

Evaluation of computer-based learning material for food chemistry education

Julia Diederer,¹ Harry Gruppen,¹ Rob Hartog² and Alphons G. J. Voragen¹

¹ Wageningen University, Department of Agrotechnology and Food Sciences, Laboratory of Food Chemistry, P.O. Box 8129, 6700 EV Wageningen, The Netherlands

² Wageningen University, Wageningen Multimedia research centre
e-mail: fons.voragen@wur.nl

Received 21 September 2004, accepted 8 April 2005

Abstract: Digital exercises were designed and developed for food chemistry education. During the design process, design requirements were described for such exercises. The exercises were evaluated in three case studies, firstly to determine whether the exercises satisfy the design requirements with respect to students' use and secondly to provide insight into the effect of the course structure and organisation on the value that the students attribute to the exercises. The results show that the exercises meet most of the design requirements. Students found the exercises clear and helpful, and most students confirmed that these exercises helped them in their preparations for their examinations. Despite this, participation in the programme was low when working on the exercises was not compulsory. The differences in evaluation results between the three studies can be explained by differences in the course structure and organisation. [*Chem. Educ. Res. Pract.*, 2005, **6** (2), 64-82]

Keywords: food chemistry; computer based learning; learning objects; evaluation; case study; design requirements; context; learning activities; learning activity preference

Introduction

The course, Food Chemistry (6 ECTS: European Credit Transfer System), is a second year course for students in their Bachelors curriculum Food Science & Technology at Wageningen University (Wageningen, The Netherlands). In this course, students acquire basic knowledge of food chemistry, i.e. qualitative and quantitative knowledge about important (bio-)chemical reactions in food and the influence of the processing conditions on these reactions, and finally, development of laboratory skills. To facilitate this, digital exercises have been designed and developed to motivate and activate the students.

For the design of digital exercises, a design (and development) process for activating and motivating digital learning material has been developed. Diederer and co-workers (2003) described in detail in this journal the design process and, in particular, design guidelines and design requirements that are based on theories and views on learning and instruction, on food chemistry subject matter and learning goals, and on user interface design. These were followed during the design and development of a set of 106 digital exercises.

Education supported by digital learning objects is relatively new, compared to education through lectures, self-study with books, group work and laboratory classes. To usefully incorporate the digital exercises into the food chemistry course, which consists of several educational activities, one needs to answer questions such as, 'what should be the sequential

Chemistry Education Research and Practice, 2005, **6** (2), 64-82

order of different educational activities such as lectures and computer classes?', 'what is the effect of the number of lectures on the need for and use of digital learning objects?' and 'what may happen when there is no extrinsic motivation for the use of the digital learning objects (e.g. when their use is not compulsory)?'

The present article deals with the evaluation of the digital exercises in three case studies, each within a different course structure and organisation. The first two were carried out in the regular educational setting of the Food Chemistry course at Wageningen University in two successive years. In this setting the use of the digital exercises was compulsory in both case studies, but the number and type of lectures accompanying the digital exercises differed. The third was performed during the Food Chemistry course at Cornell University (Ithaca, NY, USA), in which the use of the digital exercises was optional.

The aim of these studies was twofold:

1. To determine whether the learning material satisfied the design requirements on student use (Diederer et al., 2003), dealing with the influence of the context, the quality and user-friendliness of the design, the extent to which the digital exercises assist in studying the reader, and the additional value of the digital exercises.
2. To collect information about the possible relation between differences in the students' appreciation of the digital learning material and differences in course structure and organisation.

The digital exercises

In total, 106 digital exercises were designed for food chemistry education. These exercises comprise interactive questions and assignments that invite students to practise on different topics within food chemistry. A more detailed description of the digital exercises, as well as the rationales for the design decisions, has been published previously (Diederer et al., 2003). In this section a brief description is given. The content of the exercises was based on the content of the reader used during the course at Wageningen University, since the digital exercises are in the first instance designed for this course. Since a food chemistry course at another university most likely differs to some extent from that of the food chemistry course at Wageningen University, it cannot be guaranteed that all our digital exercises are equally useful to all food chemistry courses.

The digital exercises have the following seven main features:

- Exercises are grouped into sequences of at most ten exercises.
- A score for each exercise and a score overview for each sequence of exercises are incorporated into the exercises in order to motivate the students not to guess too much and to repeat an exercise with a low score (maximum score is 10; minimum can even be a negative score).
- Exercises contain diagrams, schemes, pictures and animations.
- Exercises contain examples from industry or daily life to show the usefulness of knowing and understanding typical reactions in food, e.g. lipid oxidation, Maillard reactions and enzymatic browning.
- Within each sequence of exercises the degree of difficulty is gradually increased.

- Exercises contain feedback varying from simple remarks (e.g. ‘wrong’ or ‘right’) to more detailed hints (e.g. specific feedback according to the student’s answer) as well as different kinds of additional information.
- Exercises differ in the type of questions or assignments (e.g. multiple choice, categorising, ranking) and cover different topics (e.g. about molecular structure, reactions, definitions, applications).

Exercises were developed with the software program Flash (Macromedia®) and were made accessible to the student by the web-based learning environment BlackBoard®.

Figure 1. An example of an exercise: three reactions involving quinones

An example of an exercise is shown in Figure 1. Top left the student can see that he/she is working on exercise number 4 in a sequence of five exercises. The student can click on each of the numbers to go to a specific exercise or on ‘score’ to go to the score overview. The score overview shows the student the score for each exercise within this sequence. Top right the score for the current exercise is shown. Centred at the top it is written what should be done in this exercise. In this example, students have to complete a scheme of three different reaction pathways of quinones by dragging the molecules on the right to the corresponding spot of one of the pathways on the left. At the moment of the screenshot of Figure 1 the user was dragging a quinone out of the box on the right to move it to a spot on the left. Students can ask for information, by clicking on the ‘HINT’ button. In this particular case, on using this button the effects of oxidation of phenolic components are explained through the process of tea fermentation. By reading this explanation it becomes clear which reaction pathway causes browning, formation of aromas or protein precipitation.

