Design of Virtual Experiment Environments

Sjors Verstege

Propositions

1. All laboratory classes should at least partially be replaced by virtual experiment environments (this thesis).
2. Teachers should use design principles instead of intuition when designing new learning materials (this thesis).
3. All curricula should include a course on learning psychology.
4. The answer options 'agree' and 'totally agree' in questionnaires should be renamed to avoid differences in interpretation.
5. Scientists are not teachers.
6. The COVID-19 pandemic was a blessing for laboratory education.

Propositions belonging to the thesis entitled:
Design of Virtual Experiment Environments

Sjors Verstege,
Wageningen, 3 November 2021

Design of Virtual Experiment Environments

Sjors Verstege

Thesis committee

Promotor:
Prof. Dr J.P. Vincken
Professor of Food Chemistry
Wageningen University & Research

Co-promotor:
Dr J. Diederen
Lecturer, Food Chemistry
Wageningen University & Research

Other members:
Prof. Dr M.K. Seery, University of Edinburgh, Ireland
Dr A. Bakker, Utrecht University
Prof. Dr P.J. den Brok, Wageningen University & Research
Dr M.C. Busstra, Wageningen University & Research

This research was conducted under the auspices of the Graduate School VLAG
(Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences)

Design of Virtual Experiment Environments

Sjors Verstege

Thesis
submitted in fulfilment of the requirements for the degree of doctor at
Wageningen University & Research
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 3 November 2021
at 4 p.m. in the Aula.

Sjors Verstege
Design of virtual experiment environments,
173 pages

PhD thesis, Wageningen University & Research, Wageningen, The Netherlands (2021)
With references, with summary in English

ISBN: 978-94-6395-821-9
DOI: https://doi.org/10.18174/547094

Abstract

Many of the learning outcomes that are attributed to laboratory classes can be achieved outside the laboratory, for example by student preparation. The aim of the research described in this thesis was to design virtual experiment environments (VEEs) that help students prepare for laboratory classes. Given the limited information on the design of such learning materials, the central question of this dissertation was: How to design a VEE that helps students to prepare for laboratory classes? To answer this question, I adopted the educational design research (EDR) approach. This approach is committed to develop theoretical understanding and practical solutions simultaneously, and involves three phases: i) analysis & exploration, ii) design & construction, and iii) summative evaluation. In the analysis & exploration phase, the complex educational problem and its context were defined, based on which the potential of VEEs as preparation tools for laboratory classes was established. In the design & construction phase, solutions for the complex educational problem were explored, which finally resulted in the design of four VEEs. Based on recurrent testing and formative evaluations of the VEEs, a blueprint was established that consists of three design requirements, fourteen design principles, and a design architecture. Such information in the context of VEEs is new to literature, and has the potential to help teachers and instructional designers to design high quality VEEs. In the summative evaluation phase, the designed VEEs were implemented and evaluated in the learning situation. Based on the results of the summative evaluation, the blueprint that resulted from the design & construction phase was evaluated and fine-tuned. In conclusion, the complex educational problem was solved by the practical solutions: the VEEs. The theoretical understanding acquired in the EDR process, which was captured in the blueprint, which is the answer to the central research question.

Table of contents

Chapter 1

General introduction

Contents

Welcome to my dissertation! Inside, you will find a lot of information about the design of virtual experiment environments (VEEs). Before I go into details about VEEs, let me take you back to the beginning.

1.1 Replacing laboratory classes?

Several years ago, I set off to answer the trigger question: To what degree can laboratory classes be replaced by other learning activities? This is a bit of a tricky question for three reasons. First, laboratory classes contribute to teaching students how to ‘do’ science (Seery et al., 2019). Second, many teachers enjoy being personally involved in the teaching of, and are mentally attached to ‘their’ laboratory class. Third, students generally enjoy laboratory classes. I hear you thinking: Why mess around with something that seems essential for science education, and also seems to make everyone happy?

> Before we continue, I would like to draw your attention to the definitions of laboratory education, laboratory classes, and laboratory work. *Laboratory education* is an umbrella term that encompasses i) pre-lab: student preparation for laboratory classes, ii) the laboratory classes (including the laboratory work), and iii) post-lab: all activities that may happen after completion of the laboratory work, such as processing raw data into results, results interpretation, and communication of the results (e.g. by giving a presentation or by writing a report). *Laboratory classes* include everything that happens in the physical laboratory (such as teacher explanations, laboratory work, and feedback). *Laboratory work* refers to the actual execution of protocols.

In 1924, a study was done to investigate the effect of supplementary laboratory work on exam results. The conclusion was that the retention of information was better after supplementary laboratory work than after solely studying from a textbook (Bowers, 1924). In the eighty years since, “researchers have not comprehensively examined the effects of laboratory instruction on student learning and growth in contrast to other modes of instruction, and there is insufficient data to confirm or reject convincingly many of the statements that have been made about the importance and effects of laboratory teaching” (Hofstein and Lunetta, 2004). And even today, it remains surprisingly difficult to find evidence that the time, effort, and costs invested in laboratory classes are in balance with students’ achievement of intended learning outcomes (ILOs) during laboratory education (Bretz, 2019). So, when considering to (partially) replace laboratory classes, we need to evaluate their efficiency in terms of whether students achieve the ILOs.

The relation between learning outcomes and laboratory education has been subject of several studies (e.g. Abdulwahed and Nagy (2009), Agustian (2020), and Kirschner and

Meester (1988)). In most of these studies, the (context dependent) learning outcomes are not presented. Instead, there is a large body of literature on general goals and learning domains (Agustian, 2020; Kirschner and Meester, 1988; Reid and Shah, 2007; Tamir, 1976). Examples of such general goals are: "to use laboratory skills in performing (simple) experiments", and "to interpret experimental data". Over the past years, I learned that many teachers in our university combine multiple of such general goals to define one generic learning outcome for the entire laboratory education. For example: "Design and conduct experiments and to analyse and interpret the results of these experiments". Even though this generic learning outcome is in place, I noticed that teachers do have an implicit list of ILOs in their mind. I argue that it is essential to the success of laboratory education, for both students and teachers, to make those ILOs explicit. Once those ILOs are explicit, it is possible to check whether they are aligned with the activities students actually do during the laboratory classes.

So, to provide a theoretical answer to the trigger question, the following two questions must be answered for any specific course: i) What are the ILOs of the laboratory class, and ii) which of those ILOs could be achieved outside the laboratory? As further elaborated on in chapter 2, teachers almost always select a combination of ILOs that can only be achieved in the laboratory (e.g. obtain hands-on experience), and ILOs that can be achieved outside the laboratory (e.g. interpret experimental data). In other words, to completely replace a laboratory class is generally not doable as long as we do not dispute the reported ILOs. We could, and perhaps should, place some question marks as to whether each laboratory class needs to focus on obtaining hands-on experience. Seery et al. (2019) present a curriculum model for laboratory learning, consisting of five stages. Each stage has a different focus, where only the first stage focuses on developing experimental skills, competence, and laboratory procedures (i.e., obtaining hands-on experience). So, in the early stages of laboratory learning, the focus should be on obtaining hands-on experience, and over time this focus shifts to other ILOs such as experimental design.

In conclusion, some of the ILOs that are generally attributed to laboratory classes can only be achieved in the laboratory, while others can also (and perhaps better) be achieved outside the laboratory. When it comes to *replacing* ILOs that can be achieved outside the laboratory by other learning materials, literature is not outspoken. Instead, there is a general consensus that student *preparation* prior to laboratory classes will increase the quality of knowledge construction during the laboratory classes (Bannert, 2002; Johnstone, 1997; Rutten et al., 2012).

1.2 Non-traditional labs to prepare students for laboratory classes

A large variety of learning materials such as presentations, pre-laboratory questions, and quizzes have been designed in order to help students prepare for laboratory classes (Gregory and Di Trapani, 2012). The last decade, the focus lies on non-traditional labs (NTLs), which can for example be defined as computer simulations, virtual labs, or remote labs (Alkhaldi et al., 2016; Brinson, 2015; Budai and Kuczmann, 2018; Faulconer and Gruss, 2018; Hartog et al., 2009; Heradio et al., 2016; Ma and Nickerson, 2006; Potkonjak et al., 2016; Rutten et al., 2012). The definitions of such NTLs vary among authors. Since literature does not present standard terminology for NTLs (Faulconer and Gruss, 2018), I chose to use the more general term: virtual experiment environment (VEE) in this dissertation. A VEE is defined as "Any educational resource that enables students to design and/or to carry out virtual experiments and/or to process data, and to analyse and interpret results" (Hartog et al., 2009, p. 376).

What stands out in publications about NTLs, is that they hardly provide any details on *how* the NTLs were designed. Instead, NTLs are often used as the context of another study. For example, many studies focus on the effect of a NTL on student learning: Did students learn (more), and can this learning be related to a construct such as self-regulated learning (Brydges et al., 2015)? As an educational designer, I was missing the link between established guidelines for designing instruction and the design of the NTLs. This resulted in my central research question: How to design a VEE that helps students to prepare for laboratory classes?

1.3 Educational design research approach

On my quest to find an answer to the central research question, I adopted the educational design research approach. Educational design research can be defined as "a genre of research in which the iterative development of solutions to practical and complex educational problems also provides the context for empirical investigation, which yields theoretical understanding that can inform the work of others" (McKenney and Reeves, 2019, p. 6). What sets this type of research apart from many other types of research, is that it is committed to develop theoretical understanding and practical solutions simultaneously (Bakker, 2018; McKenney and Reeves, 2019; Plomp, 2013). The educational design research approach typically consists of three phases: i) analysis & exploration, ii) design & construction, and iii) summative evaluation. These phases are described in detail in chapter 3, and are represented by the different chapters throughout this dissertation. Figure 1.1 provides a schematic representation of the educational design research approach. In the following sections, which are categorized based on the three phases of educational design research, I will shortly introduce each chapter.

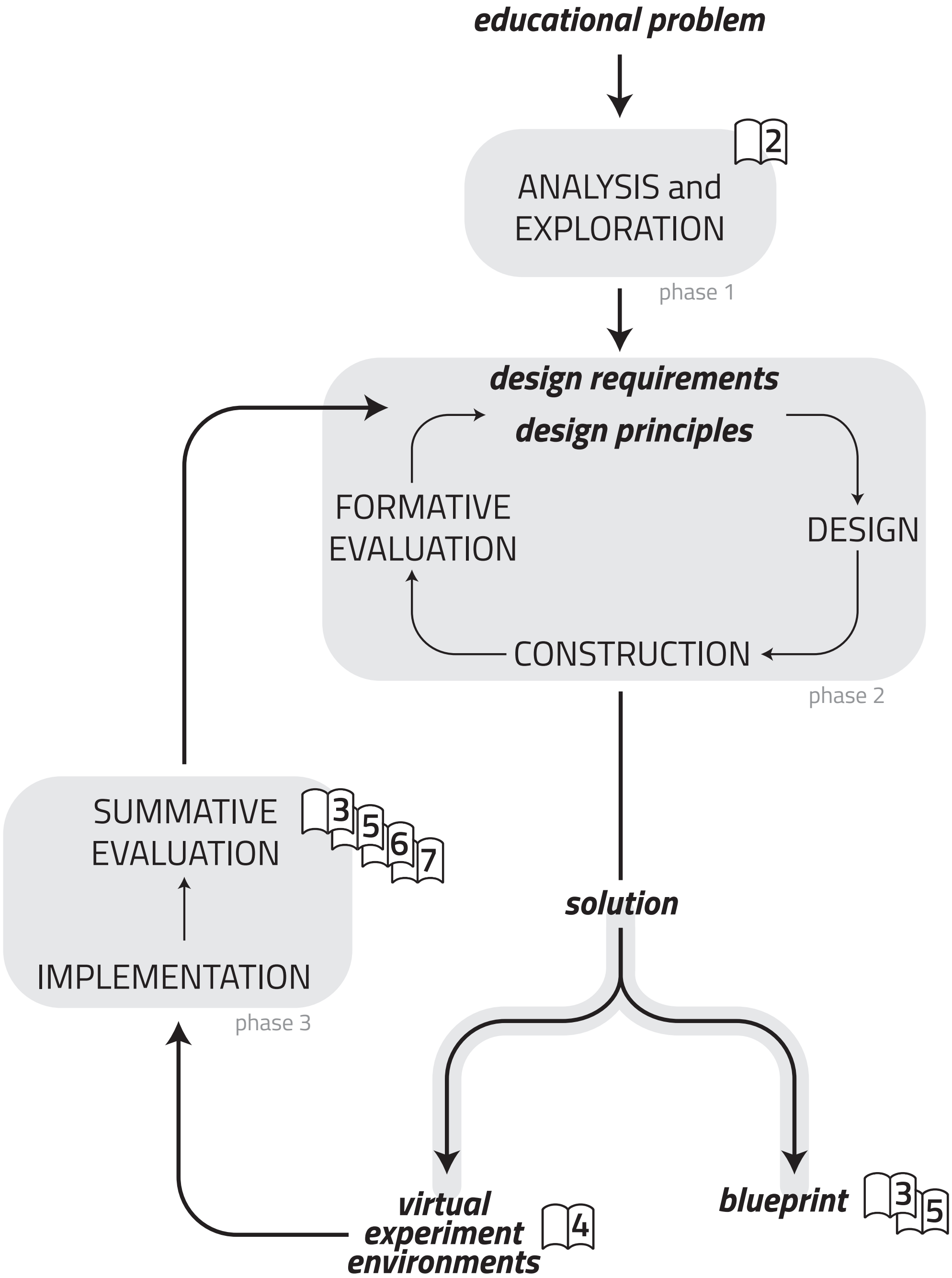

Figure 1.1: Schematic representation of the educational design research approach, with corresponding chapters in this dissertation.

1.3.1 Analysis & exploration phase

The analysis & exploration phase is represented by **chapter 2**. In this chapter, the complex educational problem and its context were defined. Stakeholders were identified, and provided with a questionnaire to get their opinion about the current state of laboratory education. Several explorative follow-up interviews were conducted to get a more complete overview. Based on the results of the questionnaire, the interviews, and literature research, it was concluded that there is potential for laboratory simulations (which will later be referred to as virtual experiment environments) as preparation tools for laboratory classes, and an initial list of design requirements was identified, which was used as the starting point for the design & construction phase.

1.3.2 Design & construction phase

The design & construction phase is represented by chapters 3, 4, and 5. In **chapter 3**, a blueprint to design virtual experiment environments is provided. Following an extensive description and overview of the educational design research approach, solutions for the complex educational problem (chapter 2) were explored, which finally resulted in the design of four virtual experiment environments (VEEs). Based on recurrent testing and formative evaluations of the VEEs, a blueprint was established that consists of three design requirements, fourteen design principles, and a design architecture. Such information in the context of VEEs is new to literature, and has the potential to help teachers and instructional designers to design high quality VEEs.

In **chapter 4**, a showcase of a VEE is provided. The user interface, the available interactions, and the feedback- and support mechanism are presented and discussed. This contextual information is limited in many publications (including my own, in chapters 6 and 7) that use an existing environment to study a construct such as self-regulated learning. The showcase, when combined with the blueprint (chapter 3), helps to place the studies presented in this thesis in the right context, and gives a complete overview that may as such provide guidance and/or inspiration to teachers and educational designers.

In **chapter 5**, an extension of the initial design (chapters 3 and 4) is introduced and evaluated. The extended design is focused on students' understanding of protocol steps, by enriching existing protocols with theoretical, practical, troubleshooting, and/or calculation questions, to form interactive protocols.

1.3.3 Summative evaluation phase

The summative evaluation phase is represented by chapters 3, 5, 6, and 7. In chapters 3 and 5, the designs were mostly evaluated based on questionnaires, so that means based on students' self-reported data. The disadvantage of self-reported data is that the assumption must be made that students are capable of reflecting on their own behaviour, are willing to answer the questions honestly, and that all students have the same interpretation of the questions. To obtain more objective information on student behaviour in the VEEs, data logging was introduced in chapters 6 and 7.

In **chapter 6**, a VEE was used as the context for exploring the relation between students' perceived level of self-regulated learning (SRL), their achieved learning outcomes in the VEE, and their logged behaviour in the VEE.

