in medicine. Our six-domain model is unique in that it combines the technical requirements from industry standards with the critical aspects of content and interaction from learning theories.

The next step is for research to validate these domains with international experts to determine if they are beneficial to postgraduate real-world e-learning. It would be interesting to know if postgraduate e-learning experts agree with our description of qualitative e-learning, based on their experience. Another important question is whether a model such as this could actually be used in practice when developing e-learning platforms. Ultimately, we would like to know whether e-learning based on our Postgraduate ME Model will reproducibly improve learner motivation.

Availability of data and materials

The dataset supporting the conclusions of this article is available in the Postgraduate ME Model repository, www.MotivateLearnApply.com

Authors’ contributions

RL participated in the study design, search, data analysis, and the learning model design. MW was the second reviewer of the search results, drafted the manuscript, and participated in the data evaluation. EN revised the manuscript, added the background data, and participated in the interpretation of results. JK designed the search strings and coordinated the search itself. FS participated in the study design, coordinated the study, was the third reviewer of the data, and revised the manuscript. All authors read and approved the final manuscript.
References


Quality indicators for learner-centered postgraduate medical e-learning

de Leeuw RA
Westerman M
Scheele F

International Journal of Medical Education
August 2017
Abstract

Objectives: The objectives of this study were to identify the needs and expectations of learners and educational experts in postgraduate medical e-learning, and to contribute to the current literature.

Methods: We performed four focus-group discussions with e-learning end-users (learners) and didactic experts. The participants were postgraduate learners with varying levels of experience, educational experts from a Dutch e-learning task group, and commercial experts from a Dutch e-learning company. Verbatim transcribed interview recordings were analyzed using King’s template analysis. The initial template was created with reference to recent literature on postgraduate medical e-learning quality indicators. The transcripts were coded, after which the emerging differences in template interpretation were discussed until a consensus was reached within the team.

Results: The final template consisted of three domains of positive e-learning influencers (motivators, learning enhancers, and real-world translation) and three domains of negatively influential parameters (barriers, learning discouragers, and poor preparation). The interpretation of the final template showed three subjects which form the basis of e-learning, namely, Motivate, Learn and Apply.

Conclusions: This study forms a basis for learning in general and could be applied to many educational instruments. Individual characteristics should be adapted to the target audience. Three subjects form the basis of, and six themes cover all items needed for, good (enough) postgraduate e-learning. Further research should be carried out with learners and real-world e-learning to validate this template.
Introduction

The benefits of medical e-learning courses are well reported and pertain to efficacy, cost-effectiveness, and interactivity. Furthermore, it is postulated that the number of available courses will continue to increase. E-learning is described as a viable solution for challenges which range from the promotion of self-directed learning to the provision of flexible learning through the continuous availability of learning opportunities. Moreover, it engages learners through collaborative learning communities and supports continuous professional development. These aspects of e-learning could be most beneficial for postgraduate and continuous medical education however, certain associated downsides must be considered.

In 2014, Cook and colleagues described the myths of e-learning and emphasized the downsides. According to the authors, e-learning is neither cheap, nor inherently more effective or even more efficient, than face-to-face learning.

Neither will it (by itself) transform education. However, the authors described one way in which it might create significant change in medical education (a disruptive innovation): “low-cost, low-tech, and instructionally-sound online learning (“good enough” instruction), represents a disruptive innovation that will soon displace high-tech, high-cost, online learning products.”

Sufficient literature exists to support the benefits of e-learning; however, we are also faced with the limitations and “myths”. The difficulty is that there is no clear definition of what constitutes “good enough” instruction. It is not even clear how to measure or define quality in postgraduate e-learning. A low-fidelity, high-quality approach, with integrated, appropriate learning theory, can successfully overcome many of the barriers to the introduction of e-learning. A few working models with quality indicators are available, but only one of these is specifically designed for postgraduate learners. In 2016 we described thirty-seven items in six domains in a quality specifications review, to provide an overview of these quality indicators. Most of these items are based on expert opinion and lack empirical evidence. Since we believe that the only way to validate a quality model is to involve the target audience and the experts, the present qualitative study will combine the knowledge of educational experts and the needs of post-graduate learners.
The aim of this study is, firstly, to further investigate and clarify quality indicators of postgraduate medical e-learning by considering them from the perspective of e-learning experts and end-users. In this study, we will use the word “learners” to mean “end-users”. Secondly, it is to identify the deeper underlying foundation of these characteristics. This study will be, to our knowledge, the first to explore the needs of learners and the opinions of experts in postgraduate medical e-learning and will, therefore, aid in the definition of “good enough” e-learning.

**Methods**

**Study design**

We decided to explore needs and opinions through focus-group discussions which were structured as per King’s template analysis. This qualitative approach aids the investigation of attitudes and beliefs, and also helps to generate new ideas and propel the development of theory. Larson and colleagues reported that an appropriate focus group is used for the following reasons: firstly, to gain clarity around people’s experiences of a given program; secondly, to generate information on participant attitudes and values; and thirdly, to add detail to previously generated information.

We paid attention to the distinct difference between group discussion and group interview. The former is an in-depth discussion between participants, with the moderator only mediating to keep participants on-topic. In an interview, in contrast, the facilitator principally focuses on individual respondents in the group, rather than allowing or encouraging discussion between the respondents. The structure covered the domains from this review, and open questions were used to elicit in-depth and personal opinions, and ideas behind the domains. All the focus-group discussions started with the sharing of experiences of good and bad e-learning.
Study participants

We used two groups of participants, as outlined in the current literature, namely, end-users of the e-learning programs (learners), and educators. It was considered that involving learners would ensure the e-learning programs became more learner-centred, while the educators could prevent a lack of learning strategies, assessment methods, and feedback mechanisms. We further divided the two groups, creating four groups in total. As research suggests that the upcoming generations (for example, “generation Y”, born between 1980 and 1990) have different learning skills and needs, we divided the learners into two groups: experienced residents (born before 1980), and less experienced postgraduates (born after 1980). In our opinion, there are also two different groups of educators: those who are in the service of a university and/or teaching hospital, and those who are in the service of a commercial company. Therefore, we formed the following groups: more experienced learners (EU); less experienced learners (LU); methodological experts (educators) (ME); and commercial experts (CE). Citations in the results section will follow these abbreviations.

The more experienced and less experienced learners were invited in person to the teaching university hospital at the VU Medical Centre, Amsterdam, the Netherlands. The methodological experts were educators invited from a national Dutch e-learning task group which we had contacted by e-mail and whose chairman put together a group of volunteers. The members of the commercial experts group were experienced designers and educators from a large e-learning company in the Netherlands, invited through the company director, who sent us an immediate response. The commercial educator group (CE) gave inputs based on their daily practice and years of designing experience, mainly from an educator’s point of view. A purposeful sampling was thus achieved to select the participants in the discussions. No financial compensation was given to any participant. The Dutch Association of Medical Education research gave ethical consent, after which all participants gave their written informed consent.

This study was conducted in the Netherlands, where residency training lasts between two and six years, depending on the discipline. Before that, the learners (EU and LU) had also received six years of undergraduate medical training. Medical educators belong to the Netherlands Association of Medical Education and have educating responsibilities at teaching hospitals and universities. E-learning is a well-developed area of commercial interest, with over 650 companies registered at the Chamber of Commerce in the Netherlands.

**Figure 1. Seven steps of template analysis by King 2012**

![Diagram of seven steps of template analysis]
Data collection method and procedure

To collect the data, we performed each focus-group discussion in the comfort of the participants’ own environment. The sessions were facilitated by RAL, lasted between forty-five and sixty minutes, and were audiotaped. We encouraged group discussion in which participants could dwell on the topics. Each new discussion allowed us to test and confirm the insights from the previous discussion. We continued conducting focus-group discussions until theoretical saturation was reached. The results of the discussions are presented using summaries and quotes.

Data analysis

Our research question is to identify additional indicators to those mentioned in the previous literature review. For this reason, we analyzed the data of the focus-group discussions using King’s template analysis. This method requires seven steps: define an original template (for example, from a previous literature research); transcribe and code the discussions; structure these into an initial template; revise that template; generate the final template; interpret the final template; and perform a quality check (see Figure 1).

Our original template is based on the six themes derived from the previously identified Postgraduate Medical E-learning Model (ME Model), i.e., preparation, design and system, communication, content, assessment, and maintenance. The recordings were transcribed verbatim and entered into f4 analysis software using the software’s workflow. After transcribing the discussions (performed by RL to gain familiarity with the data), all data were coded. The data were then re-checked to find any overlap in codes, or the addition of any new ones. From the coding of the transcripts, we created thirty-four categories that were relevant for postgraduate e-learning. These items were sorted into the original template. Items that did not fit the template were grouped into new themes. FS and MW also read different parts of the transcripts and sorted the codes into the initial coding template. Emerging differences were discussed, and the template was altered as necessary. Further team discussions resulted in the final template, and full consensus was reached.
Results

A total of four focus-group discussions, involving twenty-seven participants, took place between 1st November, 2015 and 7th April, 2016. The focus groups consisted of seven junior and six senior postgraduate learners, seven university education experts, and eight commercial educational experts. The mean age of the junior postgraduates and senior residents was 25 and 32, respectively. Overall, 88% of participants were female. During the fourth discussion, no additional categories, themes, or explanations emerged, after which information saturation was reached. Therefore, no new discussions were planned (Figure 1).

The Template Analysis

During our analysis, several considerations emerged from the thirty-four categories (Table 1). Firstly, it was necessary to consider the dynamics of positives and negatives: some items should be included to make e-learning effective, while others should not. These positives and negatives can be regarded as being continuously weighed up by the learner: “Should I participate/continue, or should I stop”. Secondly, there was a perspective change. While the original template encompassed almost all items, it did not address the needs of more emotional and cognitive elements, such as feelings of urgency, stress, and doubt. The original template was focused on creating e-learning and all the steps needed for that. After combining design, communication, content, and assessment, two new themes were added (motivators and barriers). We ended up with six themes: three positives to include (motivation to start, learning enhancers, and real-world translators), and three negatives to avoid (barriers to starting, learning discouragers, and poor preparation).
<table>
<thead>
<tr>
<th>Items</th>
<th>Themes</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feeling of importance</td>
<td>Motivation to start</td>
<td>Motivation</td>
</tr>
<tr>
<td>Rewarding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defining the kind of e-learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adding levels of learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formulating learning objectives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visualizing learning goals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of anytime, anywhere availability</td>
<td>Barriers to starting</td>
<td></td>
</tr>
<tr>
<td>Not convinced of the quality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No added value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forced to start</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not taken seriously</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offering support</td>
<td>Learning enhancers</td>
<td>Learning</td>
</tr>
<tr>
<td>Horizontal communication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertical communication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personalizing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem-based learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summarizing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeating in different formats</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Providing feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using learning activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activating and stimulating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offer knowledge overview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Too long</td>
<td>Learning</td>
<td></td>
</tr>
<tr>
<td>Too distracting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lacking user-friendliness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No navigation overview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stressing learners</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content not adapted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translating to the real world</td>
<td>Real-world translation</td>
<td>Applying</td>
</tr>
<tr>
<td>Using real-world examples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-world translators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Updating and maintaining</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowing your target audience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creating a development team</td>
<td>Preparation</td>
<td></td>
</tr>
<tr>
<td>Planning a feasible budget</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Theme One – Motivators to Start Learning

“The purpose of the e-learning should be made clear… I want to know what I am about to learn” (EU).

Both the learners and the educational experts agreed that the learner must be persuaded to start the e-learning. The first step was described by the educators as the most important: the learner should be motivated to pick up a device, or log in to the computer, and start the e-learning course. According to the educators, accessing intrinsic and/or eccentric motivation is essential.

To motivate the learner, the educator can use several methods, which were described by both end-users and educators. The experts appeared to believe that you should define the aim of e-learning (knowledge, skills, or behaviors/attitudes), formulate the learning objectives, and add levels of learning (the basic need-to-know level, a presumed known lower level, and an in-depth higher level). The learners agreed with this analysis, even suggesting that the learning goals should be visualized:

“The learning aim should be clear…the flow of the knowledge should be clear” (EU).

“You need to consider what has to be learned, how you know it has been learned...and which learning activities are needed...otherwise, you just get a lot of content” (ME).

This quote is an example of the difference between a reference work and an educational tool. E-learning should motivate, because it is an educational tool.

“The feeling of ownership, the feeling that you can control your learning path yourself, is motivating” (CE).

Motivation can be created by a strong introduction, with an emphasis on the relevance of the topic and urgency (stimulating intrinsic motivation), but also by adding rewards or a deadline, or making the e-learning compulsory (eccentric motivation).

“You must convince someone that the content is relevant for them” (CE).

Although a learner can be motivated by the e-learning program, they might still not start if the barriers are too high, as illustrated by the next section.
Theme Two: Barriers to Starting

By barriers, we mean the negatives that stand between a learner and the commencement of e-learning. When these barriers are too high, the learner might be motivated but still not start the e-learning program. One of the main benefits suggested by the learners and the e-learning experts is the ability to undertake the learning anytime and anywhere. The statements “Lack of access is a great barrier” (EU), and “E-learning is something I do when it suits my schedule” (EU) are examples of the way learners think about and experience e-learning. ‘Anytime, anywhere’ is, therefore, presumed to be important, and lack of access on these terms is a major barrier. Learners also agreed on another significant discouraging factor, namely doubt as to the quality of the e-learning program, and whether it offered added value over other didactic materials. This doubt can concern the educational quality:

“I do not have a very positive attitude (towards) e-learning, because I don’t like the idea of just clicking through texts” (LE);

or the content:

“a textbook is always accurate and complete, but e-learning can miss out important things; therefore, I prefer the textbook when it’s important” (LE).

If the learner is afraid that the e-learning course is incomplete, or when a traditional book provides more comfort in terms of quality, the learner would prefer to devote time to the traditional learning method. When the learner is in doubt, they experience “obligated e-learning” as being no longer motivating, but, rather, as a waste of their time:

“A deadline motivates… but only when I am also convinced that the e-learning is actually teaching me something new” (EU).

Aspects such as unrealistic deadlines, over-simple visuals or questions, and too much repetition with the same media cause the learner to feel they are not being taken seriously. Learners feeling that e-learning is not targeted at them, or doubting its quality, constitute barriers which may be too high.

After weighing the positives (motivators) and the negatives (barriers), it is hoped that the learner will start the e-learning program, which relates to the third and the fourth themes, as discussed below.
Theme Three: Learning Enhancers

While engaged in e-learning, the learner is again weighing the positives and negatives. ‘Learning enhancers’ are understood to mean the positives experienced by the learner during the e-learning process. This theme includes items that motivate the learner to continue, but also the didactics which make learning (more) efficient. It is the combination of these motivators and didactics that keeps the learner focused on the e-learning program, and motivated to continue and finish it. The learners identify communication as an important part of the learning process. They want to be able to communicate between learners (horizontal communication), put questions to experts (vertical communication), and, in the event of technical problems, to receive prompt support. However,

“(vertical) communication is only useful when you get a reply instantly” (LE).

Educators emphasize the difference between graduates and postgraduates when considering the amount of freedom which should be granted to the learner. We called this “personalization”, by which we mean the option for a-linear learning: the ability to move from basic to advanced knowledge, and to skip those sections with which the learner is already familiar. “It's important to give freedom to the user to navigate through the content” (CE). Given the advances in problem-based learning (PBL), all educators and learners believe that this offers the best theoretical background.

“The biggest benefit of problem-based learning...is the transfer to the workplace and therefore it is more motivating...especially for postgraduates” (ME).

To maximize the learning experience, educators believe in summaries, repetition, provision of feedback, and stimulating the learner by interactivity.

“The best way to repeat content is to wrap it in a different package every time. That way, you enable different learning styles (visual, textual, etc.) and the educator can repeat important lessons” (CE).

During the whole process, the learner needs to be activated and stimulated by using different media. “It’s nice to have questions and feedback between the texts, which keeps me better focused than just sections of text would” (LE). However, learners also emphasized the value of different formats:

“It was a short case-study, with a lot of visuals, and I had to move pictures around, which kept me busy. The content was a bit dull, but I really had the feeling I learned from it” (LE).
At the end, the learner wants to be given an overview of the knowledge gained in order to know what they have learned. As the experts stated,

“An educational tool is good when it enforces the lower-level students and raises their level of expertise” (ME).

In regards to interactivity, the experts stated that

“You want the learner to sit upright, not slouch” (ME), and

“You want to give a friendly poke in the ribs, without creating stress” (ME).

During the learning process, the many issues which may discourage the learner are regarded as the negatives of the learning enhancers, as outlined below.

**Theme Four: Learning Discouragers**

If the learning enhancers are minimal and the learning discouragers are multiplying, the learner might decide to stop before the learning aims are achieved. ‘Learning discouragers’ signifies those things that force the learner away from effective learning and may even drive them to stop. Both educators and learners gave examples of issues that can discourage to such an extent that the e-learning program is crippled. The learners identified these as a program that is too long, too distracting, lacks user-friendliness, fails to maintain learners’ interest, or even discourages them. The experts offered formulations such as

“…a long e-learning program (which) obliges learners to study things they already know (and) generates irrelevant content” (CE),

and stated that;

“E-learning is not about transferring all the expert’s knowledge, but about transferring what is needed in the daily practice of the learner” (CE).

There was some discussion about the ideal length of the e-learning course. Learners and educators say

“Videos should not take much longer than 5 minutes... a module, a maximum of 20 to 30 minutes” (LE),

and
“… people’s attention can be held for 20 minutes … our modules are about 8 to 12 minutes each” (ME);

and

“If it takes too long, I just quit” (LE).

They also emphasized the importance of a navigation overview to prevent learners becoming lost in the e-learning program. The educators tried to explain a paradox: how to motivate by offering deadlines and making the e-learning program mandatory, whilst not stressing the learner.

“Stress will overload your head and there will be no more room for learning” (ME).

Therefore, stress can come from content which is too complicated or too simple, but also from unrealistic deadlines, or modules that last too long. Finally, the content should be adapted to the e-learning platform, and should be different from text-driven mediums like textbooks:

“…when you only supply an online PowerPoint…you get a reference work and not an educational tool…there is a lack of interaction and, therefore, a lack of motivation” (ME).

For example, sentences should be shorter and simpler. When the learner finishes the e-learning program, the next step follows - real-world application. This is discussed in the next section.

Theme Five: Real-World Translators

The purpose of e-learning is to use and apply new knowledge, skills, or behaviors in the real world.

Therefore,

“you need a continuous link with the daily practice of the user” (CE).

‘Real-world translators’ is taken to mean the positives in this process of application, that is, those things in the e-learning program from which a learner needs to benefit most. The experts identified several items as important in translating e-learning to the real world. Firstly,

“The learners need to be able to recognize themselves in the content, but also in the feedback and evaluation of the e-learning course” (CE).
An illustration of this would be using enough examples which the learner recognizes from their daily work:

“When you work from a case-study or realistic situation, then you're more connected with the content and it's easier to remember and actually use the knowledge” (EU).

Secondly, e-learning can be kept connected to daily practice by being updated and maintained:

“The design…like the font…should not look very old” (LE).

It is possible, and important, to keep e-learning current, and to designate the people who are responsible for doing so. The maintenance of an e-learning program entails ensuring it works on newly developed platforms, and incorporating feedback from learners.

There are, of course, also ways to stand in the way of translating the e-learning to the real world.

“I believe a book is very limited and e-learning will be the future. You see it all around you” (LE).

When the learner cannot relate to the e-learning course, they may be motivated and even finish it, but will not use its content in their daily practice.

“You need to know and define the needs and motivation of your target audience” (CE).

Both learners and educators frequently agreed that creators of e-learning must know their target audience, its current knowledge, education level, and the things it experiences at work.

“If I give you a fourth-year science book, you will learn a lot less than an actual fourth-year student...that is because you need a reference framework” (ME).

If this consideration is not taken properly into account, learners cannot apply their e-learning.
Theme Six: Poor Preparation

It might seem strange to finish with the theme of preparation; however, poor preparation can cause harm to both real-world translation and the continuation of e-learning. Therefore, one must start with proper preparation, even if the effects of poor preparation might only manifest themselves at the end of the e-learning process.

‘Poor preparation’ is taken to mean that the e-learning program is neither created properly, nor given the attention it deserves during the creation. Two important elements of the preparation are the budget and team collaboration.

“Bad e-learning is e-learning created from one perspective without the collaboration of the target audience, content experts, education experts, and ICT staff” (CE).

An education expert helps because they

“will answer the questions of what the learning aims are, and how you know these have been acquired” (ME).

Secondly, cost-planning must be borne in mind, and creators must be aware that profits may be small or non-existent:

“I would never pay for e-learning, because once you have done it, it’s useless” (LE).

To prevent this, the educators suggested creating a dedicated development team, and careful planning of costs.

Final template interpretation

According to King’s template analysis, after the final template has been created there is room for its interpretation. Above, we have created a final template containing six themes - three positives and three negatives. From a consideration of the positives and negatives, we have distilled three aspects that are positively and negatively influenced: motivate, learn, and apply (Figure 2). This is a learner-centered construct of subjects in chronological order. In Figure 2, we visualize the process of first, motivating (balancing the motivational items and barriers); then the learning itself (balancing learning enhancers and discouragers); and, finally, the application of the e-learning content (which is only possible when it can be translated to the real world, that is, when it has not been hampered by poor preparation).
Figure 2. the final theme
Discussion

We believe the final template confers added value on the original template. After combining some of the original themes, we lifted them to a higher level of cognition and evaluation, which are what the learner uses to weigh up the e-learning process. We ended with six categories: three positives to be included (motivation to start, learning enhancers, and real-world translators) and three negatives to be avoided (barriers to starting, learning discouragers, and poor preparation). We believe that the only way to find these underlying constructs is by holding focus-group discussions. We did not find enough arguments to limit the final template to the three aspects of motivate, learn, and apply. If the template were limited to these three alone, it would lose the necessary nuances. The added value conferred by the addition of the final themes is in regard to learner-centeredness. We also believe creating broad themes, gives more freedom in the e-learning design, and more focus on the real needs. One example is the Design and System domain from the original template. One of the items is “Use software depending upon flexibility”.

This may seem straightforward; however, it does not touch on the real issue. We prefer: Does your e-learning have any barriers to starting? As a designer will have an example of these barriers, namely poor, inflexible software. However, this formulation forces an educator to think about other barriers, or even accept one barrier (poor software, for example, because they have no other options) but compensate with a powerful motivator (for example, a free mobile device which runs that software).

Another example is the content domain, one of the items in this case being the use of cognitive multimedia principles. The underlying need, however, is for “learning enhancers”: which particular aspects the educator chooses to use to improve or enhance learning does not matter. Our previous literature study showed, however, that the cognitive load principle is a particularly valuable example.

Notions from existing literature underpin the identified model. Schumacher and colleagues, for instance, describe several learning and motivational theories to construct three subjects which are analogous to our motivate, learn, and apply: the desire to learn; the ability to learn; and a context and environment for learning. Motivation has been identified as the essential component that stimulates and sustains learning behavior. While motivation may not be the main stimulus for learning, it definitely plays a critical role during the learning process. Some studies have considered individual items such as rewards (the positive aspects of rewards are to be utilized), while others give examples of themes such as barriers to e-learning. In a recent study, Reid and colleagues described several obstacles to e-learning engagement in medical students. There is an interesting overlap in the obstacles found in their study and the negatives of the template described in the current paper. Reid and
colleagues described the feeling of unfairness, which is a barrier comparable to the barrier to starting in our ‘motivation’ subject. The feeling identified by Reid and colleagues of being lost and overwhelmed could create a learning distraction and disengagement, which might be a result of poor preparation and poor user connection. The overlap between the obstacles found by Reid and the negative themes of the present study suggest, perhaps, a general working model in which different focus discussions from different research groups arrive at a similar conclusion.

**Generic or postgraduate-specific model**

As with the template model, there are very few specific items for medical postgraduates. One could argue that motivate, learn, and apply is a mantra for all (adult) learning; however, there is a difference in the specifics. As the educators state:

“The postgraduate needs to be able to scroll through subjects that are neither too simple, too boring, nor irrelevant, so he or she has the freedom needed to remain motivated” (CE).

This consideration is because

“postgraduates value flexibility more than graduates, and have more self-discipline” (ME).

Their motivation is also different; for example,

“for a medical student, a good e-learning program is one that prepares you for an exam; for a postgraduate, it is only good when it’s relevant for their daily practice” (ME).

Furthermore,

“postgraduate e-learning should be more focused on skills and learning tasks than on plain knowledge” (ME).

Researchers have applied different theories of adult learning and, although debatable, recent meta-analysis has even suggested that motivation and learning changes with age. The template may be partly generic, but others have come to a similar conclusion in the competency-based framework towards which medical education is growing, in which (postgraduate) learners drive their own education.

**Limitations**

Certain limitations must be acknowledged. The very nature of this method means that it may not always lead to the same conclusions. The interpretations of the codes, the
template, and the final subjects might be influenced by cultural values, earlier experience, and prejudice concerning the original template. The authors have used the discussions described and the international literature to broaden their view and make the model as generalized as possible. Two limitations are specific to this study, apart from the difficult distinction between generic characteristics and postgraduate specifics.

The first of these limitations is the range of items extracted from the discussions. The learners are limited by their experience, and the educators are limited by their view of medical education. Therefore, we tried to extract the underlying constructs and use the items as examples, rather than as the only possible items.

Secondly, the definition of e-learning is broad and culturally diverse, and learners and educators seem to diverge in their understanding of the meaning of the term. While learners have a clear idea about e-learning - “just learning with the computer” (LU) - educators believe that there are many different forms of e-learning, with different goals. E-learning is learning through a digital medium, but there are great differences between blended learning, hybrid learning, simulation, serious gaming, and others. These activities also have different learning goals, which may be the acquisition of new knowledge, skills, attitudes or behaviors. Our proposed themes could fit all these e-learning variations and goals. The specifics within the themes are key to the variations in e-learning goals, and target audience.
Conclusion

As noted above, low-fidelity, high-quality e-learning could be a major education disruptor; however, we need more insight into the characteristics and indicators of quality. This is the first qualitative study, to our knowledge, to incorporate the perceptions of postgraduate end-users of the quality of their e-learning. We have raised a postgraduate medical e-learning quality model to a higher level by defining the actual needs and constructs behind the detail. We concluded that three negative and three positive themes are balanced by the learner, and could be responsible for the essence of good (enough) postgraduate medical e-learning. Further research should be carried out to validate these themes, for example by asking learners to evaluate their e-learning using this model.

Acknowledgements

We wish to thank our participating postgraduates, the medical educators, and TinQwise for their time, effort, and honest inputs in the focus-group discussions. Furthermore, we wish to thank Mr. Kieran Walsh for his constructive feedback and Mr. Sha’er Ramkalup for his help with the discussions and transcripts. Finally, we wish to thank Ms. Rees for correcting the English of the draft manuscript. There was no financial support for this study.

Conflict of Interest

The authors declare that they have no conflict of interest.
References


12. Boddy C. A rose by any other name may smell as sweet but “group discussion” is not another name for a “focus group” nor should it be. Qualitative Market Research: An International Journal. 2005;8(3):248–55.


Consensus on Quality Indicators of Postgraduate Medical E-Learning: Delphi Study

De Leeuw RA
Walsch K
Westerman M
Scheele F

JMIR Medical Education
September 2018
Abstract

Background: The progressive use of e-learning in postgraduate medical education calls for useful quality indicators. Many evaluation tools exist. However, these are diversely used and their empirical foundation is often lacking.

Objective: We aimed to identify an empirically founded set of quality indicators to set the bar for “good enough” e-learning.

Methods: We performed a Delphi procedure with a group of 13 international education experts and 10 experienced users of e-learning. The questionnaire started with 57 items. These items were the result of a previous literature review and focus group study performed with experts and users. Consensus was met when a rate of agreement of more than two-thirds was achieved.

Results: In the first round, the participants accepted 37 items of the 57 as important, reached no consensus on 20, and added 15 new items. In the second round, we added the comments from the first round to the items on which there was no consensus and added the 15 new items. After this round, a total of 72 items were addressed and, of these, 37 items were accepted and 34 were rejected due to lack of consensus.

Conclusions: This study produced a list of 37 items that can form the basis of an evaluation tool to evaluate postgraduate medical e-learning. This is, to our knowledge, the first time that quality indicators for postgraduate medical e-learning have been defined and validated. The next step is to create and validate an e-learning evaluation tool from these items.
Introduction

E-learning, which also goes by many other names, is taking up a strong position in medical curricula because of its flexibility, richness, and potential for resource sharing and for high value in light of its cost [1]. E-learning is suggested as an eligible instrument for interprofessional learning [2], and Goh described e-learning not as just hype, but as a core aspect of medical education in the future [3].

However, the debate on what denotes good-quality e-learning is ongoing. More explicitly, the lack of knowledge on what constitutes good-quality e-learning has been identified as one of the main inhibitors of its usefulness [4]. Cook postulated that e-learning is not always cheaper or more efficient than traditional forms of medical education. However, he also stated that e-learning can be a very important innovation when it becomes “low-cost, low-tech, but instructionally sound ‘good enough’ online learning” [5]. The problem is that there is no useful model for “just good enough” postgraduate medical e-learning. The literature shows that there are no specific working models for this target audience [6] and that the models and tools that are used are diverse. We have previously provided a list of quality indicators [6] and tried to find the underlying constructs of which items are important and meet the needs of learners [7]. In this way, we tried to provide the categories necessary to evaluate postgraduate e-learning. Both for educators involved in postgraduate e-learning and for users themselves, it is crucial to know that e-learning is worth their investment in it. Previous research showed that users are less motivated and less eager to undertake an e-learning module when they are in doubt about its quality [7]. Furthermore, experts believe that it is necessary to know what quality features are required and expected of an e-learning course before it is created [7].

In response to this debate on what constitutes good-quality medical e-learning, we set out to provide an empirically based set of quality indicators. Thus, we performed a Delphi procedure to evaluate suggested quality indicators from the literature. To our knowledge, this study is the first international consensus by both educational experts and experienced users on quality indicators in postgraduate medical e-learning.
Methods

In this study, we performed a Delphi procedure to determine consensus on the possible quality indicators for e-learning in postgraduate medical education.

Study Design

Escaron et al describe the Delphi method as being well suited to informing health education [8]. It is based on the concept of pooled intelligence and should enhance the individual judgments and capture the collective opinion of experts [9]. We performed the Delphi digitally, facilitated by RAD, because online Delphi studies reduce costs, time, and effort [9] and are not limited by geographical boundaries. The downside is that participants have a consultative role and disagreements are hard to explore. This is even more the case when using a digital medium to communicate. To maximize the effectiveness of the Delphi, we followed the guidelines of de Villiers et al [9]. We first provided a definition of e-learning to the expert panel, then started with a questionnaire of items. After analyzing the results, we removed items without consensus, added comments on the remaining items, and, if applicable, added new items.

