Stott, A. E., & Hattingh, A. (2015). Conceptual Tutoring Software for Promoting Deep Learning: A Case Study. *Educational Technology & Society, 18* (2), 179–194.

Conceptual Tutoring Software for Promoting Deep Learning: A Case Study

Angela Stott[*] and Annemarie Hattingh

School of Education, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa // stottae@ufs.ac.za // annemarie.hattingh@uct.ac.za

[*]Corresponding author

(Submitted December 21, 2013; Revised July 15, 2014; Accepted July 25, 2014)

ABSTRACT

The paper presents a case study of the use of conceptual tutoring software to promote deep learning of the scientific concept of density among 50 final year pre-service student teachers in a natural sciences course in a South African university. Individually-paced electronic tutoring is potentially an effective way of meeting the students' varied learning needs within the limited time available for content teaching in the pedagogically focused course. Both qualitative and quantitative data were collected in the forms of engagement patterns, surveys, focus-group discussions, and pre- and post- test scores. These were used to formulate a rich description in answer to the questions "Is deep learning evident when student teachers engage with a module on density using conceptual tutoring software?," and "What design features of this module are effective in promoting deep learning?." The findings suggest that conceptual tutoring software can be effective in promoting deep learning among student teachers. The sub-step design, branched structure, and conceptual approaches used in the tutoring design, and the inclusion of a written task and a test, were found to be effective in promoting deep learning.

Keywords

Electronic tutor, Conceptual, Software, Deep learning, Science pre-service teacher education, PGCE, Natural sciences, Density

Introduction

The Postgraduate Certificate in Education (PGCE) Natural Sciences (NS) methods course is part of a single-year program which qualifies graduates to teach in grades 4 to 9 in South Africa as generalists of various subjects. Little of the class contact-time can be used to develop content knowledge since the course has a pedagogical focus. Also, the large range in content knowledge among the students makes it difficult to pitch content teaching at a level beneficial to all the students. Therefore, individually-paced electronic tutoring is potentially an effective way of meeting the students' varied learning needs. There is evidence that conceptually-oriented tutoring software can be effective in helping students to engage in deep conceptual learning (e.g. She & Liao, 2010). The existing literature on use of such software to promote deep conceptual learning, including research previously done on the particular software used in this study, iQuiz (Stott & Case, 2014), has almost exclusively been quantitative and experimental in nature. In this case study we sought to address this gap in the literature through a rich, primarily qualitative study of students' learning experiences and behavioural choices during their extended engagement with conceptual tutoring software. This occurred during a module, within the PGCE NS methods course, designed to teach the scientific concept of density. The research is guided by the questions: (1) "Is deep learning evident when student teachers engage with a module on density using conceptual tutoring software?," (2) "What design features of this module are effective in promoting deep learning?"

Literature review

Effective tutoring provides individualised formative feedback of a small enough grain-size to support a learner to remain engaged in learning for long enough to think through all the steps required for mastery (VanLehn, 2011). Grain size refers to 'the amount of reasoning required of participants between opportunities to interact' (p. 203). VanLehn calls electronic tutors (e-tutors) which offer individualised feedback of a very fine grain size sub-step intelligent tutoring systems (ITSs). A number of ITSs have been developed and researched, with generally high levels of success reported. Almost all of these ITSs provide support for learners solving problems (Mitrovic & Weerasinghe, 2009), particularly algorithmic-based problems. A few guide learners to solve more ill-structured problems (Fournier-Viger, Nkambou, & Nguifo, 2010) and a few tutor conceptual understanding. These are scarce due to the complexity associated with their creation in terms of both software programming and content which runs on the software (Mitrovic, Martin, & Suraweera, 2007).

In this study we use software which we refer to as a conceptual e-tutor, called iQuiz (iquiz.uct.ac.za), which was developed at a South African university. iQuiz can be created to provide formative feedback of a very fine grain size which differs for different learner responses, and therefore meets VanLehn (2011)'s criteria for potential high effectiveness and inclusion in the ITS category. Most authorities, however, require software to exhibit artificial intelligence (AI) to be called an ITS (e.g. Aleven, Mclaren, Sewall, & Koedinger, 2009). iQuiz does not use AI features, although it can be pre-programmed to respond differently to different user choices at each slide. iQuiz consists of a number of multiple-choice-question-, and information-, slides which are connected to one another by hyperlinks. Some of these slides form a central backbone, traversed if all questions are answered correctly. Other slides form side branches, designed to respond appropriately to a user's incorrect responses in what could be called an electronic dialogue. Content-creation is time-consuming, preferably requires advanced pedagogical and conceptual content knowledge and skill, but does not require programming knowledge and is highly flexible. Figure 1 represents part of the e-tutor. Each of the ovals refers to a slide and the connecting arrows to hyperlinks.

