Effect of improving the usability of an e-learning resource: a randomized trial

Mogamat Razeen Davids
Division of Nephrology, Department of Medicine, Stellenbosch University and Tygerberg Hospital, Cape Town, South Africa

Usuf M. E. Chikte
Department of Interdisciplinary Health Sciences, Stellenbosch University, Cape Town, South Africa

Mitchell L. Halperin
Li Ka Shing Knowledge Institute of St Michael's Hospital and Division of Nephrology, University of Toronto, Toronto, Ontario, Canada

Abstract

Optimizing the usability of e-learning materials is necessary to reduce extraneous cognitive load and maximize their potential educational impact. However, this is often neglected, especially when time and other resources are limited. We conducted a randomized trial to investigate whether a usability evaluation of our multimedia e-learning resource, followed by fixing of all problems identified, would translate into improvements in usability parameters and learning by medical residents. Two iterations of our e-learning resource [*version 1* (V1) and *version 2* (V2)] were compared. V1 was the first fully functional version and V2 was the revised version after all identified usability problems were addressed. Residents in internal medicine and anesthesiology were randomly assigned to one of the versions. Usability was evaluated by having participants complete a user satisfaction questionnaire and by recording and analyzing their interactions with the application. The effect on learning was assessed by questions designed to test the retention and transfer of knowledge. Participants reported high levels of satisfaction with both versions, with good ratings on the System Usability Scale and adjective rating scale. In contrast, analysis of video recordings revealed significant differences in the occurrence of serious usability problems between the two versions, in particular in the interactive HandsOn case with its treatment simulation, where there was a median of five serious problem instances (range: 0–50) recorded per participant for V1 and zero instances (range: 0–1) for V2 ($P < 0.001$). There were no differences in tests of retention or transfer of knowledge between the two versions. In conclusion, usability evaluation followed by a redesign of our e-learning resource resulted in significant improvements in usability. This is likely to translate into improved motivation and willingness to engage with the learning material. In this population of relatively high-knowledge participants, learning scores were similar across the two versions.

THE USABILITY OF TECHNOLOGY INTERFACES may have a major impact on learning, thus limiting the potential benefit obtained from using e-learning resources (2, 25, 31, 33, 37). We conducted a randomized trial to determine whether evaluating and optimizing the usability of a medical e-learning resource would result in improved measures of usability or learning.

The concept of usability derives from the field of human-computer interaction (HCI) and has been defined as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (1). Usability evaluation is well established in the software development industry (5, 12, 16, 22, 26, 28, 34), and there are often several cycles of testing and redesign before an application is released. This, however, is not common practice in medical education (33), where the importance of usability testing of e-learning resources is not yet widely recognized. Cost and time pressures are additional factors that may cause the evaluation of new resources to be neglected, with failure to achieve desired learning outcomes.

In the field of education, researchers have proposed guidelines for the design of e-learning resources based on cognitive load theory (CLT) (36) and the cognitive theory of multimedia learning (23, 24). These are based on a model of human cognitive architecture that views learning as involving active processing of information by working memory via separate visual and auditory channels. This system has a limited capacity. Any load that does not contribute to learning is considered extraneous and is likely to impede learning when the material is difficult and has a high intrinsic cognitive load (35). Mayer (23) has recommended several evidence-based principles to reduce extraneous cognitive load when designing multimedia learning resources. For example, according to the coherence principle, all irrelevant material should be eliminated, the signaling principle involves highlighting essential material, and the contiguity principle involves placing printed words near corresponding graphics.

There have been limited interactions to date between the fields of HCI and CLT (15). A recent review (15) reported that CLT concepts were mentioned in only 65 of >1.2 million citations in the Guide to the Computing Literature database. The two fields clearly share important concepts in that both strive to reduce extraneous cognitive load. In the case of HCI, this takes the form of usability guidelines such as "do not require the user to remember information from one screen to the next," designing for "recognition, not recall," encouraging "aesthetic and minimalist design," and "offer functionality only when needed" (27). In the case of CLT, there are instructional design principles such as the coherence, signaling, and contiguity principles. Hollender et al. (15) have proposed that the cognitive load induced by poor usability of e-learning interfaces be viewed as a specific component of extraneous cognitive load. This adds to the load resulting from poor instructional design.