E-Learning Definition

For this Delphi we chose the following, slightly adapted definition from Sangrà et al: “E-learning is 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” [10]. To simplify the discussion, we chose to talk about stand-alone, asynchronous, and distant e-learning (and not learning management systems). We provided all participants with this definition and an explanation in the introduction of the Delphi.

Expert Panel Selection

For this study, we used 2 expert groups: medical educators and end users. Medical educators are experts in the theory and practice of creating e-learning and end users know what it’s like to use the e-learning in their daily practice. A suitable expert is defined in the literature as someone who possesses the relevant knowledge and experience and whose opinions are respected by fellow workers in their field [9]. For this study, we defined an educational expert as a member of a national medical education platform (usually a university- or government-led foundation aimed at improving and validating medical education) or someone who has been published in peer-reviewed international journals on the subject of medical e-learning, and who has had
at least 3 years’ experience with medical education and e-learning development. We defined experienced postgraduate users as postgraduate residents who graduated at least 2 years ago and who have had exposure to e-learning throughout their postgraduate training.

We selected experts by means of an inquiry to the National Education Board in the Netherlands and from author contacts. We invited experienced users in the Netherlands and Great Britain because we had local contacts there. An expert panel usually consists of 15 to 30 participants, with 5 to 10 participants per category [9]. Our aim was to have 10 experts and 10 experienced users but, as we believe that educational experts have a better background in the theoretical grounding of education, we preferred to have a few more educational experts on the panel. We thus aimed for 13 experts and 10 users [9].

**Questionnaire Development**

The initial set of indicators was based on 2 previous studies and contained quality characteristics from the literature [6] (72 items) and from focus group discussions (resulting in 57 items) with both experts and end users [7] (see Figure 1). These previous studies gave a total of 57 items in 6 themes on 3 subjects: motivate, learn, and apply. The subject motivate consisted of indicators that increase the learner’s level of motivation in the theme, called starting motivators, and indicators that form a barrier to starting or finishing the e-learning, called starting barriers. The next step was the subject learn, which consisted of all pedagogical indicators that either facilitate (learning enhancers) or limit (learning discouragers) the learning experience. The subject apply was made up of indicators that help the learner to translate and apply the e-learning into their daily practice (real world translators). Finally, the theme poor preparation (6 items) consisted of indicators that help an author prepare for the creation of an e-learning resource. Items such as “Plan a feasible budget to prevent incompletion of the e-learning due to lack of funds” were not originally aimed at the end user and therefore evaluated only by the experts.

The questionnaire started with introductory text explaining the subjects, providing a definition of e-learning, and asking the experts and users to imagine e-learning that was “just good enough” and targeted at medical postgraduates. After that, the experts and end users evaluated the individual items on a 5-point Likert scale and were able to add comments [9].
After we agreed on the content of the questionnaire, we performed a pilot round with 5 participants (2 educators and 3 end users). After incorporating their feedback on the items, we invited the experts to fill out the questionnaire digitally. We started the first round with 57 items.

**Statistical Analysis**

After each round, we worked out consensus by calculating the rate of agreement: 
\[
\text{Rate of agreement} = \frac{\text{agreement} - \text{disagreement}}{\text{agreement} + \text{disagreement} + \text{indifferent}} \times 100\%.
\]
We used a rate of agreement of two-thirds to accept an item. An item was rejected when there was no consensus after 2 rounds, or when an item was rejected by a rate of agreement lower than −66% in the first round (the rate of agreement scale ranges from −100 to 100). There is no consensus in the literature regarding the best rate of agreement to be used; the range used has been between 51% and 80% [11]. We chose to use two-thirds as proposed by de Villiers et al [9].

The Ethical Review Board of the Association for Medical Education gave ethical consent (file number 475), after which all participants gave their written informed consent.
Results

We sent the first invitation emails out on March 19, 2017, and received the final response on July 20, 2017. We invited 23 experts, of whom 13 replied and participated, 9 did not reply to the invitation, and 1 did not consider himself an expert on postgraduate medical e-learning. We invited 17 experienced users, of whom 5 did not reply, 2 could not participate due to other obligations, and 10 were able to participate. In total, we had 23 participants, of whom 23 responded in both rounds. Of the participants, 13 (57%) were male. The average age of the experts was 49 years and that of the users was 31 years. The experts came from the Netherlands (n=7), Great Britain (n=3), Canada (n=2), and South Africa (n=1). They had an average of at least 3 years’ experience creating or evaluating medical e-learning and together had published 29 articles. A total of 4 were members of the Dutch Association for Medical Education expert group on e-learning. The users were Dutch (n=7) and British (n=3), and had more than 3 years’ experience as residents, and had attended on average more than 2 e-learnings during their residency.

In the first round, 37 items were accepted as important, with a rate of agreement of above two-thirds. No items were rejected, there was no consensus on 20 items, and 15 new items were added by the participants (Figure 2). In the second round, we added the comments from the first round on the items without consensus and added the 15 new items (35 items in total). We also added 3 explorative questions based on comments from the first round, exploring the usefulness of a list of indicators. Multimedia Appendix 1 shows all items, rate of agreement, and consensus.

The first explorative question was “Do you think it is possible to define a minimum and general set of criteria that can be generalized for all types of medical e-learning?” A total of 17 participants thought this was possible, 5 were not sure, and 1 thought it was too complicated. Worries about such a list of indicators included the following:

But I would be concerned that to be applicable for all types of medical e-learning it might be too general and therefore not practically useful [Medical Educator 1]  
Yes, but it’s like evidence-based medicine: you must be able to deviate with motivation. [Medical Educator 4]
The experts also raise the concern of a fast-changing definition of e-learning: the term e-learning is in a fast-changing technological world with different needs and skills for makers (and for users) and is difficult to define—without maker- or user-focused definition and context [Medical educator 1]
e-learning doesn’t mean anything in particular, tech can be used in every aspect of med-ed, and lots of different tech can be used for different purposes…. [Medical educator 5]

Participants mentioned that which form of e-learning these indicators are about is very important to explain.

The second explorative question was “Do you think a 10-question survey, like the one mentioned in the introduction, would be of added value to the current evaluation tools?” It was thought by 14 (64%) to be of added value, 7 (32%) were not sure, and 1 (4%) thought it was not of added value. Arguments were

...it would help setting priorities [Medical educator 8]

...general design principles probably will apply to e-learning as well. So why the need of a specific tool? I think there may be added value in evaluating the specific additive value of technology. But I am not sure. That’s why I am participating in this Delphi. [Medical educator 3]

The third explorative question was to explore how many items participants considered to be workable. The general opinion was “the less the better, but as much as needed”. When asked for a number, participants responded with a range of numbers from 10 to 20.

We then evaluated the remaining 35 items (see Multimedia Appendix 1). There was consensus that just 2 items should be included, 3 were rejected, and there was no consensus on the rest. After this round, a total of 72 items were addressed and, of these, 37 were accepted and 34 rejected (see Table 1).
Figure 2. Flowchart of the Delphi results.
**Table 1.** The final quality indicators. Items 32-37 are expert theme preparation items. Subject and item

<table>
<thead>
<tr>
<th>Subject</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivate</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Create a feeling of importance within the learner</td>
</tr>
<tr>
<td>2</td>
<td>Create a feeling of responsibility within the learner</td>
</tr>
<tr>
<td>3</td>
<td>Provide enough time to complete the e-learning</td>
</tr>
<tr>
<td>4</td>
<td>Define the purpose of the e-learning (knowledge, skills, and behavior or attitude)</td>
</tr>
<tr>
<td>5</td>
<td>Formulate the learning objectives and preferably visualize them</td>
</tr>
<tr>
<td>6</td>
<td>Provide an overview of all content</td>
</tr>
<tr>
<td>7</td>
<td>Prevent concerns about the quality of the content</td>
</tr>
<tr>
<td>8</td>
<td>Do not force, although obligation might be possible</td>
</tr>
<tr>
<td>9</td>
<td>Create the feeling that the learner is being taken seriously</td>
</tr>
<tr>
<td>10</td>
<td>Use a flexible platform, so that the content can be modified by the educator</td>
</tr>
<tr>
<td>11</td>
<td>Provide easy accessibility from all locations and devices</td>
</tr>
<tr>
<td>12</td>
<td>Use easy and clear navigation</td>
</tr>
<tr>
<td>13</td>
<td>Use a simple layout with a sitemap</td>
</tr>
<tr>
<td>14</td>
<td>Software should be safe and secure</td>
</tr>
<tr>
<td>15</td>
<td>Access should be fast</td>
</tr>
<tr>
<td>16</td>
<td>Make clear which device is needed and advise the learner about the skills needed</td>
</tr>
<tr>
<td>Learn</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Enable the learner to personalize the module</td>
</tr>
<tr>
<td>18</td>
<td>Allow nonlinear learning</td>
</tr>
<tr>
<td>19</td>
<td>Show what has already been achieved and what has not yet been done (progress bar)</td>
</tr>
<tr>
<td>20</td>
<td>Provide technical support</td>
</tr>
<tr>
<td>21</td>
<td>Add summaries</td>
</tr>
<tr>
<td>22</td>
<td>Give feedback</td>
</tr>
<tr>
<td>23</td>
<td>Add exercises and assignments</td>
</tr>
</tbody>
</table>
|   | Create interaction with the content  
|---|--- |
| 25 | Do not stress or frustrate the learner  
| 26 | Avoid nonadaptive content  
| 27 | Do not create too distracting a design or learning activities  

**Apply**

|   | Make the content translatable to the real world  
|---|--- |
| 28 | Update and maintain the e-learning  
| 29 | Provide sources of information and keep access available after the course is finished  
| 30 | Evaluate the e-learning after the course and collect feedback  
| 31 | Know your target audience and adapt learning objectives accordingly  
| 32 | Identify the authors at the beginning of the e-learning  
| 33 | Create a timeline with objectives and expectations of the production stage  
| 34 | Form a development team with at least 1 content expert, 1 educational expert, and 1 information technology expert, and let them all commit a certain amount of time before starting the development  
| 35 | Plan a feasible budget to prevent incompletion of the e-learning due to lack of funds  
| 36 | Consider an appropriate learning environment and learning management system  

*aExpert theme preparation items.*
Discussion

Principal Findings

We performed an international Delphi study with educational experts and experienced users that led to 37 quality indicators for postgraduate medical education. To our knowledge, this is the first list of quality indicators for postgraduate medical e-learning with an evidence-based foundation: first selecting all the indicators mentioned in the literature, then adding to this list by focus group discussions, and finally selecting the items using a Delphi.

Cook et al wrote in 2009 that internet-based learning is associated with a positive effect, but that future research should directly compare different internet-based interventions [12]. Developing peer-reviewed training and guidelines for e-learning should also be the foundation of academic e-learning [13]. However, to compare e-learning or e-education methods and to guide authors, we need to provide them with a tool. These indicators should form the basis for such an e-learning evaluation tool that can help to compare different types of education with e-learning. To evaluate the effect of e-learning in postgraduate medical education, we need a list of indicators. We believe that these indicators should be supported by experts in the field and the final end users of the e-learning resources. This study produced such a list.

After the first round of the Delphi, the experts expressed the challenges of an evaluation of this type. The term e-learning can be confusing, the added value to a landscape of many other evaluation tools might be limited, and the indicators may be too general. The term e-learning, as discussed in the introduction, is broad. However, when it is well defined, we believe it can still be a workable term. There are many quality models in the literature [14], and e-learning has been evaluated many times [15]. But these models are aimed at different target audiences, the origin of the indicators is ill defined, and the validation is limited, when present at all. The final indicators from our study are quite generic and are difficult to translate back to postgraduate learning. It could very well be that the items identified in this study are applicable to graduates or other groups of learners.
Limitations and Strengths

Potential pitfalls in Delphi studies are the imposition of preconceptions on respondents and poor techniques for summarizing and presenting the group response. We tried to limit these pitfalls by producing a simple and straightforward questionnaire. Participant selection was limited to those who responded and, by choice, from the countries of the authors’ residence. Therefore, our study lacked a certain cultural diversity, making the results possibly less generalizable. The strength of the final indicators lies in the balance of general aspects of evaluation and the specifics added when needed. We believe that the 6 themes (motivation, barriers, learning enhancers, learning discouragers, real-life translation, and poor preparation) are general enough to be applied to all kinds of e-learning.

Conclusion

Creating e-learning for postgraduates is not enough; evaluation and improvement should not be additional but mandatory to ensure maximum effect. E-learning quality indicators can be sorted into 3 groups (motivate, learn, and apply) with 5 general themes (motivators, barriers, learning enhancers, learning discouragers, and real-life translators) and a list of items that can be used in preparing e-learning resources.

This study provided a list of quality indicators for postgraduate medical e-learning. This list is unique in its evidence-based foundation and in the way that it applies broad themes with specific indicators. The most logical next step is to create and validate an evaluation tool based on these indicators.

Acknowledgments

We wish to express our greatest gratitude to the participants of the Delphi. Without their feedback, patience, and time, we could not have completed this study.

Authors’ Contributions

All authors contributed equally in the preparation of the study protocol. Experts were supplied mainly by RAD and KW, and users were mainly suggested by RAD, FS, and KW. RAD carried out Delphi facilitation, data collection, and analysis. All authors wrote and corrected the final manuscript.
<table>
<thead>
<tr>
<th>Question</th>
<th>Round 1 RoA*</th>
<th>Round 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum EDUC USERS Result</td>
<td>Sum EDUC USERS Result</td>
</tr>
<tr>
<td><strong>Theme motivators and barriers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1a. The e-learning should create a feeling of importance within the learner</td>
<td>90</td>
<td>85</td>
</tr>
<tr>
<td>Q1b. The e-learning should create a feeling of importance within the learner by encouraging them to complete it by a realistic deadline, after which the e-learning modules should no longer be accessible</td>
<td>19</td>
<td>-7</td>
</tr>
<tr>
<td>Q2. The e-learning should create a feeling of responsibility within the learner</td>
<td>67</td>
<td>57</td>
</tr>
<tr>
<td>Q3. The authors’ authority and expertise should be identified</td>
<td>43</td>
<td>50</td>
</tr>
<tr>
<td>Q4. The e-learning should provide a reward when finished</td>
<td>52</td>
<td>50</td>
</tr>
<tr>
<td>Q5. The learner should be provided with enough time to complete the e-learning</td>
<td>90</td>
<td>93</td>
</tr>
<tr>
<td>Q6. The educator should give a definition of the purpose of the e-learning (knowledge, skills and/or behaviour/attitude, defining the e-learning objectives for the next section)</td>
<td>90</td>
<td>86</td>
</tr>
<tr>
<td>Q7. The educator should formulate the learning objectives and preferably visualise them</td>
<td>95</td>
<td>92</td>
</tr>
<tr>
<td>Q8. There should be different levels of learning within the e-learning (basic, novice and advanced)</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td>Q9. The e-learning should provide an overview of all content (with details such as what you can learn, where to learn it, etc.)</td>
<td>81</td>
<td>79</td>
</tr>
<tr>
<td>Q10. Users of the e-learning (learners) should not have concerns about the quality of the e-learning content</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>Q11. The learners should not be forced to undertake the e-learning, although compulsory e-learning might be possible</td>
<td>67</td>
<td>73</td>
</tr>
<tr>
<td>Q12. The learner should have the feeling that they are being taken seriously (e.g. you should involve stakeholders, aim at the expected level of experience, not make the visuals too childish, etc.)</td>
<td>86</td>
<td>93</td>
</tr>
<tr>
<td>Question</td>
<td>Rating</td>
<td>Added</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>13.1 The availability of offline e-learning should be ensured</td>
<td>33</td>
<td>14</td>
</tr>
<tr>
<td>13.2 E-learning should be flexible, so that the content can be modified by the educator</td>
<td>76</td>
<td>71</td>
</tr>
<tr>
<td>13.3 There should be a search function</td>
<td>38</td>
<td>43</td>
</tr>
<tr>
<td>13.4 There should be easy accessibility from all locations and devices</td>
<td>62</td>
<td>43</td>
</tr>
<tr>
<td>13.5 Multi-tasking with other platforms, e.g. to play music, should be allowed</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>13.6 There should be easy and clear navigation</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>13.7 There should be a simple layout with a sitemap</td>
<td>89</td>
<td>79</td>
</tr>
<tr>
<td>13.8 Software should be safe and secure, e.g. by installing an SSL certificate (for secure communication over the internet)</td>
<td>67</td>
<td>57</td>
</tr>
<tr>
<td>13.9 Access should be fast; therefore, content should be aptly sized and formatted (smallest possible image size, etc.)</td>
<td>76</td>
<td>64</td>
</tr>
<tr>
<td>Q14. The educator should make clear what device is needed for the e-learning and the educator should advise the learner about the skills needed</td>
<td>85</td>
<td>92</td>
</tr>
<tr>
<td>Q15. The e-learning should have added value over the available learning content (offer new material or be presented in a different way)</td>
<td>60</td>
<td>46</td>
</tr>
</tbody>
</table>

**Added rows:**

- Q232. You should add the previous users' ratings of the module: **Added 30 50 0 Undecided**
- Q233. You should try to have the e-learning referred to as an important source of information by, e.g., tutors, educators, etc.: **Added -42 -27 -63 Undecided**
- Q234. Materials should include a date to indicate relevance: **Added -33 -25 -44 Undecided**
- Q235. You should formulate and communicate an "expiry date" for the e-learning, after which it should be re-evaluated and, if needed, updated: **Added -86 -83 -89 Rejected**
- Q236. A digital medium could in itself already be of added value over a book, so the content itself does not have to be: **Added -50 -30 -83 Rejected**
<table>
<thead>
<tr>
<th>Theme</th>
<th>learning enhancer</th>
<th>Q17a. The e-learning should enable the learner to personalise the module (save and continue, pre-testing, create ownership)</th>
<th>95</th>
<th>100</th>
<th>86</th>
<th>Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q17b. The e-learning should enable the learner to personalise the module by allowing a linear learning (linear is fixed from A to Z, while a linear allows users to skip parts, return to them and create their own learning route)</td>
<td>71</td>
<td>64</td>
<td>86</td>
<td>Accepted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q17c. The e-learning should enable the learner to personalise the module by showing what has already been achieved and what has not yet been done (progress bar)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>Accepted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q18. The e-learning management system should provide technical support</td>
<td>67</td>
<td>57</td>
<td>86</td>
<td>Accepted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q19. The e-learning should provide horizontal communication (communication between learners) as well as being an essential part of the learning objective</td>
<td>-15</td>
<td>-15</td>
<td>-14</td>
<td>Undecided</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Q20. The e-learning should provide vertical communication (communication between learner and educator) as well as being an essential part of the learning objective.</td>
<td>0</td>
<td>-7</td>
<td>14</td>
<td>Undecided</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Q21. The e-learning should be based on the concepts of problem-based learning as much as possible</td>
<td>45</td>
<td>15</td>
<td>100</td>
<td>Undecided</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Q22. The e-learning should provide a minimum set of contents needed to achieve the learning objectives</td>
<td>62</td>
<td>64</td>
<td>57</td>
<td>Undecided</td>
<td>-20</td>
</tr>
<tr>
<td></td>
<td>23.1 Summaries</td>
<td>71</td>
<td>84</td>
<td>86</td>
<td>Accepted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.2 Repetition of content in different media/formats</td>
<td>5</td>
<td>21</td>
<td>-29</td>
<td>Undecided</td>
<td>-5</td>
</tr>
<tr>
<td></td>
<td>23.3 Feedback</td>
<td>95</td>
<td>93</td>
<td>100</td>
<td>Accepted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.4 Exercises and assignments</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>Accepted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.5 Interaction with the content</td>
<td>81</td>
<td>93</td>
<td>57</td>
<td>Accepted</td>
<td></td>
</tr>
<tr>
<td>Q25. The e-learning should not stress or frustrate the learner (no negative feedback from a supervisor or punishment for failure, a generous time window for access)</td>
<td>68</td>
<td>58</td>
<td>86</td>
<td>Accepted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q26. Non-adaptive content should be avoided</td>
<td>95</td>
<td>92</td>
<td>100</td>
<td>Accepted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q27. The e-learning should not be too distracting in its design or learning activities (prevent distractions like irrelevant images or moving elements)</td>
<td>75</td>
<td>62</td>
<td>100</td>
<td>Accepted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q28. Multimedia use per Mayer's Cognitive Load principles of Multimedia</td>
<td>52</td>
<td>50</td>
<td>7</td>
<td>Undecided</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>23.7 Self-assessment</td>
<td>29</td>
<td>0</td>
<td>86</td>
<td>Undecided</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>23.8 Reminders by e.g. email for e.g. an assignment or unfinished module</td>
<td>48</td>
<td>43</td>
<td>57</td>
<td>Undecided</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>23.9 Activate/stimulate the learner by using fun elements, challenges or serious gaming</td>
<td>-7</td>
<td>-43</td>
<td>Undecided</td>
<td>9</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>23.10 Do you have any comments or ideas concerning a learning activity?</td>
<td>62</td>
<td>57</td>
<td>71</td>
<td>Undecided</td>
<td>-19</td>
<td>8</td>
</tr>
<tr>
<td>Q29. All presenters should be native speakers in the language or have as neutral accents as is possible. Unlike in face-2-face, misunderstandings in video presentation cannot easily be corrected. If it is not possible to find a native speaker, the video should be subtitled and proofed by native speakers</td>
<td>Added</td>
<td>-57</td>
<td>-42</td>
<td>-78</td>
<td>Rejected</td>
<td></td>
</tr>
<tr>
<td>Q28A. If possible, create a an evidence-based layout (e.g. font style), based on a guideline provided by a checklist like this one</td>
<td>Added</td>
<td>-6</td>
<td>0</td>
<td>-16</td>
<td>Undecided</td>
<td></td>
</tr>
<tr>
<td>Q249. The e-learning should be easy for students with a disability (dyslexia)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q250. Materials should record user actions for analysis</td>
<td>Added</td>
<td>-35</td>
<td>-42</td>
<td>-25</td>
<td>Undecided</td>
<td></td>
</tr>
<tr>
<td>Q251. Other learners' performance should be recorded and shown to new users to motivate them</td>
<td>Added</td>
<td>46</td>
<td>69</td>
<td>11</td>
<td>Undecided</td>
<td></td>
</tr>
<tr>
<td>Q252. Fun elements can create a childish feeling, which should be avoided.</td>
<td>Added</td>
<td>14</td>
<td>15</td>
<td>11</td>
<td>Undecided</td>
<td></td>
</tr>
</tbody>
</table>

**Theme apply**

| Q29. Make the content translatable to the real world by using examples from the daily practice of the learner and recognizable subjects, adding relevant items like time management only when applicable. | 100 | 100 | 100 | Accepted |
| Q30. Update and maintain the e-learning | 100 | 100 | 100 | Accepted |
| Q31. Provide sources of information outside the e-learning and keep access available after the course is finished | 76 | 79 | 71 | Accepted |
| Q32. Evaluate the e-learning after the course and collect feedback for the following topics: user experience, effectiveness, usability and costs. | 67 | 79 | 43 | Accepted |

**Expert theme preparation**

| Q33. Know your target audience and adapt learning objectives accordingly | 100 |  |  | Accepted |
| Q34. Identify the authors at the beginning of the e-learning. The authors should (personally) be known to the educators, and information from an unknown source should not be used | 67 |  |  | Accepted |
| Q35. Create a timeline with objectives and expectations of the production stage | 70 |  |  | Accepted |
| Q36. Form a development team with at least 1 content expert, 1 educational expert and 1 IT expert, and let them all commit a certain amount of time before starting the development | 80 |  |  | Accepted |
| Q37. Plan a feasible budget to prevent incompletion of the e-learning due to lack of funds |  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Q38. Consider an appropriate learning environment/learning management system | 100 | Accepted |  |  |

**Added new items**

| Q253. Each module should contain story telling/a narrative; It does not have to be “cases” or “PBL”; - all sorts of stories can be engaging, e.g. from different points of view (patient, doctor, relative), historical stories, stories of scientific discovery relevant to the learning | Added | -24 | -15 | -37 | Undecided |
| Q254. Use the TPACK framework ([http://www.tpack.org](http://www.tpack.org)). It is possible that, e.g., one person is both educational expert and IT expert | Added | 7 (**8**) | 25 | -16 | Undecided |
| Q255. Create a team of creators, who are also the users of the e-learning. In doing so, you will keep contact with your audience and can spread the message much more easily | Added | -43 | -38 | -50 | Undecided |

**CQ4A. Change the theme name “preparation” to “production process of e-learning”**

| Added | 22 | 29 | 14 | Undecided |

\[\text{RoA} = \text{Rate of Agreement}\]
References


12. Cook DA, Beckman TJ, Thomas KG, Thompson WG. Measuring motivational characteristics of courses: applying Keller's instructional materials motivation survey to a


A theory- and evidence-based Postgraduate Medical E-learning Development Model: nine steps to an empirical educational experience

De Leeuw RA
Scheele F
Walsh K
Westerman M

JMIR Medical Education
April 2019
Abstract

Background. Postgraduate medical digital education (PGMDE) is increasingly used and evaluated. However, evaluation focuses mainly on reaching the learning goals and little on the design. The available design models for digital education (Instructional Design (ID) models) help educators create a digital education curriculum, but none are aimed at PGMDE. Studies show the need for efficient, motivating, useful and satisfactory digital education. The present paper will enable educators to create such learning by offering an ID model founded in evidence and theory and aimed at this specific target audience.

Objectives. To create an empirical Instructional Design model for postgraduate medical digital education and to compare it with existing models used to evaluate and create PGMDE.

Methods. Previously we performed an integrative literature review, focus group discussions and a Delphi procedure to determine which building blocks for such a model would be relevant according to experts and users. This resulted in 37 relevant items. We then used those 37 items and arranged them into chronological steps. After the initial nine-step plan was created, we compared these steps with other models used in literature.

Results. The final nine steps were: 1. Describing who, why, what; 2. Selecting educational strategies; 3. Translating to the real world; 4. Choosing the technology; 5. Completing the team; 6. Planing the budget; 7. Plan the timing and timeline; 8. Implement the project and 9. Evaluate continuously. On comparing this nine-step model to other models, we found no other was as complete; nor were any of the other models aimed at PGMDE.

Conclusions. Our nine-step model is the first to be based on evidence and theory building blocks aimed at PGMDE. We have described a complete set of evidence-based steps, expanding a three-domain model of motivate, learn and apply into an ID model that can help every educator in creating efficient, motivating, useful and satisfactory PGMDE. Although certain steps are more robust and have a deeper theoretical background in current research (such as Education), others (such as Budget) have been barely touched upon and should be investigated more thoroughly in order that proper guidelines may also be provided for them.
Introduction

Medical educators have the responsibility to promote learning and create interventions and innovations to effectively help students develop proficiency in a broad spectrum of competencies [1]. One way of achieving this is by using digital education instruments, sometimes called e-learning or technology-enhanced learning. Digital education instruments can be defined as any educational intervention that is electronically mediated [2]. Some of these digital education instruments are theoretically grounded and are evidence-based [3, 4]. Studies show that digital education tools are at least as effective as other methods of training in psychomotor and non-technical skills [5] and that benefits are in unparalleled accessibility and no time or location restriction [6]. However, there is no consensus about the added value of digital education [2]. We postulate that this is part due to the fact that the focus of most studies is on the learning aim (whether the learner achieved the curriculum goals or not), whereas we believe the scope of outcomes should be broadened [7]. A recent review showed that apart from effectiveness, four other important aspects are looked at in PostGraduate Medical Digital Education (PGMDE): efficiency, motivation, usefulness and satisfaction [8]. It is obvious that digital education has to be effective as well; learners must achieve the learning goal. But when evaluating digital education, aspects apart from the learning aim should be taken into account.

The above mentioned evaluated aspects depend on the content but also on the Instructional Design (ID). In 1974, Snell becker et al. introduced the term ‘Instructional Design’ as a link between the science of how people learn and daily practice as a process for designing instruction based on empirical principles [9]. Kemp et al. describe ID as a systematic method to manage instructional process effectively so that it will ensure competent performance by students [10]. In 2002 Merrill et al. provided a very useful overview of different ID theories and models, concluding that they all share a series of first principles although no one theory or model includes all principles. Differences can be based on different theoretical insights or in the details following the first principles, depending on, for example, the target audience [11]. Several of such models are available to help experts in their quest to create, implement and evaluate a digital learning experience [12] but none is aimed at PGMDE. Most models are mainly directed towards educators, using abstract terms and theories that might not be useful for content experts with little educational experience.
Previous literature suggests that aiming an educational intervention at a specific target audience is most effective [13]. In line with this, we postulate that ID models should also be targeted as specifically as possible. This study aims at postgraduate medical professionals that work and learn after graduation. Arguments for such a specific target audience can be that adults might have different learning goals, working professionals might have specific motivational needs and medical graduates might have a unique combination of clinic work and learning by doing [14-16]. With this study we aim to provide a stepwise ID model for anyone who plans to create PGMDE, to help them cover all important steps based on theory and current evidence.

**Method**

Intervention mapping is a process for developing theory- and evidence-based health education programs [17]. Analogous to the method of this model, we used our previous work to determine quality indicators, describe a working model and compare that model to other ID models available.

**Quality Indicators**

To create a specific ID model, we started in 2016 with an integrative literature review to evaluate which indicators, determining quality in PGMDE, were already available [18]. This review was performed across a series of databases (Pubmed, IS/Web of Knowledge, Cinahl, Psychinfo and Eric) with 11,093 titles being reviewed. Ultimately, 36 relevant articles were used to gather 72 specifications which were found to be important for PGMDE. These specifications were divided into six domains, partly on the basis of the ISO-19796 standard [19]. We called this the Postgraduate Medical Elearning Model (ME-model). These domains were preparation, software design and system specifications, communication, content, assessment and maintenance.

In 2017, we discussed these 72 specifications in a series of focus group discussions with the most important stakeholders: medical education experts, postgraduate users and commercial digital education creators [20]. The aim was to select which items were most relevant and which items experts and users would add to the list. The template analysis of these interviews provided us with six domains (preparation, motivators, barriers, learning enhancers, learning discouragers and real-world translators) and 57 items. These domains gave us important insight into the main principles of PGMDE. This led to three main themes, namely motivate, learn and apply.

To determine an international consensus on the 57 items from the focus group discussion, we performed a Delphi study in 2018 [21], aiming to identify an empirically founded set of quality indicators for PGMDE. We asked a group of 13 international medical digital education experts and 10 experienced postgraduate users
to rate the 57 items, explain why they would include or exclude the items and add new items. After the first round 20 items did not achieve consensus and 15 were added. After two rounds, the Delphi produced a list of 37 indicators that we thereafter used as the basis for an ID Model. For more details about the different consensus rounds, we refer to the previously published Delhi [21].