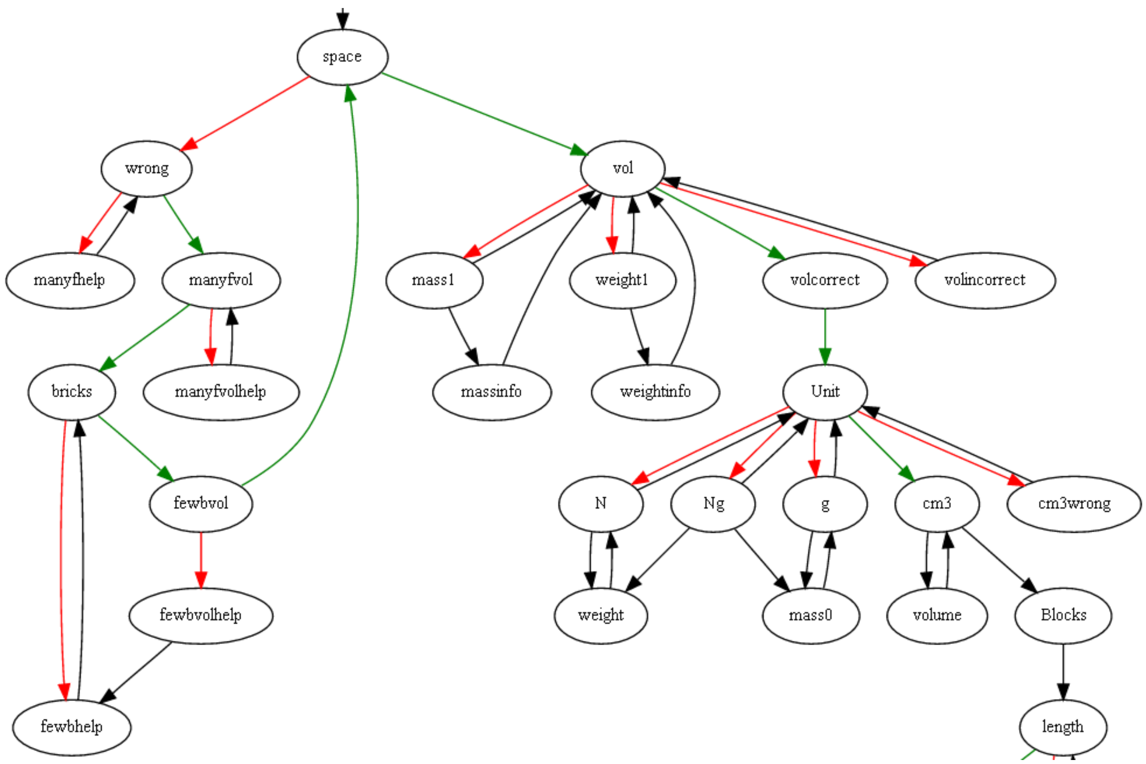


Figure 1. A representation of a part of an iQuiz e-tutor

Although previous research on iQuiz (Stott & Case, 2014) suggests that it can be an effective conceptual e-tutor, this research was done in the context of a short-duration engagement during an in-service teacher workshop. This did not provide information about how effective the software would be during more extended courses where students have greater flexibility about their engagement choices and tend to use information and communication technologies (ICT) in limited and shallow manners (Selwyn, 2007). Therefore, in this research we sought to understand how student teachers use this e-tutor as a resource in their learning for understanding.

Deep learning

Deep learning refers to learning for understanding (Schwartz, Lindgren, & Lewis, 2009). Mayer (2009)'s Theory of Multimedia Learning can be used as a lens for recognising deep learning. According to this theory, the small size and crucial role in learning, of working memory, is the limiting factor in learning. Three kinds of cognitive processing occur in working memory: Extraneous, essential, and generative processing (Mayer, 2009). For deep learning to occur, extraneous processing must be minimised, and essential and generative processing supported. Extraneous processing is cognitive activity which is not helpful to the learning process. In the context of learning with ICT, the possibility to *game the system:* Making progress within the software while avoiding learning (Muldner, Burleson, Van de Sande, & VanLehn, 2011), is a potential source of extraneous processing. During essential processing the learner selects and organises appropriate information in order to comprehend the information presented. This can be detected by the learner's performance on recall-type questions, i.e., those requiring responses similar to the presented information. During generative processing the learner integrates knowledge they have comprehended during essential processing with their prior knowledge, thus developing understanding. This can be detected by the learner's performance on transfer-type questions, i.e., those requiring learners to apply their knowledge to contexts which are not similar to the presented information.

The presence of effective self-regulation in learning is also associated with engagement in deep learning (Blom & Severiens, 2008). Self-regulated learning occurs under the students' metacognitive control and intrinsic motivation (Schraw, Olafson, & VanderVeldt, 2012). Models of self-regulated learning include an iterative internal feedback loop by which a student evaluates his or her learning and checks this evaluation against goals, rates of progress, and external feedback (Dweck, 2002). This may result in changes in affect, including self-efficacy, and behavioural choices. The more effective a student is at self-regulation, the more appropriately they are able to calibrate their self-assessment against external assessment and alter affect and behaviour toward attaining their goals (Nicol & Macfarlane-Dick, 2006). The intrinsic motivation displayed by such learners may result from a mastery-oriented epistemology (Schraw et al., 2012): The belief that the purpose of obtaining knowledge is to master a domain rather than to achieve extrinsic rewards. Mastery of the domain is associated with an increase in self-efficacy (Zimmerman, Bandura, & Martinez-Pons, 1992). Effective self-regulation is particularly critical to ensuring the sustained engagement necessary for deep learning in the context of learning using ICT, since students can end a learning session at will. An e-tutor's design features can affect the likelihood of this sustained engagement occurring or not (D'Mello, Olney, Williams, & Hays, 2012). The nature and timing of the feedback the ICT provides is particularly influential in affecting the extent to which learners engage in self-regulation, although optimal design features are still poorly understood (Narciss, 2013).

Learning the scientific concept of density

The scientific concept of density tends to be poorly understood by children and adults (Duckworth, 1986). Dawkins, Dickerson, McKinney, and Butler (2008) found that middle school student teachers also understand density poorly, their discreet pieces of knowledge about density not being integrated into a coherent understanding. A number of factors contribute to density being a difficult concept. These include the presence of conceptions alternative to the particulate model of matter (Snir, Smith, & Raz, 2003), lack of distinction between the concepts of weight and density (Smith, Snir, & Grosslight, 1992), and between substances and objects (Wiser, Smith, Asbell-Clarke, & Doubler, 2009, April), difficulties with proportional reasoning (Rowell & Dawson, 1977), the complexity and abstractness of an intensive property which cannot be observed directly and which depends on the relationship between two variables (Kang, Scharmann, & Noh, 2004), and misunderstandings arising during teaching (Xu & Clarke, 2012). We considered the topic of density an appropriate one to use within an e-tutor for the PGCE NS course, given that density is generally poorly understood and incorporates multiple important scientific concepts relevant to the curriculum these students will need to be able to teach.