Our interest is in developing learning resources to assist medical students and qualified practitioners in acquiring expertise in the diagnosis and treatment of electrolyte and acid-base disorders. This is a particularly challenging area of medicine (10). One of the resources we have developed is a web-based multimedia application called the "Electrolyte Workshop" (8). It is built in Adobe Flash and provides instruction and the opportunity for deliberate practice via an interactive treatment simulation. The content has a high intrinsic cognitive load, and we therefore attempted to minimize any extraneous load by optimizing the usability of the application. We conducted a usability evaluation of the application by testing it with typical end users (7) and followed this by conducting a

heuristic evaluation with a panel of experts (9). The information gained from these evaluations informed a comprehensive revision of our application.

This article reports on the effects of addressing the usability problems identified in our Electrolyte Workshop. Using a randomized trial, we investigated whether this had resulted in measurable improvements in usability and in improvements in learning. The reader is invited to examine the original and revised versions of the application at http://www.learnphysiology.org/sim1/ and http://www.learnphysiology.org/sim2/.

METHODS

Ethics approval for the project was granted by the Committee for Human Research of the Faculty of Medicine and Health Sciences of Stellenbosch University (project no. N08/05/158).

The e-Learning Resource

The application consists of case-based tutorials, each consisting of a series of slides, with the navigation controlled by the user. There are two main sections to the application. The first, called the WalkThru section, has cases with a "look-and-learn" approach similar to the use of worked examples in other disciplines (32). A clinical problem is presented followed by a demonstration of how an expert would analyze the data and embark on treatment. Animations illustrate changes in body fluid compartment sizes, brain cell size, and plasma Na^+ concentrations. The second section of the application, the HandsOn section, is interactive, with each case including a treatment simulation that provides the opportunity for deliberate practice. Users receive immediate feedback via on-screen text messages and animations.

Study Participants and Procedures

Residents and subspecialty trainees (fellows) were recruited from the Departments of Medicine and Anesthesiology at Stellenbosch University in Cape Town, South Africa. Participants were randomly assigned to the different versions of the two cases using a computer-generated random number sequence, blocked randomization, and stratification by discipline (internal medicine vs. anesthesiology) and seniority (residents vs. specialists who were training in subdisciplines of internal medicine or anesthesiology). Allocation concealment was ensured using sequentially numbered, opaque envelopes.

The application was loaded onto two 15-in. laptop computers, which were each equipped with a mouse and a webcam with an integrated microphone. Morae usability software was installed on each computer to facilitate the recording and analysis of testing sessions. Participants were each required to work through the allocated versions of the WalkThru and HandsOn cases. No time limits were set. After each case, participants completed a user satisfaction questionnaire and answered a set of questions designed to test learning.

Technical problems resulted in the loss of certain of the Morae recordings and, hence, the objective data on some participants. Of the 18 participants allocated to each version of the WalkThru case, we

had objective data for 17 participants in each group; of the 27 participants allocated to each version of the HandsOn case, we had objective data for 25 participants in the version 1 (V1) group and for 23 participants in the version 2 (V2) group.

Measures of Usability

There is no single best measure of usability as each measure has its pros and cons and examines a particular aspect of usability. We followed the commonly recommended approach of using multiple usability measures and collected both subjective, self-reported data as well as objective data obtained by recording and analyzing the interactions of our participants with the application.

Subjective measures.