In **chapter 7**, sequential pattern mining was used to study how students use support while doing calculations in a VEE. The use of support was related to students' learning gain and prior knowledge, with the goal to evaluate the support mechanism that is part of the blueprint (chapter 3).

In **chapter 8**, I discuss my findings from the perspectives of the university teacher, the educational designer, and the educational scientist.

Chapter 2

Exploring the potential of simulations as preparation tools for university laboratory classes

This chapter was published as:

S. Verstege, J. Van der Kolk, J. Diederen, R. J. Hartog, and H. Gruppen (2016). "Exploring the potential of simulations as preparation tools for university laboratory classes". In: *EDULEARN16 Conference*. 8th International Conference on Education and New Learning Technologies. IATED. DOI: https://doi.org/10.21125/edulearn.2016.1429

Contents

Abstract

Laboratory classes play an important role within university level curricula in natural sciences. Although widely used, there is no evidence that the time, effort and costs invested in laboratory education is in general balanced by its contribution to intended learning outcomes. In the study described in this paper, teachers (*N* = 144) from 13 universities, including 3 universities of applied sciences, in The Netherlands have participated in a survey on laboratory education. According to 53% of the teachers, students could learn more from laboratory education than they currently do. This was, among others, attributed to poor preparation by the students. Although 88% of the teachers claimed that student preparation for laboratory education is important, only 24% of all the teachers indicated that their students are adequately prepared before entering the laboratory class room. A quarter of the respondents selected the use of laboratory simulations as a suitable approach to address this problem. We concluded that there is a potential for laboratory simulations as preparation tools. Based on literature, survey results, and interviews, an initial list of design requirements for laboratory simulations was formulated. This list can be used as a starting point for the evolutionary development of laboratory simulations.

2.1 Introduction

At university level, laboratory education plays an important role in most natural sciences curricula. Since long there are serious doubts as to what students actually learn from laboratory classes (Abdulwahed and Nagy, 2009; Hawkes, 2004; Hawkins and Phelps, 2013; Kirschner and Huisman, 1998; Kirschner and Meester, 1988). In general, costs of laboratory classes in terms of facilities, equipment and efforts of staff and students are considerable, but it is surprisingly difficult to find evidence for resulting benefits in terms of achieved learning outcomes. During laboratory education, students' cognitive abilities are challenged since they have to process a lot of information. In combination with insufficient knowledge of the background theory of the experiments they are about to perform, this might lead to cognitive overload (Johnstone, 1997; Limniou and Whitehead, 2010). The cognitive load theory (CLT) is "a theory stating that human limited working-memory capacity has far-reaching implications for teaching and learning. Well-designed training systems prevent cognitive overload, decrease cognitive load that is not relevant to learning, and optimize cognitive load relevant to learning (Van Merriënboer and Kirschner, 2012)." If students obtain so much information at a specific time that the working memory is overloaded, the problem solving ability is hampered (Sweller, 1994). From CLT it follows that well prepared students or students with sufficient laboratory experience have a lower cognitive load compared to students who are poorly prepared and have little experience in a laboratory setting (Bannert, 2002; Winberg and Berg, 2007). In practice, to cope with the cognitive demanding situation, students tend to blindly follow instructions, relying on working strategies that help them to reach the 'finish' of the laboratory class (Johnstone, 1997). As a result, learning tends to be limited to hands-on experience. Careful preparation that reduces cognitive load that is not relevant for learning and helps to prevent cognitive overload during laboratory class, requires effort from both teachers and students. Teachers should prepare the pre-laboratory work as carefully as the laboratory manual itself (Johnstone and Al-Shuaili, 2001), and students should be motivated to prepare for laboratory classes.

Literature reports several activating pre-laboratory activities, such as writing synopses of experiments to be carried out, pre-laboratory discussions, designing experiments, and carrying out pre-laboratory simulation exercises (Limniou et al., 2007; Lyle and Robinson, 2002; Rollnick et al., 2001; Van der Kolk et al., 2013; Winberg and Berg, 2007). Writing synopses of the experiments, which are to be carried out in a pre-laboratory session helped students to meet the intended objectives of the laboratory, and supported the students in their practical work (Lyle and Robinson, 2002). Pre-laboratory discussions have been used to clarify misconceptions, as well as the writing of synopses. The pre-laboratory discussions were especially useful for the well prepared students. The writing of synopses resulted in increased interaction with the manuals

(Rollnick et al., 2001). Also, a web-based experiment designer to support students in designing a research strategy for laboratory education has been used. This experiment designer was highly valued by students and teachers and was used intensively (Van der Kolk et al., 2013). Simulation exercises have been developed and used to do complex equilibrium calculations in a pre-laboratory setting. The simulation was found to benefit all students (Winberg and Berg, 2007). Furthermore, a viscosity simulator was integrated in a pre-laboratory session. After this pre-laboratory session, students were able to perform the experiment in the laboratory without further guidance, and students had an improved understanding of the theory that belongs to the experiment (Limniou et al., 2007). Simulations can be very powerful tools for students to learn how a complex or dynamic system behaves (Hartog et al., 2009; Pirker and Gütl, 2015). Such simulations, along with the majority of simulations that can be found on the internet, only target one specific experiment or theory. The focus of the present paper is on the design of e-learning in the form of virtual experiment environments (VEEs) as preparation tools for laboratory classes. The additional value of a VEE over the before mentioned simulations is that it can comprise the complete process of research, which means that it is not limited to simulating one specific part of it. Hartog et al. (2009, p. 376) define a VEE as "any educational resource that enables students to design and/or to carry out virtual experiments and/or to process data, and to analyse and interpret results." In the context of this paper the term 'laboratory simulation' is used interchangeably with 'VEE'. In preparation for laboratory education, students can use a VEE to study the background theory involved in the experiment, obtain more confidence in performing the required calculations, get a clear view on the apparatus to be used and how the experiment is performed, and understand why they are doing the experiment in a wider context (Kirstein and Nordmeier, 2007; Lambourne, 2007; Mercer-Chalmers et al., 2004). The challenge is to design, develop and implement VEEs that motivate and support students to prepare for specific laboratory classes.

The aim of the research described in this paper was twofold. First, the potential of VEEs as preparation tools for university laboratory classes was explored, in particular insofar it is related to the opinions, attitudes, and experiences of teachers with respect to laboratory simulations and eventually VEEs. Second, since designing a VEE requires a clear definition of the requirements of such learning material, an initial list of design requirements for a VEE were developed. These two aims, the exploration and the description of the design requirements, form the start of the design oriented research for the development of successful VEEs in the future.

2.2 Elicitation of goals and requirements

2.2.1 Survey on laboratory education

To explore the potential of VEEs as preparation tools for university laboratory classes, a survey was distributed amongst faculty and coordinators of laboratory education at universities and universities of applied sciences in The Netherlands (N = 13). All teachers of whom public contact information was available were approached (N = 737), which led to 144 responses.

This survey had three foci: A) The learning goals of laboratory education; B) Student preparation for laboratory education; C) E-learning in relation to laboratory education.

To determine what teachers find the most important learning goals of their laboratory education (focus A), general learning goals, defined by Kirschner and Meester (1988) were given, after which teachers selected which ones were most important to their laboratory education. This leads to the following list, including the proportion (%) of teachers who selected these goals (Table 2.1):

Table 2.1: Survey results related to the learning goals of laboratory education (focus A).

#	Learning goal	Selected by
S1	The student obtains hands-on experience in the laboratory	92%
S2	The student learns to interpret data	88%
S3	The student learns to plan experiments	67%
S4	The student learns to clearly describe the experiment	58%
S5	The student remembers theory for a longer period of time	51%
S6	The student learns to use knowledge and skills in unfamiliar situations	51%
S7	The student learns to solve problems	47%
S8	The student learns to formulate hypotheses	43%

In line with foci A, B, and C, four statements were given, which teachers rated on a scale of 1 to 5. Teachers were also asked to explain their answers in follow-up questions. As shown in Table 2.2, 53% of all teachers indicated that their students could learn more from laboratory education than they currently do. From a follow-up question related to this statement, it became clear that this is mostly related to lack of time, large groups, and student attitude. 88% of the teachers reported that student preparation is important, whereas 24% of the teachers reported that their students enter the laboratory well prepared. This was attributed to the lack of need for students to prepare.

Table 2.2: Survey results (1 = disagree, 2 = somewhat disagree, 3 = agree nor disagree, 4 = somewhat agree, 5 = agree).

#	Focus	Statement	Answers (%) 1	2	3	4	5	% 4+5
S9	A	Students could learn more from my laboratory education	3	16	28	39	14	53
S10	B	Student preparation for laboratory education is important	2	2	8	29	59	88
S11	B	My students are well prepared when they enter the laboratory	4	26	46	21	3	24
S12	C	I am enthusiastic about the concept of a laboratory simulation which helps students to prepare for laboratory classes	8	9	42	32	9	41

Focus C was on e-learning in relation to laboratory education. Teachers were given 6 different types of e-learning materials, and asked to indicate which e-learning materials they found most suitable for their laboratory education. The results are shown in Tables 2.2 (S12) and 2.3. Overall, 27% and 41% of the teachers indicate that they wish to use a laboratory simulation (S18), and that they are enthusiastic about the concept of a laboratory simulation which helps students to prepare for laboratory classes (S12), respectively. As can be seen from S12, 42% of the teachers responded neutral. In a follow-up question, many teachers explained that since they did not know what to expect, they answered neutral.

Table 2.3: Survey results related to type of e-learning (Focus C).

#	Type of e-learning	Selected by
S13	E-learning that forces students to prepare	70%
S14	E-learning that reduces the work-load for teachers	36%
S15	E-learning for the design of experiments	32%
S16	Electronic lab journal	30%
S17	E-learning for registration in the lab	27%
S18	E-learning in the form of a laboratory simulation	27%

2.2.2 Explorative follow-up interviews

In the survey, teachers could indicate whether they were willing to participate in a follow up interview. As a result, in 18 explorative follow-up interviews, the potential of laboratory simulations as preparation tools was discussed in a semi-structured manner. The aim of these interviews was to gather possible design requirements for a VEE. In random order, the following points have been suggested by the teachers:

I1 The simulation should lead to time-saving for both students and teachers during the laboratory class.

I2 The simulation should incorporate all aspects of doing research, except for the 'wet' part.

I3 The simulation should stimulate students to make predictions and to critically reflect the obtained results.

I4 The simulation should contain a toolbox, from which students can select methods/tools, enabling them to design their own experiments.

I5 The simulation should include all theory behind the experiments.

I6 The simulation should allow for peer-reviewing of e.g. the experimental design.

I7 The simulation should provide accurate results based on variable inputs, initiated by student's choices, so that every student could have a unique experience.

I8 The simulation should be accessible at any time of day.

I9 The simulation should stimulate students to make carefully considered choices.

I10 The simulation should provide data and statistics for teachers.

I11 The simulation should be maintainable by teachers.

I12 The simulation should provide feedback based on student input.

I13 The simulation should be tailor-made to the specific laboratory course.

2.3 Design requirements for a virtual experiment environment

Based on literature, survey results, and interviews, an initial list of design requirements for a VEE was formed. Table 2.4 shows the specified design requirements, examples of how to evaluate whether the requirements are satisfied, and indicates on which results from the survey and interviews the design requirements are based. The following design requirements can be used to develop a VEE. After implementation of a VEE in university level education, evaluation should point out if the design requirements are met and to what extent the set of design requirements should be adjusted. When asking student opinions on whether a specific design requirement is satisfied, survey questions could be distributed, which have a five point rating scale (1=disagree, 5=agree).

Table 2.4: The initial design requirements and evaluation criteria for a VEE

#	Design requirement (the VEE should)	Examples of how to evaluate whether the requirement is satisfied	Based on
1	Motivate students to prepare for laboratory class	i) Ask whether students feel motivated to prepare for laboratory education by means of the VEE ii) Monitor how many students worked in the VEE and for how long	S5, S6
2	Prepare students for laboratory class A) In an effective way B) In an efficient way	i) Ask whether students think that they are better prepared for laboratory class due to their preparation with the VEE ii) Ask whether students feel adequately prepared for laboratory class due to their preparation with the VEE iii) Ask whether teachers experience that their students are well prepared for laboratory class iv) Ask whether students experience the VEE as efficient way of preparation	S5, S6
3	Enable students to design and plan their own experiment by means of a toolbox, and therefore provide A) A database of experiments B) Theory behind the experiments	i) Ask whether students were able to design and plan their own experiments using the experiment database ii) Ask whether students found the available theory behind the experiments adequate iii) Monitor the student use of available experiments, and monitor the use of available theory	S3, S10, I4, I5
4	Require students to make predictions, to interpret simulated data and to reflect on results critically	i) Ask whether students were required to make predictions, to interpret simulated data, and to reflect on results critically ii) Monitor how students reply to in-simulation questions regarding predictions, data interpretation, and reflection on results	S2, I3
5	Reduce unnecessary cognitive load to a minimum	Ask expert reviewers to evaluate the preparation tool with respect to this requirement	*
6	Require students to make carefully considered choices	Ask students whether they were required to make carefully considered choices	I9
7	Provide useful data and statistics for teachers	Monitor if teachers used data and statistics from the VEE	I10
8	Be maintainable by teachers	Ask whether teachers were able to maintain the VEE	I11
9	Provide feedback based on student input	Ask whether students obtained sufficient feedback based on their input	I12
10	Be tailor-made	Ask whether teachers were satisfied with this aspect of the VEE	I13
11	Imply an acceptable balance of costs and benefits	Let all stakeholders review and judge the costs and benefits	**

* Clark and Mayer, 2016
** Hartog et al., 2012

2.4 Discussion and conclusions

The aims of this research were: to explore the potential of VEEs as preparation tools for university laboratory classes and to come up with an initial list of design requirements for such a VEE. Regarding the potential of VEEs: the fact that students are often not well-prepared to enter the laboratory class seems to be a widely recognised problem (results S10 and S11). E-learning could be a means to address this problem (S13). A quarter of the respondents selected VEEs as a suitable approach. Based on these results we conclude that there is a potential for VEEs as preparation tools. It must be taken into account, that we left it to the respondents to have their own interpretation of the term 'laboratory simulations'. This was by design because of the explorative nature of this research. It might well be that the number of respondents who selected VEEs as a suitable approach reflects the percentage of early adopters and that the real potential might be larger. Regarding the initial list of design requirements for a VEE: we were able to formulate a useful list. This list can be used as a starting point for an evolutionary development of a VEE (Dybå and Dingsøyr, 2008).

One of the design requirements (#11) states that the VEE should imply an acceptable balance of costs and benefits. The costs related to the development of a dedicated VEE are spread over a smaller number of students when compared to a generic VEE, which means dedicated VEEs have a higher costs-per-student. The higher costs-per-student for development of a dedicated VEE can possibly be lowered by a) making only a part of the VEE dedicated; b) building and implementation of an environment which allows a low-cost conversion of generic VEE parts to dedicated VEE parts; and c) by sharing building blocks which are being used in different dedicated VEEs. The expected benefits are a) students could achieve learning goals that would not be achieved without preparation using the VEE; b) laboratory classes become more attractive for both students and teachers; and c) The integral costs-per-student, including the costs of laboratory classes, which might even (partially) be replaced, can probably be significantly lowered when students enter the laboratory well-prepared.

Chapter 3

Blueprint to design virtual experiment environments

This chapter was published as:

S. Verstege, J.-P. Vincken, and J. Diederen (2021). "Blueprint to design virtual experiment environments". *Computers and Education Open* 2. DOI: https://doi.org/10.1016/j.caeo.2021.100039

Contents

Abstract

Laboratory education plays an indispensable role in university level natural science curricula. To achieve all intended learning outcomes of laboratory education, students should be carefully prepared before they enter the laboratory. This can be done by means of non-traditional labs (NTLs), and more specifically, virtual experiment environments (VEEs). Although VEEs are increasingly being used in higher education, there is a lack of studies that provide actionable guidelines on how to design such VEEs. The goal of this study was to provide general design requirements, design principles, and a design architecture that can be used as a blueprint to design VEEs. To achieve our goal, we used an educational design research approach to design four VEEs, which have been evaluated in real educational contexts. Besides establishing three general design requirements, fourteen corresponding design principles, and a design architecture based on the insights gathered during the educational design research process, we also included detailed information on how they can be applied. Such information in the context of VEEs is new to literature, and has the potential to help teachers and instructional designers to design high quality VEEs.