Methodology

The case

This is a convenience case study involving 50 student teachers who all have at least an undergraduate degree and who enrolled for the Natural Science (NS) methods course in the one year professional teacher education (PGCE) qualification at a university in South Africa. This qualifies them as generalist teachers of all subjects from grades 4 to 9. Case study research is well suited to gaining a rich understanding of a particular case, from which other practitioners and researchers can extract aspects relevant to their particular cases (Merriam, 1988).

The sample

The demographic information of all the participants in this case is summarised in Table 1.

Table 1. The students' demographics and prior science education

Number of students (*n*): 50	Female: 44	Male: 6
Age: M = 26, SD = 6.1, Range: 21 – 45		
Students with prior teaching experience without a professional teaching qualification: 2		
Students who completed science related school subjects: Grade 9 Natural science (compulsory): 50 Grade 12: Physical science (physics and chemistry): 26 Life science: 35 Earth science: 23		
Students who completed science related courses at university level: Bachelor of Science honours degree: 1 Chemistry at year 1 level: 2	Master degree in Marine Biology: 1 Physics at year 1 level: 1	
Students with no university science courses: 45		
Students with Bachelor of Arts or Social Science degrees: 48		

As can be seen in Table 1, approximately half the students had exposure to physical science (PS), which includes both physics and chemistry, beyond the compulsory grade 9 school level. It is within PS that density and its related concepts are learnt, therefore the students' prior exposure to PS is most relevant to this particular study. At least 36% (18 students) ascribed their choice not to have studied PS beyond the compulsory grade 9 level to negative attitudes toward PS, e.g., "I've always hated maths and science," "My whole life I've hated science," "It still petrifies me." Remarks from students who had studied PS beyond grade 9 level regarding their experience of school PS include suggestions of: Low levels of enjoyment and long-term retention of the subject matter, e.g., "I didn't remember any of my science: I hated it"; An algorithmic focus, e.g., "I just knew how to plug numbers into the formula and manipulate the formula, because that's a lot of what we did in science"; Non-inclusive pedagogy e.g., "[I] sat at the back and no-one asked me anything," "I used to ask lots of questions, and then if they didn't answer, you just switch off, and it would be – like – really dry." Nine of the students achieved an A or B symbol, i.e., above 70%, in the grade 12 PS exit level examination. Four studied PS at tertiary level. This discussion suggests a prevalent negativity toward PS as well as a large range in prior PS knowledge among the students.

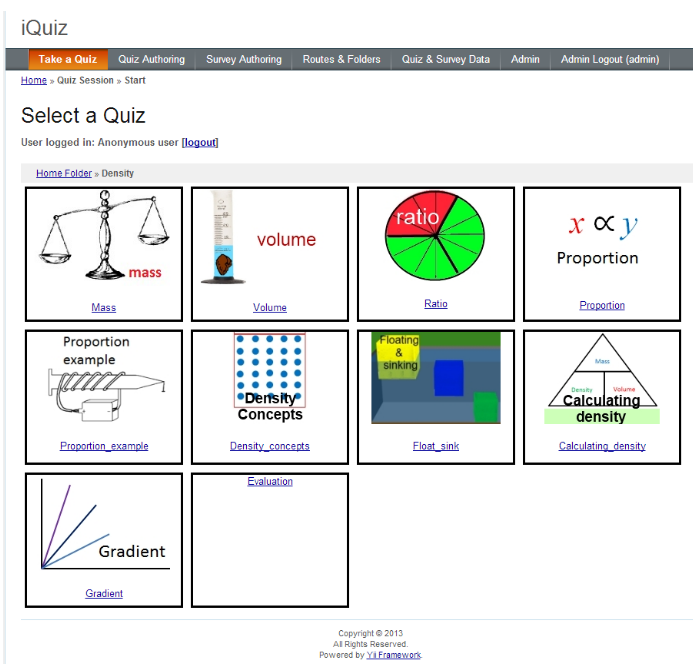


Figure 2. The menu page of the density module

Implementation of the e-tutors in an integrated pedagogical approach

Class contact time for the density module was three hours a week for four weeks. In the first week the students were introduced to the module and answered a pre-test about density, and in the last they answered a post-test. Class-time engagement with the e-tutor occurred in the remaining two contact sessions. This involved working through nine e-tutors about various aspects related to density, as shown in the menu page (see Figure 2). Each student worked individually at a computer equipped with earphones. Both researchers were on hand to provide guidance on request. Students were instructed to discuss the work with one another if they wished, however they tended to work individually most of the time. The students were also introduced to a rich problem task at the start of the module, and required to answer this, in writing, for formal assessment, by the end of the module.

Data collection

Both quantitative and qualitative data were collected in a mixed method research design (Cresswell, 2003). The design was sequential explanatory Quant–Qual–Quant in which the qualitative data collected were used to elucidate and explain quantitative findings, and direct creation of later quantitative data collection instruments (McMillan & Schumacher, 2010).

Quantitative data and instruments

The quantitative data were collected in the forms of a pre- test, post-test, biographical survey and e-tutor engagement surveys of two types: A survey answered immediately after engagement with each of the nine e-tutors, and a survey answered at the end of the entire course. All of these can be found at www.angelastott.net/academic/appendices, with samples given in the Appendix. Fourteen of the 37 marks in the pre- and 100 marks in the post- tests came from questions developed, justified and motivated by Smith, Maclin, Grosslight, and Davis (1997) and seventeen marks from the 2003 TIMSS test (Mullis et al., 2005). All the items selected from these original standardised instruments had a high degree of construct validity.

The post-test consisted of all the pre-test questions, as well as additional non-transfer questions comprising 34 marks and transfer questions comprising 29 marks. Creation of these transfer questions was informed by guidelines given by Mayer (2009) and Dawkins et al. (2008).

Creation of the nine e-tutors was informed by Mayer (2009)'s principles of multimedia design, as well as research on effective conceptual teaching of density. This includes use of dot-representation diagrams, a qualitative, conceptual focus on the component concepts of matter, mass and volume and their relationship to density, as well as development of proportional reasoning before taking an algorithmic approach (Smith et al., 1992).