Evaluation method

The digital exercises were designed according to design guidelines and design requirements based on theories of learning and instruction, on subject matter and learning goals, and on theories of user interface design. Theories of learning and instruction that were used are mainly the cognitive load theory (Sweller et al., 1998 and Kirschner, 2002), motivation of students based on the ARCS-model (Attention, Relevance, Confidence, Satisfaction; Keller, 1983) and active learning (e.g. Keyser, 2000). The sources for the different design requirements have been described extensively in the previous paper (Diederer et al., 2003). The requirements define operationally the goals of the design, and the variables, to which the requirements refer, can be measured. To be able to decide whether the learning material meets the requirements, minimum criteria need to be specified for each requirement. Some requirements need to be tested by students, some by teachers and some by specific experts as user interface designers, educational specialists or food chemistry specialists. This paper deals with the requirements that need to be tested by students, which is done with case studies.

The method of evaluation of requirements on students' use can be described as follows. The students work on the exercises and complete an evaluation questionnaire. When the results of the evaluation are equal to or higher than the criteria specified for each requirement, that requirement is judged to be satisfied. The details about how this is carried out are given later.

To investigate the influence of the course structure and organisation on the usefulness of the digital exercises for the students, the digital exercises were evaluated in different case studies.

The requirements

The requirements for the digital exercises that were evaluated in the case studies are divided into four related sets. These are listed in Table 1, together with the corresponding evaluation questions and criteria. The source for each requirement can be found in the previous paper (Diederer et al., 2003).

The four sets of requirements are:

The influence of the context on the case studies

This set of requirements takes into account the fact that the context during the case study influences the behaviour of the user (Pawson and Tilley, 1997). Results from one evaluation (case study) cannot simply be generalised. The context within each case study is defined by the prevailing conditions during the use of the learning material. Specifically for this purpose, the requirement 'The circumstances during the evaluation are taken into account' is described (requirement rE, Table 1). This requirement is not used during the design process, but is described as an *evaluation requirement*: the evaluation of the digital exercise on students' use has to meet this requirement.

The context also includes the learning activity preferences of the students. This could influence the attitude of the students to the learning material: a student who prefers to learn by working in groups or by listening will not be attracted to the digital exercises. Also, the way the student uses the digital exercises could be influenced by the learning preferences (Vermetten et al., 2002). Our learning material aims at Bachelor students enrolled in the curriculum Food Science and Technology who have a learning preference for activities such as answering questions and studying schemes, diagrams and animations.

We realize that several other circumstances could be important to describe the context during the case studies, such as the prior knowledge of the students, the quality of the teachers, prior experiences with computer supported education for both teachers and students, the gender and age of the participants, etc. An endless list of variables could be described and analysed for each study, but we chose to limit the scope of our study.

The quality and user-friendliness of the design

The second set is related to requirements that deal with the way the students perceive the design of the digital exercises. The design can be judged by criteria, such as usability, clarity, and manageability. These judgements give an indication whether students are pleased with the set up of the digital exercises.

The extent to which the digital exercises assist in studying the reader

This set of requirements is related to the main goal of the digital learning material: facilitating the learning of food chemistry. For this, students need to understand the content of the exercises, recognize the content of the exercises in the reader, which is used in the course, and feel that it helps him to remember and understand parts of the reader after completing the exercises.

The additional value of the digital exercises

The fourth set of requirements is related to the expectation that digital exercises have an additional value compared to other learning materials, such as books, and other learning activities, such as lectures and laboratory classes. The digital exercises are designed to motivate the students and require their active involvement. Every opportunity to present information in the form of a diagram, scheme or animation instead of text has been seized. Therefore, it is expected that students should learn a lot in a relative short period of time. Important for this set of requirements is whether students work on the digital exercises on their own initiative (an indication that students see their value), whether students feel they learned much and whether studying the exercises contributed to their ability to pass the final examination of the course.

The evaluation questions and criteria

In total, twenty-six questions were developed to evaluate the requirements. For the four sets of requirements there are six, ten, four and six questions respectively (Table 1). These questions were given to the students through two different questionnaires. The first is a standard course evaluation questionnaire, which is regularly used at Wageningen University. This questionnaire concerns, amongst others, the content, the organisation, the quality of the learning materials, the quality of the teachers, the perceived value of the different educational activities and the way in which information and communication technology has been used for learning support in the course. The second is a specific questionnaire developed for the evaluation of the digital exercises. In Table 1 the questions of both questionnaires are listed, from which questions qE.1 to qE.5, q1.2, qQ.1 and q3.1 refer to the questions from the standard questionnaire.

Table 1: The sets of requirements, their evaluation criteria and the corresponding questions or statements in the questionnaires.