3.1 Introduction

3.1.1 Non-traditional labs

Laboratory education plays an indispensable role in university level natural science curricula. Kirschner and Meester (1988) defined eight student-centered general learning outcomes for university level laboratory education. Three of those general learning outcomes: obtaining hands-on experience, using knowledge and skills in unfamiliar situations, and solving problems, can only fully be achieved in a laboratory class. The other five general learning outcomes: planning experiments, interpreting data, describing experiments, remembering theory for a longer period of time, and formulating hypotheses, are often considered as intended learning outcomes of laboratory education (chapter 2), although they may well be achieved outside the laboratory (Carnduff and Reid, 2003; Reid and Shah, 2007). To help students achieve those five learning outcomes, and to help students prepare for the three learning outcomes in the laboratory, part of the laboratory education should take place outside the laboratory in the form of a targeted preparation assignment (Bannert, 2002; Winberg and Berg, 2007). Targeted preparation can be done using non-traditional labs (Rutten et al., 2012).

A large variety of non-traditional labs (NTLs) such as computer simulations, virtual laboratories, and remote laboratories have been introduced, reported, and reviewed in literature (Alkhaldi et al., 2016; Brinson, 2015; Budai and Kuczmann, 2018; Faulconer and Gruss, 2018; Hartog et al., 2009; Heradio et al., 2016; Ma and Nickerson, 2006; Potkonjak et al., 2016; Rutten et al., 2012). Reported advantages of NTLs are: i) they can simulate dangerous, time-consuming, and expensive experiments, ii) they are not restricted by factors such as limited space and equipment, which makes them suitable for distance learning, and iii) they can promote learning from mistakes, without causing a significant increase in time, effort or costs. Based on these advantages, we believe that NTLs are not only suitable for stand-alone assignments, but also as assignments to prepare students for doing experiments in the laboratory.

NTLs are increasingly being used in higher education (Faulconer and Gruss, 2018), which calls for studies that provide actionable information on how to design them. Studies that focus on the design of NTLs are often categorised as design-based research, or, more specifically, educational design research. This kind of research requires collaboration between researchers and practitioners (Anderson and Shattuck, 2012; Dunn et al., 2019; Koivisto et al., 2018). Despite the increasing number of initiatives, such collaborations are difficult to establish, as it often turns out to be challenging for researchers and practitioners to share a common understanding of where a project is headed (McKenney and Reeves, 2019).

When engaging in educational design research, we distinguished three relevant categories of publications. The first category contains literature that provides detailed information on how to set up educational design research (e.g. Bakker (2018), Chen and Reeves (2019), McKenney and Reeves (2019), and Plomp (2013)). The second category provides guidelines for designing instruction. These guidelines are based on learning sciences, which developed an understanding of how technology can be used to foster learning (e.g. Clark and Mayer (2016), Mayer (2009a), and Van Merriënboer and Kirschner (2018)). The third category describes case studies based on an educational design research approach (e.g. Brahm (2017), Euler and Collenberg (2018), Koivisto et al. (2018), Lambert and Jacobsen (2019), and Thomas et al. (2019)). Such studies generally combine literature of the first two categories to make informed design choices.

Despite the value of the guidelines in the second category, we were unable to find case studies that provide detailed information on the application of such guidelines to design a NTL. Most case studies in the field of educational design research (the third category) provide limited practical information regarding the design of innovations (know how), which, when it would be combined with learning theories (know why), would provide valuable information for the design of future innovations (Bereiter, 2014; Kidron and Kali, 2017; Zheng, 2015).

3.1.2 Educational design research approach

Educational design research can be defined as "a genre of research in which the iterative development of solutions to practical and complex educational problems also provides the context for empirical investigation, which yields theoretical understanding that can inform the work of others" (McKenney and Reeves, 2019, p. 6). This type of research can be subdivided into several phases (see e.g. Bakker (2018), Branch (2009), Cronje (2013), and McKenney and Reeves (2019)). Although the exact naming and division of the phases varies slightly according to different authors, their content is similar, and we will present them as i) analysis & exploration, ii) design & construction, and iii) summative evaluation.

The analysis & exploration phase involves i) defining the complex educational problem, ii) defining the context of this problem, and iii) identifying and meeting involved stakeholders (McKenney and Reeves, 2019). This will result in a design goal, a set of contextual design constraints, and a list of design requirements. A design constraint is "a limit or restriction on the features or behaviours of the design. A proposed design is unacceptable if these limits are violated" (Dym et al., 2014, p. 7). Design requirements define the goal of a design process, and make explicit how one can observe or measure whether the design goal has been reached (Hartog et al., 2013). The design requirements can be used as a starting point for design & construction.

The design & construction phase includes i) the exploration of solutions, such as generating, considering, and checking ideas (McKenney and Reeves, 2019), ii) the mapping of solutions, in form of design principles (Van den Akker, 2013), and iii) constructing prototypes. Design principles are heuristic statements that guide the designer towards a solution for the complex educational problem (i.e. satisfying the design requirements), and can be formulated as follows: "If you want to design intervention X [for purpose/function Y in context Z], you are advised to give the intervention the following characteristics", followed by procedures, theoretical arguments and/or empirical arguments (Plomp, 2013, p. 24). Design principles provide the information needed to start constructing prototypes of the intervention (McKenney and Reeves, 2019). Recurrent testing and formative evaluations of the prototypes will contribute to a knowledge base from which the initial design principles can be complemented. When the design & construction phase has been completed, a design architecture can be extracted from the developed intervention. A design architecture is a top-level structure and a set of related overall characteristics (Hartog et al., 2013).

The summative evaluation phase proceeds upon implementation of the intervention in the learning situation. Based on the results of the summative evaluation, the design principles that were defined in the design & construction phase may be fine-tuned. Furthermore, the results will point out whether the design requirements are satisfied by the intervention.

To summarize, the educational design research approach as presented in this paper has three phases and six corresponding output classes (Table 3.1). When applying this approach to design a NTL, the output classes are either contextual or general. Contextual output classes (design goal, design constraints, and evaluation results) are specific to one learning situation, whereas the general output classes (design requirements, design principles, and design architecture) are applicable to any learning situation. Hence, the general output classes have the potential to help teachers and instructional designers to design high quality NTLs.

Table 3.1: Phases and output classes of educational design research and corresponding sections in which the output classes are discussed.

Phase	Output class	Relation to the design of a NTL	Section
Analysis & exploration	Design goal	Contextual	3.2
	Design constraints	Contextual	3.2.1
	Design requirements	General	3.2.2
Design & construction	Design principles	General	3.2.2 + 3.2.4
	Design architecture	General	3.2.3
Summative evaluation	Evaluation results	Contextual	3.3

As there is no consensus on the terminology to define categories within NTLs (Faulconer and Gruss, 2018), we categorize the NTLs that are subject of this paper as virtual experiment environments (VEEs). A VEE is defined as "Any educational resource that enables students to design and/or to carry out virtual experiments and/or to process data, and to analyse and interpret results" (Hartog et al., 2009, p. 376).

The goal of this study was to provide general design requirements, design principles, and a design architecture that teachers and instructional designers can use as a blueprint to design VEEs. The following research questions were formulated:

1. Which design requirements are relevant for VEEs?
2. Which design principles contribute to achieving the design requirements?
3. What is a suitable design architecture of VEEs?

To answer the research questions, we went through multiple cycles of the educational design research approach. In the remainder of this paper, we describe the findings of this cyclical process in a linear fashion. As indicated in Table 3.1, sections 3.2.2 through 3.2.4 provide the answers to the research questions. In section 3.3 we discuss the results of summative evaluations, which were used i) to evaluate the design requirements that were initially formulated in the analysis & exploration phase, ii) to finetune the design principles that were formulated in the design & construction phase, and iii) to visualize the corresponding design architecture.

3.2 Design of virtual experiment environments

This paper presents the results of the educational design research approach that yielded four VEEs for two different courses, which were all based on the same design goal: to design VEEs that facilitate targeted preparation for university level laboratory classes. In our context (on which we will elaborate in section 3.3.1), targeted preparation means that students: i) know the goal of the experiments and understand the principle of the methods involved, ii) know what the raw data resulting of those experiments look like, iii) are able to process the raw data into results, and iv) are able to interpret the calculated results.

3.2.1 Design constraints

Five contextual design constraints were identified in the analysis & exploration phase. First, each VEE had to be aligned with the pre-defined intended learning outcomes of the target course. Second, the amount of time in which students had to be able to complete a VEE was determined by the available time slot in the target course. Third, students had to work individually, to ensure that all students were equally well prepared, and because of initial differences in prior knowledge. Fourth, the VEEs had to be suitable for distance education. Fifth, teachers without knowledge of programming should be able to maintain and edit the VEEs.

3.2.2 Design requirements and design principles

Three design requirements (DRs) and fourteen corresponding design principles (DPs) have resulted from multiple cycles of the educational design research approach (Table 3.2). Initially, as a result of the analysis & exploration phase, we started off with eleven design requirements as introduced in chapter 2. During the educational design research processes of the four VEEs, renewed insights led to the translation of the initial design requirements into three design requirements and fourteen corresponding design principles.

Table 3.2: Design requirements (DRs) and corresponding design principles (DPs) for the design of VEEs that facilitate targeted preparation for university level laboratory classes, including codes.

Code	Design requirement	Code	Design principle
DR1	The VEE should create a positive learning experience	DP1	Provide motivational elements
		DP2	Break a continuous lesson into segments
		DP3	Use multimedia rather than text alone
		DP4	Leave out or hide information that is not required to successfully complete the learning task
DR2	The VEE should support students in achieving the intended learning outcomes	DP5	Provide pre-training in the key concepts prior to the main learning activity
		DP6	Let students design the experiment themselves
		DP7	Let students study relevant methods
		DP8	Provide students with raw data
		DP9	Let students process raw data into results
		DP10	Let students interpret the results
DR3	The VEE should enable students to complete the assignment independently	DP11	Provide procedural information
		DP12	Provide formative feedback based on student input
		DP13	Provide access to hints (one by one) that guide the students' thinking process
		DP14	Provide the opportunity for students to check their intermediate calculations

Before we discuss the theoretical and empirical arguments that support the design principles, and provide detailed information on how we applied the design principles (section 3.2.4), we will first show the design architecture of the VEEs, to give an overview.

3.2.3 Design architecture

Within the four VEEs that were all based on the same design principles (Table 3.2), we recognized a common design architecture (Figure 3.1). The design architecture was visualized in terms of a pre-experiment phase, an experiment phase, and a post-experiment phase, including five corresponding levels and the feedback and support system. Within the design architecture it is indicated which design principles (DPs) were applied to the specific parts.

In the pre-experiment phase, we discriminate three levels: the context level, the research question level, and the experimental design level. The context level is actually a preparation for the VEE, and was therefore included as 'level 0'. It may consist of questions or exercises that target to activate prior knowledge, to construct new knowledge, and introduce students to the topic of the VEE. This is especially relevant when students do not have the required prior knowledge to start working in the VEE. At the research question level, students obtain pre-defined research questions. These research questions represent the actual assignment that the student will be working on in the VEE. The first step in the VEE is to complete partially defined hypotheses. At the experimental design level, the student makes an experimental design in order to test these hypotheses. To do so, a database of methods (e.g. isolation, modification, and analytical methods) is available, which students connect to each other to form a flowchart. The methods contain questions on the protocol and on the principles behind each method. These questions have to be answered in order for the student to progress. During this process, students receive feedback on their experimental design, and feedback on the answers they have given.

As the VEEs were intended as preparation tools for laboratory classes (i.e. students will do the experiments in the laboratory classes), we chose to focus on the pre- and post-experiment phase. The experiment phase was only included to form the link between the pre- and post-experiment phase, as it provides students with raw data based on the experimental design they made.

The post-experiment phase consists of two levels: the data processing level, and the results interpretation level. At the data processing level, students process the raw data they obtained in the VEE into results (mainly by doing calculations), and enter their processed results in the VEE. During data processing, students can choose to access hints, and to let the VEE check their intermediate calculations. Finally they will receive feedback on the entered processed results. At the results interpretation level, students have to interpret and integrate the results they obtained to be able to evaluate the hypotheses they completed in the pre-experiment phase. In this level, answer-specific feedback will be provided.

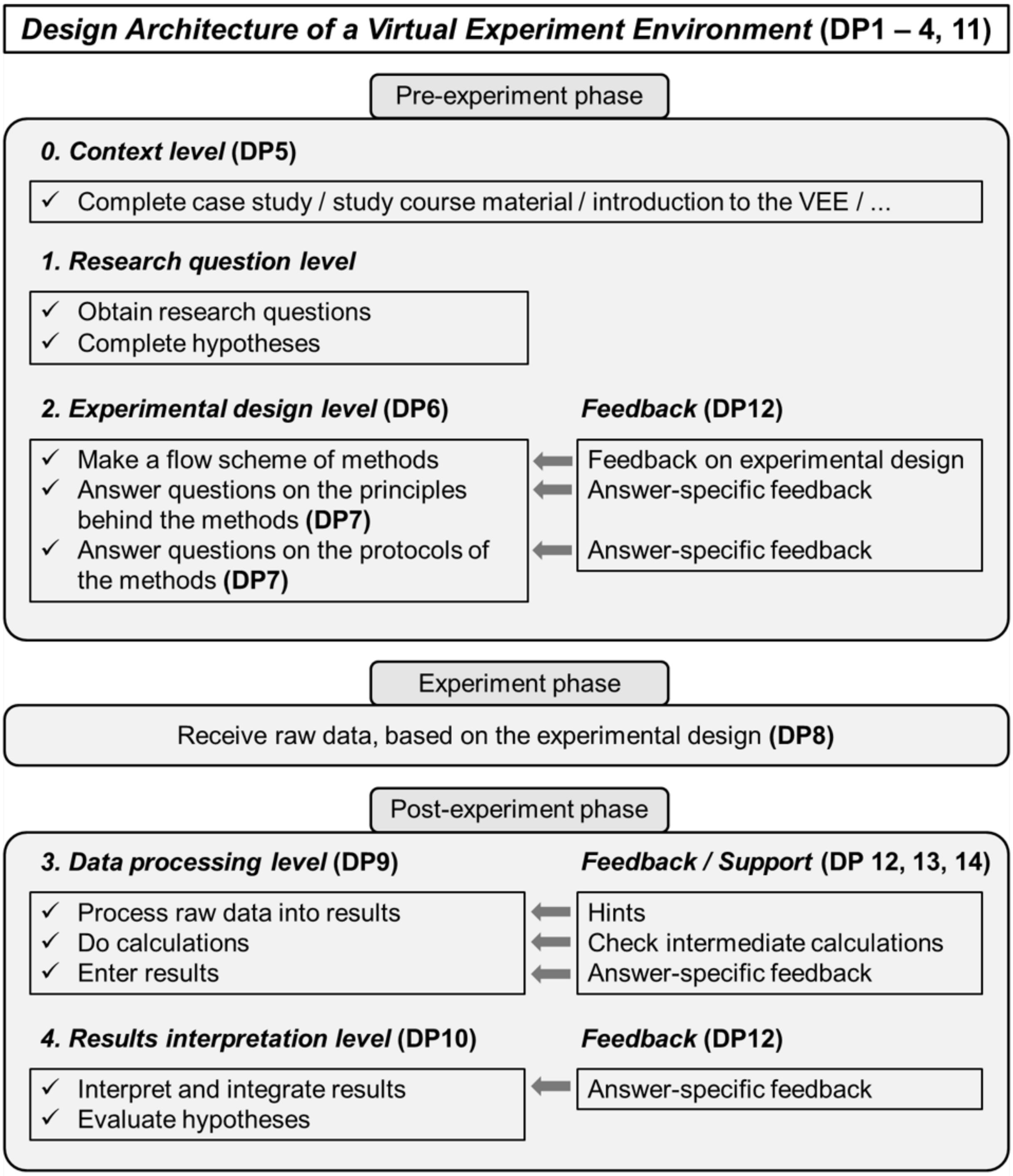

Figure 3.1: Design architecture of a virtual experiment environment (VEE), including visualization of the feedback and support system, and relation to design principles (DPs).