The working model

The above-mentioned studies provided us with three themes, six domains and 37 indicators. We then used our previous experience with creating PGMDE (for example in gynaecological ultrasound [22]) to order the items chronologically. The aim was to order them in such a way that authors can follow the steps of the model without having to go back and forth in the creation process too often. The decisions in step 1 should be reflected in step 2, not the other way around.

Comparing the model

The working model had to have two further characteristics: it had to add value to already existing models and it had to be as complete as possible. To determine the added value and to find possible missing steps, we compared the working model with seven other ID models. We chose these models because an earlier systematic review showed, that only these have been used in the evaluation or description of PGMDE [8]. The models to which we compared the steps are Kern’s six steps of curriculum development; the 4C/ID Cognitive Load principle; the ADDIE model (Analysis, Design, Development, Implementation and Evaluation); Gagne’s nine events of instruction; the ASSURE Model by Heinrich and Molenda; Merrill’s principles of instruction; and the Kemp ID model.
Results

Three stages and nine steps can be followed in chronological order to ensure that all 37 items are thought through and, when applicable, used for the creation of the digital education. All the items from these previous studies are mentioned in table 1 with the corresponding stages and steps. Stage one is motivate, learn and apply; stage two is organization; and stage three is planning. Each of these stages has been investigated, the steps explained and the original items in each listed. All steps have been summarised in Figure 1.

Table 1. Stages, Steps and principles of a postgraduate medical e-learning instructional design model.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Step</th>
<th>Principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prepare</td>
<td>1. Describing who, why, what</td>
<td>1. know your target audience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. create a feeling of importance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. convey a feeling of responsibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. take your user seriously</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. do not stress your user</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. do not force your user</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7. define goals and objectives</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8. inform the user about the goals and objectives</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9. provide an overview of all lessons to be learned</td>
</tr>
<tr>
<td>2. Selecting</td>
<td>10. feedback</td>
<td></td>
</tr>
<tr>
<td>educational</td>
<td></td>
<td></td>
</tr>
<tr>
<td>strategies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Translating</td>
<td>11. interactive elements</td>
<td></td>
</tr>
<tr>
<td>to the real-world</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Organize</td>
<td>12. summaries</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13. assessments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14. real-world translation of the content</td>
<td></td>
</tr>
<tr>
<td>2. Choosing</td>
<td>15. ease of navigation</td>
<td></td>
</tr>
<tr>
<td>the technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16. design a clear layout</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17. do not distract</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18. make adaptive content</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>choose a flexible platform</td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>easily accessible</td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>make it safe and secure</td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>have fast use and loading times</td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>allow non-linear learning</td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>personalise the learning path</td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>show progress</td>
<td></td>
</tr>
<tr>
<td>26.</td>
<td>select a learning environment</td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>inform the user about optimal use</td>
<td></td>
</tr>
<tr>
<td>28.</td>
<td>provide technical support</td>
<td></td>
</tr>
<tr>
<td>29.</td>
<td>content expert, medical educator and IT expert</td>
<td></td>
</tr>
<tr>
<td>30.</td>
<td>prevent concern about the quality</td>
<td></td>
</tr>
<tr>
<td>31.</td>
<td>identify the authors</td>
<td></td>
</tr>
<tr>
<td>32.</td>
<td>provide references and sources</td>
<td></td>
</tr>
<tr>
<td>33.</td>
<td>plan your budget</td>
<td></td>
</tr>
<tr>
<td>34.</td>
<td>create a timeline</td>
<td></td>
</tr>
<tr>
<td>35.</td>
<td>maintain</td>
<td></td>
</tr>
<tr>
<td>36.</td>
<td>update regularly</td>
<td></td>
</tr>
<tr>
<td>37.</td>
<td>evaluate</td>
<td></td>
</tr>
</tbody>
</table>
**Stage 1 - Prepare**

One of many incentives may have pushed the creator to make the digital education, for example, having been asked to do so by management, but it may also be the result of an internal motivation to share something or due to many other reasons. However, once the need is present, the first step should be to determine the aim of the digital education, how it will educate and its use for the learner. We called these domains motivate, learn and apply.

**Step 1 – Describing who, why and what**

The first step is to determine *who, why and what*, which has a direct relation with motivating the user. The *who*, or the target audience, must be defined as narrowly as possible. The more specific the definition, the better the content can be adapted. The first thing to realize is that the target audience is a digital learner who is not merely a consumer of technology, but who should realize the possibilities and potentials of digital technology and recognize the opportunity that it presents within their daily life [23]. Learner characteristics that can be used in the design should be taken into account, for example online experience, age, cultural and social context and educational culture [24]. It should, however, be kept in mind that the most important user factor is previous or existing knowledge, as this can then be properly built on [13].

When the target user has been identified, it is necessary to consider and communicate the *why*. This can be done by creating a feeling of importance for that user. When your user believes that undertaking the digital education is important, s/he will be much more determined to do so. Attributing importance also helps to convey a feeling of responsibility for not just starting but completing the digital education. These messages may be communicated when the digital education is introduced or when people are invited to take part in it. Knowing your target audience will also help to prevent discouragement. Users can be discouraged by not being taken seriously (#4) (for example by childish syntax or drawings), or by being stressed (for example by tight deadlines) or forced (for example by being obliged to do something they consider not to be useful).

Creators must then carefully consider the *what*, that is, the goal and objectives of the digital education (#7). Goals are broad or general and inform users about the aim of the whole curriculum or digital education. Objectives are specific and measurable and may include knowledge, skills and/or attitude/behavioural goals [25]. When a clear goal and objectives have been set, it is crucial that the user is informed of them and provided with an overview of all lessons to be learnt. This should be done at the beginning of the digital education, so the learner knows what to expect, but also during the digital education to keep up with expectations.
Step 2 – Selecting education strategies

The second step is to consider how the targeted user will learn and which learning strategies are to be used. This depends greatly on the objectives defined in step 1, above. Instruments that may help in this process, as described in previous PGMDE studies, are problem-based learning [26], cognitive load theory [27] and multimedia learning [28]. Which strategy is the most effective for which goal will long remain a matter of debate; however, a guiding strategy must be chosen. According to previous studies, four instruments help creators facilitate efficient learning: feedback, interactive elements, summaries and assessments.

Step 3 – Translating to the real world

The last step of the first stage is apply: translating the digital education to the real world. The user wants the digital education to be useful. This can be achieved by different means, but the digital education has to add something new to the user that s/he can actually use in daily practice. This therefore concerns not only the learning goal and objectives but also the examples used in the digital education. Questions to be considered are whether the feedback is written in a way that can be related to the users’ daily tasks; and whether assessments are not only for educational purpose but give results which may be used when they return to work the next day.

Stage 2 – organize

Completion of the first stage yields a good overview of the content of the digital education: whom you target, what they should learn, how they can learn it best and how the digital education is to be kept as close as possible to the daily practice of the user. The next step entails organizing whatever is deemed necessary for the process of creating this digital education. This may include the appropriate technology and a team to realise the plan, and the financial recourses necessary must also be considered.

Step 4 – choosing the technology

When stage one is complete, the creator will have an idea of the technological needs, that is, how the technology should enable the previously set goals to be achieved. This is highly dependent on stage 1, but certain factors are universal. The aim of the technology should always be to achieve the stated curriculum goal by making use of the attributes of the supporting features. These are ‘affordances’: features that provide a potential for action, while ‘constraints’ are those features that provide the structure of and guidance to those affordances [29]. Design elements must therefore always be borne in mind, such as ease of navigation and a layout which is clear, not too distracting and prevents non-adaptive content (content that does not change layout and...
design according to the device used). Decisions about the features should include consideration of a flexible platform which can be used on several devices and operating systems, be easily accessible, safe and secure, have fast use and loading times, allow non-linear learning, personalise the learning path and show progress. Finally, a learning environment must be selected and the user must be informed about the platform and the optimum device on which to access it, and technical support must be available.

**Step 5 – completing the team**

Most digital education creators will probably already be working as part of a team. However, once a proper insight has been gained into the content and the technology needed, the team may be supplemented. It should contain at least one content expert, one medical educator and one information technology expert. When the team is complete, its members must be asked to commit time and effort before the development is started. To prevent concern on the part of users about the quality of the digital education the identity of the authors should be clearly communicated alongside an explanation of their relevant expertise and source information should be provided.

**Step 6 – planning a budget**

To create any educational experience, a budget is necessary. This is determined by many factors. Little has been written about this and, to our knowledge, there is nothing specific for PGMDE. However, manhours, materials, licensing and technology are important topics to consider, and designers, editors, marketing, maintenance, evaluation, consultants and overhead costs must also be borne in mind. It is estimated that one hour of digital education costs about 100-160 hours to create, with an average of 18,750 USD in costs [30]. There are, however, ways to save on these costs, such as using free or low-cost recourses that already exist, making shorter courses that work on multiple devices, or using open source platforms and in-house faculty for the content [31].

**Stage 3 – create**

When the above two important stages have been completed, creators will know what they want, what is necessary to achieve their aims and who will help them. It is now necessary to plan the actual creation of the digital education and start considering what will be necessary upon its completion. At this stage, a realistic timeline should be drawn up and planning for the implementation and evaluation should begin.
Step 7 – plan the timing and timeline

It will be necessary to plan and create a timeline for the creation of the digital education to ensure the team meets that deadline. The timeline should not only be for the creation of the digital education but should be extended to consider its expiry date and the communication of that to the user, as well as the intervals at which the digital education is to be maintained and updated. These are important subjects to consider at this stage: they might force a reconsideration of the budget, and communicating these dates and planned update logs to your learners is highly recommended.

Step 8 – implement the project

Implementation can take place on several levels, but a minimum of two things must be determined. First, which factors are required for the digital education to be implemented in the existing curriculum (for example, how the learners will be invited, whether management will offer support whether any sort of marketing is necessary or there will be a public introduction). Secondly, whether enough has been done to help learners implement their newly-learned lessons in practice. (This has an overlap with real-world translation, but it is worth reconsidering how a user will actually use the digital education). It can be considered that this is the same as other change management strategies or innovation implementation methods.

Step 9 – evaluate continuously

The final step is to evaluate (#37) and implement the plan. A recent systematic review shows that PGMDE is mainly evaluated in terms of educational objective rather than design. In this review, only 4% of PGMDE studies used any form of evaluation of the curriculum design [8]. An evaluation strategy should be planned to answer the questions of what is desired, what must be evaluated and what will be done with the resulting information, given that one part of the evaluation should be evaluating the implementation strategy itself.
Existing Instructional Design models

Comparing the above nine steps to the above described other models, we found that the nine-step plan covers all the steps in other models, but that no other model covers all these steps. Table 1 contains an overview of the steps comparing the models. It shows how many items the models score per step, to see which item is scored, please see Appendix A.

*Table 2 – comparing ID models*

<table>
<thead>
<tr>
<th></th>
<th>Stage 1: prepare</th>
<th>Stage 2: organize</th>
<th>Stage 3: create</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nine-step model</td>
<td>5/5</td>
<td>3/3</td>
<td>4/4</td>
</tr>
<tr>
<td>Kern</td>
<td>5/5</td>
<td>0/3</td>
<td>2/4</td>
</tr>
<tr>
<td>4C/ID</td>
<td>3/5</td>
<td>0/3</td>
<td>0/3</td>
</tr>
<tr>
<td>ADDIE</td>
<td>5/5</td>
<td>2/3</td>
<td>3/4</td>
</tr>
<tr>
<td>Gagne</td>
<td>5/5</td>
<td>1/3</td>
<td>0/4</td>
</tr>
<tr>
<td>ASSURE</td>
<td>5/5</td>
<td>1/3</td>
<td>2/4</td>
</tr>
<tr>
<td>Merrill</td>
<td>2/5</td>
<td>1/3</td>
<td>0/4</td>
</tr>
<tr>
<td>Kemp</td>
<td>4/5</td>
<td>1/3</td>
<td>1/4</td>
</tr>
</tbody>
</table>

Kern’s six steps of curriculum development was described for the first time 17 years ago and aimed at curriculum developers responsible for the educational experience of students, residents, fellows and faculty [25]. The six steps cover most of this nine-step model (see Table 1), but Kern’s programme is not aimed at digital education. Therefore, there is little to no information on topics such as technology, budgets, updating and the team required for digital education.

The four-component instructional design model (4C/ID-model) is initiated in 1992 and aims to prescribe how to develop educational programs, which contain a mix of educational media including texts, images, speech, manipulative materials, and networked systems [27]. The 4C/ID Cognitive Load principle builds upon models of human memory and can be used to design training programmes for complex learning. The focus of this model is therefore on learning aspects and how to make learning as efficient as possible. The model does not focus on any of the other domains.
The ADDIE model was originally created to evaluate software and was first published in 1988 by Grafinger [32]. As a more generic software development model, it relates closely to the nine-step model. The five steps of the ADDIE model can be split up into smaller steps and the only thing left unconsidered by the ADDIE model is budget. Moreover, even though the design step considers educational strategies, the focus is much more on technology then learning and therefore misses domains as budget and maintain.

Gagne’s nine events of instruction came to life in its first form in 1992. It is a very complete model for learning, taking into account several learning theories, the ADDIE model, Keller’s ARSC model and evaluation instructions [33]. Although the ADDIE model refers to evaluation, the nine events of Gagne do not. Neither does the Gagne model discuss implementation, updates, team and budget.

The ASSURE model developed by Heinich, Molenda, Russell, Smaldino in 1999 and is an instructional model for planning a lesson and the technology that will enhance it [12, 34]. It consists of six steps aiming to produce more effective learning and teaching. Although the design step does consider technology, it is not aimed at digital education, with all its technological challenges. Steps such as budget, timeline and team are not included in the ASSURE model.

The first principles of instruction by Merrill is a series of five principles common to various theories aiming to promote learning [11] from 2002. The five principles focus on the learning domain almost exclusively, although technology can be considered to be covered by the demonstration principle. Domains as learning goals and educational strategies are not mentioned.

The Kemp Design Model from 2007 is the result of a number of different disciplines in ID [10]. It is distinguished by its circular approach, allowing a continuous evaluation of all steps which is more dynamic and fluid than that taken by models with a linear approach. Although it covers behavioural and cognitive approaches, it does not cover real-world translation nor technology-related domains such as budget, team, timeline, updates or implementation.
Discussion

The Postgraduate MED-model is the first ID model for PGMDE. Compared to other models it is unique in two ways. Firstly, it is based on 37 building blocks, which are evidence-based items based on three empirical studies and the collaboration of experts and experienced users. While most other models are combinations of theories and expert opinion, the nine-step model presented here combines theory with literature, expert opinions and consensus. Secondly, it is the only model that covers a wide range of steps aimed directly at digital education and postgraduate education. It can be debated whether such a model may also be used for other kinds of target audiences. We aimed to make the stages and step broad, but the 37 used indicators are quite specific. If these indicators are also applicable and which might be missing for other audiences like graduates, is not yet investigated. The broad subjects of this model on the other hand, make it very suitable for content experts with little experience of creating a curriculum. Educators may find many steps to be obvious. Even so, the aim is to cause debate within the team about each step. There might not be an optimal educational strategy for each scenario, but the use of cognitive load theory (CLT) and multimedia learning as a result from that seems useful in daily practise [27, 35, 36]. We believe that the benefit of these models is not only in the sound theory behind them, but also because they are specific enough to provide easy to follow instructional principles. Following these principles, the four mentioned instruments appear promising: feedback, interactive elements, summaries and assessments. According to the CLT, learning is done when the information in chunked, which is done in the feedback, assessment and summaries. Another way is repetition, which can be found again in feedback, summaries and assessments. Using the information actively, helps to get the information into long term memory according to CLT. This is done by using feedback, interactive elements and assessments. Therefore, these instruments do not only seem effective in literature [37, 38], but also grounded in theory.

Another promising aspect of digital education are adaptive learning environments. Unique to digital learning is that each individual can have an experienced based on her/his own needs and wished, a form of individual learning without the time and costs of one-on-one human tutoring. Digital learning allows a more intelligent system to interpret the learners previous use. It can then adapt content, non-linear learning paths, multimedia and tools to a personalized learning experience. Current studies show in increasing interest in the added value of adaptive learning environments [39, 40].
Other reviews have shown the added value of creating a curriculum with the help of learning and designing models [7]. It is clear that the planning of an educational experience is far from simply adding some online presentations and that the lack of ID leads to unanticipated and unexplained learning outcomes. Educational theory can be used to create the ID to develop effective, appealing, consistent and reliable instruction [41]. The structure of a model like this also helps to identify those points which are efficient and those that require improvement [42].

Limitations

The biggest limitation of the Postgraduate MED-model will be the ways in which an educator can interpret each step. A model like this insinuates that a curriculum may be designed by simply following a few steps. However, the whole is much more complex, and each step is worth a great deal of thought, consideration, and awareness of other theories and models. Much can be said to focus a model on a specific part of a learning experience, for example pedagogic theory. Yet we wanted to provide an overview, so educators might realize how complex digital learning is and should be. This model may be considered different from other models perhaps because the people making the latter wanted an in-depth focus on a certain subject, rather than trying to create an all-in-one solution. We do not believe this nine-step model is such a solution, although danger lies in oversimplification.

Further research

Having an overview of these nine steps reveals the gaps in the literature. While many theories and studies have been performed on the effectivity of learning [1], almost nothing is known of other subjects. Our insight into the budgets needed or expected to create digital education is rarely described. More should be written on the experience of others, for example the number of hours taken, the main costs and the personnel or team chosen to limit these. Little is also known of the ways to properly evaluate the design. Most models tell users to evaluate, but there are no validated evaluation instruments that look at the design. The same is true for implementation. We should consider how to implement the digital education into the working life of the learners, but little is known of how to do that and what may be used as outcomes for successful implementation. Implementation of digital education has an analogue with implementing innovations. There are models for the implementation of innovations like Rogers’ model of diffusion of innovation and Kotter’s eight step model [43, 44]. To our knowledge, these models have not been used for the implementation of digital education, but it seems a very interesting future research path.
Conclusion

To conclude, we believe we have described a complete set of evidence-based steps, expanding a three-domain model of motivate, learn and apply into an ID model that can help every educator in creating efficient, motivating, useful, and satisfactory PGMDE. The Postgraduate MED-model ris underpinned by aspects derived from other dominant models and should provide enough basics to start the journey of creating digital education. Much remains to be learnt, and the next most logical step would be the validation of an evaluation instrument of the digital education design.
Figure 1 – The Postgraduate MED-model

Stage 1: prepare

1. Describing who, why, what
2. Selecting educational strategies
3. Translating to the real-world

Stage 2: organize

4. Choosing the technology
5. Completing the team
6. Planning the budget

Stage 1: create

7. Plan the timing and timeline
8. Implement the project
9. Evaluate continuously
References


30. How much does it cost to create an online course [https://raccoongang.com/blog/how-much-does-it-cost-create-online-course/]


34. ASSURE MODEL [http://ed205.net/assure_model.html]


Influencing categories in the implementation of postgraduate medical e-learning: a thematic analysis

de Leeuw RA
Logger DN
Westerman M
Bretschneider J
Plomp M
Scheele F

BMC Medical Education
under review
Abstract

Background: Postgraduate medical e-learning (PGMeL) is being progressively used and evaluated. Its impact continues to grow, yet there are barriers to its implementation. Although more attention is now being paid to quality evaluation models, little has been written about the implementation of PGMeL. This study aims to determine categories and define themes influencing the successful implementation of PGMeL.

Method: We performed 10 semi-structured interviews with experienced e-learning creators, after which we carried out a thematic analysis to name and describe categories and themes.

Results: Although this was not the objective of the study, the participants stressed the importance of a definition of success. Three themes were identified containing eleven categories that influence successful implementation. We named these themes creator-, organization- and learner-dependent categories. The creator dependent categories are: the learning aim, pedagogical strategies, content expertise, evaluation and the creator’s motivational path. The organization dependent categories are management support, recourse and culture. Finally the learner dependent categories are technology, motivators/barriers and value.

Conclusion: This study shows that creator-, organization- and learner-dependent categories which can be taken into account during the creating of the PGMeL. Although creator- and learner-dependent categories are mentioned in other studies, the present study also stresses the importance of organization-dependent categories. Innovation implementation theories such as Rogers’ diffusion of innovation or Kotter’s eight steps of change management show a great overlap with these categories. Future studies can both evaluate the use of these innovation models in creating PGMeL and assess the effect of the organizational categories in greater depth.
Background

During the last decades electronic learning (e-learning) has been increasingly used in health professional education, and the number of e-learnings has rapidly increased (4, 5). Many institutions are recognizing the power and benefits of internet-based learning, and they endorse learner-centred and personalized forms of learning (4, 6). E-learning has been shown to enhance educational experience, support development, be time convenient, overcome travelling difficulties and offer flexibility (7). The increased interest can be seen throughout the continuum of medical education, specifically in postgraduate medical education (6).

Despite these benefits, studies show that barriers to the implementation of postgraduate medical e-learning (PGMeL) remain (6, 8). Implementation is described as an important step in the process of using PGMeL efficiently (9, 10). In 2001 a factor-analytic study provided a list of barriers (11) which was supported by an integrative review from 2018 (12). These reveal barriers such as time constraints, poor technical skills and inadequate infrastructure. Yet, no studies have provided categories which influence the successful implementation of PGMeL. Although certain studies address evaluation and implementation as a whole, none goes into the details of the implementation (13).

It can be argued that implementing a new innovation (in this case, an e-learning) implies change in an organization (14). Two theories on change and innovation have been used successfully to facilitate the implementation of technology in health care: Rogers’ innovation diffusion theory and Kotter’s change management model (14-16). While Rogers focusses on innovations as entities with attributes and a user decision process, Kotter describes an eight step change management process (17) (18). Both models could be useful in the implementation of PGMeL, although there seems to be no literature describing their use in this regard.

Currently, the literature emphasizes the importance of implementation, and lists barriers to this. Some innovation implementation models are available, but only from other fields of expertise. The question thus arises of which categories influence the successful implementation of PGMeL and how such categories compare to these implementation models.

The definitions used for this study are summarized in box 1.
METHOD

Study design

We performed a series of semi-structured interviews with experienced e-learning creators, after which we performed a thematic analysis in order to generate an answer to our question and categories of conditions. The semi-structured format is the most frequently used technique in qualitative healthcare studies. One of its main advantages is the interaction between interviewer and participant, which allows deeper insight into possible constructs (19). See figure 1 for the flowchart of the applied methodology.

Box 1 - Definitions

**Implementation**: the act of carrying an intention into effect, which in health research can be policies, programmes or individual practices (1) or, specifically for PGMcL, the creation, management and delivery of learning content.

**E-learning**: we used Sangrà’s definition after a 2012 Delphi: “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 knowledge, skills and/or behaviour/attitude” (2).

**Creator(s)**: the person/people who created the e-learning. This is usually a group of people with their own speciality (content, IT, education) (3).

**User**: the person who uses the e-learning to learn from it. (The term “learner” is also used.)
Legend for figure 1. The methodology for this study consists of two phases: data collection by semi-structured interviews and thematic analysis by transcribing the data, categorize categories and define themes.

Study participants

We invited experienced creators of PGMeL to participate in our study. We selected these creators from our professional network and through the national Dutch e-learning task group of the Dutch Association of Medical Education. These creators had to have experience of designing and implementing at least one PGMeL project. No financial compensation was given to any participant. The ethical board of the Dutch Association of Medical Education gave ethical consent (2018.6.3), after which all participants gave their written informed consent. The interviews were in Dutch; therefore, citations are translated into English.

Data collection method and procedure

All interviews were carried out in the Netherlands in the comfort of the participants’ own environment. The interviews lasted between 30 and 45 minutes, and were facilitated by DL and audiotaped. The interview started with a short introduction regarding the reason for the interview, followed by a few demographic questions. The interview guide (appendix 1) was developed by RL and DL and based on Jochems et al.
and Marriënboer et al (20). They grouped the implications of e-learning on pedagogy, technology and organization, which we followed in the interview guide. After each third interview the recordings were transcribed and coded. We continued interviewing experts in series of three until theoretical data saturation was reached.

Data analysis

We verbatim transcribed all interviews and performed a thematic analysis according to Braun (21). Transcribing of interviews was carried out by RL and DL. We used Atlas.ti version 8.0 for the initial coding. The reason for using thematic analysis is that it can usefully summarize key features and generate unanticipated insights, and has been proven useful for producing qualitative analysis suited to informed policy deployment (21).

To perform the data analysis in a structured method, we used the six steps proposed by Braun et al, containing:

1. Familiarizing oneself with the data
2. Generating initial codes
3. Searching for themes
4. Reviewing themes
5. Defining and naming themes and
6. Producing the report (21).

Coding and searching for themes (steps 2 and 3)

The second step was finding potential categories from the coded data. Coding by RL and DL led to 112 potential categories of which, after duplications and synonyms had been removed, 76 remained (see table 1). To search for themes, we wrote all categories on post-it notes and started arranging them on the wall. The first step was to categorize all categories that were associated with the initial themes; pedagogy, technology and organization, followed by a mind-map of all categories. This led to the final categories. The mind-map, group discussion and digital evaluation of the mind-map followed, which gave insight into the underlying constructs behind the categories which we used to define the themes (see Appendix 2).
<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>target audience</td>
<td>aim of the e-learning</td>
<td>author dependent</td>
</tr>
<tr>
<td>reaching target audience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>learning aims</td>
<td></td>
<td></td>
</tr>
<tr>
<td>knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>attitude/behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inefficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>means to reach goal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>learning methods</td>
<td>pedagogical strategies</td>
<td></td>
</tr>
<tr>
<td>separate in sections</td>
<td></td>
<td></td>
</tr>
<tr>
<td>levels of knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>interactivity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>communicate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>type of content</td>
<td>content knowledge</td>
<td></td>
</tr>
<tr>
<td>future evaluation</td>
<td>evaluation</td>
<td></td>
</tr>
<tr>
<td>improvement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>maintenance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reserve resources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>technology centered</td>
<td>motivational pathway</td>
<td></td>
</tr>
<tr>
<td>innovation driven</td>
<td></td>
<td></td>
</tr>
<tr>
<td>learner centered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>author centered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>needs driven</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Category</td>
<td>Domain</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>create commitment</td>
<td>management support</td>
<td>organization</td>
</tr>
<tr>
<td>planning</td>
<td></td>
<td>dependent</td>
</tr>
<tr>
<td>responsibilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>insufficient work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>deadlines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>manage expectations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>transparency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>integrate into daily work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>project leaders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>motivate authors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>provide support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reserve or provide time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>promote project</td>
<td></td>
<td></td>
</tr>
<tr>
<td>involve stakeholders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>include users</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICT</td>
<td>resources</td>
<td></td>
</tr>
<tr>
<td>team</td>
<td></td>
<td></td>
</tr>
<tr>
<td>knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ongoing budget</td>
<td></td>
<td></td>
</tr>
<tr>
<td>technology budget</td>
<td></td>
<td></td>
</tr>
<tr>
<td>evaluation budget</td>
<td></td>
<td></td>
</tr>
<tr>
<td>management culture</td>
<td>culture</td>
<td></td>
</tr>
<tr>
<td>user culture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>centralize</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fragmentize</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lack of structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Category</td>
<td>Domain</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>structure</td>
<td>technology</td>
<td>learner dependent</td>
</tr>
<tr>
<td>design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>infrastructure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>availability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>device variety</td>
<td></td>
<td></td>
</tr>
<tr>
<td>flexibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>layout</td>
<td></td>
<td></td>
</tr>
<tr>
<td>navigation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>attractive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>intiutive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>user friendly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>motivate team</td>
<td>motivators and barriers</td>
<td></td>
</tr>
<tr>
<td>motivate management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>motivate users</td>
<td></td>
<td></td>
</tr>
<tr>
<td>not doing it</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demotivating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ineffectiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>promote</td>
<td></td>
<td></td>
</tr>
<tr>
<td>external motivation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>duraton</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time efficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>added value</td>
<td>value</td>
<td></td>
</tr>
<tr>
<td>attractiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>efficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>satisfaction</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend: for figure 1: By combing the original codes, the codes column in this table show key words associated with those codes. It demonstrates the content associated with the categories and themes.
Results

In the period May 2018 to August 2018 we invited a total of 14 expert creators. Four creators rejected the invitation due to lack of time. Ten interviews followed, after which data saturation was reached, and no further interviews were performed. On average the experts had eight years of experience (with an interval between three and twenty-five years). We interviewed six men and four women: six content experts and four with a management task concerning e-learning, of whom five were involved in research into different categories of e-learning. Participants were working in five different University Hospitals in the Netherlands. Step one of the thematic analysis, familiarizing oneself with the data, was performed during the transcribing. Citations from the interviewees are marked as “IT”, followed by a number corresponding with that person.

Reviewing, defining and naming the themes (step 4 and 5)

From the data, we extracted eleven categories that were divided into three themes. Due to the blurred boundaries between the discussed subjects, some categories can fall within more than one theme. These categories were placed in the theme in which they, in theory, have the most influence. See figure 2 for all categories and themes.

The themes were named:

1. Creator-dependent categories
2. Organization-dependent categories
3. Learner-dependent categories
Defining success

Although this was not the objective of this study, the participants stressed the importance of a definition of successful implementation of PGMeL. During the analysis, categories arose that concerned a possible aim for success. There were four aspects to this success:

1. It was suggested that for an implementation to be successful, the e-learning should be carried out/experienced by the right group of users (the target audience). An e-learning created for residents but which is only carried out by interns is not properly implemented.

2. The e-learning should have defined learning goals, but users should be tested to see if they actually achieved those goals.

3. The users should have a satisfactory experience. The main importance of satisfaction is to provide a longer lasting learning experience and motivation for future e-learning experiences.

4. The e-learning needs continuity to maintain relevance and function as reference material for your users. To maintain relevance, the content should be updated if needed, and the technology remain accessible. An alternative is to provide a clear expiry date for the e-learning.
We defined this theme as *categories that the creators can largely influence*, in other words subjects that the creator team can discuss and make final decisions about. There were five categories within this theme:

1. **Learning aim**: what does the creator aim for the user to learn? Learning aims were grouped into knowledge, skills and attitude/behaviour.

2. **Pedagogical strategies**: the learning instruments that are chosen to achieve the learning aim. There are several ways to achieve the learning aim and there is an added value to choosing a pedagogical strategy: “using and consulting educational experts are important to reach your goals” IT6. Certain strategies seem less relevant, although several items of these strategies were mentioned, including interactivity, feedback and segmentation: “e-learning needs a sound and responsible pedagogical method, although there might be many of these” IT3.