A user satisfaction questionnaire that included the System Usability Scale (SUS) (4) was used to provide an overall measure of usability. The SUS can be used to compare different versions of a system and yields a single number (range: 0–100) with a score of 70 or greater regarded as acceptable. It is widely used, reliable, freely distributed, easy to administer, and easy to score (3, 4). We added a seven-point adjective rating scale as recommended by Bangor et al. (3). This item asked participants to rate the overall user friendliness of the application from being the worst imaginable (score of 1) through to the best imaginable (score of 7). Additional Likert-type questions asked participants to indicate whether the application increased their understanding and their confidence, whether navigation was difficult, whether they would recommend the application to others, and (for the HandsOn case) whether the simulation was realistic and engaging. The questionnaire also included two open-ended questions asking participants to comment on what they liked about the application and what they did not like or thought could be improved.

Objective measures.

Successful task completion rates and the detection of usability problems were recorded for each task as measures of effectiveness, whereas time on task and input device activity (mouse clicks and mouse movement) were recorded as measures of efficiency. The WalkThru case, the introductory slides of the HandsOn case, and the treatment simulation of the HandsOn case were each regarded as a separate task. Task completion in the WalkThru case and the introductory slides of the HandsOn case simply required that participants view all the information available. For successful completion of the simulation, participants had to treat the patient effectively and end with the summary "take home messages" slide. The severity of each usability problem detected was determined by considering the frequency, persistence, and impact of the problem (29). A serious problem is one that may cause delays or task failure for the user and that needs to be fixed before an application is released.

Measures of Learning

Eight questions related to the content of each tutorial were prepared. The first four questions tested recall, and the second four questions tested transfer. Participants were allowed 3 min/question, with each question printed on a separate sheet of paper and provided to them one at a time. Examples of the questions are shown in Table 1. The scores of the students were calculated by allocating one point

for each correct answer; no penalties were given for incorrect answers. All answers were scored independently by a specialist physician and a nephrologist and were moderated by one of the authors (M. R. Davids).

Statistical Tests

To compare scores across the two versions of the cases, the Wilcoxon rank-sum test was used for the SUS, adjective rating scale, and Likert-type questions. The *t*-test was used to compare SUS scores from the HandsOn case, as these were normally distributed. Fisher's exact test was used to compare the proportion of participants in each group with either positive or negative comments. It was also used to compare binary task completion rates and the proportion of participants encountering serious usability problems. Usability problem counts, time on task, mouse activity, and learning scores were compared using the Wilcoxon rank-sum test except where the data were normally distributed, in which case the *t*-test was used. The significance level was set at 0.05.

RESULTS

Subjective Usability Data

SUS and adjective rating scale.

The results from the SUS and adjective rating scale are shown in Table 2. Mean scores were higher for the revised version of each case, but this difference was not significant. For the WalkThru case, mean SUS scores were 84.7 for the V1 group and 87.9 for the V2 group ($P = 0.27$). Scores on the adjective rating scale were 5.8 versus 5.9 for the V1 and V2 groups, respectively ($P = 0.36$). For the HandsOn case, SUS scores were 76.6 and 81.5 ($P = 0.13$) and adjective rating scale scores were 5.4 and 5.6 ($P = 0.20$) for the V1 and V2 groups, respectively. When the WalkThru and HandsOn cases were combined, SUS scores were significantly higher for revised versions ($P = 0.03$). There was a moderate to good correlation ($r = 0.68$) between the SUS scores and those of the adjective rating scale.

Additional Likert-type questions.

The results from the additional Likert-type questions are shown in Table 3. Participants experienced navigation as more difficult in the first version of the HandsOn case compared with the revised version ($P = 0.02$). There were no other significant differences observed between the two versions of either case from this set of questions.

Open-ended questions.

There were no clear differences in the number of positive or negative comments from participants in the different groups. A selection of quotes is shown in Table 4.

Objective Usability Data

Measures of effectiveness.

TASK COMPLETION RATES.

The WalkThru case was successfully completed by all participants (n = 17 participants/group). With the more interactive HandsOn case, 18 of 25 participants successfully completed the first version, whereas 21 of 23 participants completed the second version (P = 0.09).

USABILITY PROBLEM COUNTS.