3.2.4 Application of the design principles

In this section we discuss the theoretical and empirical arguments that support the design principles that we used to satisfy the design requirements (both introduced in section 3.2.2), and we explain how those design principles were applied. For clarification of some explanations, we refer to Figure 3.2, which is located in section 3.2.5.

Design requirement 1: Creating a positive learning experience

Creating a positive learning experience (DR1) is a general design requirement, and is as such applicable to all kinds of learning materials. In order to satisfy DR1, we selected four design principles, which are generally applicable throughout the VEEs.

The first design principle (DP1) advises to provide motivational elements. Motivational elements can improve student learning by fostering cognitive processing aimed at making sense of the material (Mayer, 2014). The ARCS model of motivation identifies four motivational components (attention, relevance, confidence, and satisfaction) that should be considered in order to get and keep students motivated (Keller, 1983). Motivational elements based on the ARCS model result from i) the user-interface of the platform in which the VEEs were built (section 3.2.5), ii) the use of different types of exercises in the VEE, iii) gradually building up the difficulty as students progress in the VEE, iv) working on a relevant topic, and v) communicating the intended learning outcomes (Diederen et al., 2003).

The second design principle (DP2) advises to break a continuous lesson into segments. Key in the segmenting principle is that the learner should be able to control the time that can be spent in each segment (Clark and Mayer, 2016). Empirical research on the segmenting principle is primarily focussed on videos and narrated animations (Mayer and Pilegard, 2014), and concluded that segmenting a lesson helps students to manage their learning processes, and will lead to a better performance at transfer tests when compared to a control group (Mayer, 2009a). It was found that segmenting is especially useful when the learner is unfamiliar with the content (Ayres, 2006; Clark and Mayer, 2016). We introduced segmentation by dividing the VEE into a pre-experiment phase, an experiment phase, and a post-experiment phase, and subsequently dividing those phases into five levels, as was shown in the design architecture (section 3.2.3). Within one specific method, students gradually progress through the levels, and will not be granted access to the next level before they have completed the current level. For example, the experimental design level and the data processing level are combined in one screen (Figure 3.2e, in section 3.2.5) that consists of several tabs: 1. Background information, 2. Protocol, 3. Results, and 4. Calculations. Each tab can be opened at any time, but only one tab can be open at any moment. If a student did not reach the level of a selected tab yet (e.g. a student wants to do calculations (data processing level) before making the flowchart to obtain the required raw data (experimental design level)), it will contain feedback that directs the student to his/her current level.

The third design principle (DP3) advises to incorporate multimedia rather than text alone. Multimedia refers to the combination of text and (dynamic) graphics. Learning materials that use multimedia, on the condition that the graphics are not solely intended for decoration, keep students' attention for a longer time (Keller, 1987), and were found to promote active and deep learning (Clark and Mayer, 2016). The effect extends to transfer tests. Students who received a multimedia lesson, scored higher on transfer tests than students who received the same lesson in words alone (Mayer, 2009a). We applied this principle by introducing interpretive graphics to explain the principles behind the methods that are provided in the database. Interpretive graphics are visuals that make intangible phenomena visible and concrete (Clark and Mayer, 2016). To support the protocols of the methods, representational and transformational graphics were used. The representational graphics visualize the apparatus and materials used in the lab, and transformational graphics visualize complex protocols, by means of short clips. Making an experimental design in the form of a flowchart (Figure 3.2) naturally involves multimedia. The flowchart helps students to create an overview of how all methods are interconnected.

The fourth design principle (DP4) advises to leave out or hide information that is not required to successfully complete the learning task. "Learning is improved when interesting but irrelevant words and pictures are excluded" (Mayer, 2009a, p. 89). Many words and graphics in the VEEs have been dedicated to elaboration of the principles behind methods. Part of this information is considered to be either prior knowledge, or irrelevant to complete the learning task (and achieve the intended learning outcomes). The coherence principle states that adding extra material can hurt learning, because when learners use their limited processing capacity on extraneous material, less capacity is available for making sense of the essential content (Clark and Mayer, 2016). Following the coherence principle, elaboration should be left out. However, since the VEEs are targeted for university level education, we deem it important that students have easy access to potentially relevant and extensive background information. Therefore we made this information accessible through links labelled 'Click here for more information'.

Design requirement 2: Achieving the intended learning outcomes

Supporting students in achieving the intended learning outcomes (DR2) starts with communicating the learning outcomes to students, so that they know what they are supposed to learn from the VEE. This also points out the relevance of the learning material. Furthermore, the VEE should consist of activities (e.g. doing calculations, making choices) that support students in achieving the intended learning outcomes. In order to satisfy DR2, we selected the following six VEE-related design principles.

The fifth design principle (DP5) advises to provide pre-training in the concepts required to understand the context of the learning material prior to the main learning activity. Pre-training in the key concepts prior to exposing students to the main learning materials has a positive effect on learning in a variety of multimedia learning contexts (Clark and Mayer, 2016; Mayer, 2009a; Mayer and Pilegard, 2014). The approach we took varied for the four VEEs presented in this paper. For more contextual information about the VEEs, we refer to section 3.3. In VEE 1, students were given a preparative assignment consisting of 17 exercises (different types of closed questions) that helped students become familiar with the context of the VEE. In this preparative assignment, students cannot progress to the next question before they have correctly answered the current question. Feedback is provided to all possible answers. In VEEs 2, 3, and 4, which were designed for a different course than VEE 1, students were given a preparative assignment to become familiar with the available methods.

The sixth and seventh design principle (DP6 and DP7) advise to let students design the experiment themselves, and to let students study relevant methods. These design principles are inherent to the pre-experiment phase of the VEE. We applied those principles by giving students a database of methods that can be added to the workflow canvas (Figure 3.2b), so a flowchart can be made. To study the methods, each method contains the tabs: 1. Background information, and 2. Protocol (Figure 3.2e), in which information is given and interactive questions have to be answered. The interactive questions stimulate the student to actively look for and process the available information.

The eighth design principle (DP8) advises to provide students with raw data. This design principle is inherent to the experiment phase of the VEE. Providing students with raw data presented in the same way as in the laboratory classes, will help students to become familiar with and recognize the output that corresponds to a method. We did this by providing links to e.g. Microsoft Excel files containing raw data. The ninth and tenth design principle (DP9 and DP10) advise to let students process raw data into results, and to interpret the results. These design principles are inherent to the post-experiment phase of the VEE. Processing the raw data into results often involves doing calculations in Microsoft Excel, after which the results of those calculations have to be entered into the VEE. The feedback and support system will be discussed in the following section. Ultimately, students have to integrate and interpret the results, and accept or reject their hypotheses.

Design requirement 3: Enable students to complete the assignment independently

To enable students to complete the assignment in a VEE independently (DR3), they should be provided with support, such as feedback and hints that would otherwise be provided by a teacher. In order to satisfy DR3, we selected design principles 11, 12, 13, and 14, which together form the feedback and support system that was developed for the VEEs.

The eleventh design principle (DP11) advises to provide procedural information. To provide procedural information, "the system acts as an assistant looking over the learner's shoulder and presents procedural information precisely when a learner needs it" (Van Merriënboer and Kirschner, 2018, p. 29). We used a system that generates blinking red circles, which guide students i) towards locations in the VEEs where questions still need to be answered, or ii) towards locations where feedback is given on how they can progress within the VEE. Since the system is aware of students' actions and their progress, it provides just-in-time information that guides students through the VEEs step-by-step.

The twelfth design principle (DP12) advises to provide formative feedback based on student input. "Providing feedback is an ongoing process in which teachers communicate information to students that helps them better understand what they are to learn, what high-quality performance looks like, and what changes are necessary to improve their learning" (Dean et al., 2012, p. 3), and is among the most powerful influences on learning (Hattie, 2009; Hattie and Timperley, 2007; Neelen and Kirschner, 2020; Van der Kleij et al., 2015). A subcategory of feedback is formative feedback, which is defined as "information communicated to the learner that is intended to modify his or her thinking or behaviour to improve learning" (Shute, 2008, p. 153), and is argued to be the most effective form of giving feedback (Hattie, 2009). We incorporated formative feedback in three ways. First, a virtual professor was introduced, who can be consulted at any time. For each research question that the students have to answer, the virtual professor indicates whether the experimental design is complete to be able to answer that specific research question. Second, feedback is provided to each answer option in the questions that students have to answer about the methods. This feedback will tell the learner whether the given answer is correct, including an explanation and/or hint. Third, in calculation questions where multiple answers have to be given, feedback in the form of a green checkmark or a red cross next to the given answer will indicate which answers are correct or incorrect, respectively (e.g. Figure 3.2f). We made sure that any feedback was positioned on the screen so that learners can see the question, their answer to the question, and the feedback at the same time, to minimize split attention (Clark and Mayer, 2016).

The thirteenth design principle (DP13) advises to provide access to hints (one by one) that guide the students' thinking process. This is different from providing formative feedback (DP12), since students can choose to access the hints at any time. In other words, students do not need to make a mistake or fill in random values to trigger support. In the case of a multi-step calculation, a link containing a hint was added. Each hint ends with another link to the next hint (if available). This way, students can choose to use part of the support without the complete procedure being given away in one click.

The fourteenth design principle (DP14), which is only applicable in case of multi-step calculations, advises to provide students with the opportunity to check their intermediate calculations. This will encourage students to locate and improve upon mistakes they might make while doing calculations, without needing the help of a teacher. We did this by adding a link that leads to a table in which one or more intermediate values can be entered. For each intermediate value the student enters, we introduced a green checkmark or a red cross that appears next to the given answer (Figure 3.2f). In case students cannot find the correct value by themselves (i.e. they are stuck), they can choose to open a link with the correct value. This link will only appear when students entered an incorrect value for an intermediate calculation. We deliberately chose to provide correct intermediate values only after an incorrect intermediate value was entered by the student, to discourage gaming the system behaviour (chapter 6). In VEE 1, in the situation that students correctly entered all intermediate calculations belonging to a question, but could still not come up with the final answer, they were given the possibility to access a pdf file with the complete calculation. This pdf file was also provided upon successfully completing any calculation.

3.2.5 The VEEs

The VEEs were built in a platform called LabBuddy (Kryt b.v., Wageningen, the Netherlands). The LabBuddy platform was chosen because i) it is based on educational design research (Van der Kolk, 2013), ii) it is a web-based environment (design constraint 4), iii) knowledge of programming is not required to build, maintain, or edit a VEE (design constraint 5), iv) the platform allows for instant editing, and v) it has been used for more than a decade satisfactorily by our teachers and students. LabBuddy consists of an experiment designer and a lab manual both dedicated to educational purposes (Van der Kolk, 2013). The system was extended with simulation functionalities in the process of designing the VEEs. The degree of immersion in the LabBuddy VEEs is low, as users engage in a 2D environment. Figure 3.2 shows the interface of the LabBuddy VEEs. Students start the assignment at the virtual professor (Figure 3.2a), where they obtain research questions and have to complete partially defined hypotheses. The VEE initially contains an empty workflow canvas (Figure 3.2b shows a completed experimental design). The student has to design a suitable workflow by dragging methods onto it from the methods dashboard (Figure 3.2c and d) and making connections. When clicked on a method, its content pane will open and display four tabs, which can be expanded one at the time: 1. Background information, 2. Protocol, 3. Results, and 4. Calculations (Figure 3.2e). Once all questions in the background information tab and protocol tab of a method are correctly answered, students can progress to the results tab to receive raw data, which can subsequently be processed into results in the calculations tab (Figure 3.2f). Finally, students return to the virtual professor in order to interpret the results. For a more detailed visualization of the VEEs, a presentation of all possible user interactions, and a detailed explanation on the feedback and support system, we refer to chapter 4.

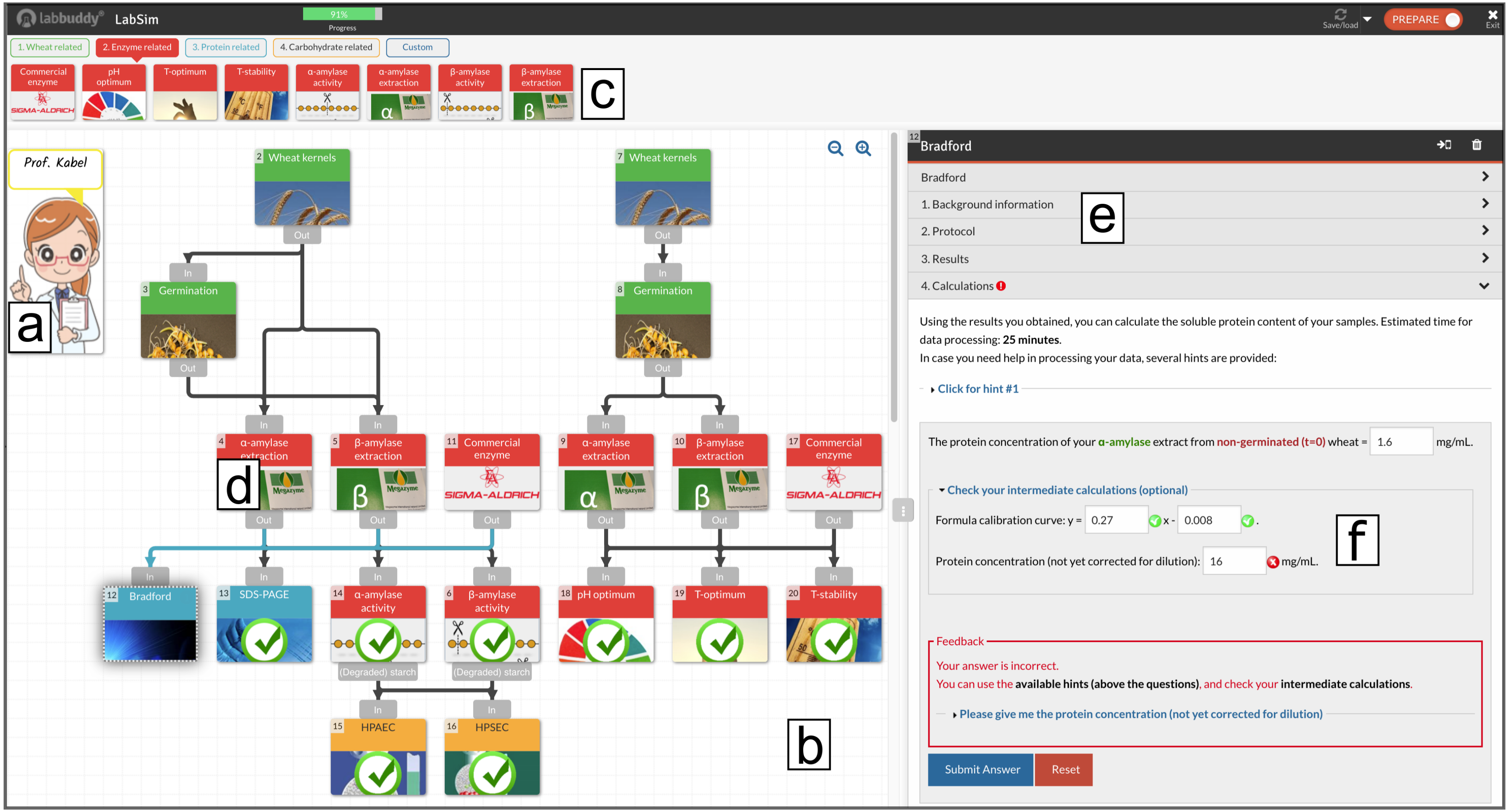

Figure 3.2: LabBuddy VEE screenshot of a nearly completed assignment: a) Virtual professor. b) Workflow canvas. c) Methods dashboard. d) Method in flowchart. e) Content of a method. f) Question in the calculations tab of a method, including feedback on the given answer.

3.3 Evaluation of virtual experiment environments

The design & construction phase resulted in four VEEs that were implemented in their target courses. Each VEE has been evaluated in terms of the design requirements (section 3.2.2).