3. **Content expert**: having at least one person with in-depth knowledge of the subject of the e-learning. Having a content expert is reported as very relevant to implementation. Having expert content knowledge available at all stages of the design and, for example, during meetings helps in discussion of relevance and importance. Moreover, although content experts can be the driving force behind the e-learning, creating PGMeL is a team effort and delegating tasks can improve efficiency: “it can help for a content expert to draft the major lines, so others can continue working on it to create an e-learning” IT5.

4. **Evaluation**: a part of the success of implementation lies in the continuity of the e-learning. Evaluation is needed to continue improving the e-learning, but also to collect alarm signals for bugs, outdated sources and lack of resources for the future.

5. **Creators motivational pathways**: the interviewees described the different ways in which they thought important decisions around creating PGMeL were initiated and motivated. We called this the “creators motivational path”. They described three different perspectives: e-learning can be technology-, learner- and creator-centred. Centredness means why the e-learning was created in the first place and what motivated the creators to make the e-learning. Technology-centred e-learning aims to use and evaluate a new technology which a creator feels should be used for teaching: “IT departments innovate faster (than education departments) and technology enhanced learning can anticipate on educational needs” IT7. Learner-centred e-learning aims to fulfil the needs of a learner, by means of technology. A potential student, or groups of students, have the need to learn...
something and a content expert searches for the right tool to achieve that goal. Creator-centred e-learning aims to fulfil the need of a creator to spread something new, or educate a certain group. This creator will then make the e-learning, after which she/he will start sharing it and search for an audience: “Authors can feel this urge (to create PGMeL), which can lead to new creations” IT2.

All three forms have their own pitfalls and possibilities (see table 2) to be considered when implementing PGMeL.

**Table 2** – pitfalls and possibilities of the motivational paths of the creators

<table>
<thead>
<tr>
<th>group</th>
<th>pitfalls</th>
<th>chances</th>
</tr>
</thead>
<tbody>
<tr>
<td>technology-centred</td>
<td>no clear learning aims</td>
<td>evaluation of new technology</td>
</tr>
<tr>
<td></td>
<td>risk of evaluation without control</td>
<td>development of next technologies</td>
</tr>
<tr>
<td></td>
<td>group</td>
<td></td>
</tr>
<tr>
<td></td>
<td>no initial user or creator support</td>
<td>investment possibilities</td>
</tr>
<tr>
<td>learner-centred</td>
<td>possible lack of innovation</td>
<td>needs come from the learner</td>
</tr>
<tr>
<td></td>
<td>need to motivate creators</td>
<td>need for improvement of available tools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>available learner commitment</td>
</tr>
<tr>
<td>creator-centred</td>
<td>risk of no learner need</td>
<td>motivated creators</td>
</tr>
<tr>
<td></td>
<td>possible lack of innovation</td>
<td>new knowledge shared</td>
</tr>
</tbody>
</table>

Legend table 2: a motivational path is the reason which motivated the e-learning creators to make the e-learning. They can be motivated by a new technology that they want to test (technology-centered), because there is a need from a learners perspective (learner-centered) or from the need of the creators themselves to share or educate a lesson (creator-centered).
Theme 2: organization-dependent categories

We defined this theme as categories largely influenced by the organization where the e-learning is being used. This theme contains three categories:

1. **Management support**: the organization contains many management aspects such as planning, having project leaders, receiving commitment from higher management, transparency and managing expectations. However, organization also means involving stakeholders, and knowing and reaching your target audience. Support from all these parts of the organization is preferred.

2. **Resources**: several topics were discussed around resources. There are non-financial resources, such as an ICT department, team support and knowledge support, and there are financial resources, which emphasize an ongoing budget for maintenance, a technology budget and an evaluation budget.

3. **Culture**: implementation also requires a certain culture, that is, a management culture to provide the needed support, but also a user culture that is willing to learn and evaluate with you.
Theme 3: learner-dependent categories

We defined the theme as *categories that are mainly influenced by the learner*, and by learner we mean the end-user of the e-learning. Three categories fall within this theme:

1. **Technology**: this was mentioned as an important aspect of the design and the infrastructure. Design elements were mentioned in the navigation, structure of the e-learning and layout. It should be attractive, intuitive and user-friendly. A proper design helps with the implementation, but the infrastructure of the e-learning is just as important: “it has to be as user-friendly as possible; nobody likes a socket that does not fit a plug” IT8. By infrastructure we mean the way the e-learning is made available, device compatibility and user flexibility (availability anytime, anywhere on any device, adjustable to the user’s personal needs). The technology is usually chosen and used by the creator and depends on the needed affordances. Technology also needs support from the organization, but it is the learner who experiences and uses it to learn. Therefore, we believe that the learner has the most influence on the added value of technology to the learning process.

2. **Motivators and barriers**: the success of your implementation will depend on how you can motivate: “to motivate (your students), you need to keep stimulating them so they (the students) will finish it” IT10. Even when motivation is high, however, barriers which might lead users to not start in the first place must be prevented. The organization and creator both have an influence on this category, but the learner has more.

3. **Value**: the learner must experience added value from the e-learning over other forms of education or already existing learning material. The added value should be clear for the user and can be found in attractiveness and efficiency: “it really needs additional value, compared to (for example), books” IT3.

Apart from the themes described above, the interviews provided two other insights: firstly, aspects of the definition of success; and secondly, different motivational paths that can be taken by the creators of the e-learning. The sections below explore both in further detail.
**Figure 2** - Themes and categories that influence the success of PGMeL implementation

Legend figure 2. Implementation success = (1) reach your target audience, (2) achieve learning goals, (3) create satisfaction and (4) provide continuity. The categories “Aim of the e-learning”, “Management support” and “Value” are concentrated within one theme. The other categories contain partially overlapping different themes, depending on the influence of creators, organization and learners.
Discussion

This study provides us with categories that influence the successful implementation of PGMeL. Eleven categories are described in three themes and although some may seem obvious, others may require further preparation to ensure proper implementation. The importance of considering and following a pedagogical strategy has been preached for a long time (12), although it is still frequently ignored (22). Technology, motivators and barriers and added value have previously been discussed as important aspects (5, 23, 24). Defining the aim of the e-learning and the evaluating the results, are also suggested in other studies (23). However, considering where your aim originates does provide new insight. Studies have stressed the importance of user-centred e-learning (25), yet this study shows that there are other ways to initiate one's learning aims, all of which present pitfalls and opportunities.

The series of organization-dependent categories is much less frequently discussed in the literature (24, 26, 27). Commitment is needed from an organizational standpoint (that is, whether management grants time, resources and merit). Management support will help to give substance to deadlines, expectations and agreements. If the organization advocates the product, this will also help with changing the culture. Financial resources are also an undervalued topic (27), and this is even more the case for the non-financial resources needed. This study aims to place more emphasis on these topics for future research.

Success in implementation

Defining success as reaching your target audience, achieving learning aims, satisfying your audience and maintaining continuity is a good starting point. However, success should be defined for each individual project. The most important message is that success should be defined before implementation. These four categories are a starting point only, on which future research can build.
PGMeL implementation as innovation implementation

Considering PGMeL in an organization as an innovation, it makes sense to use innovation implementation models. Rogers’ model of diffusion of innovation and Kotter’s eight step model have a significant overlap with the categories addressed in this study. In table 3, three aspects of Rogers’ model of diffusion of innovation are summarized: the decision steps, the attributes of innovation and the adopter categories. In the second column, the interpretation of these steps from the perspective of this study are added. The models start with an innovation, which is the e-learning. The creators (the source) will them communicate the existence of the e-learning through the organization (channel). The users will go through a decision process to determine if they accept the e-learning. This process has five stages and after stage three the user will decide whether to start the e-learning or reject it. The chances of acceptance depend on a series of attributes. Rogers’ model emphasizes the importance of the organization in terms of communication, time and social system. The attributes of relative advantage (value), compatibility (technology), complexity (adjustment to target audience), trialability (motivators and barriers) and observability (value and culture) are all mentioned in this study as well.
Table 3 – Interpretation of Rogers’ diffusion of innovation for PGMeL

<table>
<thead>
<tr>
<th>Step</th>
<th>Rogers stage</th>
<th>PGMeL equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Knowledge</td>
<td>inform users of existence</td>
</tr>
<tr>
<td>2</td>
<td>Persuasion</td>
<td>convince users of importance</td>
</tr>
<tr>
<td>3</td>
<td>Decision</td>
<td>start or reject the e-learning</td>
</tr>
<tr>
<td>4</td>
<td>Implementation</td>
<td>reach the target audience</td>
</tr>
<tr>
<td>5</td>
<td>Confirmation</td>
<td>evaluate the learning aims</td>
</tr>
<tr>
<td></td>
<td>Attributes of the innovation</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Relative advantage</td>
<td>added value</td>
</tr>
<tr>
<td>2</td>
<td>Compatibility</td>
<td>design should fit the organization</td>
</tr>
<tr>
<td>3</td>
<td>Complexity</td>
<td>adjust to target audience</td>
</tr>
<tr>
<td>4</td>
<td>Trialability</td>
<td>personalize learning path</td>
</tr>
<tr>
<td>5</td>
<td>Observability</td>
<td>results should be visible</td>
</tr>
<tr>
<td></td>
<td>Adopter categories</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Innovators</td>
<td>creator team</td>
</tr>
<tr>
<td>2</td>
<td>Early adopters</td>
<td>pilot group</td>
</tr>
<tr>
<td>3</td>
<td>Early majority</td>
<td>first users</td>
</tr>
<tr>
<td>4</td>
<td>Late majority</td>
<td>last users</td>
</tr>
<tr>
<td>5</td>
<td>Laggars</td>
<td>unwilling</td>
</tr>
</tbody>
</table>

Legend table 3: Rogers distinguishes three aspects of innovation implementation. The decision process of the user, the attributes of the innovation and the categories of users called adopters. The decision process contains five steps that the user will go through. Those steps can be translated into PGMeL, shown in the last column. The attributes should be linked to the created e-learning, making it more likely that the user will decide to accept the e-learning. Finally, defining the users/adopter categories might help identify barriers for each group.
As regards Kotter’s eight steps model, table 4 shows an interpretation of these steps within PGMeL implementation. Kotter distilled his principles of change into these eight manatory steps. Quinn et al used this model before to re-engage student into blended learning and found gaps, in particular to the support for students (28). The eight steps also only shows an overlap with a part of the categories from this study. Kotter is developed and aimed at corporate change, and emphasis importance to organizational efforts like establishing urgency, creating a coalition, communicate, empower others and create cultural changes. This study adds specific categories for education and placed more responsibility with the learner as well. Comparing the themes and categories addressed in this study with those in the two innovation models supports the importance of most categories and shows that neither model is as specific to PGMeL as the outcome of this study.

Table 4 – interpreting Kotter’s eight steps for PGMeL

<table>
<thead>
<tr>
<th>Step</th>
<th>Original description</th>
<th>PGMeL interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Establish a sense of urgency</td>
<td>Why you need this e-learning</td>
</tr>
<tr>
<td>2</td>
<td>Create a guiding coalition</td>
<td>Create your optimal team</td>
</tr>
<tr>
<td>3</td>
<td>Develop a change vision</td>
<td>What the e-learning will establish</td>
</tr>
<tr>
<td>4</td>
<td>Communicate that vision</td>
<td>Communicate these aims</td>
</tr>
<tr>
<td>5</td>
<td>Empower others to act</td>
<td>Empower users</td>
</tr>
<tr>
<td>6</td>
<td>Garner short-term wins</td>
<td>Define short-term wins</td>
</tr>
<tr>
<td>7</td>
<td>Never let up</td>
<td>Determine a strategy when resistance is shown</td>
</tr>
<tr>
<td>8</td>
<td>Incorporate changes into culture</td>
<td>Incorporate continuous, digital learning into culture</td>
</tr>
</tbody>
</table>

Legend table 4: Kotter’s eight steps can be interpreted for PGMeL, by rewriting them into questions or advice for the creators, shown in column three. Kotter’s model is focused more on the organizational aspects of implementation, therefore, takes little to no aspects of medical education into account.
Strengths and limitations

One of the strengths of this study is the overlap with existing literature. Many subjects have previously been described and every time a group of experts overlaps previous knowledge, the path we are following is reinforced. This study broadens our scope and retrieves far less discussed topics, such as the organizational aspects of PGMeL. A diverse group has been interviewed, whose variety of expertise enforces the backbone of this study. However, there are also limitations, the biggest of which is the bias in selection. It is impossible to say whether the balance between management, content and technological experience within the group of interviewees is correct, or over-emphasizes a single subject. Moreover, in studies like this, there is always the issue of cross-cultural generalizability, as all participants are from a Western culture and function and create PGMeL within these limits. Organizational aspects that are relevant in Western culture might be completely different in Eastern cultures. It may even be argued that every organization has its own sub-culture and for that reason also has different needs. To minimize that limitation, we included participants from different regions in the Netherlands, but the fact remain that they are all from the Netherlands.

Future research

Some of the categories addressed in the present study are already receiving due attention but much remains to be learnt about the organizational needs for creating PGMeL. Involving more stakeholders from management level might provide more insight into the limitations of an organization. Using models such as the innovation of diffusion provide insight as to where a specific e-learning might experience implementation barriers and which categories could be improved. Evaluating the use of these categories and models in practice would be of great value to keep adjusting and improving them. This study also provides a starting point for defining success.
Conclusion

This study shows that implementing PGMeL has creator-, organization- and learner-dependent categories. These categories influence the success of the implementation. To achieve successful implementation, you must reach your target audience, achieve learning aims and provide continuity. Categories that will help achieve that success are partly outside the creators’ influence, and partly within it. The organization-dependent categories, management support, resources and culture, might be those most in need of attention to achieve a well-earned implementation success, since relatively little is known and written about them. This study also shows that, in theory, using innovation and change management models could be very helpful, and certainly merits further research.
Appendix 1 – Interview Guide

Introduction

We met today to discuss the implementation of postgraduate medical e-learning together. By e-learning we mean every form of electronic / digital learning aimed at people who have to learn further alongside the working life. We strive today to explore what the goal is of the implementation of "postgraduate medical e-learning" (PGMeL) and what tools can be offered to achieve this goal. The interview will last about 30 minutes to an hour.

A. Interview background

• Can you tell us something about your current position and your experience with e-learning and its implementation?

B. The purpose of implementation

• In your view, what is the purpose of e-learning in a general sense?
• How does this goal relate to the implementation of e-learning?
• What do you understand by successful implementation of e-learning

C. Domain technology

• What is the role of technology in implementation?

D. Domain organization

• What is the role of an organization in the implementation?
• What do you think of the responsibilities of the organization and management in the implementation and what have you experienced or missed in the past?

E. Domain pedagogy

• What is the role of pedagogy in the implementation?

F. General remarks

Many thanks for answering these questions. Are there any general things that you think are of added value to include in the implementation of PGMeL?
Appendix 2 – Mindmap of codes
Ethics approval and consent to participate

The ethical board of the Dutch Association of Medical Education gave ethical consent (2018.6.3), after which all participants gave their written informed consent.

Competing interests

The authors declare that they have no competing interests

Authors’ contributions

RL, DL, MP and FS wrote the study protocol and methodology. RL and DL performed the interviews and initial coding. RL, DL and JB participated in the mindmap and thematic analysis. RL, MW and MP had a major contribution to the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We wish to thank especially the residents who took the time for our discussions. The users that filled out the many evaluation forms and the creators of the three evaluated e-learnings. Without these people providing insight in their needs and experience, qualitative studies like this would not be possible.
References

1. Peters D. Republished research: Implementation research: what it it and how do to it. 2014.


The Development and Validation of an instructional design evaluation survey for Postgraduate Medical E-learning

De Leeuw RA
Walsh K
Westerman M
Scheele F

Journal of Medical Internet Research
April 2019
Abstract

Background E-learning has taken a firm place in postgraduate medical education. Whereas 10 years ago it was promising, it now has a definite niche and is more than a hype: indeed, it is clearly here to stay. However, evaluating the effect of postgraduate medical e-learning (PGMeL) and, improving upon it, is can be complicated. While the learning aims of e-learning are evaluated, there are no instruments to evaluate the instructional design of PGMeL. Such an evaluation instrument may be developed by following the Association for Medical Education in Europe (AMEE) seven-step process. The first five steps of this process were previously performed by literature reviews, focus group discussion and an international Delphi.

Objectives This study will continue with step six and seven and answer the research question: is a content-validated PGMeL Evaluation Survey useful, understandable and of added value for creators of such an e-learning?

Method There are five steps in this study: 1) creating a survey from 37 items; 2) testing readability and question interpretation; 3) adjusting, rewriting and translating; 4) gathering filled out surveys from three PGMeLs; and 5) holding focus group discussions with the e-learning authors. Step five is carried out by presenting the results of the evaluations from step four, followed by a group discussion. There are three groups of participants in this study. Group A are experienced end-users of PGMeL, and participated in step two. Group B are users who undertook e-learning and were asked to fill out the survey in step four. Group C are the authors of the e-learnings described above.
**Results** From a list of 36 items we developed a postgraduate Medical E-learning Evaluation Survey (MEES). Seven residents participated in the step-two group discussion; four items were interpreted differently, three were not readable and two items were double. The items from step two were rewritten and, after adjustment, were understood correctly. The MEES was translated into Dutch and again pilot tested. All items were clear, and were understood correctly. The MEES version used for the evaluation contained three positive domains (motivation, learning enhancers and real-world translation) and two negative domains (barriers and learning discouragers), with thirty-six items in those domains, five Likert-scale questions of 1 to 10, and five open questions asking the participants to give their own comments in each domain. Three e-learnings were evaluated during the period July to November 2018. There was a total of 158 responses from a Dutch e-learning, a European OB/GYN e-learning and a surgical e-learning offered worldwide. Finally, three focus group discussions took place with a total of ten participants. The usefulness was much appreciated, understandability was good and added value was high. Four items needed additional explanation by the authors and a Creators’ Manual was created at their request.

**Conclusions** The MEES is the first survey to evaluate instructional design of PGMeL and is constructed following all seven steps of the AMEE. This study completes the design of the survey and shows its usefulness and added value to the authors. It finishes with a final, publicly available survey which includes a Creators’ Manual. We briefly discuss the number of responses needed and conclude that more is better; in the end, however, one has to work with what is available. The next steps would be to see whether improvement can be measured by using the MEES, and to continue to work on the end understandability in different languages and cultural groups.
Introduction

E-learning and distance education are a growing part of postgraduate and continuous medical education. The cost-effectiveness and logistical benefits have previously been shown [1] and whereas 10 years ago e-learning was promising, it is now part of mainstream medical education [2]. However, the overall effectiveness and added value of e-learning over “conventional” education such as face-to-face learning is debatable and results in the literature are diverse [3]. One of the problems in evaluating e-learning is the lack of a proper evaluation tool [4].

The effectiveness of e-learning can be separated into two parts: the effect of the learning aim, and the instructional design of the e-learning. By learning aim, we mean the ability of the e-learning to achieve the learning goals, usually either new knowledge, skills or attitude/behavior [5]. For example, the learning aim may be to tie a laparoscopic knot, which may be evaluated by an Objective Structured Assessment of Technical Skills (OSATS). By the instructional design we mean the functionalities (affordances) and their design [6]. In the case of e-learning, these are the design of the digital medium and affordances used to achieve the learning aim, for example the virtual reality program with interactivity, feedback and gamification to practice the knot tying.

Despite the methodological limitations, e-learning is often evaluated by comparing test results before and after it is used. Usually, the learning aim is used as the primary outcome and even when the design is also evaluated, this is generally done using instruments which are not aimed at PGMeL [7]. When the design is not included in the e-learning evaluation study, however, the question arises of how to ascertain whether the e-learning modus is suited to the learning aim (for example, virtual reality is not suited for learning to tie a knot) or if there were essential flaws in the e-learning itself (for example, the virtual reality box was poorly designed). We believe that the evaluation of the learning aim should always go together with the evaluation of the design, because they are interwoven in the final outcome.

To properly evaluate the design, however, we need an instrument that has proper content validation and is aimed at the right target audience, in our case PGMeL.
Development of a survey in medical education

The development of an evaluation instrument is complex and involves many steps [8]. In 2014 the Association for Medical Education in Europe (AMEE) published a seven-step design process for developing surveys for medical education [9]. Step one to five of this seven-step design, were previously published in two reviews [7, 10], focus group discussions [11] and an international Delphi [4] (see figure 1). The aim of this study is to proceed with step six and seven and to evaluate the results of the survey with the creators of a PGMeL. We want to know if the creators find the results helpful to improve the e-learning (usefulness), if they can understand the indicators that are used in the context of instructional design (understandability) and finally if the survey is offering them additional information over already existing evaluation methods (added value). This leads to the following research question that this study will try to answer: is a content-validated evaluation survey for PGMeL useful, understandable and of added value for the creators of PGMeL?

Legend figure 1:

1. The systematic review showed that the most used primary outcomes were satisfaction, motivation, efficiency and usefulness, and thirteen evaluation models were used [7].

2. The narrative review described 72 unique specifications which were consolidated and re-organized into a six-domain model, called the PGMeL Model [10].

3. Focus group discussions with experts and users provided a template of six domains, containing 57 relevant items used for step 5 [11].

4. Further analysis of the items from the literature, focus groups and domains led to three overall defining subjects: motivate, learn and apply [11].

5. A group of 13 international experts and 10 experienced end-users participated in a Delphi. This Delphi resulted in reducing the 57 items to 37 important and relevant items that should form the basis of the evaluation survey [4].

Step 6 and 7 are topic of this study.
Figure 1 – AMEE seven-step design process for developing surveys for medical education

<table>
<thead>
<tr>
<th>AMEE step description</th>
<th>Method of previously published steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1. Conduct a literature review</td>
<td>Literature review 1 (1) Which outcomes and evaluation methods are used to evaluate PGMeL?</td>
</tr>
<tr>
<td>Step 2. Conduct interviews and/or focus groups</td>
<td>Literature review 2 (2) which items are described to increase PGMeL quality?</td>
</tr>
<tr>
<td>Step 3. Synthesize the literature review and focus groups</td>
<td>Postgraduate Medical Elearning Model (2)</td>
</tr>
<tr>
<td>Step 4. Develop items</td>
<td>Focus group discussions (3) Which items are missing?</td>
</tr>
<tr>
<td>Step 5. Conduct expert validation</td>
<td>Motivate, Learn and Apply Model (4)</td>
</tr>
<tr>
<td>Step 6. Conduct cognitive interviews</td>
<td>57 items extracted from literature and focus groups (3)</td>
</tr>
<tr>
<td>Step 7. Conduct pilot testing</td>
<td>International Delphi with PGMeL experts: reduction to 37 relevant items (5)</td>
</tr>
<tr>
<td></td>
<td>Rewriting 37 items to a survey, taking into account comments from the Delphi: this study</td>
</tr>
<tr>
<td></td>
<td>Pilot test readability and understanding with PGMeL users: this study</td>
</tr>
<tr>
<td></td>
<td>Pilot test questionnaire and validate value with e-learning authors: this study</td>
</tr>
</tbody>
</table>
Method

Study design

To conclude the content validation this study collected evidence of response process validity to assess how participants interpreted the items (AMEE step 6), and to conduct pilot testing (AMEE step 7). To answer the research question, this study has five steps (see figure 2).

1. Create a survey draft based on 37 items of the previous Delphi [4] and address three concerns of experts in this Delphi. Firstly, the term e-learning can be confusing; secondly, the added value of another survey might be limited; and thirdly, the indicators may be too general and/or not specific enough for the evaluated e-learning [4]. We called the survey the postgraduate Medical E-learning Evaluation Survey (MEES).

2. Determine readability and understandability with experienced post-graduates by use of a focus group discussion.

3. Adjust the survey draft accordingly to the feedback of step two. The English survey was also needed in Dutch (for step four); therefore, the rewritten English survey was translated and again pilot tested for readability. Because less discussion was expected, only two (native Dutch) residents were asked to read the Dutch survey and provide feedback.

4. Use the survey, by evaluating three PGMeL modules. Contacts of several European PGMeL groups were emailed and asked to participate and the first three agreed. They were sent the participant information and the MEES. After agreeing to use the survey, they were asked to add it to the standard evaluation survey they might already have. The users were voluntarily asked to participate. The anonymous results of the surveys were sent to RDL.

5. Focus group discussion with the creators of those e-learnings about the survey results was performed. The results were presented per domain. For each domain the minimum, maximum and average score, all items and free text comments were discussed (see appendix 2). Finally, a strength-weakness analysis with scores, summary of recognized items and summary of free texts per domain was carried out. The discussion guide and short demographics questionnaire are in appendix 1.

The main methodology will be focus group discussions. We chose focus group discussions because they are an appropriate method to investigate attitudes and beliefs, and to generate new ideas [12].
Figure 2 – five steps to address step 6 and 7 of the AMEE design process and evaluate the usefulness, understandability and added value of the MEES

<table>
<thead>
<tr>
<th>Who (6)</th>
<th>Step aim</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phase A (1) Create MEES draft version</td>
<td>37 items from Delphi (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group A</td>
<td>Phase B (2) Readability and interpretation</td>
<td>Focus group with 8 residents</td>
</tr>
<tr>
<td>Group A/B</td>
<td>Phase C (3) Adjust, rewrite and translate</td>
<td>Incorporate results step 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group C</td>
<td>Phase D (4) Gather filled out surveys</td>
<td>Select 3 e-learning groups</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group D</td>
<td>Phase E (5) Evaluate results of MEES</td>
<td>3 focus group discussions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Formulate MEES final version</td>
<td>Incorporate results step 5</td>
</tr>
</tbody>
</table>
Legend figure 2:

1. **Step 1.** First draft of the MEES was based on the previous 37 items and three concerns from previous studies about a PGMeL survey like this.

2. **Step 2.** Only when there was consensus about the question or item did the discussion continue. The pilot test was not recorded to ensure anonymity and because further analysis was not needed.

3. **Step 3.** The items that were vague, poorly written or wrongly interpreted were rewritten during the discussion and the survey was translated to Dutch.

4. **Step 4.** There was a set period of 4 months in which the users (group A) were asked to fill out the MEES after completing the e-learning.

5. **Step 5.** The discussions were planned with group C for the evaluation of the results with the predefined themes.

6. Three groups are consulted for this study. Group A are experienced end-users. Group A+ is a subgroup of group A. Group B are users from three evaluated e-learnings and group C are the creators of those e-learnings.
Study participants

There are three groups of participants in this study. Group A are experienced end-users of PGMeL and participated in phase 1. These were OB/GYN residents in their fourth and fifth year at the Amsterdam University Medical Centre. A subgroup (group A+) was asked to participate with the translation from English to Dutch. Residents were invited to participate by email by RDL and could decline without any consequences or repercussions.

The second group (group B) were users who undertook an e-learning and were asked to fill out the survey in step 3. This group was asked to evaluate the e-learning they had just taken as part of the usual evaluation process. The inclusion criteria for the e-learning were: 1) an approach aiming to teach and learn, representing all or part of an educational model; 2) being based on the use of electronic media and devices as tools to improve training access, communication and interaction; 3) facilitating the adoption of new knowledge, skills and/or behavior/attitude; and 4) being targeted at postgraduates in medicine with a western cultural background. The exclusion criteria were: 1) not being able to understand, read and write in English or Dutch; and 2) a single module as part of a bigger e-learning curriculum.

The third group (group C) were the creators or authors of the e-learnings described above. For phase 4 we asked representatives tasked with the usual evaluation and improvement responsibilities of each evaluated e-learning. The Dutch Association of Medical Education research gave ethical consent (ID 2018.5.1)

E-learning groups

This study evaluated three PGMeL modules. The aim was to gather the survey outcomes to determine the usefulness, understandability and added value with the creators. The aim is not to evaluate the e-learnings themselves.

E-learning one was aimed at new doctors in a big teaching hospital in Amsterdam, the Netherlands. The e-learning aim was to train them with the local electronic patient records. It was mandatory for all new doctors to complete the e-learning, after which they were asked to fill out the evaluation. The author group did not add extra items to the survey. A total of 160 participants were asked to fill out the survey in the period of June 2018 to October 2018.

E-learning two was aimed at surgical residents and offered globally. The platform offered different surgical modules for a variety of specialties focusing on anatomy, surgical steps and pitfalls during surgery. The author group added 19 extra items. After finishing at least three modules, the users were asked to voluntarily fill out the survey.
In the period August to November 2018, 394 participants were asked to evaluate the e-learning.

E-learning three was aimed at OB/GYN residents practicing minimal invasive surgery, mainly in Europe. The e-learning is part of a certification with face-to-face and hands-on training as well. The author group added eight extra items. In the period August until the end of October 2018, about 2400 participants were asked to fill out the survey. Of these, most were older users, some of whose email addresses no longer worked. An estimated 1600 participants had recently used the e-learning and were reached by email.

Data collection and procedure

Data was collected in step two and three by focus group discussion with experienced users in the comfort of their university environment. The data from step four was collected by proving the EMSE as an online survey with a short introductory text (see appendix 3). Data for step five was collected by audio recordings after written consent was given. These focus groups discussions were also facilitated by RDL at the main offices of the e-learning groups.

Data analysis

No analysis was undertaken of the first four steps as the data were used during the steps themselves. The focus group evaluations in step five were analyzed. All interviews were verbatim transcribed and a thematic analysis was performed as per Braun [13].

The transcribing of the interviews was carried out by RDL to enable the author to familiarize himself with the data. We used Atlas.ti version 8.0 for the initial coding. Thematic analysis has been shown to usefully summarize key features and generate unanticipated insights [12, 13]. To perform the analysis, we ordered the codes by the pre-defined themes from our research question: usefulness, understandability and added value. We then reviewed the content of the themes and, if needed, redefined them as per Braun et al [13].
Results

The results are described per step in Figure 2. The initial draft and a change log of the MEES can be found online at www.MotivateLearnApply.com. The final version of the MEES is attached in appendix 3 and a Creators’ manual in appendix 4.

Step 1 – The postgraduate Medical E-learning Evaluation Survey (MEES)

To address the experts’ feedback, we added an explanation to the survey and the option for e-learning creators to add additional items. The term e-learning can be defined as per a previous Delphi and as has been used for all studies so far: “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 knowledge, skills and/or behavior/attitude” [14]. This means that all forms of electronic learning based on an educational model are e-learning.

To address the generalizability of the indicators, they are used as examples. There are, for example, many ways to be motivated. The previous steps provided nine ways to achieve this, but the creators of the e-learning might have used other strategies as well. Before the survey started, the creators of the e-learning were asked to add those indicators to each domain as examples. The domains were thus questioned using the general indicators from the literature, but also items that might be unique to that one e-learning.