Set	Requirement [#]	Questions or statements	criteria
The influence of the context	The circumstances during the evaluation are taken into account. (rE) ^{##}	qE.1: In my opinion other digital learning material (than the digital exercises) is valuable. qE.2. I am satisfied with the desk-space-facilities and computer rooms during the computer classes. qE.3. I am satisfied with the use of Blackboard. qE.4. I am satisfied with the course sequence. qE.5: Overall rate of the course on a scale of 1 (poor) to 5 (excellent).	criteria for 5-point scale*
	The exercises fulfil a need of most of the students in relation to their learning activity preference. (r12)	q12.1. I am someone who learns through (students choose maximally 3 answers from 7 learning activities).	A significant number (p<0.05) of students like to learn by the activities 'active answering questions' and 'looking at schemas/diagrams'
The quality and user-friendliness of the design	The exercises are clear. (r1)	q1.1. The exercises are clear (you know what you have to do). q1.2. The formulation of the instructions of the digital learning material was clear and understandable.	criteria for 5-point scale*
	Students perceive the quality of the learning material as good. (rQ) ^{##}	qQ.1. The quality of the digital learning material was good.	criteria for 5-point scale*
	The exercises are manageable. (Contain enough hints to work through the exercises.) (r2)	q2.1. With those exercises that I needed hints, the hints were provided. q2.2. With those exercises that hints were provided, the hints were good enough to go through the exercises.	criteria for 5-point scale*
	The exercises are fun to work on. (r5)	q5.1. The exercises are fun to work on.	criteria for 5-point scale*
	Students are motivated not to guess too much. (r10) Students are motivated to repeat an exercise when they do badly. (r11)	q10.1. I used the score as a tool that tells me how well I performed in the exercise: a low score means I did badly, a high score means I did well. q10.2. I used the score as a motivation not to guess too much (except if I did not have another choice). q11.1. When I had a low score I did the exercises again. q11.2. I did the exercises many times until my score for each exercise was close to 10.	criteria for 5-point scale*

Table 1 (continued): The sets of requirements, their evaluation criteria and the corresponding questions or statements in the questionnaires.

Set	Requirement	Questions or statements	criteria
Assistance in studying the reader	Content of an exercise is understood after completing the exercise. (r6)	q6.1. At the end of an exercise (after explanation by the computer) I understood the content of the exercise.	criteria for 5-point scale*
	Students are able to recognise the exercises in sections of the reader. (r7)	q7.1. I recognised parts of the exercises in the reader.	criteria for 5-point scale*
	Sections of the reader are easy to remember after completing the exercises. (r8)	q8.1. Parts of the reader were easy to remember after doing the exercises.	criteria for 5-point scale*
	Sections of the reader are easy to understand after completing the exercises. (r9)	q9.1. Parts of the reader were easy to understand after doing the exercises.	criteria for 5-point scale*
Additional value of the exercises	The exercises are used by students on their own initiative. (r4)	q4.1. Tell when you worked on the exercises. q4.2. Estimate how many times you worked on the exercises.	≥ 75% of the students worked outside reserved time and ≥ 75% worked on them more than once
	Students perceive the exercises as valuable. (related to r3)**	q3.1. In my opinion the digital exercises are valuable learning material.	criteria for 5-point scale*
	Students feel they learned much from doing the exercises. (r13)	q13.1. How much did you learn of the different parts of the Food Chemistry course? (Six parts are judged on 5-point scale from nothing to very much)	≥ 75% of the students answer “much” or “very much”
	Students feel that the exercises contributed to their ability to pass the final examination successfully. (r14)***	q14.1A. Estimate (in percentage) how much each part of the course contributes to your ability of doing the final examination. (Your total estimate should add up to 100%.)	Average learning efficiency is ≥ 1 (learning efficiency = %contribution / %time)
		q14.1B. The exercises contributed to my ability to answer questions in the final exam. q14.2B. The exercises are efficient for learning: I learned about food chemistry in a relative short period of time.	criteria for 5-point scale*

#: The codes of the requirements are the same codes as used in the design process (Diederer et al., 2003).

##: Requirements rE and rQ were not described during the design process. rE is an evaluation requirement (not a design requirement).

*: Average rating should be 4.0 or more AND at least 75% of the students should give a rating of 4 or 5. (5-point scale: 1=totally disagree, 2=partially disagree, 3=neutral, 4=partially agree, 5=totally agree)

** : r3 is ‘ Students see the digital exercises as a valuable addition to the reader’.

***: questions for r14 differ between case studies, see Table 2

In both questionnaires, most questions are given as statements that need to be judged on a 5-point Likert scale, ranging from 1, meaning 'strongly/totally disagree', to 5, meaning 'strongly/totally agree'. From the answers of all students an average judgement per statement can be calculated. A requirement is satisfied when the average rating for the accompanying statement is at least 4.0 *and* when at least 75% of the students rate the statement with a 4 or 5. Some requirements were not tested with 5-point scale questions, but with multiple-choice questions. Most of these requirements have also a criterion of 75% positive judgement, which is in line with the 75% criterion of 5-point Likert scale questions. See Table 1 for specific criteria.

Biases could arise from evaluation questionnaires as used for the current research. Typical method biases are for example those related to acquiescence (agree with attitude statements regardless of content), social desirability (behave in a culturally acceptable and appropriate manner), positive / negative item wording (only using positively or negatively worded statements), or common scale formats (e.g. Likert) and anchors (e.g. agree / disagree) (Podsakoff et al., 2003). It should be noted that the questionnaires are only used to compare the average attitudes of all students between case studies and to compare the average attitudes of all students with the criteria. Therefore, we think that acquiescence bias is not a big issue.

The course structure and organisation in each of the three case studies

The course structure and organization includes components and aspects such as kinds of course topics, number of course credits, number of lectures, number of scheduled computer classes, sequence order of lectures and computer classes, number of laboratory classes, staff/student ratio, relation between reader content and content of the digital exercises. Table 2 shows for each of the three case studies the relevant characteristics of course structure and organisation, and also the type of evaluation questionnaire used. Case studies I and II were both conducted during the Food Chemistry course at Wageningen University in the regular educational setting. The course concerned is a 6 ECTS course (168 study hours). Students had lectures and computer classes, and prior to the final examination in both case studies the course concluded with a laboratory class, all within a time-span of six weeks. After these six weeks there was a week without formal instruction, followed by an examination week. In both case studies I and II a reader, covering the course content, was available to the students. After the final examination the students were asked to fill in the two questionnaires.

With respect to the students in studies I and II, one important difference could be expected between these two groups of students: their assumed prior knowledge of food chemistry when they started working on the digital exercises. Students in case study II had fewer lectures than students in case study I and worked on the exercises on the same day as the lectures were given, which gave them no time to process newly acquired information. In fact, students in case study II were exposed to new information during the exercises. Information, which students in case study I already heard of during the lectures.