First, we want to provide an overview of differences between the VEEs in terms of their characteristics (Table 3.3). These characteristics give insight in the similarities and differences between the four VEEs, and help to interpret the evaluation results. For example, the size and complexity of VEE 1 is much higher when compared to VEEs 2, 3, and 4, as it provides many more data points (in the form of raw data, obtained in the '3. Results' section of a method), requires students to give more answers, provides more hints, and has a much higher number of checkable intermediate values. Furthermore, in VEE 3, students could not check intermediate values, as it contained only one-step calculations, and VEE 4 provides double the number of hints compared to VEEs 2 and 3, which reflects that students have to do more complicated calculations.

Table 3.3: Characteristics of the evaluated VEEs.

	VEE 1	VEE 2	VEE 3	VEE 4
General topic	Enzymology	Carbohydrates	Lipids	Proteins
Average time needed to complete (hours)	7.5	1.5	1.5	2
Number of available chemical methods	14	10	7	10
Number of data points provided	310	57	40	88
Number of answers to be given	240	71	53	61
Number of available hints	31	10	12	22
Number of checkable intermediate values	100	12	0	5

3.3.1 Participants

Participant information of all evaluations is shown in Table 3.4.

VEE 1 was evaluated as part of the master level course Enzymology for Food and Biorefinery (168 study hours). We were provided with the unforeseen opportunity to evaluate VEE 1 in the situation where the whole course was taught online. Therefore, we included an extra evaluation: VEE 1 online. All students had previously obtained a bachelor degree in natural sciences: 49% from the Netherlands, and 51% from a university outside the Netherlands. All students were enrolled in a master study programme, being Food technology (80% of the students in VEE 1, and 64% of the students in VEE 1 online), Biotechnology (19% and 28% of the students), or other (1% and 8% of the students). Each master study programme is subdivided in several specializations. The prior knowledge of students was diverse because of the aforementioned differences in prior education, master study programme, and specialization.

Table 3.4: Participant information of all VEE evaluations.

	VEE 1	VEE 1 online	VEE 2	VEE 3	VEE 4
Course level	master	master	bachelor	bachelor	bachelor
Number of students	81	65	81	76	75
Age	22.9 (SD = 1.9)	23.0 (SD = 2.0)	19.9 (SD = 1.1)	20.2 (SD = 1.7)	20.7 (SD = 3.1)
Gender	female: 64%	female: 46%	female: 67%	female: 49%	female: 61%
	male: 36%	male: 54%	male: 33%	male: 51%	male: 39%
Nationalities	Dutch: 52%	Dutch: 48%	Dutch: 75%	Dutch: 82%	Dutch: 76%
	Chinese: 21%	Chinese: 17%	Singaporean: 12%	Singaporean: 11%	Singaporean: 13%
	Other: 27%	Other: 35%	Other: 13%	Other: 7%	Other: 11%
Study programme*	MFT: 80%	MFT: 64%	BFT: 68%	BFT: 67%	BFT: 63%
	MBT: 19%	MBT: 28%	BBT: 12%	BBT: 15%	BBT: 8%
	Other: 1%	Other: 8%	Other: 20%	Other: 18%	Other: 29%

* MFT = master Food Technology; MBT = master Biotechnology; BFT = bachelor Food Technology; BBT = bachelor Biotechnology.

VEEs 2, 3, and 4 were evaluated as part of the bachelor level course Food Chemistry (168 study hours). All students had previously obtained a high school (or similar) degree: 86% in the Netherlands, and 14% outside the Netherlands. The participants of the evaluations of VEEs 2, 3, and 4 formed a more homogeneous group than the participants of the evaluations of VEE 1, since there is a big overlap in the curricula of the bachelor Food Technology (63 – 68% of the students participating in VEEs 2, 3, or 4) and the bachelor Biotechnology (8 – 15% of the students) up to the point where this course is taught. Students that were part of an 'other' study programme (18 – 29% of the students) include a wide range of students (exchange students, master level students, or students with a second major), whose prior knowledge was known to a limited degree. Each VEE prepares students for one of the three main topics of the laboratory classes. Students were assigned to the VEE corresponding to their lab class topic.

3.3.2 Procedure

All student activities related to the evaluation of the four VEEs are shown in Table 3.5. The activities coded A1 – A6 correspond to the evaluation of VEE 1, the activities coded A2 – A4 correspond to the evaluation of VEE 1 online, and the activities coded B1 – B4 correspond to the evaluation of VEEs 2, 3, and 4. All VEEs were scheduled on the day(s) prior to the start of the laboratory classes as a compulsory activity. Students voluntarily completed the questionnaires (A4 and B3) and the pre- and post-knowledge tests (A1 and A5), while they were aware that the anonymized results could be used for course innovations.

Table 3.5: Student activities related to the evaluation of VEE 1 (A1 – A6), VEE 1 online (A2 – A4), and VEEs 2, 3, and 4 (B1 – B4), with corresponding measurements, timing, and time indication.

Code	Student activitiy	Measurement	Day*	Duration
A1	Pre-knowledge test	Content-specific prior knowledge	1	30 min
A2	Preparative assignment	N/A	1	1 h
A3	VEE	Student behaviour**	1-3	7.5 h
A4	Questionnaire	Evaluation of design requirements	2 or 3	5 min
A5	Post-knowledge test	Content-specific knowledge	8	30 min
A6	Laboratory classes	N/A	8-12, 15-19	27 h
B1	Preparation assignment	N/A	1	3 h
B2	VEE	Student behaviour**	2	1.5-2 h
B3	Questionnaire	Evaluation of design requirements	2	5 min
B4	Laboratory classes	N/A	5-9, 12-15	20-36 h

* Running days relative to the start of the experiment. There were no course activities in between the VEE and the post-knowledge test in the evaluation of VEE 1.

** This has been elaborated in chapter 6.

Evaluation procedure of VEE 1

To measure students' achievement of the intended learning outcomes of VEE 1, we used the same method as described in chapter 6. Students were given 30 minutes to complete a pre-knowledge test (A1) on paper. The pre-knowledge test consisted of 13 open questions, covering learning outcomes of the VEE. Each correctly answered question was rewarded with one point, yielding a possible range of scores for students between zero and 13 points. The assessment of the pre-knowledge test was done by a domain-related expert, who used a detailed key including how many points should be awarded for specific answers. A randomly selected sample (N = 10) was checked by another domain-related expert to make sure that the assessment was accurate and reproducible. The scoring by both experts was identical. Most of the students (N = 77) completed the pre-knowledge test, while four students were absent that day. The scores of this test were used as a measurement of their content-specific prior knowledge.

After the pre-knowledge test, students started working on the preparation questions (A2), and consequently on the VEE (A3). Students worked in the VEE for an average of 7.5 hours. Computer rooms were available and three carefully instructed supervisors were present to answer questions of students. Upon completion of the VEE, students were asked to fill in a questionnaire (see Appendix) that included questions and statements for evaluation of the design requirements (A4). All questions had clearly defined answer options, based on Thalheimer (2016), in order to rule out interpretation of answer options. They were fully written out, and coded as follows: A = (Almost) never, B = Less than half of the time, C = Half of the time, D = Most of the time, E = All the time, and, if applicable, F = I did not use this option. When using questions in the questionnaire would result in odd answer options (e.g. I learned a lot from the VEE half of the time), we replaced the question with a statement. All statements were provided with a 5-point Likert scale (1 = disagree, to 5 = agree).

Just before the start of the laboratory classes, a post-knowledge test was conducted (A5) to determine the learning outcomes of the VEE. This test was identical to the pre-knowledge test. The scores were used as a measure of the content-specific knowledge the students (N = 79) had at the start of the laboratory class. There were no other course activities in between the pre- and post-knowledge test, other than the virtual experiments in the VEE. Students were not aware that a post-knowledge test would follow, to prevent targeted studying for it. The validity of these knowledge tests was confirmed by three subject matter experts, including the main teacher of the course.

Evaluation procedure of VEE 1 online

The evaluation procedure of VEE 1 online was to a large extent determined by unforeseen circumstances, due to which the whole course was taught online. Given the time limitation, we were unable to conduct the pre- and post-knowledge tests in such a way that we could reliably compare the scores to the scores found in the evaluation of VEE 1. Nevertheless, we were able to provide students with the questionnaire (see Appendix) to evaluate the design requirements (Table 3.5, A4). Furthermore, this situation provided us with the ultimate opportunity to test whether students would be able to complete the assignment in the VEE independently (DR3) when supervisors and students are not together while working in the VEE.

Evaluation procedure of VEEs 2, 3, and 4

Students started off with a preparation assignment (Table 3.5, B1), in which they made an experimental set-up for the actual lab classes, after which they started working in VEE 2, 3, or 4 (B2). Students worked an average time of 1.5 hours in VEEs 2 or 3, or an average time of 2 hours in VEE 4. Upon completion of a VEE, students were asked to fill in a questionnaire (see Appendix), that was almost identical to the questionnaire that was used in the evaluation of VEE 1 (B3).

3.3.3 Evaluation: were the design requirements satisfied?

This section will focus on the results of the questionnaires (Table 3.5, A4 and B3) that were used to evaluate whether the design requirements were satisfied. The questionnaire results of all four VEEs are presented in Table 3.6.

When comparing the questionnaire results of the four VEEs, we observed a high degree of similarity between the answers given to the same questions. This suggests that using the same design architecture based on the same design requirements and design principles (sections 3.2.2 – 3.2.4) creates similar student experiences and student appreciation when applied in different contexts.

In terms of DR1 (the VEE should create a positive learning experience), we found that all four VEEs were capable of capturing the students' attention (question 1). What stands out is that students felt less confident (question 2) while working in VEE 4, especially compared to VEEs 2 and 3. During supervision it became clear that students experienced VEE 4 as difficult, which is also reflected in the VEE's characteristics, as students were provided with more data points and double the amount of hints compared to VEEs 2 and 3. Ninety percent of the students reported that most or all of the content of the VEEs was clear (question 3). The information obtained through questions 4.1 – 6.3 is highly valuable during formative evaluations in the design & construction phase, as it may direct the fine-tuning of the VEEs. For example, in questions 4.1 – 4.3, we see that the degree to which students are challenged by the different parts of the VEEs varies from being challenged all the time, to being challenged (almost) never. The VEEs were not intended to be highly challenging, as part of the content should be prior knowledge. However, we found big differences in student scores at the pre-knowledge test in the evaluation of VEE 1. These differences can be explained by the heterogeneity of the

Table 3.6: Questionnaire results of VEEs 1 – 4. The results are shown per VEE and per evaluation question (questionnaire in Appendix). The values represent the percentage of students who selected the corresponding answer option. Shading was used to visualize the distribution of the results. The answer options to all questions were coded as follows: A = (Almost) never, B = Less than half of the time, C = Half of the time, D = Most of the time, E = All the time, and, if applicable, F = I did not use this option. All statements were provided with a 5-point Likert scale (1 = disagree, to 5 = agree).

DR1	Positive learning experience	**VEE 1** (*N = 81*)					
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	
1	Capture attention	0	1	7	49	43	
2	Feel confident	0	5	14	65	16	
3	Clear content	0	0	8	62	30	
4.1	Challenged - experimental design	4	15	19	41	21	
4.2	Challenged - answering questions	4	10	24	37	25	
4.3	Challenged - processing data	4	18	23	29	26	
5.1	Useful - experimental design	3	1	7	42	47	
5.2	Useful - answering questions	0	3	13	37	47	
5.3	Useful - processing data	1	8	13	34	44	
6.1	Enjoy - experimental design	1	4	22	40	33	
6.2	Enjoy - answering questions	0	13	16	42	29	
6.3	Enjoy - processing data	5	17	18	28	32	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
7	User-friendly	0	3	8	29	60	
DR2	Achieve learning outcomes						
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	
8	Helped to understand experiments	0	0	7	66	27	
9	Confidence in processing data	0	4	9	63	24	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
10	Learned a lot	0	1	13	31	55	
11	I think I am well prepared	0	1	13	52	34	
DR3	Complete independently						
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	**F**
12	Sufficient support to complete	0	3	12	58	27	
13	Ability to process data	0	1	13	70	16	
14	Useful - hints	0	0	7	56	37	0
15	Useful - intermediate calculations	5	7	21	33	22	12
16	Consulted other students	54	27	7	7	5	
17	Checked pdf with calculations	26	29	25	16	4	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
18	Pdf was useful	2	7	23	29	39	

DR1	Positive learning experience	**VEE 1 online** (*N = 65*)					
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	
1	Capture attention	1	0	8	48	43	
2	Feel confident	2	8	26	52	12	
3	Clear content	0	2	20	60	18	
4.1	Challenged - experimental design	3	8	27	51	11	
4.2	Challenged - answering questions	0	5	32	49	14	
4.3	Challenged - processing data	0	12	39	32	17	
5.1	Useful - experimental design	2	0	8	41	49	
5.2	Useful - answering questions	2	2	9	44	43	
5.3	Useful - processing data	2	3	17	32	46	
6.1	Enjoy - experimental design	3	5	23	49	20	
6.2	Enjoy - answering questions	5	3	29	45	18	
6.3	Enjoy - processing data	4	14	37	28	17	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
7	User-friendly	2	6	15	28	49	
DR2	Achieve learning outcomes						
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	
8	Helped to understand experiments	0	1	8	63	28	
9	Confidence in processing data	0	2	9	52	37	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
10	Learned a lot	1	3	8	45	43	
11	I think I am well prepared	1	2	9	43	45	
DR3	Complete independently						
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	**F**
12	Sufficient support to complete	0	14	15	43	28	
13	Ability to process data	0	3	11	60	26	
14	Useful - hints	0	0	17	49	34	0
15	Useful - intermediate calculations	1	5	19	34	37	4
16	Consulted other students	60	20	12	6	2	
17	Checked pdf with calculations	20	25	18	31	6	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
18	Pdf was useful	0	6	17	38	39	

DR1	Positive learning experience	**VEE 2** (*N = 81*)					
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	
1	Capture attention	0	1	6	63	30	
2	Feel confident	0	0	6	68	26	
3	Clear content	0	1	3	60	36	
4.1	Challenged - experimental design	7	27	30	27	9	
4.2	Challenged - answering questions	1	21	38	31	9	
4.3	Challenged - processing data	12	27	35	17	9	
5.1	Useful - experimental design	1	4	15	28	52	
5.2	Useful - answering questions	0	5	17	30	48	
5.3	Useful - processing data	0	10	13	36	41	
6.1	Enjoy - experimental design	3	6	16	43	32	
6.2	Enjoy - answering questions	2	9	25	44	20	
6.3	Enjoy - processing data	1	11	30	38	20	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
7	User-friendly	0	1	5	31	63	
DR2	Achieve learning outcomes						
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	
8	Helped to understand experiments	0	5	12	51	32	
9	Confidence in processing data	0	0	9	68	23	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
10	Learned a lot	0	1	21	44	34	
11	I think I am well prepared	1	4	15	61	19	
DR3	Complete independently						
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	**F**
12	Sufficient support to complete	0	1	4	52	43	
13	Ability to process data	0	0	6	52	42	
14	Useful - hints	0	0	4	44	41	11
15	Useful - intermediate calculations	0	2	5	25	44	24
16	Consulted other students	78	16	5	1	0	
17	Checked pdf with calculations						
Statement	*Short description*						
18	Pdf was useful						

DR1	Positive learning experience	**VEE 3** (*N = 76*)					
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	
1	Capture attention	0	1	4	63	32	
2	Feel confident	0	0	9	67	24	
3	Clear content	0	0	3	59	38	
4.1	Challenged - experimental design	17	27	24	29	3	
4.2	Challenged - answering questions	1	17	30	45	7	
4.3	Challenged - processing data	13	14	29	36	8	
5.1	Useful - experimental design	0	4	12	37	47	
5.2	Useful - answering questions	0	0	16	46	38	
5.3	Useful - processing data	1	3	10	33	53	
6.1	Enjoy - experimental design	0	5	24	43	28	
6.2	Enjoy - answering questions	1	11	38	36	14	
6.3	Enjoy - processing data	3	10	33	35	19	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
7	User-friendly	0	1	8	28	63	
DR2	Achieve learning outcomes						
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	
8	Helped to understand experiments	0	2	20	58	20	
9	Confidence in processing data	0	0	15	67	18	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
10	Learned a lot	0	3	20	52	25	
11	I think I am well prepared	0	5	18	53	24	
DR3	Complete independently						
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	**F**
12	Sufficient support to complete	0	0	1	55	44	
13	Ability to process data	0	0	5	51	44	
14	Useful - hints	0	1	9	45	42	3
15	Useful - intermediate calculations						
16	Consulted other students	64	20	8	4	4	
17	Checked pdf with calculations						
Statement	*Short description*						
18	Pdf was useful						

DR1	Positive learning experience	**VEE 4** (*N = 75*)					
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	
1	Capture attention	0	4	7	56	33	
2	Feel confident	1	7	28	52	12	
3	Clear content	1	1	14	68	16	
4.1	Challenged - experimental design	7	15	31	32	15	
4.2	Challenged - answering questions	1	8	23	47	21	
4.3	Challenged - processing data	9	25	16	34	16	
5.1	Useful - experimental design	0	5	7	37	51	
5.2	Useful - answering questions	1	3	7	45	44	
5.3	Useful - processing data	0	5	11	52	32	
6.1	Enjoy - experimental design	1	7	17	47	28	
6.2	Enjoy - answering questions	1	12	35	44	8	
6.3	Enjoy - processing data	5	8	32	43	12	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
7	User-friendly	1	3	8	40	48	
DR2	Achieve learning outcomes						
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	
8	Helped to understand experiments	0	3	25	53	19	
9	Confidence in processing data	0	1	24	60	15	
Statement	*Short description*	**1**	**2**	**3**	**4**	**5**	
10	Learned a lot	3	0	17	49	31	
11	I think I am well prepared	0	6	29	49	16	
DR3	Complete independently						
Question	*Short description*	**A**	**B**	**C**	**D**	**E**	**F**
12	Sufficient support to complete	0	0	30	53	17	
13	Ability to process data	0	3	24	58	15	
14	Useful - hints	1	3	16	49	28	3
15	Useful - intermediate calculations	1	1	10	32	47	9
16	Consulted other students	60	18	9	8	5	
17	Checked pdf with calculations						
Statement	*Short description*						
18	Pdf was useful						

group of participants, and this might cause the differences in reported challenge. Despite the differences in reported challenge, almost all students experienced the VEEs as useful (question 5). The user-friendliness (statement 7) scored an average of 4.4 out of 5 over the four VEEs. Based on these results, we concluded that the VEEs create a positive learning experience, and that the VEEs satisfied DR1.