The MEES contains questions in five domains, each of which starts with a Likert-scale of 1 to 10. These are followed by the indicators from the previous steps and those added by the e-learning creators depending on the aims of the specific e-learning. Finally, there is an open question about the domain. The domains are motivators, barriers, learning enhancers, learning discouragers and real-world translators. The MEES therefore, contains 10 questions, 36 examples and five open questions. Figure 3 shows the relation between the domains and table 1 lists all 35 examples with a short explanation of the purpose.
Figure 3 – the domains and structure of the MEES

Legend figure 3: this figure shows the five domains of the MEES. Each domain starts with a 1-10 Likert-scale question, followed by the general domain items from this study and the added specific items from the creators and finished by an open question.
Table 1 – the 36 items from the final MEES with a short explanation of their meaning from a creators point of view.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Original item</th>
<th>Short explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivate</td>
<td>1 I felt this e-learning was important</td>
<td>Creating a feeling of importance is very important for the user. The challenge is to convey to your users that the learning aims are important for their work and personal development</td>
</tr>
<tr>
<td></td>
<td>2 I felt it was my responsibility to undertake this e-learning</td>
<td>Along with importance, your user needs to feel responsible for the learning aim as well. This can be done by emphasis on the importance, but also by, for example, rewarding or giving responsibility for an outcome</td>
</tr>
<tr>
<td></td>
<td>3 I had enough time to complete the e-learning</td>
<td>Proving time to do the e-learning seems contra to “anytime, anywhere” learning, but it does give the learning the feeling of priority from a management level</td>
</tr>
<tr>
<td></td>
<td>4 I had a good understanding of the general purpose of the e-learning</td>
<td>The general purpose is the learning aim: knowledge, skills or attitude/behavior. It should be very clear to the user what they gain from finishing the e-learning</td>
</tr>
<tr>
<td></td>
<td>5 The e-learning objectives (for each educational section) were clear to me</td>
<td>When an e-learning is separated into different sections / chapters, make sure you communicate what the learning objectives are for each section</td>
</tr>
<tr>
<td></td>
<td>6 There was a clear overview of all content</td>
<td>Providing an overview of all lessons, objectives and options gives the user the possibility to manage expectations and if possible, create their own learning process</td>
</tr>
<tr>
<td></td>
<td>7 I knew how to navigate to the content</td>
<td>Navigation is an important part of the User Interface and should be very clear for the user so they can find content easily and go back and forward through the content</td>
</tr>
<tr>
<td></td>
<td>8 I felt comfortable with the quality / truthfulness of the content</td>
<td>Trust is important when learning. If the user has doubts about the truthfulness or quality, it will limit the working memory used for learning. Trust can be gained by the transparency of the creators, referring to recent literature, etc.</td>
</tr>
<tr>
<td></td>
<td>9 I was able to undertake this e-learning without being forced</td>
<td>Forcing a user to undertake an e-learning is the opposite of motivating them. If force or even blackmail is needed, the user will feel resentment, which kills motivation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>10</td>
<td>I felt taken seriously as an adult learning</td>
<td>Taking the learning seriously means avoiding childish illustrations and/or examples, and aiming at the level of experience means that you take into account what the user already knows to prevent repetition of basic knowledge.</td>
</tr>
<tr>
<td>11</td>
<td>The e-learning was aimed at my level of experience</td>
<td>Making your e-learning too easy will decrease motivation, while making it too complicated will make users learn less. This is why knowing the background knowledge of the target audience is of great importance.</td>
</tr>
<tr>
<td><strong>Barriers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>I was not able to create my own learning path to my own needs</td>
<td>This questions the difference between synchronized and asynchronized learning paths. Creating your own learning path means the option to test and skip already known sections or to go from A to C and then to B.</td>
</tr>
<tr>
<td>13</td>
<td>The e-learning was not easily accessible at my location or with my device</td>
<td>Accessing the e-learning should ideally be possible from every device and location so consider, for example, internet speeds in foreign countries. If access is not possible, consider helping your users get the right device.</td>
</tr>
<tr>
<td>14</td>
<td>The navigation did not make sense to me</td>
<td>Good navigation is helpful but poor navigation will not only limit an e-learning, but make it impossible to finish. Make sure your users can follow all steps without using their cognitive load for navigation.</td>
</tr>
<tr>
<td>15</td>
<td>The layout of the e-learning was too complicated</td>
<td>Navigation and layout are both important aspects of the User Interface. The less cognitive energy is used for the learning environment, the more can be used for the learning itself.</td>
</tr>
<tr>
<td>16</td>
<td>There was no instrument to help me navigate the e-learning (for example a sitemap)</td>
<td>Even if the navigation is of a high standard, it is still very helpful to have some instrument that gives an overview of all content and helps direct users where they want to be.</td>
</tr>
<tr>
<td>17</td>
<td>I had worries about the security and safety of the e-learning, regarding my personal information</td>
<td>Worries about quality are mentioned in item 8, but worries about security and privacy are also relevant in many countries, and may even have a legal aspect.</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>Recommendation</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>18</td>
<td>The e-learning was slow and took too long to load</td>
<td>Fast and logical use of the e-learning is also an important aspect of the User Experience. Waiting on affordances or loading frustrates and distracts, and should be minimized.</td>
</tr>
<tr>
<td>19</td>
<td>I did not know which devices the e-learning was compatible with and I might have used the wrong one</td>
<td>If your e-learning has specific needs, for example a specific operating system such as iOS, you need to clearly state that at the beginning. Try to prevent users from experiencing your e-learning in a wholly different way than planned because they use the wrong device.</td>
</tr>
<tr>
<td>20</td>
<td>The e-learning was too long</td>
<td>The duration should have been specified. Duration of videos, sections and the e-learning overall are taken together as one item. If there are, for example, longer videos, their duration can be added as a separate item.</td>
</tr>
<tr>
<td>21</td>
<td>The e-learning did not divide the content into proper sections</td>
<td>Learning and memory theories suggest that learning has a limited time span. Sectioning or chunking is a very effective way to help users through a bigger e-learning.</td>
</tr>
<tr>
<td>Learning enhancers 22</td>
<td>I could personalize the e-learning (for example by saving and continuing, filling out questionnaires and getting my personal score, etc)</td>
<td>Personalizing a learning experience allows the user to know how they are doing and follow a preferred method and path. The more personal/specific such things as feedback are, the more the user will gain. This is a very important motivator as well.</td>
</tr>
<tr>
<td>23</td>
<td>I could create my own learning path, and was not forced to follow the directed path (for example by skipping parts or returning to previous sections if needed)</td>
<td>This questioned the difference between synchronized and asynchronized learning paths. Creating your own learning path means the option to test and skip already known sections or to go from A to C and then to B.</td>
</tr>
<tr>
<td>24</td>
<td>I had an idea of the progress I had made and what was left to do (for example by a progress bar)</td>
<td>When learning, it's important to manage expectations. Knowing what is already done and what is left to do is an important affordance of, for example, a book, and should preferably be available in an e-learning as well.</td>
</tr>
<tr>
<td>25</td>
<td>If needed, I had access to technical support</td>
<td>To minimize the effort spent on technical aspects rather than learning, providing support as fast as possible will prevent users from stopping learning</td>
</tr>
<tr>
<td>26</td>
<td>The e-learning provided summaries where needed</td>
<td>Learning theory suggests that summaries support learning by offering repetition of content in a new format and allow chunking of the bigger picture</td>
</tr>
<tr>
<td>27</td>
<td>The e-learning provided feedback on my answers</td>
<td>Learning theory also suggests that learning is more effective when based on previous experience and knowledge, and providing feedback helps the user to make connections between new knowledge and their mistaken or correct assumptions</td>
</tr>
<tr>
<td>28</td>
<td>There were exercises and/or assignments in the e-learning</td>
<td>Learning theory suggests that actively using new knowledge will help it to go from working memory to long-term memory. Therefore, exercises or assignments help the transfer of the learning aim to long term memory</td>
</tr>
<tr>
<td>29</td>
<td>I could interact with the content of the e-learning (for example questions, exercises or other interactivities)</td>
<td>Interaction is another example of actively using the content, helping users learn more efficiently</td>
</tr>
<tr>
<td>Learning discouragers</td>
<td>30</td>
<td>I got stressed or frustrated by the e-learning, for whatever reason</td>
</tr>
<tr>
<td>31</td>
<td>The content was not able to adapt to my device when needed (for example, the e-learning should work on a mobile device but the icons were much too small for that)</td>
<td>Non-adaptable content can cause frustration and degrade the User Experience, again moving energy away from learning and towards technical aspects</td>
</tr>
<tr>
<td>32</td>
<td>The e-learning design and visuals were too distracting for me</td>
<td>Multimedia learning provides a theory and guidelines for how to use the combination of visuals and auditory stimuli effectively. Distraction should always be prevented</td>
</tr>
<tr>
<td>Real-world translators</td>
<td>33</td>
<td>The e-learning content and examples are translatable to my daily real-world work</td>
</tr>
<tr>
<td>------------------------</td>
<td>----</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>The e-learning seems up-to-date and properly maintained</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>The e-learning provided sources for the information which were also accessible after finishing it</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>Besides this questionnaire, the e-learning was evaluated on topics like user experience, effectiveness, usability and/or costs</td>
</tr>
</tbody>
</table>

*Legend for table 1: this table shows the questions distilled from the items of the previous studies. The questions are sorted by domain and are the same in this table as in the final MEES. The fourth column provides a short explanation for the creators about the relevance or interpretation of the questions.*
Step 2: readability and item interpretation.

Eight residents were asked to participate, all of whom agreed, but one was unable to come to the discussion on time. The discussion lasted 65 minutes and took place in May 2017 at the Amsterdam University Medical Centre, the Netherlands. Four items were interpreted differently than intended, three items were not readable and two items were in two domains. Details can be found in the change log online. Overall, the domains and questions were well understood.

Step 3: adjust, rewrite and translate.

The items from phase one were rewritten. After they had been adjusted, they were understood correctly. After finishing the English MEES, RDL translated the survey into Dutch. Two native Dutch residents, again at the Amsterdam University Medical Centre, the Netherlands, read the survey in May 2017. All items were clear and were understood correctly. No other changes were made. The MEES version used for the evaluation contained three positive domains (motivation, learning enhancers and real-world translation) and two negative domains (barriers and learning discouragers), with thirty-six items in those domains, five Likert-scale questions of 1 to 10 and five open questions asking the participants their own comments in each domain. See figure 3 for an overview of the MEES survey.

Step 4: gathering filled out surveys.

Details of the three evaluated e-learnings are summarized in table 2 and the scores per domain are in table 3. All evaluations took place between July and the end of November 2018, after which we concluded the evaluation.

Table 2: demographics of the three evaluated e-learnings

<table>
<thead>
<tr>
<th>Target audience</th>
<th>Addition items</th>
<th>Evaluation surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>location</td>
<td>function</td>
<td>invitations</td>
</tr>
<tr>
<td>the Netherlands</td>
<td>medical staff</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>OB/GYN residents</td>
<td>19</td>
</tr>
</tbody>
</table>

Legend: this table shows that the MEES has been used internationally with a response rate of 7 to 10%. For more details please find Appendix B.
Table 3: mean scores per domain of the three evaluated e-learnings

<table>
<thead>
<tr>
<th>E-learning</th>
<th>Motivators</th>
<th>Barriers</th>
<th>Learning enhancers</th>
<th>Real world discouragers</th>
<th>Real world translators</th>
</tr>
</thead>
<tbody>
<tr>
<td>No 1</td>
<td>7.3</td>
<td>3.8</td>
<td>6.8</td>
<td>4.1</td>
<td>7.5</td>
</tr>
<tr>
<td>No 2</td>
<td>8.6</td>
<td>3.8</td>
<td>8.2</td>
<td>3.6</td>
<td>8</td>
</tr>
<tr>
<td>No 3</td>
<td>9</td>
<td>3</td>
<td>9</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

Legend: this table shows the mean scores per domain. The higher the score in the positive domains “motivators”, “learning enhancers” and “real world translators” the better. In the negative domains “barriers” and “learning discouragers”, a lower score is better. Note the discrimination between the positive and negative domains.

In total there were 77 free text comments (see appendix 2 for all comments). The main positive comments concerned the availability and the added value of the e-learning to local education. “Unfortunately, surgical skills at my university are not well taught due to the large number of residents. There is also lack of standardization in teaching. I was happy to find a fun way to learn the best, standard way to perform common gynecological procedures and no longer rely on sketchy YouTube videos.”

On the negative side, users complained mostly about technical barriers as long loading times, login problems, software crashes and non-functioning affordances as a search function. “It sometimes took too long to load despite good internet connection, and I have often been forced to abandon a procedure due to this” and “Video streaming, is a serious limitation. There is need for video downloads”. Another frequent complain was about the language barriers such as poor English or videos with hard to understand speakers “Difficulté de langue. Je ferais effort d’apprendre”.

Step 5: focus group discussions with the e-learning creators.

Three focus group discussions took place, one with each e-learning creator group at their main office in November and December 2018. The average age of the ten participants was 51, the experience of creating e-learning was 0-5 years, and there were content, didactic and technical experts at the interviews. Eighty percent had experience with previous evaluations, but only 10% had used any formal evaluation methods. The three subjects of discussion are now described.
Usefulness: All participants described the results as very useful.

“Grouping the results into positive (domains) and negative (domains), resulted in a clear overview of what we need to keep and what to improve” – 1B

All groups said that the option to add items specific to an e-learning increases the usefulness. It provides feedback on the additional items that are considered important for the creator group. The first group regretted not adding any items themselves and, seeing the results now, said they would have added them.

Understandability: going over all the items one by one, some were not clear to the creators. Even though they understood the question, they did not know how to interpret it from a creator’s point of view. These items (10, 12 and 36) are in appendix 4, table 1 with a short explanation in the second column. The general advice was to have an explanation manual for the creators of each item which could be consulted in the event of misunderstanding.

“It would be nice to have a short explanation per item” – 2C

There were also worries about the form in which the results were presented. All the filled-out surveys were presented by RDL and summarized by him as well.

“How much time did it take you to formulate the results like this, and can this be done by us as well?” – 3A

Added value: although the participants had experience of evaluating e-learning, only one had used a formal evaluation method (although it is unknown which one). There were three subjects that added value to the MEES, the first of which was the domains. Using the domains gave the creators a structure that they did not have with other, informal, evaluation methods. Secondly, the items provided concrete examples of ‘do’s’ and ‘don’ts’, which inspired the changes needed in an update. Thirdly, using a formal method gave the creators a feeling of importance and allowed them to formalize the needs for improvement. The creators believed that using a formal method would allow them to more easily convince management and increase the commercial benefits of the e-learning.

“This really helps to improve the commercial value of our e-learning” - 2A

The five steps of this study provided a first draft of the MEES, an adjusted version (appendix 3), evaluation data from three PGMeL modules (appendix 2) and the thematic analysis of three focus group discussions with the creators of these modules.
Discussion

To our knowledge, the MEES is the first survey designed to evaluate PGMeL. Content validation has been completed and this study completes all seven steps described by the AMEE. We set out to investigate the usefulness, understandability and added value of this survey with the creators of three PGMeL’s. This study shows that the MEES is very useful, that understandability is clear with the help of a creator’s manual (appendix 4) and the MEES is of added value in connection with the structured domains and validated nature of the survey. Although the MEES has reached a second stage, two questions remain.

First of all, how do we know the items have been correctly understood by all the participants? The pilot evaluation was carried out with Dutch participants, who seemed to understand the questions in a group session. That might be generalizable to Dutch residents. However, the second and third evaluated e-learning involved European and worldwide users who might not have interpreted the questions correctly. To then interpret an evaluation, it is important to at least consider the content integration and evaluate the equity pedagogy [15]. Content integration means using content that will illustrate the same examples in a variety of cultures. Equity pedagogy exists when the curriculum is modified to accommodate and facilitate the academic achievements of a diverse group. Although debatable, using Hofstede’s cultural dimensions can also help to place extreme high or extreme low scores in a cultural light [16]. For example, in a culture with high power distance (in other words, where power and inequality are fundamental facets of the society), users might be tempted to express gratitude by providing extreme positive feedback.

Secondly, how many filled out surveys are needed for a proper evaluation? As seen above, the average reply was around 8%; thus, 92% of users are missed when reply is made voluntary. This can, of course, be increased by, for example, asking respondents to fill out a survey before supplying a needed certificate. This might, however, raise an ethical dilemma in a research setting like this. Other ways to assemble a representative group include purposeful sampling [17] but the question remains as to how many is enough. The conclusion is that the more respondents, the better the data, but the reality is that researchers can only work with the data made available to them.

On the other hand, although missing 92% of users might seem to indicate a failed evaluation, the question of who is most likely to fill out the survey is a valid one. Hu et al, for example, show that users who provide feedback are most likely to be either very satisfied or very disappointed [18], and thus may be exactly the respondents required; that is, it may be that the ‘middle group’ yield less information of use to the evaluation.
Strengths and limitations

Following all steps of the AMEE guide and peer publishing them might be the biggest strength of the MEES. Questionnaire validation is a complex and diverse field of expertise and this study takes an important next step. An international Delphi in 2010 (COSMIN study) helped to structure the diverse terminology within this field by providing this definition: “validity is the degree to which an instrument truly measures the constructs it proposes to measure” [8]. Many types of validity can be evaluated, and validation is a continuous process. A shortlist, partly based on Tsang et al, of forms of validation is given in table 4 with an explanation of their applicability to the MEES [19]. It can be seen that face and content validity are completed and that construct, criterion and predictive validity leave room for future research. Because structural and concurrent validity are not applicable, factor analysis is deliberately not performed.

Another strength lies in the involvement of end-users. By including experienced residents in the focus group discussions, Delphi and pilot evaluation, we believe the MEES is truly a user-centered method of evaluation as it evaluates not only theoretical items, but also the subjects that matters to the user. The numbers from the Likert-scale questions do not offer enough insight, but the free texts fields do add knowledge about the users’ needs and wishes. Most of the comments concerned the e-learning affordances and the technical execution (for example interactive videos’ that load too slow). This emphasis the importance of evaluating not only the content of the e-learning, but the instructional design as a whole.

The biggest limitation of the MEES is also addressed above. It is not possible to know what feedback has been missed from those users who did not fill out the form. This might be reduced by making an evaluation mandatory, but it is impossible to predict how motivated users will be to provide useful feedback. We therefore believe that the MEES should always be accompanied by in-depth focus group evaluations with the users. Not only will these provide missed feedback, but they will also allow researchers to find out why some items are recognized or missed. Proper evaluation should never contain only an online survey.
Table 4 – forms of validation of a survey in relation to the MEES

<table>
<thead>
<tr>
<th>Name</th>
<th>Explanation</th>
<th>Relation to MEES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face validity</td>
<td>Whether the instrument is understandable and relevant</td>
<td>Checked in this study by asking the PGMeL creators about understandability and added value</td>
</tr>
<tr>
<td>Content validity</td>
<td>Whether the instrument measures the most important aspects of a concept that it is designed to evaluate</td>
<td>Checked in three previous studies by means of a review [10], focus groups [11] and an international Delphi [4]</td>
</tr>
<tr>
<td>Construct validity</td>
<td>The degree to which the instrument’s scores relate to other measures in a manner that is consistent with an a priori hypothesis concerning the concepts being measured</td>
<td>Awaiting future validation: the construct of the MEES is predicting efficiency and effectiveness by evaluating the experience of affordances to determine the quality of the instructional design</td>
</tr>
<tr>
<td>Criterion validity</td>
<td>How well one measure predicts the outcome of another</td>
<td>Awaiting future validation: MEES might predict the satisfaction or learning aim transference of the PGMeL</td>
</tr>
<tr>
<td>Predictive validity</td>
<td>The instrument’s ability to predict future test results</td>
<td>Awaiting future validation: as regards the MEES, this is in line with the criterion validity and can be checked in the future</td>
</tr>
<tr>
<td>Structural validity</td>
<td>Whether all items in a scale or subscale measure the same concept of the dimensionality of the instrument</td>
<td>Not applicable; this can be done by factor analysis but assumes that a scale or subscale is highly correlated, which might not be the case in the MEES</td>
</tr>
<tr>
<td>Known-group validity</td>
<td>The ability to be sensitive to differences between groups of users that may be expected to score differently in the predicted direction</td>
<td>Not applicable; this can be done by comparing PGMeL designed for certain cultural groups, but the assumptions are too complex to use this in practice</td>
</tr>
<tr>
<td>Concurrent validity</td>
<td>The association of the instrument with accepted standards</td>
<td>Not applicable; there is no golden standard of evaluating PGMeL</td>
</tr>
</tbody>
</table>

Legend for table 4: this table shows a shortlist of forms of validation. Each method of validation has its purpose, strengths and limitations. The face and content validation of the MEES are presented in this study, the structural, known-group and concurrent validation are not applicable and the construct, criterion and predictive validity are possible future validation steps.
Future research

Validation is a never ending story and it is necessary to continue collecting validity evidence [9]. We believe three steps should be taken. First, as part of the construct and criterion validation, it would be very interesting to see if the MEES can be used to measure improvement by taking a PGMeL, evaluating it with the MEES and a focus group discussion, adjusting it accordingly and re-evaluating to see if the second evaluation is better. This would provide insight into the actual benefit for future learning. The second step would be to evaluate readability and understandability in different languages. To this end, the survey should be translated into other languages, and those new translations pilot tested for readability and understandability. The third step is to evaluate the understandability and reliability of the survey within different (sub) cultures. Evaluating how different cultural groups interpret digital evaluation and the questions can provide insight into the way the creators should use the results of the survey.
Multimedia Appendix 1: focus group discussion guide

Introduction

We met today to look at the results of the previously completed questionnaire. You already had these results before and the goal today is to review these results once again and then evaluate together which aspects were of value to you. It is not about the evaluation outcomes themselves. So, we are not going to look at the strengths and improvements of your e-learning. We will look at the strengths and weaknesses of the evaluation survey itself, for you as a creator. We will do this by category: motivation, learning and applicability.

A. Interview background

- Would you please fill out the demographic questionnaire below?

B. Domain motivation

- After this brief overview of the results, did you find these results useful?
- Did you understand all the items questioned?
- Which aspects come from the EMSE are not of added value?
- Which aspects would you miss in the evaluation of this domain?

C. Learn domain

- After this brief overview of the results, did you find these results useful?
- Did you understand all the items questioned?
- Which aspects come from the EMSE are not of added value?
- Which aspects would you miss in the evaluation of this domain?

D. Domain applicability

- After this brief overview of the results, did you find these results useful?
- Did you understand all the items questioned?
- Which aspects come from the EMSE are not of added value?
- Which aspects would you miss in the evaluation of this domain?
E. General comments

- Many thanks for answering these questions. Are there any general things that you think can improve MEES in any way?

Demographic questions

1. What is your age? ______________________

2. What is your responsibility in the e-learning creation process?
   A. Content expert
   B. Didactic / education expert
   C. Technical expert
   D. Other ______________________

2. How many years of experience do you have with e-learning?
   A. 0-5 years
   B. 5-10 years
   C. > 10 years

3. Do you have experience in the formal evaluation of e-learning?
   A. Yes
   B. No

4. If yes, how many e-learnings did you evaluate approximately?
   A. 0-5
   B. 5-10
   C. >10

5. If yes, did you use any validated evaluation instruments?
   A. Yes
   B. No
   C. If yes, which one? ______________________
Multimedia Appendix 2: e-learning survey outcomes

E-learning 1: twenty-three additional comments were given in the open questions. Eight items of the positive domains were recognized by at least half the participants and none of the items in the negative domains.

E-learning 2: Most of the residents were from general surgery (n=19) and from the Netherlands (n=8) and Indonesia (n=4). Ten items from the positive domains and one from the negative were recognized. Twenty-four remarks were left in the free text fields.

E-learning 3: Ten items in the positive domains were recognized and none in the negative domains. Thirty free text comments were left.

Table 1 shows all scores per e-learning and per domain. Table 2 provides all the original free text comments and can be found online at www.MotivateLearnApply.com

Table 1: scores per e-learning, per domain
Chapter 8

Learning enhancers

Min  Mean  Max

Group 1  Group 2  Group 3

Learning barriers

Min  Mean  Max

Group 1  Group 2  Group 3

Real World Translator

Min  Mean  Max

Group 1  Group 2  Group 3
Dear e-learning creator,

This survey is aiming to evaluate the quality of the instructional design of postgraduate medical e-learning by evaluating the experience of the affordances of your e-learning, aiming to predict the efficiency and effectiveness. There are 10 questions, 36 examples and 5 open questions already in the survey. You are asked to add examples, depending on your learning aims and unique affordances. Please find a detailed description of the way to use this survey and how to interpret all examples in the attached “Creator manual”. An example text for your users and a short summary example text for the documentation of the evaluation can also be found in this manual.

The Evaluation Survey for postgraduate Medical E-learning (MEES) version 1.0-2019

[please add your own introduction text here, or find an example in the creators’ manual]

1. On a scale of 1-10, how motivated were you to start the e-learning?

Unmotivated Motivated

[1 2 3 4 5 6 7 8 9 10]

2. Please select which of the following items motivated you (multiple options).

☐ I felt this e-learning was important
☐ I felt it was my responsibility to do this e-learning
☐ I had enough time to do the e-learning
☐ I had a good understanding of the general purpose of the e-learning
☐ The e-learning objectives (for each educational section) were clear to me
☐ There was a clear overview of all content
☐ I knew how to navigate to the content
☐ I felt comfortable with the quality / truthfulness of the content
☐ I was able to do this e-learning unforced
☐ I felt taken seriously as an adult learning
☐ The e-learning was aimed at my level of experience
☐ ADDED EXAMPLE(S)
☐ Please write down any other item(s) that motivated you:
3. On a scale of 1-10, did you experience any barriers to starting the e-learning?

<table>
<thead>
<tr>
<th>Stimulants</th>
<th>Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

4. Please select which of the following items you experienced as a barrier to starting the e-learning.

- [ ] I was not able to create my own learning path to my own needs
- [ ] The e-learning was not easily accessible at my location or with my device
- [ ] The navigation did not make sense to me
- [ ] The layout of the e-learning was too complicated
- [ ] There was no instrument to help me navigate the e-learning (for example a sitemap)
- [ ] I had worries about the security and safety of the e-learning, regarding my personal information
- [ ] The e-learning was slow and took too long to load
- [ ] I did not know which devices the e-learning was compatible with and I might have used the wrong one
- [ ] The e-learning was too long
- [ ] The e-learning did not divide the content into proper sections
- [ ] ADDED EXAMPLE(S)
- [ ] Other item(s) that I experienced as a barrier:

5. On a scale of 1-10, how educative was the e-learning for you?

<table>
<thead>
<tr>
<th>Uninformative</th>
<th>Educative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
6. Please select which of the following 8 items helped you to learn and remember or add your own.

- [ ] I could personalize the e-learning (for example by saving and continuing, filling out questionnaires and getting my personal score, etc)
- [ ] I could create my own learning path, and was not forced to follow the directed path (for example by skipping parts or returning to previous sections if needed)
- [ ] I had an idea of the progress I had made and what was left to do (for example by a progress bar)
- [ ] If needed, I had access to technical support
- [ ] The e-learning provided summaries where needed
- [ ] The e-learning provided feedback on my answers
- [ ] There were exercises and/or assignments in the e-learning
- [ ] I could interact with the content of the e-learning (for example questions, exercises or other interactivities)

ADDED EXAMPLE(S)

Other item(s) that helped me learn and remember:

7. On a scale of 1-10, did you experience limitations in the e-learning that prevented you learning?

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>[10]</td>
</tr>
<tr>
<td>[2]</td>
<td></td>
</tr>
<tr>
<td>[3]</td>
<td></td>
</tr>
<tr>
<td>[4]</td>
<td></td>
</tr>
<tr>
<td>[5]</td>
<td></td>
</tr>
<tr>
<td>[6]</td>
<td></td>
</tr>
<tr>
<td>[7]</td>
<td></td>
</tr>
<tr>
<td>[8]</td>
<td></td>
</tr>
<tr>
<td>[9]</td>
<td></td>
</tr>
</tbody>
</table>

8. Please select which of the following 3 limitations you experienced or add your own.

- [ ] I got stressed or frustrated by the e-learning for any reason
- [ ] The content was not able to adapt to my device when needed (for example, the e-learning should work on a mobile device, but the icons were way too small for that)
- [ ] The e-learning design and visuals were too distracting for me

ADDED EXAMPLE(S)

Other limitation(s) in my learning from this e-learning:
9. On a scale of 1-10, how do you think you can apply the newly learned knowledge, skills or attitude to your daily work?

Irrelevant  Applicable

10. Please select which of the following 4 items helped you to apply the e-learning, or add your own.

☐ The e-learning content and examples are translatable to my daily real-world work
☐ The e-learning seems up-to-date and maintained
☐ The e-learning provided sources for the information which were also accessible after finishing it
☐ Besides this questionnaire, the e-learning was evaluated on topics like user experience, effectiveness, usability and/or costs
☐ ADDED EXAMPLE(S)
☐ Another way this e-learning helped me apply newly learned knowledge, skills or attitude to my daily work:
Multimedia Appendix 4: creators’ manual

This is a Postgraduate Medical e-learning (PGMeL) Creators’ Manual which should be used with the Evaluation Survey for Postgraduate Medical E-learning (MEES).

This survey is publicly available and was published in 2019 in XXX, with the title “The development and validation of an Evaluation Survey for Postgraduate Medical E-learning”. It can be freely used, and all questions may be addressed to the corresponding creator of the article.

The aim of the survey is to evaluate the quality of the instructional design of PGMeL by evaluating the experience of the affordances of an e-learning with the aim of predicting its efficiency and effectiveness.

Table of content

1. Using the survey
2. Definitions
3. All items explained
4. Using the results
5. Publishing the results
Using the survey

Before digitalizing the survey, you should first consider each domain for your own e-learning. Each domain is described below and will contain three question types: first, a Likert-scale of 0-10; then, a series of literature-based examples which your users may or may not recognize. Because each e-learning is unique, you are asked to think about the unique characteristics or functionalities of yours. Try to place them in the domains and add them to the list of examples. Each domain ends with an open question, asking the learners what they think can be added or removed from the list.

When you have done that, we advise you to digitalize the survey and add it to the end of your e-learning. It can be made mandatory, which might elicit many neutral and unmotivated responses, or voluntary, which will give you fewer and usually more polarized opinions. Our experience is that the latter provides more useful feedback.
Definitions

To understand the explanations of the survey, we have provided you below first with a list of used definitions and then with definitions of the used domains.

**The user;** the person using and learning from your e-learning

**The creator;** the person, or more usually the team, that created the e-learning. (We did not use the word ‘author’ to prevent confusion with the author of a manuscript)

**User interface;** all components of the e-learning that make interaction between human and hardware/software possible, for example a menu

**User experience;** the relation between the e-learning and the user: which emotions and feelings – for example, pleasure or frustration – they experienced during the use of your e-learning

**Affordances;** the functionalities and options that your e-learning provides to the user in order to make the experience as efficient and effective as possible, for example feedback on incorrect questions

**Instructional design;** the way your e-learning is designed; which educational strategies are used; and which affordances are chosen and executed in order to make the e-learning

**Cognitive load;** a part of the cognitive load theory, a learning theory that provides a model of how people learn. Many aspects of the evaluation are based on this theory and further reading is advised, for example *Cognitive load theory in health professional education: design principles and strategies*, van Merriënboer et al. *Medical Education* 2010: 44: 85–93

The domains

**Motivators:** motivators are aspects of the e-learning that will help the user start the e-learning and continue it to the end.