Responses were captured by the software as the students engaged with the e-tutors. This engagement data includes times spent and answers given. After engagement with each of the nine e-tutors the students answered six survey questions, each of which had quantitative and qualitative components. The quantitative components were seven-item Likert scale questions. Each item was given a description, with lower extents being assigned to lower item values. The survey questions were developed specifically for this research by the two researchers, guided by the research questions.

Qualitative data and instruments

An open-ended survey item formed the qualitative component of each survey question. Additionally, seven focus-group interviews were conducted and audio recorded. Interview protocols were informed by the survey questions. These further informed the design of the final evaluation survey with the aim of detecting the students' overall e-tutor learning experience.

Data analysis

The quantitative data analysis included analysis of the participants' pre- and post- test scores. Statistically significant differences in performance for the questions common between these two tests was determined by a p value < 0.05 for the ANOVA test performed.

The qualitative data analysis was done through an iterative process of engagement with, and manipulation of, the data guided by the research questions. This was done by the two researchers for data generated by the focus group interviews and open items in the respective surveys. An initial inter-rater agreement of 85% was achieved for the two raters on the qualitative category codes. Discrepancies between the interpretations of the two raters led to dialogue until consensus was reached. This guided subsequent analysis. Manipulation techniques used to aid analysis include summarisation of salient observations, coding of survey responses and focus-group transcriptions, and calculation and graphing of various indices from the quantitative data collected. Counts of the frequency of responses are given in brackets in the text as some indication of their relative weighting.

A multi-method type of triangulation design (Cohen, Manion, & Morrison, 2003) was used to generate data from various sources using multiple instruments to study the phenomena of deep learning and design features that enhance deep learning. The various data sources have been evaluated against one another by the two raters, who were the researchers, to enhance the validity of the findings.

Results

We now give a rich description organised according to our research questions and our theoretical framework.

Deep learning

To answer our first research question, regarding whether deep learning was evident as the students engaged with the module, we analysed the data for evidence of engagement in essential and generative processing, and self-regulated learning under intrinsic, mastery-oriented motivation. We also evaluated the extent to which students underwent extraneous cognitive processing, and the effect this had on learning depth.

Essential processing

As explained in the Mayer (2009) theoretical framework, essential processing can be detected by a learner's performance on recall-type questions, i.e., those requiring responses similar to the presented information. Pre-post means for identical questions (Table 2) rose from 65% to 81%, with 40 of the 50 students making at least some improvement, 14 improving by over 25% and the mean improvement being 15%. This improvement was statistically significant, $F(1,98) = 25.99$, $p < 0.05$. It should be noted that the students received no feedback on the pre-test and had no access to either their answers or the question paper until after they had written the post-test, administered nine months after the pre-test. We therefore consider the pre-post score changes for the questions common between the two tests to be valid indicators of essential processing, rather than of superficial recall of answers due to prior exposure to the questions.

Table 2. Statistics for questions common to both pre- and post- tests

n = 50	Mean (%)	Standard deviation	p value
Pre-test	65	17.4	$1.67 \times 10^{-06*}$
Post-test	81	12.2	

As can be seen in Figure 3, all 50 students obtained at least 50% for the non-transfer questions asked in the post-test, with 46 obtaining at least 70%. Some of these questions were identical to the pre-test and others were unique to the post-test but similar to the information which had been presented in the course. The mean post-test score obtained for these questions was 84%. These findings suggest that almost all the students did engage in considerable essential processing. This is necessary, but insufficient, for generative processing.

Generative processing

Generative processing, as explained in the theoretical framework, is synonymous to deep learning. The following comment suggests that heavy work-loads in the PGCE course tend to discourage generative processing: "I'm just like churning out the work: I don't have any energy or time to actually do it properly." However, all the students were able to score at least some marks on post-test transfer questions, with the mean score being 67%. Figure 3 shows the score distribution for the transfer questions, with those for the non-transfer questions included for comparison. For convenience, we define scores below 50% on transfer questions as evidence that low, 50% to 69% moderate, 70% to 79% high, and 80% to 100% very high, levels of generative processing had occurred. Based on this definition, 42 of the 50 students underwent moderate to very high levels of generative processing. This is consistent with the students' self-reported seven-item Likert ratings regarding the extent to which they underwent deep learning. The mean rating of 4,5 falls between *a moderate amount* (12/42 students) and *quite a lot* (15).