In terms of DR2 (the VEE should support students in achieving the intended learning outcomes), for all four VEEs, 83% of the students indicated that the VEE helped them to understand most or all of the experiments (question 8), and 85% of the students indicated that they are confident that they can successfully process most or all of the data they will obtain during the laboratory classes (question 9). Students had the feeling they learned a lot (statement 10) and that they were well prepared for the laboratory classes (statement 11) with average scores of 4.2 and 4.0 respectively. For VEE 1, a paired-samples t-test was conducted to compare the scores of the pre- and post-knowledge tests. In line with the questionnaire results, there was a significant difference in the scores of the pre-knowledge test ($M = 5.2$, $SD = 2.7$) and the post-knowledge test ($M = 8.5$, $SD = 2.0$); $t(75) = 11.5$, $p < 0.001$. These results indicate that the VEE helped students to prepare for laboratory education, increasing the average score on the knowledge tests from 5.2 to 8.5 points out of 13, which would likely have contributed to their knowledge construction during the laboratory class (Johnstone, 1997). It is worth noting that the knowledge tests focused on procedural knowledge solely, as this was the most important learning outcome of the VEE. We did not test all learning outcomes, so as a result we might have underestimated the learning effect of the VEE. Based on these results, we concluded that the VEEs support students in achieving the intended learning outcomes, and that the VEEs satisfied DR2.

In terms of DR3 (the VEE should enable students to complete the assignment independently), 85% of the students indicated that there was sufficient support to complete the VEE all the time or most of the time (question 12), and 87% of the students indicated that they were able to process most or all of the data of the virtual experiments by themselves (question 13). However, for VEE 4, the scores of questions 12 and 13 were slightly lower. We established that VEE 4 is slightly more difficult, and that we included more feedback and support for this VEE to compensate for the higher difficulty. The available hints were used by 97% of the students, and of those students, 86% found most or all of the hints useful (question 14). The option to check intermediate calculations was used by 90% of the students, and of those students, 78% found this option useful most or all of the time (question 15). From these results it was clear that the majority of the students used the support and found it useful. The majority of the students did not consult other students while working in the VEE (question 16). This indicates that students were able to complete the assignment in the VEE independently, hence supporting DR3. When comparing VEE 1 online with VEE 1, we found similar results in terms of consulting other students. This suggests that when students are in need of help, they will get into contact with peers, regardless of their working environment. Overall, we concluded that the VEEs enabled students to complete the assignment independently, and that the VEEs satisfied DR3.

When comparing the evaluation results of VEE 1 and VEE 1 online, we found a high degree of similarity. There were a few minor differences, such as that students felt slightly less confident, found the content a little less clear, found the interface slightly less user-friendly, and experienced a little less support in VEE 1 online. We think that these minor differences may be caused by the absence of supervisors and fellow students in the same room. Although a similar number of students indicated that they consulted other students while working in VEE 1 online, and supervisors could be consulted online, students were less likely to have a look at the screens of others, and the supervisors received only few questions.

3.4 Limitations

In this research, we used an educational design research approach to design four VEEs. These four VEEs were the context in our search for general design requirements, design principles, and a design architecture that can be used as a blueprint to design VEEs. These general output classes are well-grounded, based on literature, practical experience, and evaluations. However, we acknowledge that there are some limitations to this study.

First, we have not measured the effect of the better student preparation as a result of the VEEs during the actual laboratory classes. The measurement of this effect falls outside the scope of this study, as to our best knowledge, there is no simple or fast method to do so. For the moment, it is our conviction that a good preparation increases the potential for students to learn during laboratory classes (Johnstone, 1997).

Second, we have not been able to reliably quantify the learning effect of VEEs 2, 3, and 4. In the initial evaluation procedure, pre- and post-knowledge tests were included, but we found that many bachelor level students (as opposed to the master level students) were unwilling to voluntarily and seriously complete the pre- and post-knowledge test. Based on this behaviour, we judged the outcomes of these tests as unreliable. As a result, our conclusions regarding DR2 (achievement of intended learning outcomes) for VEEs 2, 3, and 4 were based on self-reported data.

Third, other than the pre- and post-knowledge tests used in the context of VEE 1, self-reported data was used to learn about student behaviour and their experiences in the VEEs. The disadvantage of self-reported data is that the assumption has to be made that students are capable of reflecting on their own behaviour, are willing to answer the questions honestly, and that all students have the same interpretation of the questions. To obtain more objective results, we are currently investigating the opportunity of educational data mining. Educational data mining uses logging data of students' interaction with an online environment, and is commonly used to detect patterns of student behaviour in an online environment, and to evaluate the relation between these patterns and students' learning outcomes (Aldowah et al., 2019). As such, educational data mining may provide insight in which features of a VEE contribute to learning, and which features are detrimental to learning (Aldowah et al., 2019; Hung and Zhang, 2008).

The VEEs in this study were designed for university level courses in the fields of food chemistry and food enzymology. In the evaluation we have shown that the VEEs were successful, as they satisfied the design requirements. However, we have not yet proven that our general output classes (i.e. design requirements, design principles, design architecture) will function in other levels and other fields of education. In terms of education level, we believe that extensive laboratory education is required in order to reach the full potential of VEEs. As extensive laboratory education is uncommon to primary and secondary education, the potential of VEEs will be highest for tertiary education, regardless of the level. In terms of fields of education, we believe that the general output classes will be applicable to courses in any field, as long as the course includes laboratory education.

3.5 Conclusions

The goal of this study was to provide general design requirements, design principles, and a design architecture that can be used as a blueprint to design VEEs. In order to do so, we followed an educational design research approach to design four VEEs, which were evaluated multiple times in real educational contexts. After the summative evaluation of each VEE, we finetuned the design requirements (research question 1) and corresponding design principles (research question 2), which have evolved throughout the multiple cycles of the educational design research approach, and we visualized a common design architecture (research question 3) based on the design principles.

We established three general design requirements that are applicable for VEEs: the VEE should i) create a positive learning experience, ii) support students in achieving the intended learning outcomes, and iii) enable students to complete the assignment independently. We concluded that these design requirements are general, and as such applicable to any NTL. Based on summative evaluations, we concluded that the design requirements were satisfied by each VEE. In order to satisfy these design requirements, we used fourteen design principles that were based on theoretical arguments, empirical arguments, and/or (formative) evaluations. Besides formulating these design principles, we also provide detailed information on how we applied those principles, so that their practical relevance is clear, and so that they can easily be adopted by teachers and instructional designers in the process of designing VEEs. Once we completed multiple cycles of the educational design research approach, we were able to extract a common design architecture from the four VEEs, which includes three phases (pre-experiment phase, experiment phase, and post-experiment phase), five levels (context level, research question level, experimental design level, data processing level, and results interpretation level), and the feedback and support system.

In our study, we combined literature on how to set up educational design research with literature that provides guidelines for designing instruction, in order to gain and report practical experience regarding the design of VEEs. The blueprint that resulted from these efforts does not only provide general design guidelines, design principles, and a design architecture, but it also includes detailed information on how they can be applied. Such information in the context of VEEs is new to literature, and has the potential to help teachers and instructional designers to design high quality VEEs.

Appendix

Design requirements (DRs) with corresponding questions (Q) and statements (S) used in the questionnaire.

DR1	**The VEE should create a positive learning experience**
Q1	Did the VEE capture your attention?
Q2	Did you feel confident while working in the VEE?
Q3	How clear was the content of the VEE?
Q4	Indicate for each part of the VEE to what degree you felt challenged.
Q5	Indicate for each part of the VEE how useful you think it is.
Q6	Indicate for each part of the VEE how much you enjoyed it.
S7	The VEE is user-friendly.
DR2	**The VEE should support students in achieving the intended learning outcomes**
Q8	To what degree did the VEE help you to understand the experiments?
Q9	To what degree are you confident that you can successfully process experimental data that you will obtain in the lab class?
S10	I learned a lot from the VEE.
S11	I think I am well-prepared for the lab class now that I finished the VEE.
DR3	**The VEE should enable students to complete the assignment independently**
Q12	Is the amount of feedback/support given within the VEE sufficient to complete the VEE without help from supervisors or other students?
Q13	How able were you to process the data of the experiments in the VEE without help of supervisors or other students?
Q14	How useful were the hints provided with the calculation questions?
Q15	How useful was the option to check your intermediate calculations?
Q16	Did you consult other students while working in the VEE?
Q17	How often did you check the pdf with calculations that you received after successfully completing a calculation?*
S18	The pdf with calculations that I received after successfully completing a calculation was useful.*

* These questions only apply to the evaluation of VEE 1 and VEE 1 online.

Chapter 4

Virtual experiment environment: a showcase of a preparation tool for laboratory classes

This chapter was published as:

S. Verstege and J. Diederen (2019). "Virtual experiment environment: A showcase of a preparation tool for laboratory classes". In: *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), 1406–1416

Contents

Abstract

Laboratory education plays an indispensable role in the conveyance of tacit knowledge, and is therefore an important part of natural science education. The success of laboratory education in terms of learning outcomes depends amongst others on the ability of students to cope with the extraneous cognitive load in the laboratory. This extraneous cognitive load can be reduced through careful student preparation. The aim of this paper is to present teachers and educational designers with an extensive example of a virtual experiment environment (VEE) that was developed to facilitate student preparation for laboratory classes. The VEE contains a self-directed learning task through which students can actively learn and apply concepts related to the laboratory classes. Students progress through the VEE by making an experimental design, obtaining raw data, processing raw data into results, and drawing conclusions. The user interface, the available interactions, and the feedback- and support mechanism in the VEE are presented and discussed.

4.1 Introduction

Laboratory classes are an important part of most curricula of natural science programs, because they support students to develop scientific skills and understand background theory of experiments (Alkhaldi et al., 2016; Bruck and Towns, 2013; Carnduff and Reid, 2003; Johnstone and Al-Shuaili, 2001; Reid and Shah, 2007; Winberg and Berg, 2007). To make optimal use of the learning potential of laboratory education, students should be carefully prepared when starting with laboratory work (Bannert, 2002; Crandall et al., 2015; Winberg and Berg, 2007). In order to prepare students for laboratory education, many non-traditional labs such as computer simulations, virtual laboratories, and remote laboratories have been introduced (for reviews, see Alkhaldi et al. (2016), Heradio et al. (2016), Potkonjak et al. (2016), and Rutten et al. (2012)), and were proven to be effective in terms of learning outcomes (Brinson, 2015; Rutten et al., 2012).

In this paper we will focus on the enrichment of a laboratory activity by means of a computer simulation. Computer simulations are programs that contain a model of a system (natural or artificial; e.g. equipment) or a process (De Jong and Van Joolingen, 1998). Computer simulations that can be used to prepare students for laboratory classes can be referred to as Virtual Experiment Environments (VEEs). A VEE is defined as "Any educational resource that enables students to design and/or to carry out virtual experiments and/or to process data, and to analyse and interpret results" (Hartog et al., 2009, p. 376). Besides functioning as a preparation tool, VEEs can also be used as a stand-alone.

The design of a VEE starts with a design challenge, which is to facilitate students in achieving Intended Learning Outcomes (ILOs). The definition of a VEE allows the designer a huge design space, or in other words, a huge number of candidate solutions for the design challenge. The aim of this paper is to present a showcase of a design solution that was based on educational theories (e.g. Clark and Mayer (2016) and Mayer (2009a)), which led to a successful VEE in terms of facilitating students to achieve the ILOs.

In the showcase of the VEE, we will describe all its elements in detail, and: i) visualise the user interface, ii) present all available interactions, and explain why these interactions have been introduced, and iii) present the feedback- and support mechanism that is available to users throughout the learning task. With this showcase, teachers and educational developers are provided with an extensive example of how a VEE can be built up.

4.2 The LabBuddy VEE

The LabBuddy VEE has been developed as part of an educational design research project, targeted as a preparation tool for the laboratory classes of a master level course in the field of enzymology, taught at Wageningen University. The LabBuddy VEE is an extended version of a previously developed web-based experiment designer and web lab manual (Van der Kolk et al., 2012), nowadays known as LabBuddy (Kryt b.v., Wageningen, the Netherlands) (Van der Kolk et al., 2013).

The VEE comprises the complete research cycle, focusing on experimental design, processing of raw data, and interpretation of results. The VEE requires users to 'perform' a chemical experiment consisting of 38 sub-experiments, such as isolations, modifications, and analyses (e.g. extraction of enzymes from a raw material, and determination of the extracted enzyme's activity) that will all contribute to achievement of the ILOs. The degree of immersion in the VEE is low, as users engage in a 2D environment, in which a time scale has been chosen so that it takes no time to perform an experiment. As a result, the user saves approximately 80 hours of time, which means that the learning task in the VEE can be completed in within eight hours. Upon completion of the learning task, users should be fully prepared for the laboratory classes on a conceptual level.

4.2.1 Preparative assignment

Being knowledgeable about the context of the VEE before starting to work in the VEE is of utmost importance. The goal of this VEE is not to learn the subject of the context, but to learn about how to do science. Users need an understanding of the context to be able to make choices for their experimental design, to process raw data into results, and to interpret their results.

The VEE is preceded by a preparative assignment that was introduced to activate prior knowledge and/or construct new knowledge to ensure users have sufficient background on the context of the VEE. This should result in a lower extraneous cognitive load while working in the VEE, so that more cognitive processing capacity can be directed towards essential processing (Mayer, 2009c). The preparative assignment consists out of 17 questions that have to be answered in the given order. Users cannot progress to the next question before they have correctly answered the current question. The feedback that is provided during this preparative assignment, depends on the answers of the users.

4.2.2 Overview of the complete user interface

In Figure 4.1, we provide an overview of the user interface of the VEE. In the following sections all elements will be described and visualized systematically.