**Barriers:** the opposite of the motivators domain; things that will prevent the user from starting or finishing the e-learning. There is a scale with motivators on one hand and barriers on the other.

**Learning enhancers:** affordances that will, based on learning theories, increase the efficiency of the learning process of the user.

**Learning discouragers** aspects that will distract or demotivate the user, or wrongly guide their learning.

**Real-world translators** aspects of the e-learning that will help the user to apply the new knowledge, skills and/or behaviour in their daily work.
Using the results

After gathering a sufficient sample of filled out surveys, there are many ways to use them. One way of doing so is by creating a strength and weakness analysis.

1. Gather the highest, lowest and mean score. Having outliers can be interpreted as a good thing, as it means you have the opinions of very satisfied but also dissatisfied users.

2. The strengths are in the motivation, learning enhancers and real-world translation domains. Items that are recognized by under 50% of users might need more attention. Use the free text comments as inspiration for further development.

3. The weaknesses are in the barriers and learning discouragers domains. When more than 50% of users do recognize an item, it might need to be removed. Use the free text to determine what you recognize, as well, and try preventing such items in the next version.

An example of such a strength and weakness analysis can be found online at www.MotivateLearnApply.com
Publishing the results

We encourage creators to publish the results of the MEES. This will show the users that their effort is actually being put to good use, which will motivate future evaluations. It also shows that you are serious in wanting to improve and continue the e-learning, which will increase the trust the users have in it.

Publishing can mean publishing the mean scores, examples recognized by at least 50% and a short plan of improvement for future versions. Ideally, you should indicate a schedule for when this will take place, so users know when to return for a new version as well.

The use of the MEES can be justified by the following text:

“We used a validated questionnaire called the Evaluation Survey for Postgraduate Medical E-learning (MEES) to evaluate our e-learning. This survey is developed specifically for PGMeL and is based on five publications in peer reviewed medical journals. The version we used is published in Journal of Medical Internet Research. April 2019, with the title “The development and validation of an instructional design evaluation survey for postgraduate medical e-learning” by de Leeuw et al.

The aim of the survey is to evaluate the quality of the instructional design of PGMeL by evaluating the experience of the affordances of the e-learning, and aiming to predict the efficiency and effectiveness. For further reading about this survey, we refer to the above-mentioned paper.”
References


Digital vs face-to-face information provision in patient counseling for prenatal screening: a non-inferiority randomized controlled trial

Prenatal diagnosis
May 2019
Abstract

Objective: to evaluate face-to-face information provision in patient counselling for prenatal screening compared to two forms of digital information provision, namely non-interactive instructional video or interactive video.

Method: We performed a prospective, non-inferiority, cluster-randomized controlled trial comparing face-to-face (usual care) to two forms of digital information provision (intervention) in counselling for prenatal screening. This study was performed in the Amsterdam UMC, the Netherlands in 2017, and included women in the first trimester of pregnancy. Main outcomes were knowledge gained by the patient and counselling duration. We performed a non-inferiority analysis.

Results: 141 women were included, randomized and analysed. The baseline characteristics were comparable. The intervention group was non-inferior compared to the control group regarding the level of satisfaction. The knowledge grade difference was higher after using intervention and the duration was significantly longer in the face-to-face group at 23 minutes versus 16 minutes. The addition of interaction with the video made no difference in any of the outcomes.

Conclusion: adding an instructional video to patient counselling is of added value to improve patient knowledge and shorten time consumption of the counsellor, therefore saving costs. But this form of counselling maintains the same level of satisfaction.
Introduction

Prenatal screening was developed in the 1970s, as the result of medical innovations, such as invasive prenatal diagnostic tests and obstetrical ultrasound. Since then, prenatal screening has been a routinely offered medical procedure. This has led to new challenges, among which is the counselling needed for prenatal screening. Comprehensive prenatal counselling is complicated; it should be personalized and provide sufficient information to promote autonomous decision-making. Earlier research has shown not only the importance of counselling, but also an increase in inter-counsellor variation and unpredictable time consumption, depending on the complexity of the subject.

A counselling consultation can be divided into two parts: information provision and the counselling itself. Since the counselling is complex and can have different requirements in different countries, such as face-to-face contact, provision of information is where new innovations may be of help. In a digital age, patients increasingly need digital information provision and even digital counselling. Previous studies show that websites provide added-value beyond face-to-face information and recommend online resources. Yet they also show that it’s difficult and confusing to find information online. A review from Marokasis et al in 2016 describe most of the added to consistent information and the extended length of consultation.

There is yet another evolving area of innovation, which is the use of interactive electronic media to facilitate teaching and learning. Better educated patients tend to be healthier and e-learning contributes to increased knowledge, improves satisfaction with the consent process and consultation and is shown to be effective in improving both physical and mental health outcomes. It is unknown whether digital information provision (instructional video with or without interactivity) can be as satisfying as information provided face-to-face in prenatal counselling during screening for prenatal testing. There is evidence that instructional videos do not have to include interactive elements to be effective. However, evidence in the field of prenatal screening counselling is lacking.

In this study, we analysed whether an instructional video is non-inferior to face-to-face information provision with regard to patient satisfaction. We also evaluated if video can be beneficial in regard to learning effect and duration of counselling. Moreover, we aimed to evaluate whether the addition of interactive elements to the instructional video is of added value.
Method

Trial design and participants

We performed a prospective, non-inferiority, cluster-randomized controlled trial comparing face-to-face (control group) to digital (intervention group) information provision before counselling for prenatal screening. All eligible patients were invited in the outpatient clinic of the Amsterdam University Medical Centers, VUmc University Amsterdam, the Netherlands. Each consultation consisted of two parts: information provision, followed by face-to-face, personal counselling. The control group received usual care, meaning a single consultation of face-to-face information provision and counselling. The intervention group was randomized between information provision by means of an instructional or interactive video before they continued to face-to-face, personal counselling. Both groups had the same face-to-face component. The counsellors (n=5) were experienced and trained according to the national guidelines (both midwives and doctors follow the same training and perform at least 50 counselling sessions per year), to inform the participants of the same content as that given in the video.

In the trial we included healthy women in the first trimester of their pregnancy, aged above 18 years, who had requested counselling for prenatal screening. The participants referred for an indication of increased risk of a chromosome abnormality in pregnancy were excluded. After filling out the informed consent form, the participants were given a short demographics questionnaire; they also completed a knowledge pre-test, satisfaction questionnaire afterwards and knowledge post-test directly after the counselling. The counsellors were blinded for the intervention randomization (instructional or interactive video) outcome and received a satisfaction questionnaire after the counselling. Counsellors timed the duration of their total consultation. The participant watched or interacted with the video in a private waiting room immediately before the counselling. Women were asked to participate when they were eighteen years or older, spoke the Dutch language and came for routine prenatal screening counselling. They were excluded when they had an increased risk on chromosomal abnormalities. Thirty-four women did not meet the inclusion criteria, mainly due to insufficient Dutch language skills. One hundred participants were lost in follow-up at the six-week questionnaire, after which we decided not to use the related data for further interpretation. See Figure 1 for the inclusion flow chart.
Figure 1 – Inclusion flow-chart
Information provision

Information provision and the counselling are usually followed by a well-informed decision on the part of the participant. All groups received both the same information and counselling in one consultation. The standardized information provision consisted of general information about the basic understanding of prenatal screening, the screening options and the possible consequences of a negative or positive test result. The aim of the information provision was to educate the participant to enable her to make a proper, well-informed decision during the counselling. After the counselling, we wished the participants to have insight into the following subjects: prevalence of trisomies in the general Dutch population; which chromosomal anomalies are tested for and how; which screening methods are offered; the difference between screening and diagnosing; the difference between invasive and non-invasive testing; and, finally, the limitations of ultrasound screening. The video was based on a previously used group consultation presentation.

Both the information provision and the counselling are needed to decide whether to undergo prenatal screening. Therefore, we carried out a pre-test before the information provision and a post-test after the counselling. The key difference between the interactive and passive video were the mandatory questions that pauses the interactive video. We added a progress bar, four pauses with written information, ten multiple choice questions and five stop/rewind popups elements to the video. Please find the instructional and interactive video here: http://eprint.elephantelearning.com/video-page/

Randomization

Counselling took place two days per week. Due to logistics in the outpatient clinic, it was not possible to randomize all subjects every day and we had random days of usual care and days of video information provision. On the video provision days, we computer randomized immediately before the counselling between an instructional (non-interactive) and an interactive video. The allocation was 1:1 for the control group and the intervention group and, again, 1:1 for the instructional video and the interactive video group. Participants were given an appointment at the desk for either of those days randomly (the administrative personnel were blinded). All counsellors were blinded for the randomization outcome.
Outcomes

Our primary outcome was the level of participant satisfaction after the entire consultation. Satisfaction was preferred over knowledge as primary outcome, because in case of counselling, it’s more important to make a well informed, satisfying decision then to gain knowledge of the subject. Satisfaction is determined by a previously validated, adjusted and translated Genetic Counselling Satisfaction scale. The secondary outcomes were the duration of the face-to-face counselling which followed, the blinded counsellors’ satisfaction with the consultation on a scale of 1 to 5 and the participant knowledge score before and after the counselling. Knowledge was evaluated by means of a seven-question test based on the content of the information provided.

Statistical analysis

Data collection was conducted using Research Survey as a data management system and data analysis and reporting were carried out according to the CONSORT guideline with IBM SPSS as statistical analysis software. A descriptive table for the baseline characteristics is reported and primary and secondary outcomes were analysed on a noninferiority basis. Statistical analysis was conducted using SPSS for Windows version 10.0 on PC computer. None of the results were normally distributed; thus, no transformations were applied. The main outcomes of the survey were compared using the Aspin-Welch test. Difference within groups was analysed using the Wilcoxon Signed Rank test. The analysis was conducted in two phases. Phase one was a comparison between the main groups and phase two compared the intervention subgroups. Fisher’s test on binary outcomes could not be conducted since one of the groups had a 100% score. We calculated that a sample size of 160 woman (40 participants per video group and 80 in the control group) would be needed, with a risk of type I error of 5% and a power of 80%, to show statistically significant difference in satisfaction. We hypothesised that the group difference could be around 18%, based on earlier participant education studies (on premature birth). We added 25% lost in follow-up, therefore, we needed 200 participants in total.
Results

Between August 2017 and December 2017, 203 women were approached and asked to participate. We stopped including participants when the sample size was reached. One hundred and sixty-two women consented and five were lost in follow-up for the post-counselling questionnaire. These five participants (all in the control group) did not leave their post-counselling questionnaire and were not available anymore afterwards. We lost 10 participants in the control group and six in the intervention group because the counsellor did not fill out their questionnaire about these participants. We were able to analyse 67 women in the control group, 36 in the informative video group and 38 in the interactive video group for the primary outcomes (Figure 2, CONSORT flow chart). We formulated higher education as a completed bachelor or master’s degree. We asked the participants to score their presumed knowledge on a scale of 1 to 10 (self-assessed knowledge score). Overall, almost one third of the participants were primiparous, less than half had previous experience with prenatal screening, almost 80% were higher educated and more than half had a positive attitude towards prenatal screening.

Figure 2 - CONSORT flow-chart
The baseline characteristics were comparable between the groups, with the exception of the positive attitude toward screening; more women were positive in the intervention group towards screening. Participants were more positive towards screening in the intervention group (Table 1). There were no differences between the instructional video and interactive video groups (Table 2). Although there were more participants in the intervention group with previous experience with prenatal screening.

Table 1. Demographics usual care group and intervention

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Intervention group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=77</td>
<td>n=80</td>
</tr>
<tr>
<td>Maternal Age (average) at inclusion</td>
<td>33,6 ± 4,5</td>
<td>35,1 ± 4,1</td>
</tr>
<tr>
<td>Married or cohabiting (% yes)</td>
<td>76 (98,7%)</td>
<td>75 (93,8%)</td>
</tr>
<tr>
<td>Education (% higher educated)</td>
<td>58 (75,4%)</td>
<td>65 (81,2%)</td>
</tr>
<tr>
<td>Multi para (% yes)</td>
<td>55 (71,4%)</td>
<td>56 (70%)</td>
</tr>
<tr>
<td>Experience with prenatal screening (% yes)</td>
<td>25 (45,5%)</td>
<td>34 (60,7%)</td>
</tr>
<tr>
<td>Attitude towards screening (% positive)</td>
<td>34 (44,2%)</td>
<td>57 (71,3%)</td>
</tr>
<tr>
<td>Companion at consultation (% yes)</td>
<td>56 (72,7%)</td>
<td>51 (63,7%)</td>
</tr>
<tr>
<td>Education of companion (% higher educated)</td>
<td>36 (64,3%)</td>
<td>33 (64,7%)</td>
</tr>
<tr>
<td>Self-assessed knowledge score (average on 1-10)</td>
<td>5,62 ± 2,463</td>
<td>6,87 ± 1,580</td>
</tr>
<tr>
<td>Religious (% yes)</td>
<td>30 (39%)</td>
<td>24 (30%)</td>
</tr>
</tbody>
</table>
### Table 2. Demographics of the intervention group

<table>
<thead>
<tr>
<th></th>
<th>Intervention group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>instructional video</td>
</tr>
<tr>
<td><strong>n=40</strong></td>
<td></td>
</tr>
<tr>
<td>Maternal Age (average)</td>
<td>35.6 ± 4.7</td>
</tr>
<tr>
<td>Married or cohabiting (% yes)</td>
<td>38 (95%)</td>
</tr>
<tr>
<td>Education (% higher educated)</td>
<td>37 (92.5%)</td>
</tr>
<tr>
<td>Multi para (% yes)</td>
<td>26 (65%)</td>
</tr>
<tr>
<td>Experience with prenatal screening (% yes)</td>
<td>17 (65,4%)</td>
</tr>
<tr>
<td>Attitude towards screening (% positive)</td>
<td>31 (77,5%)</td>
</tr>
<tr>
<td>Companion at consultation (% yes)</td>
<td>25 (62,5%)</td>
</tr>
<tr>
<td>Education of companion (% higher educated)</td>
<td>21 (84%)</td>
</tr>
<tr>
<td>Self-assessed knowledge score (average on 1-10)</td>
<td>7,13 ± 1,418</td>
</tr>
<tr>
<td>Religious (% yes)</td>
<td>10 (25%)</td>
</tr>
</tbody>
</table>

### Primary and secondary outcomes

There was no difference between the control group and the intervention group in the satisfaction scale, nor among the subgroups of the Counselling Satisfaction scale, 3.91 (CI 3.38-4.42) and 3.93 (CI 3.53-4.33) P = 0.88, respectively (Table 3). In both groups over 95% of the participants were satisfactorily informed (95.6% control group, 100% intervention group). We asked participants if they already made a choice considering the prenatal screening before and after the counselling and 5.9% changed their opinion in the control group, while 7.2% changed their opinion in the intervention group (almost 94% whom changed their mind, changed from wanting to not wanting screening).

The duration of the counselling was 7 minutes shorter in the intervention (P = 0.00). After counselling, the knowledge grade increased in both groups. However, the difference between pre- and post-counselling knowledge grade was significantly greater in the intervention group. The knowledge score before the counselling was not significantly different between the control and intervention group. Counsellor satisfaction after the counselling was not significantly different.
Table 3. Primary outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Control</th>
<th>Intervention</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant satisfaction</td>
<td>n=68</td>
<td>n=69</td>
<td></td>
</tr>
<tr>
<td>Genetic Counselling Satisfaction scale</td>
<td>3.9 ± 0.5</td>
<td>3.9 ± 0.4</td>
<td>0.88</td>
</tr>
<tr>
<td>Satisfactorily informed (% yes)</td>
<td>95.6</td>
<td>100</td>
<td>nnvt</td>
</tr>
<tr>
<td>Knowledge grade before counselling</td>
<td>5.7</td>
<td>6.1</td>
<td>0.11</td>
</tr>
<tr>
<td>Knowledge grade difference pre/post test</td>
<td>0.91</td>
<td>2.07</td>
<td>0.00</td>
</tr>
<tr>
<td>Counsellor outcomes</td>
<td>n=58</td>
<td>n=63</td>
<td></td>
</tr>
<tr>
<td>Duration of counselling (average minutes)</td>
<td>23.0 ± 6.6</td>
<td>16.3 ± 7.4</td>
<td>0.00</td>
</tr>
<tr>
<td>Counsellor satisfaction (scale 1-10, average)</td>
<td>8.0 ± 1.2</td>
<td>7.7 ± 0.9</td>
<td>172</td>
</tr>
<tr>
<td>Counsellor satisfaction (Likert 1-5, average)</td>
<td>4.2 ± 0.7</td>
<td>4.1 ± 0.7</td>
<td>393</td>
</tr>
<tr>
<td>Well-considered decision (scale 1-5, average)</td>
<td>4.4 ± 0.7</td>
<td>4.3 ± 0.08</td>
<td>282</td>
</tr>
</tbody>
</table>

Subgroup analysis instructional video versus interactive video

Comparing the instructional video group with the interactive video group reveals no significant difference for any of the outcomes (Table 4). Satisfaction was comparable, as was counselling duration, knowledge grade before and after the counselling and counsellor satisfaction.
**Table 4. sub-analysis**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Passive video</th>
<th>Interactive video</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant satisfaction</td>
<td>n=33</td>
<td>n=36</td>
<td></td>
</tr>
<tr>
<td>Genetic Counselling Satisfaction scale</td>
<td>4,0 ± 0,5</td>
<td>3,9 ± 0,4</td>
<td>406</td>
</tr>
<tr>
<td>Satisfactorily informed (% yes)</td>
<td>100</td>
<td>100</td>
<td>nnvt</td>
</tr>
<tr>
<td>Knowledge grade before counselling</td>
<td>6,07</td>
<td>6,14</td>
<td>0.85</td>
</tr>
<tr>
<td>Knowledge grade difference pre/post test</td>
<td>2,1 ± 1,7</td>
<td>2,1 ± 1,6</td>
<td>987</td>
</tr>
<tr>
<td>Follow-up 4-6 weeks</td>
<td>n=15</td>
<td>n=19</td>
<td></td>
</tr>
<tr>
<td>Satisfaction (scale 1-5, average)</td>
<td>3,5 ± 1,2</td>
<td>3,5 ± 0,7</td>
<td>984</td>
</tr>
<tr>
<td>Well-considered decision (scale 1-5, average)</td>
<td>3,9 ± 1,3</td>
<td>4,0 ± 0,6</td>
<td>716</td>
</tr>
<tr>
<td>Counsellor outcomes</td>
<td>n=29</td>
<td>n=34</td>
<td></td>
</tr>
<tr>
<td>Duration of counselling (average minutes)</td>
<td>17,1 ± 7,5</td>
<td>15,6 ± 7,3</td>
<td>432</td>
</tr>
<tr>
<td>Counsellor satisfaction (Likert 1-5, average)</td>
<td>4,1 ± 0,7</td>
<td>4,1 ± 0,7</td>
<td>0.81</td>
</tr>
<tr>
<td>Well-considered decision (scale 1-5, average)</td>
<td>4,4 ± 0,8</td>
<td>4,2 ± 0,8</td>
<td>226</td>
</tr>
</tbody>
</table>
Discussion

Main findings

Using digital information provision as part of the counselling for prenatal screening proves to be just as satisfying as face-to-face information provision. The counsellors spent an average of seven minutes less per consultation and the participants using digital information provision performed better on the knowledge test. Adding interactivity to the video had similar results to the non-interactive video.

Strengths and limitations

This study shows that the use of video in patient information provision before prenatal screening counselling is acceptable for patients and has a potential benefit for the counsellor. When evaluating an instructional instrument, it is crucial to have a control group and, if possible, to evaluate an educational variation as well. This study did both. Furthermore, the study reached the necessary sample size to evaluate our primary outcome: participant satisfaction. Finally, randomization was ensured, given the comparable baseline characteristics. There were also limitations. First of all, we were not able to randomize by participant; however, we randomized days between the control group and the intervention. This could cause a selection bias, although we blinded the administration and participants, and it did not affect the baseline characteristics. Although the baseline characteristics did show that the study population is very homogenous and overall very well educated. This will influence the generalizability of the results. Secondly, the primary outcome, satisfaction is, due to its subjective nature, a difficult outcome to measure. It could very well be that participants filled out a satisfaction scale more positively when they had invested more time in watching a video. Our secondary outcome knowledge was not corrected for a possible test re-test bias, which could have caused an increase in knowledge, but should be expected to be equal on both groups. Thirdly, we were not able to have a proper six-week follow-up and even had a loss in the counsellor questionnaire in follow-up. We believe this is due to the fact that participants, and even counsellors, did not have enough commitment for this study to invest more time for the follow-up. Finally, we did not properly pilot-test to evaluate the video and the interactive version before using them. Therefore, improving the design of the intervention (by using the feedback from a pilot) could have improved the outcomes of the video and, especially, of the interactive video.
### Interpretation

Patient counselling is becoming increasingly important in shared decision making and this study shows that digital media can make it more time- and cost-effective. The benefit of using an instructional video is that each participant will get the same information. Face-to-face information provision never guarantees this. This could be the reason that the intervention group performed better on the knowledge test. Of course, the counselling of patients who had already been informed by video was shorter than that given to the patients in the control group. This is the most logical consequence of adding the intervention to the counselling. It is good to realize, that prenatal counseling is complex, and the needs are manifold, but digital information provision does improve knowledge, shortens the counselling and remains the same level of satisfaction. Adding an instructional video to the consultation made it more time-consuming for the participants, with equal satisfaction and more knowledge gain. Although the counsellor was equally satisfied, it seems that their satisfaction is not influenced by the duration of the counselling. These 7 minutes per counselling can save a lot of money. In the Netherlands we have 170.000 woman per year that get pregnant. If we would counsel all these woman, 7 minutes shorter, at a 150 euro/hour fee, we save 2,9 million euro per year in the Netherlands. The knowledge gain is in line with other studies that show the added value of computer-assisted instruction in improving patients’ knowledge and satisfaction. Yee et al show that an interactive computer aid can convey the relevant information about genetic screening and diagnostic concepts, although they did not evaluate this within the context of counselling and their focus was on knowledge only.

The fact that adding interactivity to the video did not improve the experience is also in line with other studies showing inconsistent results from adding interactivity. A recent study by Logan et all shows that an instructional video benefits most from segmenting and least by adding practice and inserting pauses. The difference between our instructional and interactive video mainly contains segmenting, practices and pauses. The lack of effect is therefore in line with Logan et all. We believe future studies will have additional value when evaluating more of these variations. Another reason could be the already fulfilled engagement. A theory behind the added value of interactivity is the added engagement of the learner towards the topic. Perhaps the emotional investment in the topic and the “just-in-time” manner of the video, already engaged the participant, after which the interactivity is of little added value. Counselling will always need a face-to-face element, due to legislation and ethical considerations, but optimizing the patient’s prior knowledge, understanding and attitude can improve counselling. Finding the right format will be the challenge. Whether this will be interactive information aids, virtual reality question-and-answer environments or any other variant of digital information provision should be further evaluated. What should also be addressed are the actual implementation challenges of an intervention like this. Can people watch the video at home? Will they all actually
watch it? How can the counsellor prevent an unexpected delay when people have not watched it? These are questions that can be evaluated by a follow-up implementation study.

**Conclusion**

Our study shows that adding an instructional video to patient counselling is of added value to the time consumption of the counsellor, save costs, improves patient knowledge but maintains the satisfaction of both patient and counsellor. Adding interactivity to the instructional video did not change these effects. There is still much to learn about patient counselling and education. Other methods should be evaluated, other educational strategies can be used, and further evaluation studies should always include an instructional design evaluation as well.

What’s already known about this topic?

- **Counselling for prenatal screening is a complex process containing education, information and evaluation in order to make a well-considered decision**

- **Counselling for prenatal screening has an increase in inter-doctor variation and unpredictable time consumption.**

What does this study add?

- **Digital information provision added to face-to-face counselling shortens the counsellors time significantly without decreasing satisfaction and even improving knowledge**

- **Shortening the counsellors time consumption can be a very cost-effective way of saving time or increasing patient care**

- **Adding interactivity to patient information provision does not improve knowledge or satisfaction**
Contribution to authorship

RDL, PCB and CJG conceptualised the trial. Together with SFH, AMS and FS they drafted the protocol, which was revised by all authors. RDL, SFH and AMS included all participants. RDL, SFH and JH performed the statistical analysis. All authors read and approved the final manuscript.

Details of ethical approval

The study protocol was approved by the medical ethical committee of the VU University, Amsterdam, the Netherlands with file ID 398 on August 2th 2017.

Funding

There was no funding for this study.
References


24. van der Vleuten CPM, Driessen EW. What would happen to education if we take education evidence seriously? Perspectives on Medical Education 2014;3:222-32.
General discussion

English Summary

Nederlandse samenvatting

Dankwoord

Curriculum vitae
Background

In this thesis, we set out to identify which Instructional Design (ID) elements are important to give e-learning added value over face-to-face learning. In the last years, e-learning has taken an increasingly prominent place in medical education, making the evaluation and awareness of ID even more important [1]. The focus of added value may be on the learning aims or the ID of the e-learning [2]. It is notable that even in 2019, after over 30 years of digital forms of education, it remains debatable what the added value of e-learning is [3].

To gain a better understanding of those aspects that can increase or decrease the efficiency and effectiveness of postgraduate medical e-learning (PGMeL), we performed literature reviews to identify which evaluation instruments and outcomes are used. We then identified a series of 72 indicators, which was reduced to 37 items which are potentially adding value to PGMeL, and proposed how these indicators might be used in its design, implementation, and evaluation.

In this chapter we will provide a reiteration of the main finding by answering the central research questions, and give our perspective on the current situation of PGMeL based on these answers. We will discuss the strengths and limitations of this thesis and conclude with practical implications, future research, and our perspective on the future of digital learning.

Main findings and answers to the central research questions

This thesis posed five central research questions, introduced in chapter one. Each question will be explored and answered by providing the main findings of this thesis. Please see figure one for an overview of these chapters.
Figure 1 – an overview of three domains with 37 items leading to an ID model, implementation factors, and an evaluation survey.

Legend for figure 1: three domains (Motivate, Learn, and Apply, chapter four) contain 37 items (chapter five). These items form the basis for an ID model (chapter six), an evaluation instrument (chapter seven), and a series of 11 implementation influencers (chapter eight).
1. What instruments or outcomes are currently used to evaluate postgraduate medical e-learning?

In chapter two, we performed a thematic analysis of 418 studies that showed that the learning aims of PGMeL are knowledge, skills, and attitude/behavior. The review showed that the most frequently used secondary outcomes were satisfaction, motivation, efficiency, and usefulness. Twelve instruments were identified that measure a specific outcome (for example, laparoscopic skills) and only 4% of the studies used any form of e-learning design or educational method for either the creation or evaluation of PGMeL. None of these methods were specific to postgraduate medical education. Four educational models are most frequently mentioned. Two are aimed at digital education in general, namely Gagne’s instructional design and Mayer’s multimedia learning. The other two are aimed at medical education, but not at digital education per se: Kirkpatrick’s hierarchy and Kern’s six-step curriculum development.

Although the literature reveals that the use of ID models and educational theories in e-learning is fundamental [4-6], only 4% of the studies included in the review given in chapter two actually describe the use of such models. A review in 2016, looking at medical education in general, found that 7.2% (18/251) of papers described ID principles [7]. It could very well be that most published papers just did not explain which model they used for the ID, or that most PGMeL is based on local expertise without any reflection on learning theories. This does not mean that those e-learning modules are poorly designed; we just do not know how they are designed. There are two pitfalls in this trend. First of all, it causes doubt as to quality. If the modules had been designed differently, perhaps they would have been more effective or efficient. Secondly, because we have no insight into the design methods, we cannot compare design models with effectiveness. If those 418 studies had explained their design process, we could have evaluated which method led to more efficient and effective e-learning. Papers that do describe the design, for example the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model, find it very useful and of great importance [8]. Chapter two emphasizes the need to gain more insight into which indicators can be used to develop such a model.
2. Which indicators are known in the current literature that can set best practice norms for the creation of postgraduate medical e-learning?

To answer the question above, an integrative literature review (chapter three of this thesis) showed that over 36 papers, 72 unique indicators are mentioned. These indicators were organized in six domains, showing that only one domain contains the content of the e-learning (content), while the others focus on the ID (preparation, design/system, communication, assessment, and maintenance). The model was based partly on a technology ISO standard, but also on CLT and multimedia principles.

A paradox now arises. This study shows that content is only a small part of e-learning as a whole, while other studies show that content is the most evaluated topic [3, 7, 9]. Although other reviews have focused on technology [10], software [11], assessment [12], and communication [13], they have not provided a diverse list of different domains. The 72 indicators from chapter three are now identified, but it is not yet known which are more or less relevant to creating an ID model or evaluation instrument. We can now use these items to weight their relevance with the main stakeholders and either remove indicators or add new ones, aiming to compile a list of indicators that can be used for evaluation or ID.
3. Which of these indicators are acknowledged and which are missing according to the main stakeholders, namely users, educational experts, and content creators?

The previous question led us to 72 unique items, but we did not yet know their relevance. To evaluate this, in chapter four we asked users, didactic experts, and expert creators what they found to be the most relevant factors. The focus group discussions provided us with 34 items. King’s template analysis led to three subjects (motivate, learn, and apply), containing six domains (motivators, learning enhancers, real world translators, barriers, learning discouragers, and poor preparation). To determine relevance even further, we then set out to perform an international Delphi using those 34 items in chapter five. After adding and removing some items, the Delphi provided us with 37 items in the three domains: motivate, learn, and apply. These 37 items are now the results of a literature study, focus group discussions with residents, didactic experts, and experienced creators, and a Delphi with international residents and medical e-learning experts.

It is, to our knowledge, the first time that a series of items has been initiated from three studies, taking into account several stakeholders and focusing on PGMeL. Other models use comparable items in their ID models. The ADDIE model [14], Kern’s curriculum design model [15], and Gagne’s nine events of instruction [16] come close to covering the same domains but they are all based on expert opinions and reviews only. Moreover, they do not focus on postgraduate medical education or cover the extent of all 37 items. We now know which 37 items are relevant for PGMeL. Therefore, we can use them for a theory- and evidence-based ID model and evaluation instrument.
4. How can these indicators be used to design and evaluate postgraduate medical e-learning?