As expected, participants encountered very few usability problems with the two versions of the WalkThru case. In total, five serious problem instances were recorded. These were encountered by five different participants: four participants from the V1 group and one participants from the V2 group (P = 0.17). With the interactive HandsOn case, serious usability problems were encountered by 22 of 25 participants in the V1 group as opposed to 2 of 23 participants in the V2 group (P < 0.001). The median number of serious problem instances recorded per participant was five (range: 0–50) for the V1 group and zero (range: 0–1) for the V2 group (P < 0.001). When these separate problem instances were consolidated into distinct usability problems for each participant, the median problem count was two (range: 0–4) for the V1 group and zero (range: 0–1) for the V2 group (P < 0.001). Of the 25 participants in the V1 group of the HandsOn case, 2 participants expressed frustration and 3 participants asked for help while using the application. There were no such events recorded in the V2 group.

Measures of efficiency.

TIME ON TASK.

Participants spent similar amounts of time on the two versions of each case. Mean times for the V1 and V2 groups were 11.8 ± 4.9 versus 12.7 ± 4.6 min for the WalkThru case (P = 0.57) and 19.2 ± 18.4 versus 18.4 ± 19.0 min for the HandsOn case (P = 0.68).

MOUSE ACTIVITY.

Mouse activity was similar for the two versions except for a higher click count in the V1 group versus the V2 group of the interactive HandsOn case. For the WalkThru case, click counts for the V1 group versus the V2 group were 29.2 ± 14.4 versus 25.5 ± 15.0 clicks (P = 0.89) and mouse movement was 20,193 ± 29,978 versus 30,251 ± 27,082 pixels (P = 0.05). For the HandsOn case, click counts for the V1 group versus the V2 group were 142.6 ± 79.6 versus 89.0 ± 36.4 clicks (P = 0.008) and mouse movement was 73,259 ± 37,681 versus 66,724 ± 49,444 pixels (P = 0.29).

Measures of Learning

Tests of recall and transfer.

For the WalkThru case, recall test scores were 17.2 ± 2.6 and 16.6 ± 2.6 for the V1 and V2 groups (P = 0.16); for the HandsOn case, scores were 20.4 ± 5.0 and 21.0 ± 4.0 for the V1 and V2 groups (P =

0.58). For the WalkThru case, transfer test scores were 7.0 ± 3.1 and 6.9 ± 2.5 for the V1 and V2 groups ($P = 0.91$); for the HandsOn case, scores were 7.4 ± 3.2 and 6.6 ± 2.5 for the V1 and V2 groups ($P = 0.31$).

DISCUSSION

A thorough evaluation followed by an extensive revision of our application resulted in measurable improvements in usability, in particular with regard to the HandsOn case with its interactive treatment simulation. The most striking finding was the large number of serious usability problems participants encountered in the original version of the HandsOn case compared with very few in the revised version. Nearly all the participants in the V1 group were affected but only two participants in the V2 group, suggesting that we had succeeded in eliminating most of the serious usability problems. Task completion rates and user satisfaction scores were also higher for the V2 group, although these were not statistically significant.

Expressions of frustration and requests for help were documented for participants in the V1 group but not in the V2 group.

We observed an interesting disconnect between subjective and objective measures of usability. Participants awarded high SUS and adjective rating scale scores to both versions of each case, even to the original version of the HandsOn case where many serious usability problems were encountered. This phenomenon has been noted previously (3, 7) and underlines the importance of not relying only on subjective measures of usability when evaluating e-learning resources or programs.

The improvements in usability were not accompanied by differences in learning, with scores on tests of retention and transfer being similar between the groups. A possible reason for the lack of impact on learning measures might be that our participants, all practicing clinicians, were not novices with regard to the subject area. All had received instruction on electrolyte and acid-base disorders as undergraduate students, some had received additional instruction in the course of their postgraduate training, and all of them had at least some experience in managing patients with these conditions. High-knowledge learners obtain less benefit when learning materials are designed to reduce cognitive load and, in some cases, may even suffer a decrease in performance, a phenomenon called the expertise-reversal effect (17). Another reason for the absence of a learning effect might be that our application implemented the segmenting principle (23) by allowing participants to control the navigation. This breaks the lesson into user-paced segments and is likely to minimize the negative impact of any extraneous load caused by poor usability.