The course structure and organisation of case study III was different from the other two. This was carried out at Cornell University (Ithaca, NY, USA), with students taking the Food Chemistry course (3-credit course, comparable with 4.5-6 ECTS). For study purposes an extensive set of lecture notes, which was almost like a textbook, was given to the students. The content of these lecture notes were in some aspects different from that of the reader used at Wageningen University: topics such as water and food additives were included, while topics such as enzymes or phenolic components received less attention. Another important difference in course organisation was that the students' use of the digital exercises was optional, while in case

studies I and II the students were obliged to participate in computer classes in which they had to work with the digital learning material. After the final examination, in case study III the students were also asked to fill in a specific questionnaire. The questions were identical to the questions in the specific evaluation questionnaire used at Wageningen University (questions from Table 1), with some minor changes related to the course structure and organisation at Cornell University. As shown in Table 2, the questions related to requirement r14 were different in each case study.

TABLE 2: The course structure and organisation in each of the three case studies.

Item	Case study I	Case study II	Case study III
Course & Institute	Food Chemistry, Wageningen University	Food Chemistry, Wageningen University	Food Chemistry, Cornell University
Course topics	proteins, lipids, carbohydrates, phenolic components, enzymes	proteins, lipids, carbohydrates, phenolic components, enzymes	water, food colloids, proteins, lipids, carbohydrates, food additives
ECTS* (for course)	6 ECTS	6 ECTS	4.5-6.0 ECTS **
course code	FCH 20806	FCH 20806	FD SC 417
Computer classes (h) for digital exercises	16 h	15 h	Not scheduled
Computer classes (h) for other learning material	16 h	16 h	Not scheduled
Lectures: -introductory -regular -response	- 0 h - 16 h - 0 h	- 5 h - 0 h - 4 h	42 h of lectures.
Laboratory classes (h)	40 h	40 h	Not included (separate class)
order of lectures and digital exercises	1. lectures 2. computer classes	1. 1-h introd. lecture + 3-h computer class repeated 5 times 2. response lectures***	not specified: use of exercises was voluntary
Relation of the content of the reader / lecture notes to the content of the digital exercises	directly related	directly related	only indirectly related
Number of students	34 students	30 students	25 students
Type of questionnaire	Both standard and specific.	Both standard and specific.	Only a specific.
Questions for r14 (Table 1)	<i>q14.1A</i> and <i>q14.2A</i>	<i>q14.1B</i> and <i>q14.2A</i>	<i>q14.1B</i> and <i>q14.2B</i>

*: ECTS is short for the European Credit Transfer System: system of awarding credits for courses according to the criterion of students' workload per course (1ECTS equals 28 study hours).

** : This is a 3-credit course: 3 contact hours plus 6 to 9 self-study hours per week for 14 weeks (a semester), which makes 126 to 168 hours or 4.5 to 6 ECTS.

***: Response lectures are lectures in which students could ask specific questions or additional explanation on various topics.

Results and discussion

The results of the case studies are described and discussed on basis of the four sets of requirements as defined before (Table 1).

Requirements set 1: The influence of the context on the case studies

Requirements rE ‘the circumstances during the evaluation are taken into account’ and r12 ‘the exercises fulfil a need of most of the students in relation to their learning style’ are related to the influence of the context. To evaluate the influence of the context on the results of the case studies, questions related to the circumstances during the evaluation (qE.1 to qE.5) and about preferred learning activities (q12.1) were asked to the students.

The appreciation of the circumstances during the case studies

Table 3 gives the results of the answers to questions qE.1 to qE.5. These are part of the standard course evaluation questionnaire and were, therefore, not answered by the students in case study III. The responses to these questions indicate that the students in case study I liked the circumstances of the course better than the students in study II. In case study II the circumstances of the course, such as desk-space facilities and the course sequence, are viewed less favourably. It should be noted that the facilities were actually the same in both case studies. We have no satisfactory explanation for the difference. The lower appreciation of course sequence could negatively influence the response of the students to evaluation questions related to the digital exercises.

Table 3: Average response to the questions qE.1 to qE.5 in case studies I (n=28) and II (n=32-41).

Questions	Case study I		Case study II	
	% 4+5	Average response (n=28)	% 4+5	Average response (n=30-41*)
qE.1: In my opinion the digital exercises for calculations on reactions are valuable learning material.	74	4.0	64	3.7
qE.2: I am satisfied with the desk-space-facilities and computer rooms during the computer classes.	92	4.5	64	3.8
qE.3: I am satisfied with the use of Blackboard.	83	4.4	-	-
qE.4: I am satisfied with the course organisation: 1h lecture, 3h digital exercises, 1h response lecture.	-	-	48	3.6
qE.5: Overall rate for the course on a scale of 1 (poor) to 5 (excellent)	81	3.9	71	3.8

Answers are grouped in percentage of students who responded with a 4 or 5 (1=totally disagree, 2=partly disagree, 3=neutral, 4=partly agree, 5=totally agree) and the average response.

*: 30 students followed the course. The final examination was attended by these students and by some students who did a resit examination, and therefore more students filled in the standard course evaluation questionnaire

The learning activity preferences

To evaluate the learning activity preference of the students in the three studies, students were asked to choose up to three options, which correspond with the activities through which they prefer to learn, from a list of the seven main learning activities in the food chemistry course of studies I and II. Their responses give an indication whether the students in the studies will like the computer-based exercises. Desired learning activity preferences for these exercises are firstly

the activity ‘active answering questions’ and secondly the activity ‘looking at schemes/diagrams’. Those students who prefer other learning activities to these two are likely to favour computerised exercises less. It should be mentioned that learning activity preference can be influenced by experience (Vermetten et al., 1999); students who find particular activities successful may come to prefer them.