The 37 items described above were used to create an ID model (chapter six) and an evaluation survey (chapter eight). Chapter six provides an eight-step model in which all 37 items are used to guide a digital educator in his/her path in designing PGMeL. The purpose of the model is not to make the creation of PGMeL a step-by-step experience. Creating an educational instrument is much too complex for this. However, it helps creators to consider all 37 relevant items and decide whether or not they are applicable to their e-learning. The eight steps are proposed building blocks, aiming at guiding creators rather than leading them. The eight steps are: (1) who, why, what; (2) educate; (3) real-world translation; (4) technology; (5) team; (6) budget; and (7) time and timeline and (8) evaluate. Chapter six compares this eight-step model with other available models, but none contain the same items, as concluded when answering the previous question.

Using the same 37 items, chapter eight followed the AMEE (Association for Medical Education in Europe) seven-step process in creating an evaluation instrument for medical education. In five methodological steps, a survey is validated that we called the postgraduate Medical E-learning Evaluation Survey (MEES) in order to provide the current e-learning landscape with a validated evaluation instrument. The MEES was thought to be useful, understandable, and of added value according to the creators of three e-learning modules. This is in line with a paper by Ruggeri et al who proposed an evaluation model for e-learning in health care [17]; however, this did not lead to a validated questionnaire, and nor is it aimed at postgraduate education. Specific questionnaires are available, but they look at a certain outcome (for example, laparoscopic skills [18] or the experience of stress [19]). The MEES allows digital educators to add a new dimension to their evaluation, namely, the evaluation of the ID. This enables them to improve the ID when needed, but also provides insight as to possible reasons for disappointing results in the effectiveness and efficiency of their e-learning.
5. When this e-learning is created, which factors influence implementation and how can the team prepare for successful implementation?

After designing PGMeL, implementation needs to follow. A series of 10 semi-structured interviews with experienced e-learning creators demonstrated that 11 factors influence the implementation of PGMeL (described in chapter seven). These factors were divided into three groups, named after the people who influence them the most, thus being creator-, organization-, and learner-dependent factors. We compared these factors with two different innovation models, namely Rogers’ diffusion of innovation and Kotter’s eight steps of change management. There was a great overlap between them, but the factors of chapter seven remain unique for PGMeL.

The emphasis on organization-dependent factors is in line with O’Doherty et al, who suggest that one of the key barriers to implementation is the absence of institutional strategies [1]. The negative attitudes of all involved and lack of support create a serious problem. Chapter seven supports the importance of these organizational factors. Using the results of chapter seven, the creators of PGMeL can properly prepare the implementation phase and, it is hoped, prevent barriers like those which O’Doherty et al describe.

Chapter nine demonstrates the effect of singling out just one of the items: interactivity. When interactivity is added to patient education for a complicated subject such as prenatal diagnosis, there is no effect. There was a significant effect on the duration of consultation and therefore on costs, but only when comparing face-to-face with digital information provision. It makes little difference what form digital information is provided in, which is in line with other studies.

Chapter nine shows that the effect of the 37 items is expected to be highest when they are combined, and negligible when evaluated one by one. Fiorella et al also provided this insight, when focusing on video only [20]. We believe this also supports the argument that the ID should be a fundamental part of e-learning and needs specific attention in the evaluation of PGMeL.
Perspective on postgraduate medical e-learning

After answering the five research questions in the above paragraphs, we now postulate a perspective of PGMeL in 2019. PGMeL has been used for over 30 years in different forms. The focus of evaluation and development has been on the effect of the primary learning aim: knowledge, skills, and/or attitude/behavior. The conclusions are still contradictory. This thesis shows that to date, little attention has been paid to addressing ID in the evaluation. It also shows that ID is a very important factor influencing the efficiency and effectivity of e-learning.

We believe that we can now accept three statements:

1. E-learning is not the future: it is now, and it will only increase;

2. E-learning is, if applied properly, non-inferior to other forms of education, yet has not been proven superior; and

3. ID is fundamental in the evaluation of the effectiveness and efficiency of e-learning and should be theory-based, aimed at the most specific target audience possible.

We know e-learning will have an effect on people. Asking if adults can learn from e-learning is, in our opinion, no longer a relevant question. Standalone e-learning, blended learning, and flipped classroom education are receiving the attention they need [9, 21] and we postulate that this should become a variation of the standard, not an experiment. The current literature shows us that the time has come to evaluate e-learning not by primary learning aim alone. Adults learn from almost anything they are given, whether digital, analog, or face-to-face. This thesis shows that of the three subjects that form the basis of PGMeL (motivate, learn, and apply), the subject of ‘learn’ is the most studied and most evaluated, while ‘motivate’ and ‘apply’, are much less so, or even not evaluated at all.

We also know that targeting an audience as narrowly as possible will likely increase the effect on learning [22, 23]. Thus, we should not make an e-learning for graduates and give it to postgraduates to learn and evaluate. Following a sound ID model will lead creators to consider the target audience as specifically as possible. We believe that ID and the evaluating of any e-learning can be seen as a tree: the main branches are the more generic elements, that is, motivate, learn, and apply, while the leaves should be specific for the target audience. (See figure 2.)
Figure 2 – the tree of PGMeL design and evaluation

Legend for figure 2: the roots of the tree are built on experts, the science of this thesis, and learner needs. The tree divides into three branches: motivate, learn, and apply. The first two then provide room for barriers and motivators on one branch and learning enhancers and discouragers on the other. Finally, the leaves illustrate all the items considered by this thesis. The strength of this tree is that it can be applied to different target audiences simply by changing the leaves, preferably based on learning theories.
As a final part of our perspective on PGMeL, we would like to look back at the learning theories discussed in chapter one. As shown in chapter two, CLT and multimedia learning have been successfully used in PGMeL [24, 25]. Reed et al. used adult learning theory to distill instructional principles such as self-directed learning, practice sessions, and case-based discussions [26] and, although it seems that learning theories will keep evolving, these three theories are helpful in creating PGMeL. This thesis, on the other hand, suggests a new insight that is not mentioned in these theories: the aspect of lifestyle. Many people are familiar with learning styles, such as, inter alia, being a visualizer vs verbalizer or visual vs haptic learner. According to Cook, these learning styles are vague and empirical evidence is nonexistent to guide IDs [27]. What they do offer is the insight that there are benefits in stimulating learners in different ways, which is in accordance with multimedia learning. They also suggest an individual preference, although we believe they are mistaken in the direction they follow in this regard: in our opinion, this personal preference is due not to the sensory input (visual vs auditory), but to lifestyle. A surgical resident who is also mother of three children will certainly have different needs, motivators, and learning enhancers, to those of a younger, internal medicine resident who lives on campus. Therefore, we believe that learning theories should take into account the variations of individual lifestyles and the needs resulting from them.

Strengths and limitations of the thesis

There are three strengths and three limitations of this thesis, discussed below.

The first strength is the foundation used of the 37 ID items. The list of items is based on two reviews, a focus group study, and an international Delphi panel. The selection of these items from the original 72 was carried out in rigorous fashion, proving a high level of content validation. This methodology has been suggested as the ‘gold standard’ for content validation [28], which makes this list the first content-validated list of ID elements ever for PGMeL.

The second strength is the fact that we involved users from the beginning. E-learning is thought to be most effective when it is learner-centered [29]. This thesis involved learners as important stakeholders, not only in weighing the relevance of the first 72 items from chapter three (learners were part of the focus groups and made up half of the Delphi participants) but also in the evaluation of understandability (in the MEES validation), and among the implementation influencers (the learner-dependent factors). By involving the learners, in this case postgraduate residents, we aimed to attribute as much importance to the learners’ needs and wishes as to the expert opinions of creators and educational experts.

The third strength is that we aimed at a very specific targeted audience, namely medical postgraduates. This allowed us to focus on a group of adult learners who have
a comparable background, are within a comparable age segment, and have a great overlap in living style and learning needs. Not only does the previous literature suggest that a focused target audience will increase efficiency; it also allowed us to involve this group as an influential stakeholder, as described in the previous paragraph.

However, there are also limitations. The experts of the Delphi also mentioned the biggest of these: is all this really of added value to the current landscape of education literature? Would it not be possible simply to use models from other fields on PGMeL? We have attempted to show that the current literature rarely evaluates ID, and that creating an evaluation survey for PGMeL is of added value. On the other hand, ‘motivate, learn, and apply’ is very generic, and could be applied to almost any target audience. While starting with a specific target audience, therefore, we ended with three groups and thus may have lost the initial specificity.

A second limitation could be the cultural interpretation of the ID items. One great value of e-learning is the option to learn anytime, anywhere. This also allows an international target audience to be easily reached. The qualitative studies from this thesis were carried out with Western experts and residents exclusively. Therefore, we have no insight into the cross-cultural generalizability of our findings. Studies have shown that evaluation instruments can differ between cultures, and that a survey should not only be translated but also re-validated [30]. This means that in an MEES being used for an e-learning with a worldwide audience (as in chapter eight) a Dutch resident might give a different response from an Indonesian resident, even if both had similar experience, or vice versa. In an ideal situation, the validation steps from chapter eight should be carried out separately in each culture before they are used in the evaluation of an international e-learning.

Finally, it can be argued that this whole thesis is just an academic exercise, unrelated to the daily creation of PGMeL. The qualitative approach to the final 37 items might appear abstract. However, we have tried to bring these abstract items as close as possible to the creation process. The eight-step design model provides concrete steps, aimed at creators. The MEES provides a list of all items with a short explanation of the interpretation of each, aimed at PGMeL creators. In these chapters we have tried to give the 37 items a practical implication, creating theory- and evidence-based PGMeL.
Practical implications and future research

The practical implications of this thesis are found in chapter six (eight-step ID model), seven (11 implementation factors), and eight (MEES). Let us presume that a content expert has the need to share a certain skill. This person has developed a new surgical technique and believes it is important to share that technique. One way would be to write a step-by-step tutorial and publish it in a peer-reviewed journal, but our expert thinks that this skill is more suited to e-learning. This will be a creator-centered e-learning (see chapter eight). The first thing the creator needs is to design the e-learning. The eight-step model presented in chapter six can provide all the steps that should be considered. Some steps might not be applicable to this specific e-learning, but the steps will help to start the design. After assembling a team, creating the content, and defining learning aims and the target audience, chapter seven will help with the implementation. Here, the creator is confronted with the need for the organization to support the e-learning and reconsider the needed resources and cultural changes. The e-learning might even benefit from preparing the target audience for the importance of the e-learning, following Rogers’ diffusion of innovation model [31]. Chapter nine has shown that using interactivity alone is not enough to increase the educational value, so multiple strategies should be set. Finally, when the e-learning is implemented and used, the MEES can help the creator to evaluate it. Not only does the creator want to know if the learning aims are met, but also if the ID needs improvement and if the affordances are experienced in the way they were planned. This closes the loop, and the feedback can be used for version 2 of the e-learning, improving efficiency and effectiveness.
Having discussed the current position of PGMeL and how this thesis can have practical implications, we need to consider future research. As described above, there is enough evidence to show that e-learning can help adults learn. Therefore, we believe that evaluating the primary learning aim is important but should no longer be the main question. It is important to assure that the aims are met, but the added value of e-learning will lie elsewhere. There remains much to learn about the ‘motivate’ and ‘apply’ domains. A digital environment can greatly increase the motivational value of a learning experience [32]. Studies have been carried out using variants of digital learning instruments such as gamification, virtual reality, augmented reality, and simulation [33], although the outcomes of these studies are most frequently, still, the learning aim (see chapter two). We believe that the primary outcome of these studies should change. Learning aims remain important, but motivation or the ability to apply the new knowledge/skills/attitude/behavior in the real world should become the primary outcome: those are the outcomes that matter now. Future research could contain evaluation studies of new educational instruments, by evaluating cohorts or performing qualitative evaluation studies and randomized trials. Qualitative evaluation and theory-grounded methodologies can also focus on the development and evaluation of learning theories focused on digital learning and resident lifestyle and ID methods in usability and effectiveness.
Another important aspect is costs. Costs are rarely taken into account, and little is written about the development costs of PGMeL [34]. It is important to obtain a better insight into which aspects of PGMeL are most costly, which costs relate to one-time development, what maintenance costs are to be expected, and how many users are needed to break even or be more cost-effective than face-to-face learning. There is little to no transparency about this aspect of PGMeL, although studies are suggesting that a low-fidelity, high-quality e-learning might be the path to follow [35, 36].

Finally, since validation is a continuous process, both the MEES and the eight-step model should be further validated. There is a great opportunity to evaluate whether or not the MEES scores are correlated with satisfaction and achieving the learning aim (criterion validity), but also to see if the MEES can predict these outcomes from a pilot evaluation (predictive validity). There are other ways to continue the validation of the MEES, but we believe these two are relatively easy and of great value as regards the impact of the survey. The eight-step model should be evaluated on its usefulness for less experienced creators and its predictive value for an effective and efficient PGMeL module.
The future of e-learning

In chapter one we introduced the effect of the internet on e-learning. We are currently experiencing “Web 3.0”, and e-learning is benefitting from this intelligent and semantic web [37]. The use of virtual reality and A.I. is rising in fundamental topics such as anatomy [38]. Using e-learning in Massive Open Online Courses is becoming more a standard then a mere experiment [39]. However, this is the position in 2019: where will it go from here? The work conducted for this thesis provides a possible insight into what is to come. Although nobody can be sure, theorization is nevertheless possible and useful.

Wartmen and Combs said that medical education should move beyond foundational biomedical and clinical sciences and should teach the doctors of the future how to use intelligent tools such as large data sets, machine learning, and robots [40]. We strongly agree. The future of digital learning is not if we will use digital media, as we already do, but what we will teach.

The future of digital learning might not even be aimed at residents, but at machines. Machine learning is a method whereby humans teach computers how to solve clinical problems, and is already being used and successfully evaluated [41]. Machine learning still implies that the teacher (human) is teaching the machine, but the question arises as to what happens if the student surpasses the teacher. Deep learning is a methodology whereby we merely feed the machine data, and the machine learns by itself. After this we, as humans, can either simply use the new data (black box scenario) or learn from the machine which model it has developed (white box scenario) [42].

Nobody knows what will transpire in the future. However, Web 5.0 can provide us with an entirely new area of possibilities. The sensory, emotive aspects of Web 5.0 will allow computers to know how you feel, adapt accordingly, and transfer emotions back, as a human does. Taking this development into account, it may be that the future involves the following three scenarios:

1. Computers are used for facts, calculations, and associations. They have access to an infinite amount of data and can associate better than humans. Computers will diagnose for us. Patients will come to the doctor with a diagnosis which is more certain than a human doctor could make, to discuss the emotional consequences and treatment.

2. Computers can reflect emotions and mental support. Once we are accustomed to interacting with electronic devices (as we are now with Siri and Google home), we will be comfortable to share more and more. Computers are always there for you, have all the time in the world, and reflect your emotions better.
than doctors working to tight 15-minute schedules. The computer will take over the consultation. Patients will come to the doctor only to discuss treatment.

3. Computers gain trust over humans who are prone to error. Once we are comfortable with computers being an essential part of our health care system, we trust their judgement. Furthermore, just as we believe that self-driving cars are much safer than human-controlled machines, we start to distrust human doctors. Autonomous robots are less likely to err, and medication is produced and delivered without any mistakes. Computers have now taken over treatment.

The future of digital learning is now. We need to start the debate on what we are going to teach our future doctors about the digital world, followed by how we are going to teach them.

Final remarks

Let us return to the resident described at the beginning of chapter one. What would her learning path look like in 2050? Clearly nobody knows, but we can postulate. The laparoscopy guru had taken the time to create an e-learning about the retroperitoneum. The last 100 hysterectomy videos were uploaded into a database and the deep learning program distilled the most important steps and pitfalls. The resident experiences the most important parts of the 100 videos, selected by the computer, in a simulator aimed at experienced gynecology residents. After not only viewing the videos, but experiencing and participating in the surgeries, she travels to the experts for her tutorial. When she arrives, she has already performed 50 virtual surgeries, passed three tests, and been evaluated by the computer system to be at 80% of the efficiency of the experts. Given these numbers, she is allowed to experience three days of live surgery, observing very little difference between the virtual lessons and the real world. Satisfied with her learned skills, she now feels comfortable performing the surgery herself.

On the other hand, perhaps she went there merely to see the autonomous DaVinci robot perform the surgery.
References


42. Zhang L, Tan J, Han D, Zhu H: From machine learning to deep learning: progress in 
1680-1685
English Summary
English Summary

E-learning is becoming a mainstream part of PGMeL. Yet the results of evaluation studies are contradictory in terms of efficiency and effectiveness. This thesis is concerned with the question of which factors influence the success of PGMeL. PGMeL is frequently evaluated on learning aims (knowledge, skills, or attitude/behavior), but little attention is given to the ID of the e-learning. The ID is the link between the science of how people learn and daily practice as a process for designing instruction based on empirical principles. In other words, it is the way the e-learning is designed, which technical aspects are used, which pedagogical elements are incorporated, etc. The features of e-learning that provide potential for action are called affordances, and it is these affordances which are the result of the ID, as focused on in this thesis.

Chapter 1 describes the current landscape of PGMeL evaluation studies. It takes the reader along a path introducing the development of the ‘e’ of e-learning (the history of the internet and its influence on e-learning) and the most prominent learning models used in PGMeL. This chapter does not aim to provide in-depth insight into learning and all its psychological models, but introduces the basics of CLT, multimedia learning, and adult learning theory. These three models help to give insight into the theoretical rationale behind certain ID choices in PGMeL. The development of the internet has had a great influence on possible affordances in e-learning. Aspects such as communication with peers, which would be beneficial based on adult learning theory, became possible because of the evolution of the internet. After covering these elements of e-learning in the introduction, we present the five research questions that this thesis attempts to answer. The first concerns which instruments, or outcomes, are currently used to evaluate PHMeL. The insight gained into means of evaluation allows us to ask ourselves which indicators are known in the current literature that can set best practice norms for PGMeL. These indicators from the literature further raise the question of which are acknowledged and which are missing according to the main stakeholders, namely users, educational experts, and content creators. Having created a list of indicators, we then ask how they can be used to evaluate and design PGMeL. Finally, when this e-learning is created, we ask which factors influence implementation, and how a team can prepare for successful implementation.

Chapter 2 described a systematic review aiming to identify and compare the outcomes and methods used to evaluate PGMeL. The initial search identified 5,973 articles, of which 418 were used for our analysis. The performed thematic analysis showed that the most frequently used learning aims of PGMeL are knowledge, skills, and attitude/behavior. Twelve instruments were used to evaluate a specific outcome, such as laparoscopic surgery skills. Only 4% (19/418) of the papers used an ID model
or theory for the creation or evaluation of the e-learning. The most frequently used models were Kirkpatrick’s hierarchy, Gagne’s instructional design, the Heidelberg inventory, Kern’s curriculum development steps, and scales based on CLT. Chapter two provides short introductions to all these models, none of which is specifically aimed at medical postgraduates. This study shows that, apart from the learning aim, the most evaluated aspects are satisfaction, motivation, efficiency, and usefulness. Design models are rarely used and, when they are, are not aimed at PGMeL. This shows that the current literature has not yet reached any form of consensus about which aspects of PGMeL should be evaluated. There seems a great need for an evaluation tool that evaluates ID instead of learning aims, and which is properly constructed and validated.

Chapter 3 reports on an integrative literature review of the current literature performed with the aim of seeking out and identifying quality indicators for PGMeL. The search resulted in 11,093 papers, and a selection based on titles, abstracts, and full texts resulted in 36 papers being used for the analysis. From these papers, 72 unique indicators originated. These indicators were organized into six domains: content, preparation, design/system, communication, assessment, and maintenance. We called this model the Postgraduate Medical E-learning model (postgraduate ME model), which is partially based on the ISO-19796 standard and drew on cognitive load principles. Although most evaluation studies focus on the content of PGMeL, this study shows that content is only a part of the educational experience. The five other domains focus on ID, placing emphasis on the importance of evaluating this aspect as well.

Chapter 4 further explores the needs and expectations of learners, educational experts in postgraduate medical education, and commercial e-learning designers. Three focus group discussions with these were performed, and the verbatim transcribed recordings were analyzed using King’s template analysis. Initially, 34 items arose, which were placed into an initial template, based on the ME model of chapter three. The final template consisted of three domains of positive influencers (motivators, learning enhancers, and real-world translation) and three negative parameters (barriers, learning discouragers, and poor preparation). The interpretation of these domains showed three general subjects which form the basis of PGMeL: motivate, learn, and apply. These domains provide a foundation for educational tools, and the individual categories can be adapted to fit the target audience. So far, however, these individual categories had not been validated for PGMeL and these domains had not been proven useful in the daily practice of creating e-learning.
Chapter 5 continues with the 34 items from chapter three. We performed a Delphi procedure with a group of 13 international educational experts and 10 experienced users. The Delphi started with 57 items as a result of the literature study (chapter two) and focus group discussions (chapter four). Consensus was reached when a rate of agreement of more than two thirds was achieved. After two rounds, 72 items were addressed, of which 37 were accepted. These items were divided into the same three domains as chapter four: motivate, learn, and apply. The 37 items from this chapter could now be used to create an ID model and evaluation instrument for PGMeL.

Chapter 6 aims to create an empirical ID model for PGMeL and compare it with the existing evaluation models discussed in chapter two. Analogous to the intervention mapping model, we arranged the 37 items from chapter five into eight chronological steps, these being proposed building blocks which aim to guide creators, not lead them. The eight steps are (1) who, why, what; (2) educate; (3) real-world translation; (4) technology; (5) team; (6) budget; and (7) time and timeline and (8) evaluation. When comparing these steps to other models such as the ADDIE, 4C/ID, Kerns model, Gagne’s nine events, ASSURE, and Merrill and Kemp’s model, no other was as complete, and neither were any other models aimed at medical postgraduates. Chapter six presented the first evidence- and theory-based ID model aimed at PGMeL. Although certain steps are more robust and have a deeper theoretical background in current research (such as Education), others (such as Budget) have been barely touched upon and should be investigated more thoroughly in order that proper guidelines may also be provided for them.

Chapter 7 discussed the next part of ID, namely implementation. We performed a series of 10 semi-structured interviews with experienced e-learning creators, after which we carried out a thematic analysis to name and describe categories and themes. Although this was not the objective of the study, the participants stressed the importance of a definition of ‘success’. Associated with this definition were: reaching your target audience, achieving learning aims, satisfying your audience, and maintaining continuity. The thematic analysis revealed 11 categories, which were divided into three groups and named after the people who influence them the most, thus being creator-, organization-, and learner-dependent factors. The first theme (creator-dependent factors) contained the categories of learning aim, pedagogical strategies, content expertise, evaluation, and motivational pathway. The second (organization-dependent factors) encompassed management support, resources, and culture. The last (learner-dependent factors) consisted of technology, motivators/barriers, and value. We compared these factors with two different innovation models (Rogers’ diffusion of innovation and Kotter’s eight steps of change management) and found an (incomplete) overlap between them; however, the factors of this chapter remain unique to PGMeL. Future studies can both evaluate the use of these
innovation models in creating PGMcL and assess the effect of the organizational categories in greater depth.

Chapter 8 follows the AMEE (Association for Medical Education in Europe) seven-step process in creating an evaluation instrument for medical education. In five steps, this study aimed to create and validate a survey that we called the MEES. The first step was creating the survey from the 37 items from chapter five, followed by testing readability and question interpretation. The third step was adjusting, rewriting, and translating the survey. This was followed by gathering filled-out surveys from three international e-learning modules, after which we finally evaluated the usefulness, understandability, and added value of the survey by focus group discussion with the e-learning creators. A total of 158 responses led to three focus group discussions with a total of ten participants. The usefulness of the MEES was much appreciated, understandability was good, and added value was high. Four items needed additional explanation by the authors, and a Creators’ Manual was created at their request. We briefly discussed the number of responses needed and concluded that more is better; ultimately, however, one has to work with what is available. The next steps would be to see whether improvement can be measured by using the MEES, and to continue to work on the end understandability in different languages and cultural groups.

Chapter 9 demonstrates the effect of evaluating just one item of the 37 items of chapter five, namely interactivity. This chapter aims to evaluate face-to-face information provision in patient counselling for prenatal screening compared to interactive and non-interactive video information provision. We performed a prospective, non-inferiority, cluster-randomized, controlled trial comparing those three groups. One hundred and forty-one women were included, randomized, and analyzed. The baseline characteristics were comparable. The intervention group (digital information provision) was non-inferior compared to the control group (face-to-face) in satisfaction. The knowledge grade was significantly increased in the intervention group and the duration of the counselling following the information provision was significantly shorter, at 23 minutes versus 16 minutes, to the benefit of the video group. This also implies a cost-benefit from this type of information provision. When we compared the interactive with the non-interactive video group, there was no difference in outcomes. This is in line with other studies which suggest that instructional videos benefit most from a combination of segmenting, practices, and pauses. This chapter demonstrated that the added value of e-learning will most probably be found not in single affordances, but in the combination of many.
Chapter 10 combines the previous research findings to answer the research questions from chapter one. Each question is answered, and the answer is put into the perspective of this thesis or the current literature. Placing the findings from this thesis in the perspective of the current view on PGMeL, we propose that e-learning is not the future, but it is now, and will only increase; furthermore, it must now be accepted that e-learning, if designed properly, is non-inferior to other forms of education and that ID is fundamental in the future evaluation of the effectiveness and efficiency of PGMeL. PGMeL should be theory-based and aimed at the most specific target audience available. This thesis builds on the fundamentals of motivate, learn, and apply, which should be complemented by specific affordances based on learning theory and aimed at the right target audience. We discuss the effect of learning theories on these affordances and suggest that the development of learning theory should take into account the lifestyle of the learner.

Subsequently, the strengths and limitations of this thesis are discussed. The strengths lie in the rigorous methodology which led to the identification of 37 items in chapter five. Other strengths are the involvement of learners from the beginning of this exploration, and being aimed at a specific target audience: medical postgraduates. The limitations pertain to the relatively generic nature of motivate, learn, and apply, and the lack of cross-cultural validation. E-learning is very easily spread, but there is no evidence as to whether the findings of this thesis are also applicable in non-Western cultures. The practical implications of this thesis can be found in chapter six (eight-step ID model), chapter seven (11 implementation factors), and chapter eight (MEES). We hope that any medical educator who plans to create an e-learning aimed at medical postgraduates can use these chapters to design their e-learning, create it, and evaluate their initial design. Finally, we discuss the future of e-learning, discussing less how we will make use of digital media (as we already do) and more what we will teach in an exponentially increased world of digitalization and automatization.
Nederlandse samenvatting
Nederlandse samenvatting

E-learning is een toenemend belangrijk onderdeel van de opleiding tot medisch specialist. Echter zijn de resultaten van evaluaties over de effectiviteit van e-learning tegenstrijdig. E-learning wordt vaak vergeleken met andere vormen van onderwijs en de uitkomsten van deze onderzoeken zijn in tegenspraak met elkaar. Dit proefschrift gaat over de vraag welke factoren het succes van e-learning voor medische vervolgonderwijs beïnvloeden. E-learning wordt vaak geëvalueerd op leerdoelen (kennis, vaardigheden of houding/gedrag) en er wordt weinig aandacht besteed aan het instructieontwerp van de e-learning. Het instructieontwerp is de link tussen de wetenschap van hoe mensen leren en de dagelijkse praktijk van het ontwerpen van e-learning op basis van empirische leerprincipes. Met andere woorden: de manier waarop de e-learning er uitziet, welke technische aspecten worden gebruikt, welke pedagogische elementen zijn opgenomen, enzovoort. E-learning heeft bepaalde mogelijkheden voor actie (bijvoorbeeld communicatie, interactie, etc). Deze mogelijkheden worden affordances genoemd. Het zijn deze affordances die het resultaat zijn van het instructieontwerp, waar we ons in dit proefschrift op zullen richten.

Hoofdstuk 1 beschrijft het huidige landschap van onderzoek over e-learning in de medische vervolgonderwijs. Het neemt de lezer mee op een pad dat de ontwikkeling van de "e" van e-learning (de geschiedenis van het internet en de invloed van e-learning) introduceert. Daarna gaat het verder met een korte samenvatting van de meest prominente leermodellen die in e-learning worden gebruikt. Dit hoofdstuk is niet bedoeld om een compleet inzicht te geven in de theorie van het leren en al zijn psychologische modellen, maar het introduceert de basis van de "Cognitive Load Theory", multimedia leren en “adult learning". Deze drie modellen helpen om inzicht te geven in de theoretische grondgedachte achter bepaalde ontwerp keuzes in e-learning. De ontwikkeling van internet heeft een grote invloed op de mogelijke voordelen van e-learning. Affordances zoals communicatie met leeftijdgenoten, die gunstig zou zijn op basis van “adult learning” theorie, werd mogelijk door de evolutie van het internet. Na de introductie van het internet en leertheorieën, presenteren we de vijf onderzoeksvragen die dit proefschrift zal proberen te beantwoorden. De eerste vraag is welke instrumenten of uitkomsten momenteel worden gebruikt om e-learning voor de medische vervolgonderwijs te evalueren. Nadat we inzicht hebben in de verschillende manieren van evaluatie, vragen we ons ten tweede af, welke indicatoren in de huidige literatuur bekend zijn om de kwaliteit van deze vorm van e-learning te bepalen. Ten derde vragen we ons af welke indicatoren worden erkend en welke er ontbreken volgens de belangrijkste belanghebbenden, namelijk: gebruikers, onderwijskundigen en de makers van e-learning. De vierde vraag is hoe deze indicatoren gebruikt kunnen worden in het ontwerp en evaluatie. En tot slot, wanneer
Hoe wordt deze e-learning gecreëerd, welke factoren beïnvloeden de implementatie en hoe kan het team zich voorbereiden op een succesvolle implementatie?

Hoofdstuk 2 beschrijft een systematische review, gericht op het identificeren en vergelijken van de uitkomsten en methoden die gebruikt worden om e-learning te evalueren. De zoekopdracht identificeerde 5.973 artikelen, waarvan 418 werden gebruikt voor onze analyse. Uit de thematische analyse bleek dat de meest gebruikte leerdoelen (1) kennis, (2) vaardigheden en (3) houding/gedrag zijn. Twaalf soorten vragenlijsten werden gebruikt om een specifieke uitkomst te evalueren, zoals vaardigheden op het gebied van laparoscopische chirurgie. Slechts 4% (19/418) van de artikelen gebruikte een instructieontwerp of theorie voor het maken of evalueren van de e-learning. De meest gebruikte modellen waren de Kirkpatricks-hiërarchie, Gagne instructieontwerp, de inventaris van Heidelberg, de ontwikkelingsstappen van Kern's curriculum en enkele schalen op basis van de “Cognitive Load Theory”. Hoofdstuk twee biedt korte introducties in al deze modellen. Geen van hen is specifiek gericht op medische vervolgopleiding. Deze studie laat zien dat, afgezien van het leerdoel; tevredenheid, motivatie, efficiëntie en bruikbaarheid de meest geëvalueerde aspecten zijn. Instructieontwerpen worden zelden gebruikt en zijn niet gericht op de medische vervolgopleiding. Het laat zien dat er in de huidige literatuur nog geen consensus is bereikt over welke aspecten van e-learning geëvalueerd zouden moeten worden. Er lijkt een grote behoefte te bestaan aan een gevalideerde evaluatievragenlijst die het instructieontwerp evalueert, in plaats van leerdoelen.