While some researchers have reported significant learning effects from optimizing usability (2, 25), others have not observed differences but have found improvements in efficiency, satisfaction, or motivation. These effects are important in the light of the alarmingly high dropout rate from e-learning courses (38). Highly motivated and self-regulated learners are more likely to persist and succeed in e-learning environments, and optimizing usability can make an important contribution to their satisfaction and motivation. A study (13) among medical undergraduate students found that perceived quality of the e-learning program was an important determinant of their attitudes toward

computer-based learning. In other studies, better usability resulted in improved task completion rates and less time on task (18) and in increased self-regulation by learners (21). Levy (20) found satisfaction to be a key indicator in the completion of online courses, whereas Zaharias and Poylymenakou (38) reported a strong relationship between learners' perceptions of system usability and their motivation to learn.

Traditional usability goals usually involve designing for effective and rapid task completion; however, systems that are less efficient to use or more difficult to learn may sometimes have a positive influence on motivation or learning. In a recent study (11), students who used disfluent learning materials with harder-to-read fonts had improved retention of the content compared with control students. This inclusion of "desirable difficulties" in their learning materials appeared to promote deeper processing and thereby improved learning.

Our study is a real-world example of the benefit of optimizing the usability of e-learning resources for medical education. The study participants were representative of our primary target audience, and the lack of a learning effect with these relatively high-knowledge learners is not surprising. As we also intend to use our Electrolyte Workshop for teaching undergraduate students, followup studies could investigate whether improved usability may translate into better learning for these novice learners. Compared with their senior colleagues, they are more likely to experience the content matter as having a high intrinsic cognitive load and should therefore be more sensitive to the addition of an extraneous load imposed by poor usability.

E-Learning has now become part of the medical education mainstream, with increasing investments in developing e-learning materials, modules, and programs. We would recommend that the usability of these resources be evaluated and optimized as a matter of routine. An iterative development process should be followed, with usability evaluation beginning early and involving both subjective and objective methods. Educators need to be aware that any existing digital divide will be widened by educational software that is poorly designed and that improving usability will lead to accessibility for a wider range of learners (6). Optimizing usability may therefore contribute to improved rates of persistence and success in e-learning environments.

Future research should examine the effect of optimizing usability and cognitive load on learning in learners who are novices regarding the subject matter, especially when the material to be learnt is complex, and in learners from the wrong side of the digital divide. A wider range of measures to evaluate the user experience will increasingly be used, including measures such as engagement, motivation, aesthetics, fun, and pleasure (14, 19, 30, 38).

In conclusion, the adoption of a design-test-redesign approach led to significant improvements in the usability of our multimedia e-learning application. This is likely to result in improved motivation and engagement with the learning resource and increases the chances of achieving desired educational outcomes. We support the recommendation that the development of e-learning materials should integrate user-centered technology design with learner-centered instructional design (15). The process should be iterative and focused on optimizing usability as well as on implementing principles of good instructional design based on CLT.

GRANTS

This work was supported by a Doctoral Fellowship Award from Stellenbosch University's Faculty of Medicine and Health Sciences.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: M.R.D., U.M.E.C., and M.L.H. conception and design of research; M.R.D. performed experiments; M.R.D. analyzed data; M.R.D. and U.M.E.C. interpreted results of experiments; M.R.D. drafted manuscript; M.R.D., U.M.E.C., and M.L.H. edited and revised manuscript; M.R.D., U.M.E.C., and M.L.H. approved final version of manuscript.

ACKNOWLEDGMENTS

The authors thank Althea Goosen for the valuable assistance with data collection.