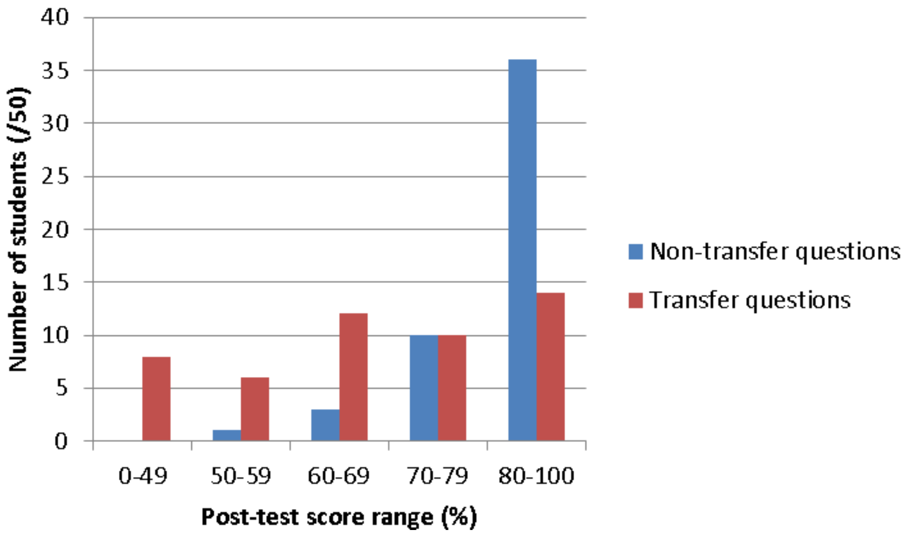


Figure 3. Post-test score ranges for transfer and non-transfer questions

Self-regulation

The students appear to have responded to the responsibility of deciding their own learning pace by making choices appropriate to their particular learning needs, i.e., by undergoing self-regulated learning. Evidence to support this includes the large ranges in the time-engagement data which seems to correspond to the large range in prior knowledge within the class. For example, the time to complete a single e-tutor ranges from a minimum of 30 seconds, for a student with good prior knowledge, to a maximum of 53,5 minutes during class and 124,3 minutes out of class time, for students with poor prior knowledge. Some students voluntarily spent considerable time engaging with the software. The course requirement was a single engagement with each of the nine e-tutors. However, only 14 of the 50 students did not repeat any of the e-tutors and the mean number of accesses per student was 15.3. One student repeated a single e-tutor nine times. Seven of the students repeated each of the nine e-tutors, but it was more common (15) for the students to only repeat those for topics they felt they had not yet mastered. Other evidence of self-regulation includes remarks (3) by stronger students that they self-directed their learning toward pedagogy since this, rather than content, was their individual learning need: "Not a deep learning in me per se, but an alternative way of viewing the concept of density from a learner's mind frame and processing it from that point of view."

Motivation

The test clearly provided a strong extrinsic motivation for the students to engage with the software beyond the minimum requirements of the course. This is evidenced by the spike in incidents of out of class e-tutor access in the weekend (46 on Saturday, 132 on Sunday) and Monday morning (33) before the test. Nine of the students indicated the test as their main motivation. Sixteen said that although this was initially their motivation, as they engaged with the software a desire to master the concepts became a stronger motivation. Eight claimed they were driven primarily by intrinsic mastery-oriented motivation throughout the module:

> "I didn't remember there was a test… I just saw it something we should learnt to teach. Because also, I'm not that confident with science concepts and explaining them in a class, and having been in a school now where I had to do a natural sciences lesson, it motivates you to try and understand."

This student refers to learning for understanding (deep learning) being linked to a conceptual approach and intrinsic motivation. The following student generalises the association between a conceptual approach and enjoyment:

> "iQuiz you're forced to think through each of the steps, and you're forced to make sure you understand each of the concepts before you can build on it. That's good, because otherwise … I don't think you'll really enjoy science if it was just like: Here's a formula, use it, plug it in."

One of the students referred to improvements in her levels of self-efficacy regarding knowledge of the content as additional motivation for re-engagement with the e-tutors. During this re-engagement she found that she understood the work and so progressed rapidly through the e-tutors with few errors, increasing her confidence and enjoyment: "I [did] them all about five times. Even Sunday night, and I came in early … And I got it right, and it's like: Yes! Correct!" This student's improvement in self-efficacy was not an isolated incidence: The pre-engagement ratings of confidence to explain content had a 3.3 mean, where 1 referred to *not at all confident* (106 / 548 responses) and 7 to *extremely confident* (14). The mean post-engagement rating was 4.7, with 406 responses indicating a rise in confidence due to engagement with the software.

Extraneous cognitive processing

Extraneous processing is cognitive activity which impedes the learning process according to the Mayer (2009) framework. Data used to evaluate the extent to which the students underwent extraneous cognitive processing includes: A Likert-scale item and open-ended question, in the survey at the end of each e-tutor, about the extent to which the student clicked without thinking and reasons for this; Examination of engagement data for students who revisited slides, since this is an indication of the software having rerouted the student *in circles*, which may have resulted from attempts to game the system.

Out of 546 survey responses regarding clicking without thinking, 332 were *not at all*, assigned a rating of 1, and none were *all / almost all the time*, assigned 7. The mean rating, 1.9, corresponds closely to the option *almost never* (82). Most (217/368) of the reasons given suggest a mastery-oriented epistemology favourable for deep learning. For example:

> "I've been through a lot of bad ones where to me it was just pointless and I went through it just click click click. But when I started this immediately I realized… the information they've given me it's actually making me want to go in depth and really involve myself 100% in it. I think it's … the way the program is structured – it made me want to understand."

Of the 368 responses about clicking without thinking: 35 referred to the intervention being easy. For some this was a reason for clicking without thinking (22 responses): "I felt I knew the answers and was just going through the motions of answering" and for some a reason for not clicking without thinking (13). Other explanations for clicking without thinking include: Being in a hurry (9), boredom or tiredness (9), doing so by accident (8), lapses in concentration (8), a false sense of knowing the content (5), technical and trialing issues (6), and not knowing how to answer (22): "When I didn't know what the answer was I simply guessed and thus just clicked to get to the next slide." However, deep learning appears to have resulted despite this, in some cases, due to the repetitive and sub-step designs of the software: "Because it still gives you the explanation of why it is correct," "It's really small steps so even if you guess it's almost as if it's a progression to the next one."

We considered 20 slide-revisits or fewer reasonable and so not indicative of an attempt to game the system. Almost all (732/764) the times a student accessed an e-tutor, fewer than 20 slide revisits were recorded, with 343 involving zero, and 269 one to five, revisits. The 32 cases of slide revisits greater than 20 were examined in detail. For 11 of these the students reported little to no attempts to game the system, and gave positive associated remarks. For example, a student who made 86 slide-revisits, spending 53 minutes on the e-tutor as a whole, remarked that it "actually challenged me to think critically and reason things out first before pressing the click button." Even for the remaining 17 cases, in which students reported attempting to game the system, the majority gave positive ratings for experiencing the e-tutor as helpful (mean = 4.8/7) and enjoying it (mean = 5/7). A student who revisited 80 slides, spending 2.5 hours on the e-tutor, found it "excellent." One student admitted trying to game the system for the first e-tutor, but reported that its design had deterred her from repeating this:

> "Natural dislike for this subject so habit to try click through it to get it over! But it not letting me continue forced me to think about it."