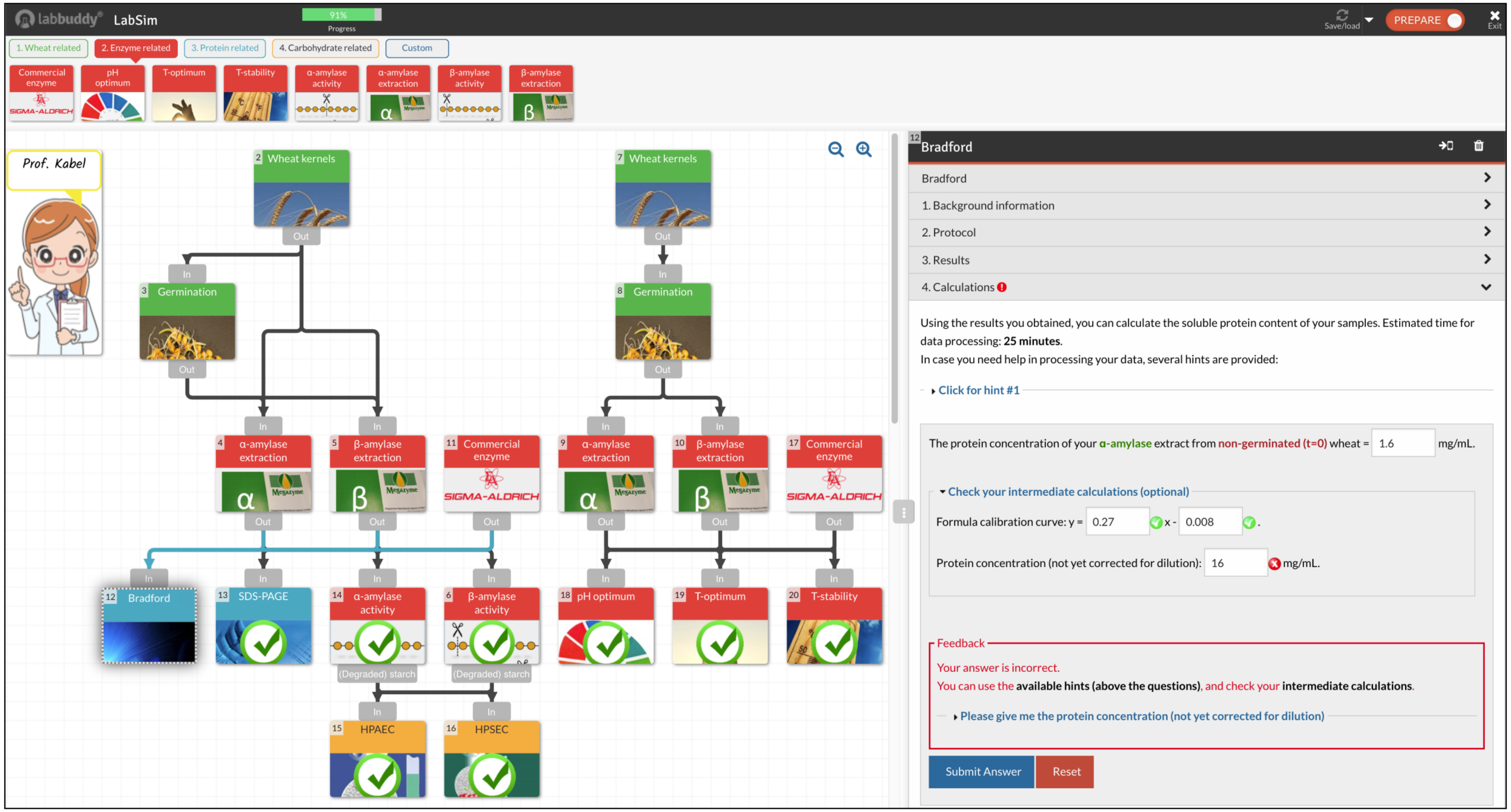

Figure 4.1: LabBuddy VEE screenshot. Overview of complete user interface.

4.2.3 General guidance

Users will encounter 73 questions throughout the VEE that have to be answered in order to progress. The questions were introduced to support users in achieving the ILOs. At any time, when a question needs answering, a red circle appears and starts blinking in the element in which the question is located (e.g. Figure 4.2a). When the corresponding element has been clicked on, the tab in which the question is located will also have a blinking red circle (e.g. Figure 4.2b). Once the question has been answered correctly, the blinking red circle will disappear. This guidance was introduced for two reasons. First, to inform users that a new question has been unlocked and needs answering, and second, to guide users through the VEE and prevent them from getting lost in search for how to progress. As the learning task in the VEE takes several hours to complete, a progress bar was introduced, which is visible at any time. The progress bar indicates the user's progress from 0 to 100%.

4.2.4 User activities

Obtain research questions and formulate hypotheses

Research normally starts with a research question, so also in our VEE, the users will start with six research questions that have to be answered to complete the VEE. The research questions are provided by the virtual professor (Figure 4.2).

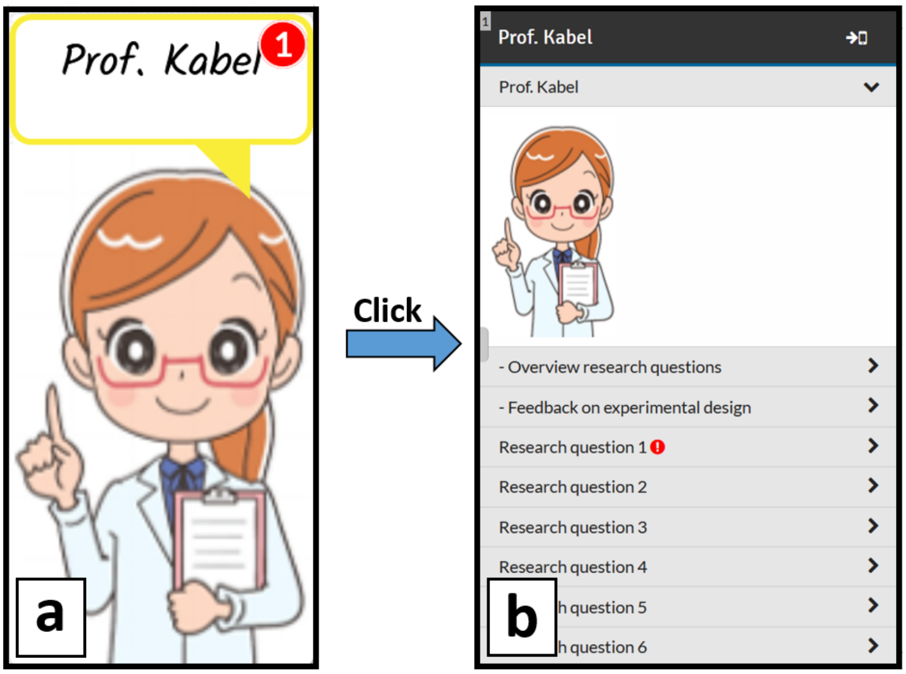

Figure 4.2: LabBuddy VEE screenshots. a) Virtual professor. b) Content of the virtual professor.

Users have to complete hypotheses belonging to each research question. Formulating hypotheses is an integral part of the research cycle. Since being able to formulate hypotheses is not an ILO of the VEE, pre-defined hypotheses with gaps (Figure 4.3a) were incorporated to make users familiar with the research questions, and trigger them to think about their expectations.

Just like in any other question within the VEE, users click on the submit answer button, which will trigger answer-specific pre-defined feedback, that will be displayed above the submit answer button. In case the question was answered correctly, the feedback will colour green (Figure 4.3b). In case a question is answered incorrectly, the feedback will colour red (Figure 4.3c), and will require users to change their answer to this question. Generally, when a question is answered incorrectly, constructive feedback and/or guidance is provided, which will be elaborated on in the following sections.

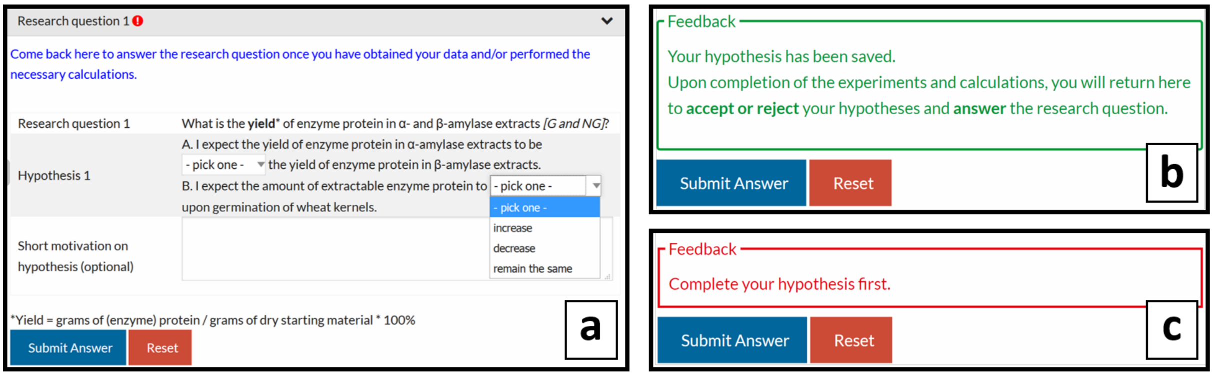

Figure 4.3: LabBuddy VEE screenshots. a) Example of a pre-defined hypothesis with gaps. b) 'Positive' feedback in green, given when a question is answered correctly. c) 'Negative' feedback in red, given when a question is answered incorrectly or incompletely.

Make an experimental design

Based on the research questions, users make an experimental design in the form of a workflow. To make this workflow, users are provided with a chemical methods dashboard (Figure 4.4a) that contains all raw materials, sample preparation methods (e.g. isolation or modification), and analysis methods, that are referred to as 'chemical methods' in this paper. A detailed description of a chemical method is given in the following section. In the chemical methods dashboard, the chemical methods are subdivided in categories, which can be accessed by clicking on them. The workflow canvas (Figure 4.4b) is the area to which chemical methods can be dragged. Chemical methods have an in- and/or out-port, that can be connected to another, by means of drag and drop, to form a workflow.

Users can add the same chemical methods to the workflow canvas multiple times. This functionality was introduced to allow users to shape their own thoughts about the

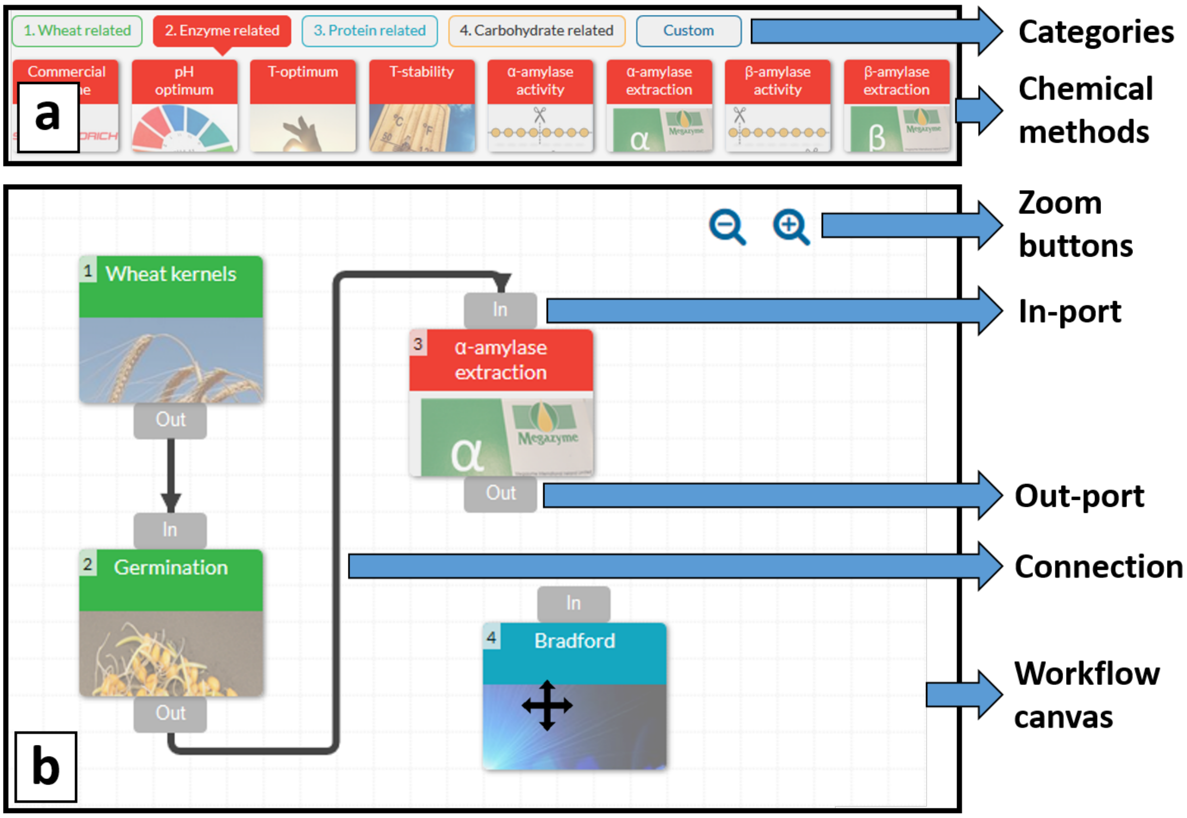

Figure 4.4: LabBuddy VEE screenshots. a) Chemical methods dashboard, including annotation. b) Partially filled in workflow canvas, including annotation.

experimental design without being restricted by the VEE. As a result, there are multiple ways in which the same experimental design can be built. Users can for example connect several chemical methods to one analysis method (e.g. Bradford) (Figure 4.5a), or alternatively they can connect each chemical method to a separate copy of the same analysis method (Figure 4.5b). The consequence is that in the case of Figure 4.5a, the user will receive all results within the analysis method, whereas in the case of Figure 4.5b, the results are spread over the three analysis methods.

Users can open the 'feedback on experimental design' tab in the virtual professor, in which the users are informed which research questions can be answered with their current experimental design (Figure 4.6a). In case there are mistakes in the experimental design, feedback will be provided, which is indicated with a blinking red circle appearing in the 'results tab' (that will be introduced in the next section) of the incorrectly connected chemical method (Figure 4.6b). If a chemical method is not connected to the prerequisite chemical method(s), feedback will be provided, but it will not be indicated with a blinking red circle (Figure 4.6c). The choice for not indicating this feedback by a blinking red circle was made to prevent an overload of feedback while users are in the process of making the experimental design.

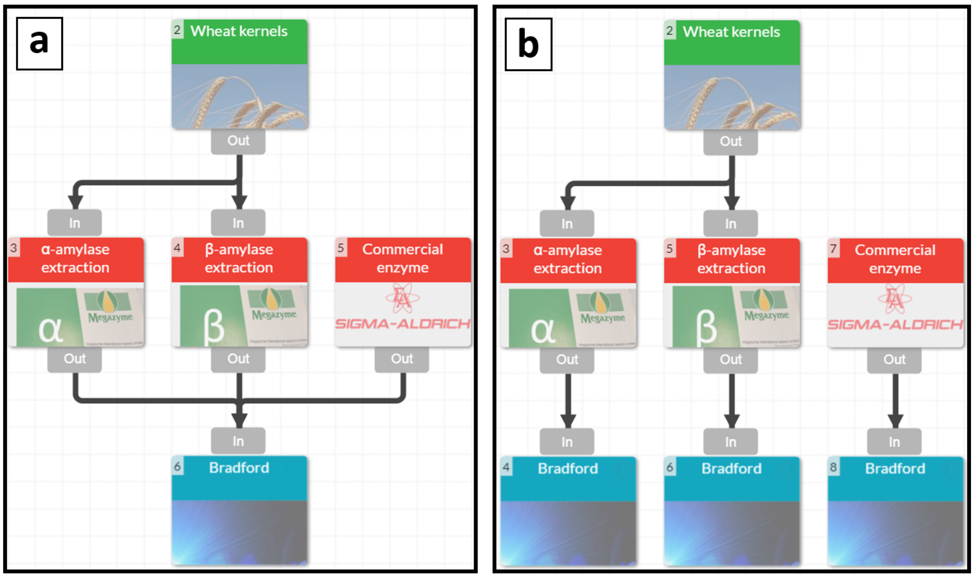

Figure 4.5: LabBuddy VEE screenshots. Two experimental designs (a and b), leading to the same results.