Table 1. Measures of learning

Questions
Tests of retention
How did Suzie develop severe acute hyponatremia? Write down all the factors mentioned in the case that could have contributed.
Describe the major body fluid compartments in healthy individuals with respect to their volumes.
How would we know that antidiuretic hormone is acting on the kidney?
Write down all the case data you can remember. If you don't know the number you can simply indicate whether a parameter was normal (N), increased (↑), or decreased (↓).
Tests of transfer
"Runners hyponatremia" related to water overload may occur with long-distance races. You are advising the medical support team of next year's Two Oceans Ultramarathon. List all possible "risk factors" that could identify runners with a greater likelihood of developing acute hyponatremia during the race.
An athlete has a seizure at the end of a long-distance race. His plasma Na^+ concentration is 125 mmol/l. He is given 200 ml of 3% saline over 30 min. However, the followup plasma Na^+ concentration is 124 mmol/l and there is no clinical improvement. List the possible reasons why the plasma Na^+ concentration did not rise in response to treatment.
How much water would a 72-kg woman have to take in (and retain) to drop her plasma Na^+ concentration from 140 to 126 mmol/l? Show your calculations.
A 90-kg male patient developed acute hyponatremia from psychogenic polydipsia. You want to raise his plasma Na^+ concentration rapidly from 121 to 126 mmol/l. How many millimoles of Na^+ need to be administered? Show your calculations.

Table 1. Examples of questions designed to test the recall of information and questions to test the transfer of problem solving ability are shown. These are related to the WalkThru case.

Table 2. Scores for the two versions of each case and for both cases combined

	System Usability Scale			Adjective Rating Scale		
	WalkThru case	**HandsOn case**	**Both cases***	**WalkThru case**	**HandsOn case**	**Both cases**
V1 group	84.7 ± 12.0	76.6 ± 18.2	79.8 ± 16.4	5.8 ± 0.5	5.4 ± 0.8	5.5 ± 0.7
V2 group	88.0 ± 14.0	81.5 ± 12.9	84.1 ± 13.6	5.9 ± 0.5	5.6 ± 0.6	5.7 ± 0.6

Table 2. Values are means ± SD. V1 and V2, versions 1 and 2, respectively. The only significant difference observed was for System Usability Scale scores for both cases combined (* P = 0.03).

Table 3. Answers to additional Likert-type questions

		Item Description					
	Total Number of Participants	**Increased my understanding**	**Increased confidence**	**Navigation difficult***	**Would recommend**	**Simulation realistic**	**Simulation engaging**
WalkThru case							
V1 group	18	14	14	0	16		
V2 group	18	13	14	2	15		
HandsOn case							
V1 group	27	21	19	8	23	20	19
V2 group	27	21	20	1	23	22	23

Table3. Positive responses to the question items (i.e., agree and strongly agree) were combined. The only significant difference observed was for the item on navigation for the HandsOn case (* $P = 0.02$).

Table 4. Selection of responses to open-ended questions

Participants Liked the Following	Participants Did Not Like or Thought the Following Could Be Improved
WalkThru case	
"Bright, very good visuals. Clear smooth integration. Visuals and words coupled well together. Makes a sometimes daunting subject approachable/fun." (V1 group)	"Font very small!" (V1 group)
"Giving you a case, explain the treatment in a stepwise, easy to understand fashion. Also not too much detail." (V1 group)	"Perhaps a bit 'wordy' in places. Less paragraphs and more bullets/points perhaps." (V1 group)
"Contemporary example. Flows like a story–easier to remember the facts." (V2 group)	"Suzie's 'blinking eyes' distracted from the text on the last slide." (V1 group)
"Visually pleasing. Simple yet clear message. Useful animations that demonstrate the concept well." (V2 group)	"Too many different things to look @ at one time → gets distracting–I tended to ignore the graphics and just read the text." (V2 group)
HandsOn case	
"Excellent the way it responds and gives feedback. Take home messages are good too!" (V1 group)	"I did not clearly follow the last management steps. Do electrolyte and fluid administration and IV steroid use all impact on the outcome of the case simulation? Are all these maneuvers considered by the computer?" (V1 group)
"Being able to play around and see the effect of treatments administered." (V1 group)	"I think the way to use treatment options needs to be a bit explained before use." (V1 group)
"Animation again was excellent → seeing the consequences immediately of certain therapies was excellent." (V2 group)	"Lab data hidden (only found it after 5 minutes)." (V1 group)
"Real life case and can see what actually will happen if you give certain amount of fluids and sodium." (V2 group)	"Took me a while to figure out SBP was systolic blood pressure." (V1 group)
"Initially I was not thinking of the actual solution, rather fooling around with slides to see what would happen to brain if I give inappropriate therapy." (V2 group)	"Should add the appropriate management at the end of the case, as a teaching tool." (V1 group)
	"Have to drag slider; doesn't work if click at a certain point." (V2 group)
	"I would have liked a model answer with explanation." (V2 group)
	"When answering doesn't indicate which part of the answer was wrong → a bit frustrating." (V2 group)