Effective design features

In the final evaluation 12 students responded to a question about design features which they found helpful in promoting deep learning. The features they identified were: Allowing them to work at their own pace (2), requiring them to think deeply (4), being visual, real and relevant (2), interactive (2), and interesting (2). During the focus group interviews we found additional evidence for each of these views. Some of this evidence is discussed below. Additionally, there was general agreement when some students referred to the value of the conceptual approach, as opposed to the very algorithmic approach they had experienced in school science, for example:

> "I feel like in high school, yes, I remember the concept, but I feel like what I was given there was a recipe on how to calculate density or volume but I was never taught where it comes from – like the foundation. So now the iQuiz is like an eye-opener because now I understand volume, I understand density and I understand where it comes from."

During the post-course focus group interviews there was also general agreement about the value of all the components of the module: The e-tutor, the rich problem task (*Archimedes task*), and the test, for promoting deep learning. For example, a student described how her initial engagement with the e-tutor led her to a basic understanding of the concepts so that she was able to understand the requirements of the Archimedes task as well as information she voluntarily accessed on the Internet. She was then able to use all this learning in answering the Archimedes task, developing understanding in the process:

> "[For the task] you just have to pull [the concepts] together. So the process of engaging with the [task] really made me understand density… And it's just that process of grappling with information to reach understanding."

Only six of the 42 students who answered the final evaluation indicated that they would have preferred traditional lectures. Reasons for preferring the e-tutor to traditional lectures include being able to work at their own pace (17): "By doing it by yourself you can go at your own pace. If it is done in class and it is gone through slowly people could get bored," enhanced interactivity (7): "We really got to engage with the work more so than we would have in lectures," and the safety of answering without fear of ridicule (1): "It was very easy to understand and was done in a very neutral way with no fear of doing the wrong thing." These responses were echoed in the focus group interviews by students with both extensive and limited science backgrounds:

> "That was the biggest advantage for me: The fact that you could work through it at your own pace, because there were a couple of things I felt you could just go through easily and there were other things I knew if I needed to I could go back and redo it."
>
> "For the first time it wasn't a proper chore"… "And I watched every video and wrote my notes."…"I love that I can go at my own pace. And I thought: Imagine if you taught this in class. I wouldn't have listened. I would have just done what everyone else does and I would never have understood and I could never have gone back and done it."

The students were asked, in both survey questions and interviews, to comment on how the e-tutors' sub-step design and branched structure affected their learning experience. Evidence for the contribution of each of these design features in guiding the learners toward deep learning has already been given. The majority of responses (436/542) suggested that most students found the sub-step design effective in promoting deep learning. However, particularly students with higher prior content knowledge found the sub-step design tedious and frustrating (46 responses) and some (15 responses) ascribed their attempts to game the system to this. The branched and potentially repetitive structure of the e-tutors was a cause of frustration in some cases since a participant may be re-routed, possibly multiple times, to previously incorrectly answered questions: [I tried to game the system] “Because I got a bit tired of going around in circles and repeating some things.” This frustration and attempt to game the system were sometimes temporary and the repetition alerted the students to their need for learning, resulting in deeper learning. For example:

> “I think I went around thrice and I’m thinking: There’s something wrong here. And then I re-read … the question and I thought, okay but then it said this and this and that and then I sat and I stared at the computer and I was like really thinking and then it came to me after you know – wait a minute – it’s actually … A, and when I went ‘A’ and it was right, I was like ‘Oh that’s why!’ and then I rethought the process through.”

Conclusion, discussion, limitations and implications

We provide the following assertions in summary of the findings and discussion, and in answer to our research questions:

- Most of the students used the e-tutor engaged in a manner which suggests moderate to high levels of deep learning. This is suggested by evidence for limited extraneous processing associated with a lack of clarity, evidence for essential processing by most of the students, and evidence for generative processing, metacognitive control, and intrinsic mastery-oriented motivation among many of the students.
- Effective design features of the module include test and problem-based task inclusion, the sub-step design, branched structure, and conceptual approach of the electronic tutoring software.

In some cases the frustration which the students felt as the software rerouted them to previously incorrectly answered slides led to deep learning. This appears to be because the extraneous processing associated with this frustration made the student aware of their need to learn. This appears to contradict Mayer (2009)’s view of extraneous processing not serving an instructional goal. Instead, it appears to correspond to Schwartz et al. (2009)’s findings that problematisation of learning, associated with some extraneous processing, can motivate engagement in deep learning. We believe that extraneous cognitive processing which results from a lack of clarity does not aid deep learning. However, extraneous cognitive processing which results from discrepance between the learners’ conceptual structures and the presented information, requiring the learner to recalibrate their self-assessment (Nicol & Macfarlane-Dick, 2006) and realise the need to undergo accommodation, does aid deep learning. It appears that most of the extraneous processing the students underwent as they engaged with the e-tutor was of the latter kind.

The value of the e-tutors’ sub-step design in promoting deep learning is consistent with VanLehn (2011)’s thesis that fine-grained formative feedback is crucial to a tutor’s success. The value of the conceptual approach to deep learning is consistent with Paul and Elder (2008)’s view that understanding is developed through conceptual, rather than algorithmic, engagement with subject matter. The value of inclusion of a rich problem task to accompany engagement with the e-tutors, for promoting deep learning, is consistent with work by Stott (2008) and Van Loggerenberg-Hattingh (2003).

As a convenience case study, the sample used is understood not to represent any particular population. It is possible that the PGCE students at the institution where the research was conducted are not similar to PGCE students in other institutions. These students’ access to the relevant resources for e-tutors to be effective are likely to be higher than for students in less developed parts of the world, particularly in the rest of Africa. This includes access to high speed broadband and computers both on campus and at home. Also, since the iQuiz e-tutor is conceptual in focus, it requires greater content and pedagogical knowledge, and skill and time to populate with appropriate content, than the more commonly available algorithmic ITSs (Graesser, Conley, & Olney, 2012). The resulting limitations to scalability may reduce the viability of the solutions presented here. However, as pointed out by VanLehn (2011), the reusability and effectiveness of such conceptual e-tutors may make their high initial costs worthwhile.

This study provides valuable insight into potential ways to improve teaching and learning in general using conceptual tutoring software to promote deep learning. It also adds to the sparse literature available on in-depth mainly qualitative research about student engagement with, and effective design of, a somewhat extended learning module involving a conceptual e-tutor.