If the students pick at random three learning activities out of seven, then each activity would get picked by 43% of the students. * So, if an activity is chosen by more than 43% of the students, this activity is favoured by this group of students, and vice versa. To determine whether an activity is significantly in favour, the p-value is calculated for a few activities (binomial test, test proportion=0.43, SPSS 10.0). If $p < 0.05$ then the proportion of students that choose a learning activity is significantly higher than the test proportion.

Table 4: The learning activity preference of the students in the three case studies (question q12.1). Students could choose at most three activities. For each activity the percentage of students that has chosen that activity are given.

Activities: I am someone who learns through ...	Case study I (n=34)		Case study II (n=26**)		Case study III (n=16**)	
	%	p-value*	%	p-value*	%	p-value*
A: active answering questions	71	0.001	56	0.18	31	-
B: looking at schemes / diagrams / animations	50	0.26	30	-	56	0.21
C: (active) reading	56	0.09	59	0.09	50	0.37
D: listening (for example a lecture)	44	-	37	-	44	-
E: producing schemes / diagrams / animations	35	-	48	0.45	50	0.37
F: explanation of my questions by others (teacher/student)	32	-	26	-	25	-
G: group discussion / working in a group	18	-	26	-	25	-

*: Calculated with binomial test, test proportion=0.43.

** : Not all students had filled in this questionnaire.

It was noticed for all case studies that those students who learn by active answering questions also like to learn by looking at visuals, (active) reading and producing visuals (e.g. pictures, schemes) (data not given). Those who did not choose active answering questions like to listen (attending lectures) and like to learn by (active) reading. For this group of students also “explanation of my questions by others” is a frequently chosen option (data not shown).

Requirements set 2: The quality and user-friendliness of the design of the exercises

It is assumed that, in the perception of the students, the quality of the design of the learning material is related to characteristics such as usability, ability to motivate, clarity, manageability and ability to enjoy. Results from the questionnaire related to six requirements on how students perceive the design of the learning material are listed in Table 5.

According to the criteria for 5-point scale questions the four requirements r1, rQ, r2, and r5 are met, which means that the exercises are clear (r1), the exercises are manageable (r2), the exercises are fun to work on (r5), and the students perceive the quality of the learning material as

* Calculated with hypergeometric distribution (random selection, without repetition): $f(A) = \frac{\binom{1}{1}\binom{6}{2}}{\binom{7}{3}} = \frac{\left(\frac{1!}{1!*0!}\right)\left(\frac{6!}{2!*4!}\right)}{\left(\frac{7!}{3!*4!}\right)} = 0.43$

good (rQ). This gives an indication whether the design and user-friendliness of the exercises is of an adequate quality for the students in each of the three case studies.

Feedback or hints were provided to make the exercises manageable. The results (Table 5) from the questionnaire for the two questions that are related to hints (q2.1 and q2.2), show that the hints are applied correctly in the digital exercises for the students in case studies I and II. The lower valuation of the hints by the students in case study III is probably related to the difference in course content. The content of the lectures and lecture notes at Cornell University varies from that of the digital exercises in such a way that these students could need some more hints to be able to answer the questions than the students at Wageningen University or possibly some different hints, since their course content was different, but there are no data that could demonstrate that these students asked for more hints.

Table 5: Results for the questions related to the quality of the design for case study I, II and III.

r*	Question	Case study I		Case study II		Case study III	
		% 4/5	average	% 4/5	average	% 4/5	average
r1	q1.1: exercises are clear (you know what you have to do)	82	4.3	83	4.2	81	4.3
	q1.2: The formulation of the instructions of the digital exercises was clear and understandable.	90	4.3	84	4.2	-	-
rQ	qQ.1: The quality of the digital learning material was good.	95	4.4	84	4.2	-	-
r2	q2.1: hints are provided when needed	75	3.8	83	4.1	63	3.6
	q2.2: the hints are useful	85	4.0	83	4.0	69	3.7
r5	q5.1: exercises are fun	88	4.2	73	4.1	81	4.0
r10	q10.1: score tells about performance	85	4.0	50	3.4	63	3.6
	q10.2: score is motivation against guessing	44	2.9	53	3.3	38	2.8
r11	q11.1: low score is motivation to repeat exercise	79	3.9	50	3.3	67	3.4
	q11.2: repeat exercises until maximum score is achieved	38	2.5	28	2.1	47	2.8

Given are the average rating and the percentage of students that rated the question with a 4 or 5 (answers on a 5-point scale: 1=totally disagree, 2=partly disagree, 3=neutral, 4=partly agree, 5=totally agree).

*: r=requirement

The relation between the score and extrinsic motivation

Each exercise is designed with a score, as described before. This score is incorporated in the exercises to motivate students: to discourage guessing and to motivate students to repeat an exercise when getting a low score. The results in Table 5 show that in case study I, the students **did** recognise this score as a performance-grader (q10.1), while students in case studies II and III, did not really recognize the score as a performance-grader. The same counts for the question q11.1. In case study I, around 79% of the students agreed with the statement “*the score is a motivation to repeat an exercise when the score is low*”. For these students, the score could be recognised as a successful tool to motivate them to perform well, but in case studies II and III fewer students agreed with this statement.

Two factors may explain the lower value of the score in case study II as compared to the score in case study I. Firstly tables 3 and 4 show that in case study II the contextual variables such as desktop facilities and computer rooms respectively learning activity preferences have lower values. Secondly the knowledge of the students prior to doing the exercises was most

likely very much less in case study II because there were considerably fewer lectures. Due to a lower prior knowledge, students made more mistakes, resulting in a lower score. A low score is not an indication for these students that they performed poorly, but maybe it is an indication for these students that they did not have enough knowledge yet. Students in case study I maybe felt they should have enough knowledge to perform well and therefore they **did** see the score as a performance grader. Students in case study III worked on the exercises of their own free will and could, therefore, not care much about a score or performance.