Table 4. Shown are verbatim quotes from participants followed by the version group.

REFERENCES

1. Abran A, Khelifi A, Suryn W, Seffah A. *Usability meanings and interpretations in ISO standards. Software Qual J 11: 325–338, 2003.*
2. Avouris NM, Dimitracopoulou A, Daskalaki S, Tselios NK . *Evaluation of distance-learning environments: Impact of usability on student performance. Int J Educ Telecommum 7: 355–378, 2001.*
3. Bangor A, Kortum PT, Miller JT . *An empirical evaluation of the system usability scale. Int J Hum Comput Interact 24: 574–594, 2008.*
4. Brooke J . *SUS: a "quick and dirty" usability scale. In: Usability Evaluation in Industry, edited by Jordan PW, Thomas B, Weerdmeester BA, McClelland IL. London: Taylor & Francis, 1996, p. 189–194.*
5. Bygstad B, Ghinea G, Brevik E . *Software development methods and usability: perspectives from a survey in the software industry in Norway. Interact Comput 20: 375–385, 2008.*
6. Chalmers PA . *The role of cognitive theory in human-computer interface. Comput Hum Behav 19: 593–607, 2003.*
7. Davids MR, Chikte U, Grimmer-Somers K, Halperin ML . *Usability testing of a multimedia e-learning resource for electrolyte and acid-base disorders. Br J Educ Technol; doi:doi:10.1111/bjet.12042.*
8. Davids MR, Chikte UME, Halperin ML. *Development and evaluation of a multimedia e-learning resource for electrolyte and acid-base disorders. Adv Physiol Educ 35: 295–306, 2011.*
9. Davids MR, Chikte UME, Halperin ML. *An efficient approach to improve the usability of e-learning resources: the role of heuristic evaluation. Adv Physiol Educ 37: 242–248, 2013.*
10. Dawson-Saunders B, Feltovich PJ, Coulson RL, Steward DE. *A survey of medical school teachers to identify basic biomedical concepts medical students should understand. Acad Med 65: 448–454, 1990.*
11. Diemand-Yauman C, Oppenheimer DM, Vaughan EB. *Fortune favors the bold (and the italicized): effects of disfluency on educational outcomes. Cognition 118: 111–115, 2011.*
12. Gould JD, Lewis C. *Designing for usability: key principles and what designers think. Commun ACM 28: 300–311, 1985.*
13. Hahne AK, Benndorf R, Frey P, Herzig S. *Attitude towards computer-based learning: determinants as revealed by a controlled interventional study. Med Educ 39: 935–943, 2005.*
14. Hancock PA, Pepe AA, Murphy LL. *Hedonomics: The power of positive and pleasurable ergonomics. Ergonom Design 13: 8–14, 2005.*
15. Hollender N, Hofmann C, Deneke M, Schmitz B. *Integrating cognitive load theory and concepts of human-computer interaction. Comput Hum Behav 26: 1278–1288, 2010.*
16. Holzinger A, Errath M, Searle G, Thurnher B, Slany W. *From extreme programming and usability engineering to extreme usability in software engineering education (XP+ UE→ XU). In: Proceedings of the 29th Annual International Computer Software and Applications Conference. Washington, DC: IEEE Computer Society, 2005, p. 169–172.*
17. Kalyuga S, Ayres P, Chandler P, Sweller J. *The expertise reversal effect. Educ Psychol 38: 23–31, 2003.*
18. Kanuka H, Szabo M. *Conducting research on visual design and learning: pitfalls and promises. Can J Learn Technol 27: 105–123, 1999.*
19. Khalid HM. *Embracing diversity in user needs for affective design. Appl Ergonom 37: 409–418, 2006.*
20. Levy Y. *Comparing dropouts and persistence in e-learning courses. Comput Educ 48: 185–204, 2007.*
21. Liaw SS, Huang HM. *Perceived satisfaction, perceived usefulness and interactive learning environments as predictors to self-regulation in e-learning environments. Comput Educ 60: 14–24, 2013.*
22. Mao JY, Vredenburg K, Smith PW, Carey T. *The state of user-centered design practice. Commun ACM 48: 105–109, 2005.*