References

Aleven, V., Mclaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education, 19*(2), 105-154.

Blom, S., & Severiens, S. (2008). Engagement in self-regulated deep learning of successful immigrant and non-immigrant students in inner city schools. *European Journal of Psychology Education, 1*, 41-45.

Cohen, L., Manion, L., & Morrison, K. (2003). *Research methods in education* (5th ed.). London, UK: Routledge Falmer.

Cresswell, J. W. (2003). *Research design: Qualitative, quantitative and mixed methods approaches* (2nd ed.). London, UK: SAGE Publications.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International journal of human-computer studies, 70*(5), 377-398.

Dawkins, K. R., Dickerson, D. L., McKinney, S. E., & Butler, S. (2008). Teaching density to middle school students: Preservice science teachers' content knowledge and pedagogical practices. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 82*(1), 21-26.

Duckworth, E. (1986). Teaching as research. *Harvard Educational Review, 56*(4), 481-496.

Dweck, C. S. (2002). Beliefs that make smart people dumb. In R. J. Sternberg (Ed.), *Why smart people do stupid things*. New Haven, CT: Yale University Press.

Fournier-Viger, P., Nkambou, R., & Nguifo, E. M. (2010). Building intelligent tutoring systems for ill-defined domains. *Advances in intelligent tutoring systems* (pp. 81-101). Berlin Heidelberg, Germany: Springer.

Graesser, A. C., Conley, M. W., & Olney, A. (2012). *Intelligent tutoring systems.* APA handbook of educational psychology. Washington, DC: American Psychological Association.

Kang, S., Scharmann, L. C., & Noh, T. (2004). Reexamining the role of cognitive conflict in science concept learning. *Research in Science Education, 34*(1), 71-96.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Santa Barbara, CA: Oxford University Press.

McMillan, J. H., & Schumacher, S. (2010). *Research in Education: Evidence-based inquiry* (7th ed.). Boston, MA: Pearson.

Merriam, S. B. (1988). *Case study research in education*. San Fransisco, CA: Jossey-Bass.

Mitrovic, A., Martin, B., & Suraweera, P. (2007). Intelligent tutors for all: The constraint-based approach. *IEEE Intelligent Systems, 4*, 38-45.

Mitrovic, A., & Weerasinghe, A. (2009). Revisiting iIll-definedness and the consequences for ITSs. In V. Dinatrova et al. (Eds.), *Artificial intelligence in education: Building learning systems that care from knowledge representation to affective modelling* (pp. 375-382). Amsterdam, Netherlands: IOS Press.

Muldner, K., Burleson, W., Van de Sande, B., & VanLehn, K. (2011). An analysis of students' gaming behaviors in an intelligent tutoring system: Predictors and impacts. *User modeling and user-adapted interaction, 21*(1-2), 99-135.

Mullis, I. V., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A., & Erberber, E. (2005). *TIMSS 2007 Assessment Frameworks*. Retrieved from ERIC database. (ED494654).

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the interactive tutoring feedback model. *Digital Education Review, 23*, 7-26.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education, 31*(2), 199-218.

Paul, R., & Elder, L. (2008). *Critical & creative thinking*. Dillon Beach, CA: The Foundation for Critical Thinking.

Rowell, J., & Dawson, C. (1977). Teaching about floating and sinking: An attempt to link cognitive psychology with classroom practice. *Science Education, 61*(2), 243-251.

Schraw, G., Olafson, L., & VanderVeldt, M. (2012). Fostering critical awareness of teachers' epistemological and ontological beliefs. *Personal Epistemology and Teacher Education, 61*, 149-164.

Schwartz, D. L., Lindgren, R., & Lewis, S. (2009). Constructivism in an age of non-constructivist assessments. In S. T. T. M. Duffy (Ed.), *Constructivist Instruction: Success or Failure?* (pp. 34-61). New York, NY: Routledge.

Selwyn, N. (2007). The use of computer technology in university teaching and learning: A critical perspective. *Journal of Computer Assisted Learning, 23*(2), 83-94.

She, H.-C., & Liao, Y.-W. (2010). Bridging scientific reasoning and conceptual change through adaptive web-based learning. *Journal of Research in Science Teaching, 47*(1), 91-119.

Smith, C., Maclin, D., Grosslight, L., & Davis, H. (1997). Teaching for understanding: A comparison of two approaches to teaching students about matter and density. *Cognition and Instruction, 15*(3), 317-393.

Smith, C., Snir, J., & Grosslight, L. (1992). Using conceptual models to facilitate conceptual change: The case of weight-density differentiation. *Cognition and Instruction, 9*(3), 221-283.

Snir, J., Smith, C. L., & Raz, G. (2003). Linking phenomena with competing underlying models: A software tool for introducing students to the particulate model of matter. *Science Education, 87*(6), 794-830.

Stott, A. E. (2008). *Promotion of critical thinking in school physical science* (Unpublished doctoral dissertation). University of KwaZulu- Natal, Durban, South Africa.

Stott, A. E., & Case, J. M. (2014). Electronic tutoring as a tool for promoting conceptual change: A case study of in-service science teacher workshops. *African Journal of Research in Mathematics, Science and Technology Education, 18*(2), 139-150. doi: 10.1080/10288457.2014.912832

Van Loggerenberg-Hattingh, A. (2003). Examining learning achievement and experiences of science learners in a problem-based learning environment. *South African Journal of Education, 23*(1), 52-57.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221.

Wiser, M., Smith, C. L., Asbell-Clarke, J., & Doubler, S. (2009, April). Developing and refining a learning progression for matter: The inquiry project: Grades 3-5. In C. S. (Chair), *Developing and refining a learning progression for matter from pre-K to grade*. Symposium conducted at the meeting of the American Educational Research Association, San Diego, CA.

Xu, L., & Clarke, D. (2012). Student difficulties in learning density: A distributed cognition perspective. *Research in Science Education, 42*(4), 769-789.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29*(3), 663-676.