From the answers on the questions q10.2. and q11.2. in all three studies it is clear that the score itself was not a good motivator for students not to guess too much or to work on the exercises until they did not make any mistakes at all (Table 5). It can be concluded that the motivating factor of the score was quite variable per student: some students were motivated by the score and some were not. Just incorporating a score is certainly insufficient to induce an extrinsic motivation for the students to perform well.

Requirements set 3: The extent to which the exercises assist in studying the reader

The reader for the course Food Chemistry at Wageningen University contains a range of relevant chemical concepts. Since students have difficulties to recollect the facts from the reader, the digital exercises were designed in order to stimulate students to work actively with the subject matter. In other words, the digital exercises were designed in such a way that working on the exercises helps students to study the reader. Requirements r6 ‘content of the exercises is understood’, r7 ‘students are able to recognise the exercises in sections of the reader’, r8 ‘sections of the reader are easy to remember after completing the exercises’ and r9 ‘sections of the reader are easy to understand after completing the exercises’ are related to this objective. In Table 6 the results from the questionnaire for the questions related to these four requirements are listed.

Table 6: Results for the questions related to the assistance in studying the reader for each case study.

r*	Question	Case study I		Case study II		Case study III	
		% 4 / 5	average	% 4 / 5	average	% 4 / 5	average
r6	q6.1: content is understood	75	4.0	75	4.1	88	4.3
r7	q7.1: exercises are recognised in reader	76	4.2	73	3.9	75	3.8
r8	q8.1: exercises make reader easy to remember	76	4.0	72	3.9	56	3.6
r9	q9.1: exercises make reader easy to understand	76	4.2	67	3.8	63	3.7

Given are the average rating and the percentage of students that rated the question with a 4 or 5 (answers on a 5-point scale: 1=totally disagree, 2=partly disagree, 3=neutral, 4=partly agree, 5= totally agree).

*: r=requirement

Because it is very likely that the students in case study II did have less relevant prior knowledge when they started on the exercises their learning task will have been larger: they came across more new chemical concepts during working on the digital exercises than the students in case study I. Apart from this the factors mentioned earlier such as less satisfactory desktop facilities in case study II and differences in learning activity preferences may also have been relevant.

Requirements set 4: The additional value of the digital exercises

The digital exercises are designed for the students as an addition to lectures and self-study activities (reading and studying the readers) in order to give them the possibility to work actively with the subject matter. Therefore, it is important to know whether students value the digital exercises as a useful addition. The requirements r3 'students see the digital exercises as a valuable addition', r4 'the exercises are used by students on their own initiative', r13 'students feel they learned much from doing the exercises' and r14 'students feel that the exercises contributed to the ability to successfully pass the final examination' are defined during the design process to ensure that the digital exercises are a valuable addition to the students.

The answers to question q3.1 confirmed that students saw the digital exercises as a valuable addition to the range of learning materials available. In the standard questionnaire students judged all parts of the course, i.e. the reader, the laboratory classes, the teachers and the computer classes. The judgement of the students for the digital exercises can, therefore, be compared with the judgement for other parts, of which the results for the reader, the laboratory part, and the teachers are also presented in Table 7. In this perspective, the judgement about the digital exercises was very high in case study I and high in case study II, indicating that students strongly valued the digital exercises.

Table 7: Average responses for the standard course evaluation questionnaire in case studies I (n=28) and II (n=30-41).

part of the course	Case study I		Case study II	
	% 4+5	Average response	% 4+5	Average response
question q3.1: In my opinion the digital exercises are valuable learning material.	91	4.6	84	4.1
judgement for reader		3.3		3.5
judgement for laboratory classes		3.8		3.7
judgement for teachers (on average)		3.8		3.7

Answers are grouped in percentage of students who responded with a 4 or 5 (1=disagree, 2=partly disagree, 3=neutral, 4=partly agree, 5=agree) and the average response.

From the answers on questions q4.1 and q4.2 for case studies I and II (Table 8), it seems fair to conclude that requirement r4 'students work on the exercises on their own initiative' was met, since 70% of the students worked on the exercises outside the computer classes and 80% of the students worked on some or all exercises more than once. Since these students worked on learning material on their own initiative (i.e. outside the computer classes), they must have seen additional value in the exercises.

The results from case study III show clearly that when the digital exercises were optional, students were less inclined to do them. Some other results strengthen this conclusion. From the twenty-five students that were enrolled in this course, five students did not use the digital exercises at all. From the twenty students who used the exercises, 25% did not enter the website for a second time. Furthermore, from the sixteen students who filled in the questionnaire 56% worked on all subjects, 13% worked only on the three main subjects (proteins, carbohydrates and lipids) and 31% worked only on the subject lipids. This preference for lipids was probably because the teacher asked them specifically to work on these lipid exercises. Although students in case study III clearly spent less time on the digital exercises, their opinion about the additional value of the digital exercises was rather positive, as will be discussed later on.

Table 8: Results for the questions q4.1 and q4.2 related to requirement r4 ‘students work on the exercises on their own initiative’ for case studies I, II and III.

question	possible answers	case study I	case study II	case study III
q4.1: Tell when you worked on the exercises.	During computer classes	100%	100%	-
	Outside classes	70%	70%	-
q4.2: Estimate how many times you worked on the exercises.	some exercises, once only	9%	10%	38%
	All exercises, once only	12%	10%	31%
	All at least once, some more often	47%	47%	13%
	All more than once	32%	33%	19%

The additional value of the digital exercises is also related to how much students learn from the exercises and whether students feel able to successfully pass the final exam with what they learned from the exercises. For the eight different parts of the course (1. lectures, 2. digital exercises, 3. other digital learning materials, 4. reader, 5. studying own lecture notes, 6. studying notes from digital exercises, 7. laboratory classes, 8. writing report of laboratory class experiments) students were asked to tell how much they learned from these parts (question q13.1). For each part they could choose between “very much”, “much”, “reasonable amount”, “little” and “nothing”. In case studies I and II from “doing the digital exercises”, 68% and 78% of the students learned much or very much, respectively (Table 9). For requirement r13 ‘students feel they learned much from doing the exercises’, the criterion is that 75% of the students should have learned much or very much, which means that r13 is not totally met in case study I, but it is in case study II.