Hoofdstuk 3 rapporteert een “integrative review” dat de huidige literatuur doorzoekt om kwaliteitsindicatoren voor e-learning te identificeren. De zoekopdracht resulteerde in 11.093 artikelen en selectie op basis van titels, samenvattingen en volledige teksten resulteerde in 36 artikelen die voor de analyse werden gebruikt. Uit deze artikelen werden 72 unieke indicatoren gehaald. Deze indicatoren waren georganiseerd in zes domeinen: inhoud, voorbereiding, ontwerp, communicatie, beoordeling en onderhoud. We noemden dit model het “Postgraduate Medical E-learning-model”, dat gedeeltelijk gebaseerd is op de ISO-19796-standaard en berust op principes van “Cognitive Load Theory”. Hoewel de meeste evaluatiestudies zich richten op de inhoud van e-learning, toont deze studie aan dat inhoud slechts een deel van de onderwijservaring is. De vijf andere domeinen richten zich op instructieontwerp, waarbij de nadruk wordt gelegd op het belang van het evalueren van deze aspecten.
Hoofdstuk 4 onderzoekt wat de behoefte en verwachtingen van medisch specialisten in opleiding, onderwijsexperts en commerciële e-learning ontwerpers zijn. Dit is onderzocht door drie focusgroep discussies met deze groepen uit te voeren. De opnames werden uitgeschreven en geanalyseerd middels een methode die King’s template-analyse heet. Aanvankelijk werden er 34 items gevonden die belangrijk werden geacht voor e-learning. Deze zijn in een originele sjabloon werden geplaatst, welke is gebaseerd op het ME-model van hoofdstuk drie. Het uiteindelijke sjabloon bestond uit drie domeinen van positieve invloed (motivatoren, leerversterkers en vertaling naar de echte wereld) en drie met negatieve invloed (barrières, leerontduikers en slechte voorbereiding). De interpretatie van deze domeinen toonde drie algemene onderwerpen die de basis vormen van e-learning: motiveren, leren en toepassen. Deze domeinen vormen een fundament voor educatieve hulpmiddelen en de individuele categorieën kunnen worden aangepast aan de doelgroep. Maar tot nu toe zijn deze individuele categorieën nog niet gevalideerd voor e-learning voor de medische vervolgopleiding en deze domeinen zijn nog niet nuttig bewezen in de dagelijkse praktijk.

Hoofdstuk 5 gaat verder met de 34 items uit hoofdstuk vier. We voerden een Delphi-procedure uit met een groep van 13 internationale onderwijsexperts en 10 ervaren gebruikers. De Delphi startte met 57 items als resultaat van de literatuurstudie (hoofdstuk drie) en focusgroep discussies (hoofdstuk vier). Consensus werd bereikt bij een mate van overeenstemming van meer dan twee derde. Na twee rondes waren 72 items beoordeeld en werden er 37 items geaccepteerd. Deze items waren verdeeld in dezelfde drie domeinen uit hoofdstuk vier; motiveren, leren en toepassen. De 37 items uit dit hoofdstuk kunnen nu worden gebruikt om een instructiemodel en evaluatie-instrument voor e-learning in de medische vervolgopleiding te maken.

Hoofdstuk 6 was gericht op het creëren van een empirisch, instructief ontwerpmodel voor e-learning in de medische vervolgopleiding. Vervolgens wordt dit vergeleken met bestaande evaluatiemodellen uit hoofdstuk twee. Analoog aan het “intervention mapping model”, rangschikten we de 37 items uit hoofdstuk vijf in acht chronologische stappen. De acht stappen zijn bouwstenen met als doel de makers van e-learning slechts te sturen en niet te leiden. De acht stappen zijn (1) wie, waarom, wat, (2) opleiden, (3) vertalen naar de echte wereld, (4) technologie, (5) team, (6) budget (7) tijd en tijdslijn en (8) evalueren. Bij het vergelijken van deze stappen met andere modellen, (de ADDIE, 4C/ID, Kerns-model, Gagne negen events, ASSURE, Merrill en Kemps-model) was geen enkel ander zo compleet, en waren er ook geen andere modellen gericht op de medische vervolgopleiding. Hoofdstuk zes presenteert het eerste op bewijs- en theorie gebaseerde instructiemodel, gericht op e-learning voor de medische vervolgopleiding. Hoewel bepaalde stappen robuuster zijn en een diepere theoretische achtergrond hebben in de huidige literatuur (zoals onderwijs), zijn
anderen (zoals de begroting) nauwelijks onderzocht en moeten ze grondiger worden bekijken.

**Hoofdstuk 7** bespreekt het volgende deel van het instructieontwerp, namelijk de implementatie. We voerden een reeks van 10 semigestructureerde interviews uit met ervaren e-learning makers. Hierna hebben we een thematische analyse uitgevoerd om categorieën en thema's te benoemen en te beschrijven. Hoewel dit niet het doel van het onderzoek was, benadrukten de deelnemers het belang van een definitie van "succes". Bij deze definitie horen: het bereiken van de doelgroep en leerdoelen, het tevreden stellen van de doelgroep en het handhaven van de continuïteit. De thematische analyse onthulde 11 categorieën. Deze categorieën waren onderscheiden in drie groepen, genoemd naar de mensen die deze het meest beïnvloedden: maker, organisatie en leerling-afhankelijke factoren. Het eerste thema (maker-afhankelijke factoren) bevatte de categorieën leerdoelen, pedagogische strategieën, inhoudelijke expertise, evaluatie en motivatie van de maker. De tweede (organisatie-afhankelijke factoren) omvatten managementondersteuning, middelen en cultuur. De laatste (leerling-afhankelijke factoren) bestaan uit technologie, motivatie/barrières en waarde. We hebben deze factoren vergeleken met twee verschillende innovatiemodellen ("Rogers diffusion of innovation" en de acht stappen van Kotter voor verandermanagement) en vonden een (onvolledige) overlap. Echter blijven de factoren in dit hoofdstuk uniek voor e-learning voor de medische vervolgopleiding. Toekomstige studies zouden het gebruik van deze innovatiemodellen kunnen evalueren en het effect van de organisatorische categorieën grondiger beoordelen.

**Hoofdstuk 8** volgde het zeven stappen proces van de AMEE (Association for Medical Education in Europe) om een evaluatie-instrument voor e-learning in de medische vervolgopleiding te creëren. In vijf stappen was deze studie gericht op het creëren en valideren van een vragenlijst die we de postdoctorale Medische Elearning Evaluation Survey (MEES) noemden. De eerste stap was het maken van de vragenlijst uit de 37 items uit hoofdstuk vijf, gevolgd door het testen van de leesbaarheid en vraginterpretatie. De derde stap was het aanpassen, herschrijven en vertalen van de vragenlijst. Dit werd gevolgd door het verzamelen van ingevulde vragenlijsten van drie internationale e-learning modules. Hierna zijn we het nut, de begrijpelijkheid en de toegevoegde waarde van deze vragenlijst gaan evalueren door focusgroep discussies met de makers van e-learning. Er waren in totaal 158 reacties, waarna drie focusgroep discussies met in totaal tien deelnemers volgde. Het nut van de MEES werd zeer gewaardeerd, de begrijpelijkheid was goed en de toegevoegde waarde was hoog. Vier items hadden aanvullende uitleg van de makers nodig en op hun verzoek werd een Creators Manual gemaakt. We bespreken hier het aantal benodigde antwoorden en concluderen dat meer beter is, maar uiteindelijk moet men werken met wat beschikbaar is. De volgende stappen zouden zijn om te zien of verbetering kan worden
gemeten met behulp van de MEES en om te blijven werken aan de uiteindelijke begrijpelijkheid in verschillende talen en culturele groepen.

**Hoofdstuk 9** toont het effect van het evalueren van slechts één item van de 37 items van hoofdstuk vijf, namelijk interactiviteit. Dit hoofdstuk was gericht op het evalueren van persoonlijke informatieverstrekking bij patiëntenvoorlichting voor prenatale screening, in vergelijking met interactieve en niet-interactieve video-informatieverstrekking. We voerden een prospectieve, non-inferiority, cluster-gerandomiseerde, gecontroleerde studie uit waarin deze drie groepen werden vergeleken. Er werden 141 vrouwen geïncludeerd, gerandomiseerd en geanalyseerd. De karakteristieken van alle groepen waren vergelijkbaar. De interventiegroep (video informatieverstrekking) was qua tevredenheid niet-inferieur in vergelijking met de controlegroep (persoonlijke informatieverstrekking). Kennis was significant verhoogd in de interventiegroep en de duur van de voorlichting na de video informatieverstrekking was aanzienlijk korter, namelijk 23 minuten vergeleken met 16 minuten ten voordele van de videogroep. Dit impliceert ook een kostenvoordeel van video informatieverstrekking. Toen we het interactieve video vergeleken met de niet-interactieve videogroep, was er geen verschil in uitkomsten. Dit komt overeen met andere onderzoeken die suggereren dat instructievideo's het meest profiteren van een combinatie van segmentering, oefeningen en pauzes en niet alleen interactiviteit. Dit hoofdstuk laat zien dat de toegevoegde waarde van e-learning hoogstwaarschijnlijk niet in de enkele affordances ligt, maar in de combinatie van vele.

**Hoofdstuk 10** combineert de onderzoeksresultaten van de vorige hoofdstukken om de onderzoeksfragen uit hoofdstuk één te beantwoorden. Elke vraag wordt beantwoord en het antwoord wordt in het perspectief van dit proefschrift of de huidige literatuur geplaatst. Als we de bevindingen van dit proefschrift binnen huidige visie op e-learning voor de medische vervolgopleiding zetten, stellen we een aantal dingen voor. Het is tijd om te accepteren dat e-learning niet iets voor de toekomst is. E-learning is al onderdeel van het dagelijks onderwijs en het gebruik zal alleen maar toenemen. Het is ook tijd om te accepteren dat e-learning, indien op de juiste manier ontworpen, niet onderdoet voor andere vormen van onderwijs en dat instructieontwerp fundamenteel is voor toekomstig evaluatieonderzoek. E-learning moet op leer theorie zijn gebaseerd en gericht zijn op de meest specifieke doelgroep die beschikbaar is. Dit proefschrift bouwt voort op de fundamentele Motiveren, Leren en Toepassen, die moeten worden ingevuld door specifieke uitingen op basis van leertheorie en gericht op de juiste doelgroep. We bespreken het effect van leertheorieën op deze affordances en suggereren dat de ontwikkeling van de leertheorie rekening moet houden met de levensstijl van de leerling. Vervolgens worden de sterke en zwakke punten van dit proefschrift besproken. De sterke punten lagen in de rigoureuze methodologie om te komen tot de 37 items uit hoofdstuk vijf. Andere sterke punten zijn de betrokkenheid van medisch specialisten in opleiding vanaf het begin van dit proefschrift en het feit dat
alles gericht is op de specifieke doelgroep; de medische vervolgopleidingen. De beperkingen hebben betrekking op het relatieve generieke kenmerk van Motiveren, Leren en Toepassen en het ontbreken van interculturele validatie. Omdat e-learning heel gemakkelijk verspreid wordt, is er geen bewijs of de bevindingen van dit proefschrift ook toepasbaar zijn in niet-westerse culturen. De praktische implicaties van dit proefschrift zijn te vinden in hoofdstuk zes (acht stappen instructie onderwerp model), hoofdstuk zeven (elf implementatiefactoren) en hoofdstuk acht (Medical E-learning Evaluation Survey). We hopen dat elke medische onderwijzer die van plan is om een e-learning gericht op medische vervolgopleiding te maken, deze hoofdstukken kan gebruiken. Het kan helpen om hun e-learning te ontwerpen, het te implementeren en hun oorspronkelijke instructieontwerp te evalueren. Ten slotte bespreken we de toekomst van e-learning, waarbij we minder in gaan hoe we digitale media zullen gebruiken, omdat we dat al doen, maar meer wat we zullen leren in een exponentieel groeiende wereld van digitalisering en automatisering.
Dankwoord
Dankwoord

Tijdens mijn co-schap gynaecologie in het (toenmalige) AMC te Amsterdam werd ik betoverd door de magie van de verloskunde. Er was zelden een co-schap waar je zo veel voor iemand kon betekenen, met zo weinig specifieke vaardigheden. Ik zal nooit vergeten dat ik als co-assistent naast een niet Nederlandssprekende vluchteling zat. Zij alleen probeerde ze te bevallen met alle emotionele stress en gemengde gevoelens door haar groepsverkrachting negen maanden geleden. Als je simpelweg door mee te zuchten, een hand vast te houden en erbij te blijven een verschil kan maken, dan voel je dat je goed zit. Toen ik aan de toenmalige opleider zei “Ik wil gynaecoloog worden hoe doe ik dat?”, antwoordde hij: “onderzoek is de sleutel. Wat vind je leuk behalve de gynaecologie?”. Lesgeven was mij met de paplepel ingegoten dus het antwoord was duidelijk “onderwijs”. “Dan moet je naar Fedde!” was zijn antwoord. En zo geschiedde.

Als oudste co-assistent in het (toenmalige) SLAZ heb ik Fedde leren kennen. De implementatie van een nieuw EPD stuurde mijn onderzoek richting ICT-innovaties. Het vak werd steeds kleurrijker en de verpleging leerde me de basis van de verloskunde. Als je dan een e-learning krijgt die begint met “deze borstvoeding e-learning duurt 10 uur”, ga je weer twijfelen. Uiteindelijk heeft dit geleid tot de behoefte om onderwijs en ICT te combineren tot een inspirerende omgeving, waar tot mijn verbazing, nog relatief weinig over geschreven was. In 2012 ben ik begonnen met het proefschrift dat u nu voor u heeft en ik ben nog lang niet klaar met het ontwikkelen en evalueren van onderwijsinnovaties.

Rubeena; uiteindelijk is het allemaal bij jou begonnen, nog ver voordat ik Fedde leerde kennen. Jij bent mijn muze, mijn spiegel en de liefde van mijn leven. Jij bent er als ik aan mezelf twijfel, jij bent er wanneer ik overmoedig word, wanneer relativering nodig is, wanneer de wereld oneerlijk is, wanneer het nodig is om mezelf te leren kennen… Eigenlijk ben je er altijd. Het schrijven van een proefschrift naast de dagelijkse kliniek van de opleiding en als staflid is een offer waar vooral jij het slachtoffer van bent geweest. Maar je hebt me nooit een onvertogen woord gegeven en me altijd ondersteund en geholpen. Zonder jou was dit nooit af gekomen en zonder jou zou ik nooit zijn geworden wie ik ben.
Fedde; je bent een voorbeeld als mens, opleider, promotor en mentor. Je abstracte antwoorden, je commentaar (wat soms niet meer was dan een cirkel), maar vooral de vrijheid die je me gaf, was precies wat ik nodig had. Wanneer ik als een vlinder van A naar B vlieg, breng jij me terug naar het hier en nu. Het grootste deel van onze gesprekken hadden niet met mijn onderzoek te maken, maar dingen over medische politiek, verborgen agenda’s, tuchtrecht en van alles wat ons vak zo breed en interessant maakt. Je bent een enorme inspiratiebron en ik hoop nog veel samen met je te kunnen doen.

Michiel; zonder jouw geduld was dit nooit gelukt. Je advies “maar dat schrijf je niet zo op” was elke keer opnieuw een schot in de roos. Wanneer ik de korte, pragmatische route wilde nemen, bracht jij me terug naar de theorie en het “waarom”. De basis voor mijn academisch denken komt bij jou vandaan, een vaardigheid waar ik dagelijks gebruik van maak en vaak aan je woorden denk.

Hans; na Fedde was je mijn introductie in academische afwegingen van complexe steekspelen en de kracht van het netwerk. Je nuchtere en pragmatische aanpak maakte je een zeldzaam fijne opleider op OK en je hebt me niet alleen geholpen een meer academische carrière na te streven, maar je stond ook aan de basis van mijn meer geavanceerde laparoscopische vaardigheden en ik verwijs dagelijks naar je wijze lessen en inzichten.

Judith; ik heb nog nooit iemand gezien met zo een eindeloze energie en passie voor het vak. Ik weet zeker, dat dankzij jou, de benign gynaecologie de nodige professionalisering zal krijgen en het is een eer om daar een bijdrage aan te leveren. Je bent altijd mijn sparringpartner wanneer iets lastig is, je durft beslissingen te nemen en de verantwoordelijkheid hiervoor te dragen. Je zal me altijd stimuleren om te blijven innoveren en jouw energieniveau na te streven.

Wouter; mijn ultieme kamergenoot. Jouw zachtaardigheid en eindeloos geduld geven altijd ruimte voor professioneel en persoonlijk overleg. Alle gezichten die bij ons in de kamer langskomen voor je, stralen bewondering en plezier uit als ze je zien. Ik hoop ook zo met mijn collegae en patienten om te kunnen gaan. Ik ben trots om je een collega, kamergenoot en vriend te noemen.

Freek en Nora; “gendermaatjes”, jullie maken het eiland van de gynaeccologische transgenderzorg groter en kleurrijker. In een doolhof van zorgpaden, besprekingen en overvolle agenda’s proberen we deze prachtige zorg vorm te geven en het naar een hoger academisch niveau te brengen. Mogen we nog veel samen blijven doen.
Sabiene en Anneloes; jullie geduld en incasseringsvermogen is top. Ondanks de chaos van mijn organisatie vaardigheden, bleven jullie geduldig, mensen lastig vallen, nieuwe werkprocessen verzinnen en vooral; volhouden. Ik ben blij dat al jullie moeite, geleid hebben tot twee mooie publicaties.

Christanne; vanaf het begin in de VU was je mijn mentor en steun in ook rumoerige tijden. Je hebt een stijl van leidinggeven waarbij je zelf een als rolmodel functioneert, maar tegelijk de groep voor je uit laat lopen. Dit is voor mij prachtig om zo te zien. Je onvoorwaardelijke steun en geloof in mijn kunnen heeft me in elk moment van de opleiding en daarna het gevoel gegeven dat ik er nooit alleen voor sta.

Velja en Roel; vanuit de VPG en endometriose zijn jullie mijn inspiratie als mens en professional. Jullie kunnen als geen ander verschillende manieren demonstreren in het omgaan met patiënten. Jullie staan altijd klaar voor overleg en advies, zowel in de privésfeer als professioneel.

Verpleging; het is veel te onpersoonlijk om zo een belangrijke groep mensen naar hun functie te noemen. Maar het zijn te veel namen om per persoon te noemen. Zonder jullie zou ik ons prachtige vak nooit eigen kunnen maken. Jullie leren mij dagelijks hoe je mens moet zijn en blijven. Het helpt me nuchter te blijven en een teamspeler te blijven.

Alle AIOS; zonder jullie hulp, het invullen van eindeloze vragenlijsten, het geven van jullie meningen en ideeën had ik niets om dit proefschrift mee te vullen. Ik zal me blijven inzetten om ons onderwijs te innoveren, immers blijf je altijd een (ex)AIOS.

Asaf; als eigenaar van ExpertCollege wil ik je nogmaals bedanken. Je uitgesproken bewondering van dit proefschrift is voor mij een inspiratie om door te gaan en ik zie ons mooie dingen samen doen in de toekomst.

Sian; your literacy, sense of grammar and language improved almost each chapter of this thesis. Therefore, you might be one of the few that actually read the whole book. Thank you so much for this.

Kieran; although we never met face-to-face, your supportive thinking and emphasis on budget was of incredible added value. Thanks to you we kept going back to keeping the e-learning user-centered. I truly hope to work with you in the future.

Eric and Danielle; my interest in e-learning all started in Dhaka, with our COTS program. It showed the possibilities of sharing knowledge digitally. I really hope to get part of my focus back to the less fortunate of the world and use innovations to decrease the barriers of proper education. Although, let’s be honest, everybody liked the pocket cards best.
Farida; er is niemand die zoveel in mij zag als jij. Je bent de belichaming van mijn scholing en zonder jou was dit allemaal niet gelukt. Jouw geloof in mijn kunnen en de trots die je laat zien elke keer als je naar me kijkt is een van de beste motivators die ik heb.

Joyce; centraal voor mijn periode op de middelbare school. De manier waarop je altijd aandacht probeert te krijgen voor leerproblemen (zoals dyslectie of dislexie) maakt je een van de belangrijkste “influencers” in onderwijs die ik ken. Ik hoop ooit voor het onderwijs te kunnen betekenen, wat jij lokaal hebt kunnen bewerkstelligen.

Raaf en Irish; uiteraard zijn verantwoordelijk voor Rubeena, waarvoor ik jullie eeuwig dankbaar ben. Jullie zijn ook altijd een oase van rust en ontspanning in drukke tijden en zonder jullie hadden wij nooit de rust kunnen pakken die nodig is om daarna weer te kunnen pieken.

Tariq; brother in IDEACT crime. Ik bewonder je rust en kalmte. Ondanks frustratie en tegenslagen, blijf je vooruit kijken. Er zijn niet veel mensen die met zo veel gemak, grote moeilijke beslissingen kunnen nemen en de consequenties dragen met opgeheven hoofd, zoals jij dat kan. Hopelijk kunnen we samen nog mooie avonturen beleven, zowel privé als professioneel.

Gijs; voorbeeld van loyaliteit en relativering. Je hebt mij het inzicht gegeven, hoe ongelofelijk belangrijk loyaliteit is. Een ander laten zien dat je er altijd voor die persoon bent is voor mij een karaktereigenschap welke ik altijd zal proberen na te streven.

Imre; je bent er altijd om te sparren over de kunst van het zijn en het anders zijn op het podium. De energie die je kan stoppen in projecten, waar andere de zin en onzin niet van inzien, helpt mij om ook verder te gaan waar andere zouden stoppen. Jouw geloof in je projecten zijn een direct voorbeeld voor mijn eigen vertrouwen in mijn projecten zoals dit proefschrift.

Lida; moeder van al mijn ideeën, mijn ziel en mijn inspiratie. Ik weet zeker dat je trots geweest zou zijn dat ik je sporen in de zorg heb gevolgd. Ik mis je altijd.

Herman/senior; zonder jou was het niets om mee te beginnen. Je onvoorwaardelijke steun geeft mij het gevoel dat er altijd iemand is om op terug te vallen. Ik heb met niemand zoveel meegemaakt en dat heeft onze band ongelofelijk sterk en hecht gemaakt. Je bent mijn voorbeeld voor onvoorwaardelijke liefde en zeldzame creativiteit. Je vooruitstrevendheid motiveert mij om er nooit bij neer te gaan zitten en iets te accepteren. Je zelfkritische persoonlijkheid heeft geleid tot het feit dat je altijd open staat voor persoonlijke verandering en verbetering. Dit is een academische vaardigheid waar ik dagelijks gebruik van probeer te maken. Verander maar vooral niet.
Curriculum vitae
Publications

1. **de Leeuw RA**, Westerman M, Walsch K, Scheele F.
   The Development and Validation of an instructional design evaluation survey for Postgraduate Medical E-learning (Preprint).
   JMIR 2019

2. **de Leeuw RA**, Walsch K, Scheele F, Westerman M.
   A theory- and evidence-based Postgraduate Medical E-learning Development Model: nine steps to an empirical educational experience (Preprint).
   JMIR Medical education 2019

   The journal of sexual medicine. 4, 2019

4. SI Stegwee, IPM Jordans, LF van der Voet, MY Bongers, CJM de Groot, CB Lambalk, **de Leeuw RA**, et al
   Single- versus double-layer closure of the caesarean (uterine) scar in the prevention of gynaecological symptoms in relation to niche development – the 2Close study: a multicentre randomised controlled trial.
   BMC pregnancy and childbirth 19 (1) 2019

   Digital vs face-to-face information provision in patient counselling for prenatal screening: a noninferiority randomized controlled trial.
   Prenatal diagnosis 2019

6. Inge PM Jordans, **de Leeuw RA**, Sanne I Stegwee, Nazar N Amso, Pere N Barri Soldevila, Thierry van den Bosch, Tom Bourne, Hans AM Bröllmann, Oliver Donnez, Margret Ducholm, Wouter JK Hеченkamp, Nicole Jastrow, Davor Jurkovic, Roy Mashiach, Osama Najj, Isabelle Streuli, Dirk Timmerman, Lucet F van der Voet, Judith AF Huirne.
   Niche definition and guidance for detailed niche evaluation
   Acta obstetricia et gynecologica Scandinavica 2019

   How We Evaluate Postgraduate Medical E-Learning: Systematic Review.
   JMIR Medical Education 1 (5) 2019
Sonographic examination of uterine niche in non-pregnant women: a modified Delphi procedure.  
Ultrasound in Obstetrics & Gynecology 53 (1) 2019

9. Angelo B Hooker, de Leeuw RA, Peter M van de Ven, Hans AM Brölmann, Judith AF Huirne.  
Reproductive performance after the application of hyaluronic acid gel after dilation and curettage in women who have experienced at least one previous curettage: long-term results of a multicenter prospective randomized trial.  
Fertility and sterility 110 (7) 2018

10. Thierry Van den Bosch, Annefleur M de Bruijn, de Leeuw RA, Margit Dueholm, Caterina Exacoustos, Lil Valentin, Tom Bourne, D Timmerman, Judith AF Huirne.  
A sonographic classification and reporting system for diagnosing adenomyosis.  
Ultrasound in Obstetrics & Gynecology, 2018

Is a mHealth application for pregnancy related work advice usable and useful? A mixed method approach. (Preprint).  
JMIR mHealth and uHealth 2018;

12. Kim WM van Delft, de Leeuw RA.  
How to attract talented juniors to urogynaecology.  
International urogynecology journal 29 (3) 2018

13. de Leeuw RA, Kieran Walsh, Michiel Westerman, Fedde Scheele. Consensus on Quality Indicators of Postgraduate Medical E-Learning: Delphi Study  
JMIR medical education 4 (1) 2018

European Journal of Obstetrics & Gynecology and Reproductive Biology 219, 2017

*Does application of auto-cross linker hyaluronic acid in women following D & C for miscarriage with at least one D & C in history improve reproductive outcome?*

*Human reproduction* 32, 2017

16. Angelo B. Hooker, **de Leeuw RA** Peter M. van de Ven, Erica A. Bakkum, Andreas L. Thurkow, Niels EA Vogel, Huib AAM van Vliet, Marlies Y. Bongers, Mark H. Emanuel, Annelies EM Verdonkschot, Hans AM Brölmann, Judith AF Huirne. *Prevalence of intrauterine adhesions after the application of hyaluronic acid gel after dilatation and curettage in women with at least one previous curettage: short-term outcomes of a multicenter, prospective randomized controlled trial.*

*Fertility and sterility* 5 (107) 2017

17. van der Voet LF, Huirne JAF, Hehenkamp WJK, Jordans IPM, **de Leeuw RA**, Stegwee S.

*Echoscopische beoordeling van niches.*

*Reproductieve geneeskunde, gynaecologie en obstetrie* 2017

18. Hehenkamp WJK, Groenman FA, Huirne JAF, **de Leeuw RA**, Brolmann HAM, Slager E.

*Niet-chirurgische myoom behandelingen anno 2017*

*Reproductieve geneeskunde, gynaecologie en obstetrie* 2017

19. Hooker AB, Brolmann HAMN, Huirne JAF, **de Leeuw RA**

*Zijn intra-uterine adhesie na verwijdering van zwangerschapsresten te voorkomen?*

*Reproductieve geneeskunde, gynaecologie en obstetrie* 2017

20. **de Leeuw RA**, M. Westerdam, Scheele F.

*Quality indicators for Learner-Centered Postgraduate medical eLearning*

*International Journal of Medical Education, 8, 2017*


*Prevention of Adhesions Post Abortion (PAPA studie); a randomised controlled trial evaluating application of hyaluronic acid following D&C in women with at least one D&C in history.*

*Gynecological Surgery* 13 (1) 2016
   Quality specifications in postgraduate medical e-learning– an integrative review leading to a
   Medical E-learning (M.E.) Model.
   BMC Medical Education. 16 (1) 2016

   A Randomised Controlled Trial of Consent Procedures for the Use of Residual Tissues for Medical Research: Preferences of and Implications for Patients, Research and Clinical Practice.
   PLoS One. 2016 Mar 30;11(3)

24. de Leeuw RA, Leung EY.
   A report from #BlueJC: 'Mesh or no mesh' for pelvic organ prolapse?
   BJOG. 2014 Jan;121(1): 132

25. de Leeuw RA, Leung EY.
   A report from #BlueJC: how should we manage antepartum haemorrhage of unknown origin?
   BJOG. 2014 Dec;121(13):1757

   Clinical Infectious Diseases 2009; 49:323–4

   Cholera outbreak training and shigellosis program.
   Dhaka Bangladesh ICDDR,B
About Robert

Robert de Leeuw is born in 1982 in Numansdorp, a small village near Rotterdam. Grown up without siblings he moved around South and North Holland, and settled in Hoorn. After graduating from high school, he went to Amsterdam to find his education. After going to the Amsterdam school of performance arts as a technician, he decided to go to university. Starting with psychology at the University of Amsterdam, he then choose for medical training at the Amsterdam Medical Center in Amsterdam in 2002.

In 2003 Robert founded IDEACT, a company aiming to make IDEa’s into ACTivities. He continued as a sound and light engineer during his medical training and added photography and videography as part of IDEACT. During his stay in Bangladesh, Robert collaborated with the ICDDR,B and John Hopkins University to create a virtual hospital for a Cholera Outbreak Training and Shigellosis program (COTS Program). This was his first design and production of postgraduate medical e-learning. In time, he added website development to the IDEACT spectrum and continued creating smartphone applications, datamanagement systems (Research Survey) and medical e-learning (Elephant Elearning) until today. He finished medical training in 2009 and worked in the Sint Lucas Andreas Hospital as a resident.

In 2011 Robert started his speciality training within obstetrics and gynaecology at the VU University Hospital Amsterdam. During his speciality training he continued his research on postgraduate medical e-learning, of which you are reading the end result. In 2013 he started the InterActive Publishing Foundation, being an open access, interactive, medical, distance education foundation.

In 2017 he finished his speciality training and started working at the VU University Hospital (now the Amsterdam University Medical Center) as a attending gynaecologist, specialised in minimal invasive surgery, with special interest for niches, myoma’s, endometriosis and adenomyosis. He continues to develop and evaluate different innovations in medical education and research methodologies.