Appendix

Sample questions from:

Pre-test (repeated in post-test)

Question 1

(5)

Look at the pairs of sugar-water mixtures alongside.
Circle the correct option in each case.

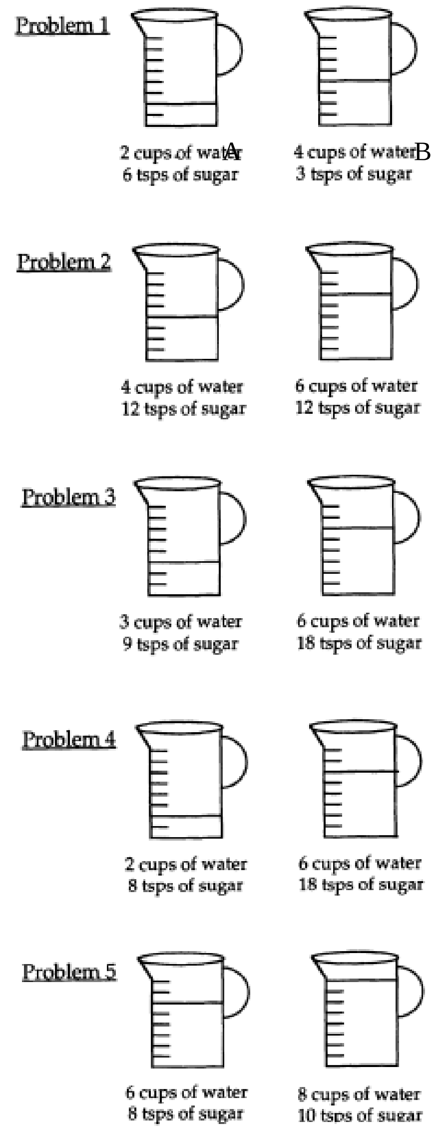
Problem 1:
The sweeter mixture is:
[A / B / both A and B are the same sweetness].

Problem 2:
The sweeter mixture is:
[A / B / both A and B are the same sweetness].

Problem 3:
The sweeter mixture is:
[A / B / both A and B are the same sweetness].

Problem 4:
The sweeter mixture is:
[A / B / both A and B are the same sweetness].

Problem 5:
The sweeter mixture is:
[A / B / both A and B are the same sweetness].



Question adapted from "Teaching for understanding: A comparison of two approaches to teaching students about matter and density" by C. Smith, D. Maclin, L. Grosslight and H. Davis, 1997, *Cognition and Instruction, 15*(3), p. 391. Copyright 1997 by Taylor & Francis, Ltd. Adapted with permission.

Post-test

QUESTION 17

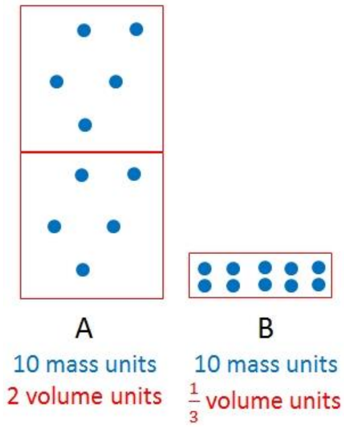
15 Marks

You are a Natural Sciences teacher. You are teaching your class about density. Explain how you would answer a learner who makes each of the following statements. In your answer you need to say whether the learner's statement is correct or not, and why this is so. This must be done in a way that helps all the learners in the class to understand density better. You can make use of examples and / or pictures to help make your explanation clearer.

17.1. "It is possible for a big object to be lighter than a smaller object. "

17.2. "Block A is 1 kg. Block B is 0,5 kg. If Block A floats, B will definitely float too."

17.3. "For the diagram alongside, B is denser than A because B's mass is squashed into a smaller space than A's. But this might not mean that B's particles are squashed together more than A's."



Biographical survey

Permission to collect data

If you give permission for us to use this data, this will be anonymous and every effort will be made that no harm will come to you from this. Your name is only asked in the questionnaire for logistical purposes. It will never appear in any publication. Do you give us permission to use this data?

☐ yes
☐ no

Yourself

Surname

Name

Age

Gender:

☐ male
☐ female

School science learning

Did you complete grade 12 in South Africa?

☐ yes

☐ no

Did you take natural science up to grade 9?

☐ yes
☐ no

Give your approximate grade 9 natural science mark

☐ A
☐ B
☐ C
☐ D
☐ E or below

e-tutor engagement survey answered after each e-tutor engagement

Helpfulness

How helpful was this tutoring session?

☐ 1 A waste of time
☐ 2 Hardly helpful
☐ 3 Slightly helpful
☐ 4 Fairly helpful
☐ 5 Helpful
☐ 6 Very helpful

Why was this tutoring session helpful / not helpful?

Meaningless clicking

How much did you click links without thinking?

☐ 1 Not at all
☐ 2 Almost never
☐ 3 A few times
☐ 4 A moderate number of times
☐ 5 Fairly many times
☐ 6 A lot
☐ 7 All / almost all of the time

Why did you / did you not click without thinking?

Final e-tutor engagement survey

How you engaged with iQuiz

Did you take notes while you engaged with iQuiz?

☐ 1 Never
☐ 2 I took a few very short notes
☐ 3 I took notes for a fair amount of the work
☐ 4 I took notes for much of the work
☐ 5 I took notes for all of the work

Why did you / did you not take notes while engaging with iQuiz?

What was your main motivation for engaging with the density iQuizzes?

- ☐ 1 Having to tick off the list
- ☐ 2 The test
- ☐ 3 Wanting to learn the knowledge
- ☐ 4 Initially the test / ticking the list and later wanting to learn the knowledge
- ☐ 5 Other

If you say ‘other’, please say what this was

Why was this your main motivation for engaging with iQuiz?

Did you repeat any of the iQuizzes?

- ☐ 1 None
- ☐ 2 Yes, I randomly repeated some
- ☐ 3 Yes, I purposefully repeated those on topics which I felt I needed to work more on
- ☐ 4 Yes, I repeated all of them