23. Mayer RE. *Applying the science of learning to medical education. Med Educ 44: 543–549, 2010.*
24. Mayer RE. *Cognitive theory of multimedia learning. In: The Cambridge Handbook of Multimedia Learning, edited by Mayer RE. New York: Cambridge Univ. Press, 2005, p. 31–48.*
25. Meiselwitz G, Sadera W. *Investigating the connection between usability and learning outcomes in online learning environments. J Online Learn Teach 4: 9, 2008.*
26. Myers BA, Rosson MB. *Survey on user interface programming. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Monterey, CA: ACM, 1992, p. 195–202.*
27. Nielsen J. *Nielsen Norman Group. 10 Usability Heuristics for User Interface Design (online). http://www.nngroup.com/articles/ten-usability-heuristics/ [27 January 2014].*
28. Nielsen J. *Iterative user-interface design. Computer 26: 32–41, 1993.*
29. Nielsen J. *Nielson Norman Group. Severity Ratings for Usability Problems (online). http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/ [27 January 2014].*
30. O'Brien HL, Toms EG. *The development and evaluation of a survey to measure user engagement. J Am Soc Inform Sci Technol 61: 50–69, 2010.*
31. Oviatt S. *Human-centered design meets cognitive load theory: designing interfaces that help people think. In: Proceedings of the 14th ACM International Conference on Multimedia. Santa Barbara, CA: ACM, 2006, p. 871–880.*
32. Renkl A. *The worked-out examples principle in multimedia learning. In: The Cambridge Handbook of Multimedia Learning, edited by Mayer RE. Cambridge, UK: Cambridge Univ. Press, 2005, p. 229–245.*
33. Sandars J. *The importance of usability testing to allow e-learning to reach its potential for medical education. Educ Prim Care 21: 6–8, 2010.*
34. Sohaib O, Khan K. *Integrating usability engineering and agile software development: a literature review. In: International Conference on Computer Design and Applications. Qinhuangdao, China: IEEE, 2010, p. 32–38.*
35. Sweller J, Chandler P. *Why Some Material Is Difficult to Learn (online). http://ammonwiemers.com/IdetPortfolio/articles/Instructional%20Design%20Theory/Why%20Some%20Material%20is%20Difficult%20to%20Learn.pdf [27 January 2014].*
36. Sweller J, van Merrienboer JJG, Paas F. *Cognitive architecture and instructional design. Educ Psychol Rev 10: 251–296, 1998.*
37. Zaharias P. *Usability in the context of e-learning. Int J Technol Hum Interact 5: 37–59, 2009.*
38. Zaharias P, Poylymenakou A. *Developing a usability evaluation method for e-learning applications: beyond functional usability. Int J Hum Comput Interact 25: 75–98, 2009.*