Table 9: Results for the part digital exercises for the question q13.1 ‘How much did you learn from the digital exercises?’.

Question	Answer	Percentage of students (%)		
		Case study I	Case study II	Case study III
Q13.1: How much did you learn from the digital exercises?	Nothing or little	3	7	12
	Reasonable amount	29	15	38
	Much	41	52	25
	Very much	27	26	25

Although the data are not separately shown in a table, it is interesting to note that for the eight different parts of the course in case study I the part ‘reading the reader’ scored the highest (76% said to have learned much or very much), followed by the digital exercises (68%). In case study II this is the other way around: the digital exercises scored the highest (78%) and the reader scored lower (59%). This difference is remarkable, especially since the time spent on each part is comparable for the two case studies.

Again, the differences between case studies I and II can be explained by the fact that students in case study II had a lower prior knowledge, when starting on the exercises, than the students in case study I. In general students in case study II had to learn more during working on the digital exercises. Students in case study I already learned about various topics during the lectures, hence the information in the digital exercises was not totally new to them.

When comparing the results for question q13.1 of the first two case studies with the results of case study III (Table 9) it is obvious that the students from Cornell University learned less

Chemistry Education Research and Practice, 2005, 6 (2), 64-82

from the exercises than the students in case studies I and II, which was to be expected. The content of the food chemistry course at Cornell University is somewhat different from the content of some of the digital exercises, which means that not all exercises are applicable to the situation of these students. In addition, the students worked on the digital exercises of their own accord, which means that most students did not work rigorously on the exercises (see also the results in Table 8). Still, 50% of the students in case study 2 said to have learned “much” or “very much”, which is a reasonably good result, in spite of the differences.

Learning efficiency

In case study I, the students were asked to estimate how different parts of the food chemistry course contributed to the ability to pass the final exam (q14.1A) and to estimate how much time they spent on each part (q14.2A). For the first question, students graded for the eight parts of the course (Table 10) the percentage they felt that each part contribution to their capability to pass the final examination, taking into account that the total contribution for these eight parts should add up to 100%. Students also indicated how much time (hours) they spent on each part.

From the estimated contribution and estimated time for each part a learning efficiency can be calculated. Efficiency in general is defined as the ratio of the output to the input of any system. Learning efficiency could, therefore, be defined as the ratio of *the percentage contribution of a part to the ability to pass the final exam to the percentage time spent on that part*. The learning efficiency for each part was calculated for each student with the following formula:

$$\text{learning efficiency}_{(\text{part})} = \frac{\% \text{ contribution}}{\% \text{ time}} = \frac{\% \text{ contribution}_{(\text{part})}}{\frac{\text{time}_{(\text{part})}}{\text{time}_{(\text{total})}}}$$

In this formula:

contribution_(part) = contribution of a part to the capability to pass the final examination estimated by the student (%)
 time_(part) = time spent on a part estimated by the student (hours)
 time_(total) = total time spent for all parts as estimated by the student (hours)

When the learning efficiency is 1 the percentage time and the percentage contribution are equal. The higher the number the more efficiently the time is spent. The average estimated contributions and average of the total time spent on the eight parts, according to the students, are listed in Table 10, together with the average learning efficiencies. The learning efficiency is calculated for each student and then averaged. All averages have very large standard deviations.

The learning efficiency of a certain part can only be compared with other parts of the course if they have common educational goals. The first four parts in Table 10 are related to activities that are offered by the course. Of these four parts, the digital exercises and the readers have the same learning goals. The lectures are more intended for introduction and the quantitative exercises specifically deal with quantitative understanding of chemical reactions in food. The quantitative exercises are based on a problem based learning paradigm. A description of these exercises falls outside the scope of this article. The second two parts, lecture notes and notes from the exercises, are learning activities that students produce themselves. The last two parts are related to the laboratory classes. These have separate learning goals, mainly laboratory technique and report writing. The skills obtained during the laboratory classes are not examined during the

final examination and, therefore, the laboratory classes have a low learning efficiency with respect to the contribution to the final examination (Table 10).

Table 10: The average estimated contribution to the final examination, the average estimated percentage of time spent and average learning efficiency calculated per student for case study I (n=27).

Parts	Average % contribution	Average % time [#]	Average learning efficiency [#]	% students with learning efficiency ≥ 1
Attending lectures	11	14	0.8	19
Completing quantitative exercises	9	8	1.5	56
Doing the digital exercises	21	11	1.9	64
Reading the reader	30	25	1.4	75
Studying own lecture notes	5	2	4.7	100
Studying own notes from exercises	7	2	5.3	93
Doing the practical part in the laboratory*	9	28	0.4	10
Writing the report for the practical part in laboratory*	7	9	1.1	41

[#]: The average percentage time and average efficiency are both calculated by first calculating this number for each student and then averaging this over all students.

*: The laboratory classes have their own learning goals, which are not examined during the final examination. Separate grades are credited to the laboratory part.

For quite a lot of the students (64%), an efficiency rating of the exercises is higher than 1 (average is 1.9). This indicates that requirement r14 is met, since for that the average learning efficiency should at be 1 or more. Moreover, when comparing the learning efficiency for the exercises with reading the reader, the exercises score well. Therefore, with an average learning efficiency of 1.9 and an average contribution of 21%, it can be concluded that the digital exercises certainly have an additional value for the students.

In case studies II and III students were not asked to rate the contribution. Instead, students were asked to rate the statement “*The exercises contribute to my ability to answer questions from the final exam*” on a 5-point scale. Unfortunately, students were not asked to rate this statement for the other parts of the course. Results are shown in Table 11, which contains the results of case studies II and III for the questions related to requirement r14.

In case study III we were only interested in the efficiency of the exercises, since other parts of the course (e.g. lectures and readers) were outside our influence. Of the students that filled in the questionnaire 50% agreed that the exercises contributed to their ability to answer questions from the final exam (Table 11). So, although the content of the exercises is derived from the reader of the course at Wageningen University, still half the students at the corresponding course at Cornell University could answer questions of their own examination with what they learned in the digital exercises.

In case study III 75% of the students agreed with the statement “*The exercises are efficient for learning: I learned about Food Chemistry in a relative short period of time.*” (q14.2, Table 11). This result shows that although the content of courses are not equal, the students learned in an efficient way about Food Chemistry. So, for the exercises it seems that when content is not identical to the content of the course the usefulness for examination is not 100%, but they can still help students learn about food chemistry. This implies that the digital exercises can be a valuable aid for food chemistry courses outside Wageningen University.

Table 11: Results for the questions related to requirement r14. The questions differ in the different case studies. There were open questions, multiple-choice questions and questions on a 5-point scale.

Question / Statement	Answers	
	Case study II	Case study III
The exercises contribute to my ability to answer questions in the final exam.	3.5 (60%)*	3.3 (50%)*
Estimate the time spend on the digital exercises.	Average 15 hours	-
The exercises are efficient for learning: I learned about Food Chemistry in a relative short period of time.	-	3.9 (75% agreed)*

*: Average of answers on a 5-point scale (1=totally disagree, 2=partly disagree, 3=neutral, 4=partly agree, 5=totally agree), with between brackets the percentage of students that chose 4 or 5 as an answer.

Case studies I and III show that the exercises were rather efficient for learning according to the students, but this efficiency as perceived by the students is not a good measure of the ability to pass the final examination successfully. Based on the case study results it will not be possible to provide evidence for the effectiveness requirements.

Conclusions

Despite the differences between the two case studies at Wageningen University and the case study at Cornell University it is clear that the digital exercises are useful to varying degrees in at least three courses with different structure and organisation. Based on the results of the evaluation process carried out in the three case studies it is concluded that most design requirements, that were described during the design process, are met. Therefore, we will continue to apply the design process for food chemistry digital learning material as described in Diederer et al., (2003). The design process and the evaluation process together are an example of a design oriented research approach in chemistry education. It has been shown in this paper that this research approach can result in useful digital learning material and in a satisfactory and clearly defined starting point for improvements of education. We expect that this research approach can also be used for development and investigation of other digital learning materials.

The evaluation results show that students' appreciation of the same learning material is different when applied in a different course structure and/or organisation. Therefore, we have carefully described the situation of the case studies, and this is regarded as crucial for an accurate interpretation of the results. From the interpretation of the results, the following preliminary conclusions are described for further research on the effect of the on students' appreciation.

First, when a course consists of several learning activities, the sequential order of these activities influences what students learn from each activity. This is explained by the following. The digital exercises seem to assist the students in studying the reader when the students have already gained some knowledge about food chemistry by attending the lectures. In contrast, students seem to learn more from the digital exercises *without* first attending several lectures. It is reasonable to assume that the knowledge of students prior to starting with the digital exercises will be influenced by previous exposure to relevant information in lectures. Also students who have less relevant prior knowledge invest more time in completing the exercises and thus are likely to learn more from working with the exercises. A second conclusion is about the usefulness of presenting a performance score to the student who does the exercises. This relates the prior knowledge of the student on the topic of the exercise to the value, which a student

attributes to the presentation of his performance score in that exercise. It states that low prior knowledge results in a low value attributed by the student to the function that presents a performance score to the student. And third, it is concluded that when the type of activities that are incorporated within the learning material matches with the learning activity preference, the student is more likely to use this learning material and vice versa. This is derived from the differences in use of the digital learning material between the case studies, which are, amongst others, explained by the idea that students will use approaches they like.

Finally the results of case study III are in keeping with the quite general belief that using digital learning material as basis for a self-study activity, which is not explicitly scheduled, will result in a low use of this digital material, even when the digital material is regarded useful.

Acknowledgements: We wish to thank the students and faculty from the Department of Food Chemistry of the Cornell Institute of Food Science, Cornell University for their participation, and Professor Dennis D. Miller in particular for his cooperation in this study.

References

- Diederer J., Gruppen H., Hartog R., Moerland G. and Voragen A.G.J., (2003), Design of activating digital learning material for food chemistry education, *Chemistry Education: Research and Practice*, **4**, 353-371.
- Keller J. M., (1983), *Development and use of the ARCS model of motivational design*, Twente Univ. of Technology, Enschede, The Netherlands.
- Keyser M. W., (2000), Active learning and cooperative learning: understanding the difference and using both styles effectively, *Research Strategies*, **17**, 35-44.
- Kirschner P. A., (2002), Cognitive load theory: implications of cognitive load theory on the design of learning, *Learning and Instruction*, **12**, 1-10.
- Pawson R. and Tilley N., (1997), *Realistic evaluation*, SAGE Publications, London.
- Podsakoff P.M., MacKenzie S.B., Lee J. and Podsakoff N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies, *Journal of Applied Psychology*, **88**, 879-903.
- Sweller J., Merriënboer J.J.G. and, Paas F.G.W.C., (1998), Cognitive architecture and instructional design, *Educational Psychology Review*, **10**, 251-296
- Vermetten Y.J., Lodewijks H.G. and, Vermunt, J.D. (1999), Consistency and variability of learning strategies in different university courses, *Higher Education*, **37**, 1-21.
- Vermetten Y.J., Vermunt J.D. and Lodewijks H.G. (2002), Powerful learning environments? How university students differ in their response to instructional measures, *Learning and Instruction*, **12**, 263-284.