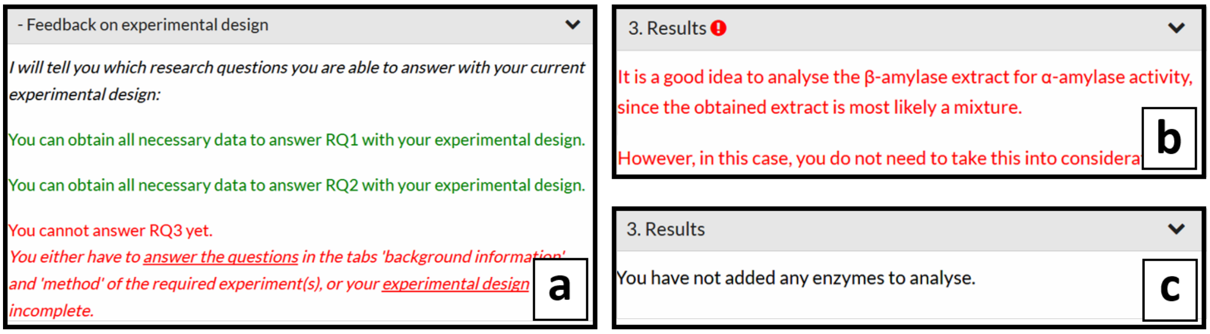

Figure 4.6: LabBuddy VEE screenshots. Feedback on experimental design, regarding a) possible completion of research questions, b) incorrectly connected chemical methods, c) missing prerequisite starting materials.

Open a chemical method

There are three types of chemical methods: i) raw materials, ii) sample preparation methods, and iii) analysis methods. When users open a chemical method, a content dashboard will appear. The content dashboard can contain the following four tabs, depending on the type of chemical method (Figure 4.7a): 1. Background information, 2. Method, 3. Results, 4. Calculations. Each tab will be discussed and visualized separately in the following sections. When users have completed an analysis method (i.e. answered all questions, obtained all data, and processed all data correctly), the appearance of this analysis method will change, as a green 'check' will appear (Figure 4.7b).

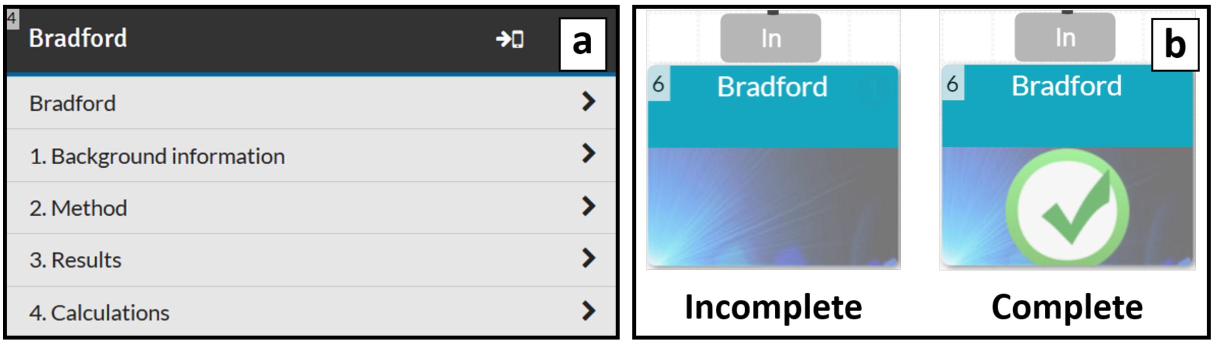

Figure 4.7: LabBuddy VEE screenshots. a) Content dashboard of chemical methods. b) Appearance of an incomplete and a complete chemical method.

The background information tab The background information tab contains a short description and explanation of the chemical method, followed by one or more questions about the background information. Users have to answer all questions correctly before any results belonging to that chemical method can be unlocked (in tab 3. Results). The questions were introduced to make sure the users understand the principle behind the chemical method. In case they already knew, users can easily answer the question, and if not, the answer can be found in the background information. Users can access more detailed information on the chemical method by opening a collapsed frame. The collapsible frames were introduced to segment the information, hence reducing the user's cognitive load (Mayer, 2009b). For complex methods, the more detailed information is extensive, and consists of not only text, but also visuals to explain mechanisms.

The method tab In the method tab users can study the protocol of the selected chemical method. When appropriate, the protocol starts off with a short introductory clip (Figure 4.8a). The protocol itself shows all practical steps, and is displayed using the integrated 'web lab manual' tool that was developed by LabBuddy (Figure 4.8b). The protocol in the VEE is similar to the protocol of the laboratory classes, to maximize transfer. Questions were included to ensure the user's understanding of the protocol. For example: users have to calculate the dilutions of a calibration curve, or choose which fraction to continue with after centrifugation.

The results tab In the results tab, users will obtain feedback on their experimental design (Figures 4.6b and c), and depending on the completeness of the experimental design, they will receive the results. Results are usually raw data of the analysis. Results can be presented in text (Figure 4.9a), or as a link to a Microsoft PowerPoint or Excel file (Figure 4.9b). An example of raw data provided in an Excel file is shown in Figure 4.9c. The user is notified when he/she obtained all results belonging to one analysis method (Figure 4.9b).

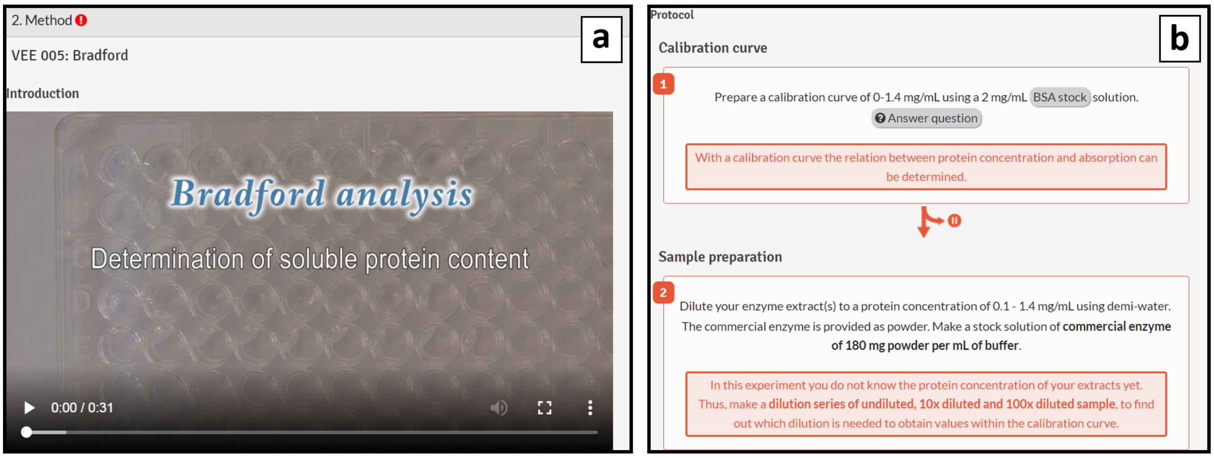

Figure 4.8: LabBuddy VEE screenshots. a) Short introduction clip placed in the introduction of the method tab. b) Example of two protocol steps, including an 'answer question' button that has to be clicked by the user to open the question.

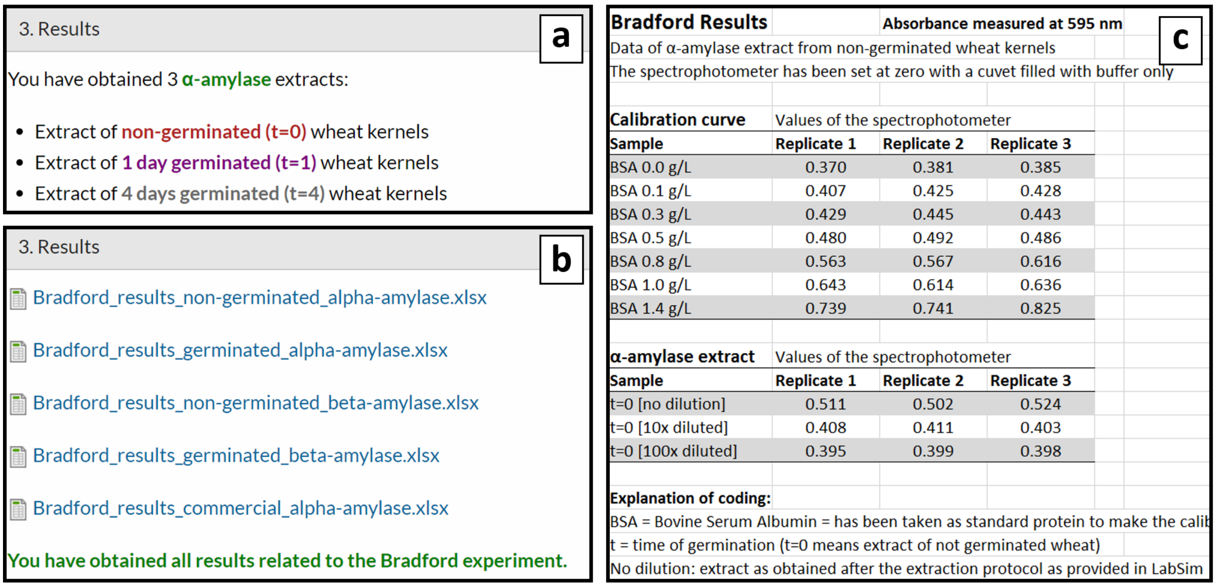

Figure 4.9: LabBuddy VEE screenshots. a) Results presented in text. b) Links to Excel files. c) An example of raw data provided in an Excel file.

The calculations tab Most analysis methods yield raw data that needs to be processed, often by doing calculations. Therefore the calculation tab contains calculation questions for which the answers are the results of the user's calculations. Within the calculation questions, a range of answers are accepted to make up for small deviations as a result of rounding. To make sure all users are able to complete the calculations, a feedback and support system was introduced. The feedback and support system consists of three elements: i) hints, ii) feedback on which answers are (in)correct, iii) the option to check intermediate calculations, including targeted feedback.

At any time, the user can access hints that guide through every step of the calculation. Hints are provided in collapsible frames. When the user opens the first hint, only the first step of the calculation will be given. Inside the frame of the first hint, the second hint is presented in another collapsed frame (Figure 4.10a). This way, users can progress or check their approach step by step. Depending on the nature of the calculation, up to five hints are available per calculation, and 36 hints in total for the whole VEE.

Each value the user submits in a calculation question, will be checked by the VEE, which will in turn provide feedback by means of a green check for a correct value, or a red cross for an incorrect value (Figure 4.10b). This is a first step to help the user identify his/her mistake. However, the mistake could be the result of a misunderstanding that the user might not be able to identify without further support or the help of a peer or teacher. Since users leave the VEE to do their calculations in Microsoft Excel, the calculation process is a black box, which makes it extremely hard to monitor and identify mistakes. The challenge was to implement a feedback- and support mechanism to help users to identify their own mistakes, without giving away the answers too easily, since that would lead to gaming the system behaviour (chapter 6). The result was a 'check your intermediate calculations (optional)' collapsible frame. In this collapsible, one or more intermediate calculations can be checked (Figure 4.10c). The provided feedback is similar as before, with one addition for incorrect values. In case of one or more incorrect values, one or more collapsibles will appear which, when opened, will give the correct value. Now that the user knows in which part of the calculation the mistake was made, the user can choose to search for it, or alternatively open the collapsible that will provide the correct value needed for further calculations. This way, the user is in control of his/her own learning process. When a user did all intermediate calculations correctly, but still has a mistake in the final answer, a link to the complete calculation will be provided. The same link will also be provided when the final answer is correct, to make sure that users who did not make use of the feedback- and support mechanism also have the option to check their calculations. A complete overview of a typical calculation tab is given in Figure 4.10d.

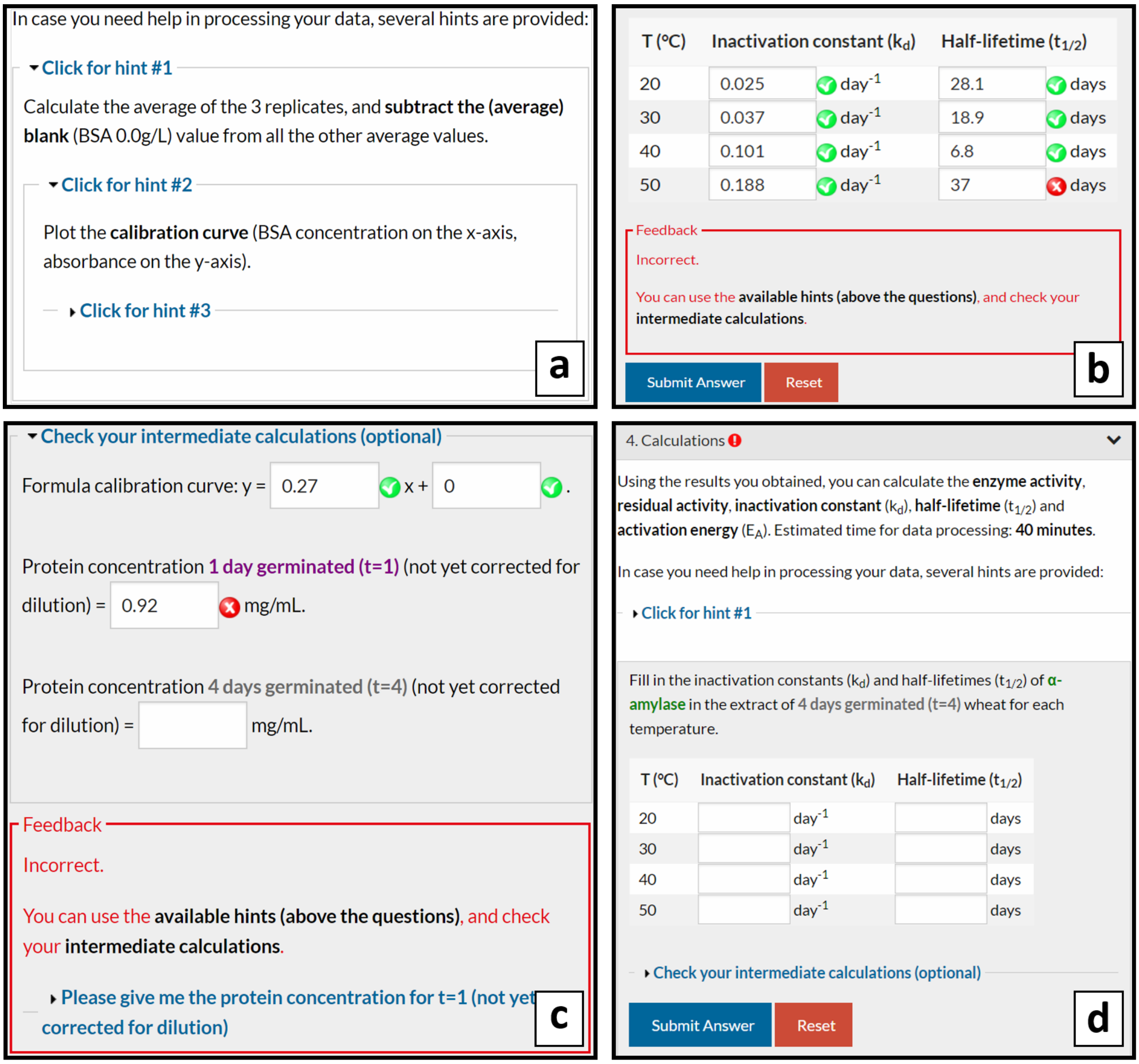

Figure 4.10: LabBuddy VEE screenshots. a) Support through collapsible hints. b) Feedback on a calculation question. c) Content of a 'Check your intermediate calculations (optional)' collapsible, including feedback. d) Example of a typical calculations tab.

Answer research questions and evaluate hypotheses

Based on the nature of the research question, users have to interpret their (calculated) results, integrate (calculated) results of different experiments, and/or do follow up calculations. Those results have to be submitted at the virtual professor, in the same tabs where users initially formulated their hypotheses (Figures 4.3a and 4.11). The hypotheses have to be accepted or rejected to complete the research cycle.

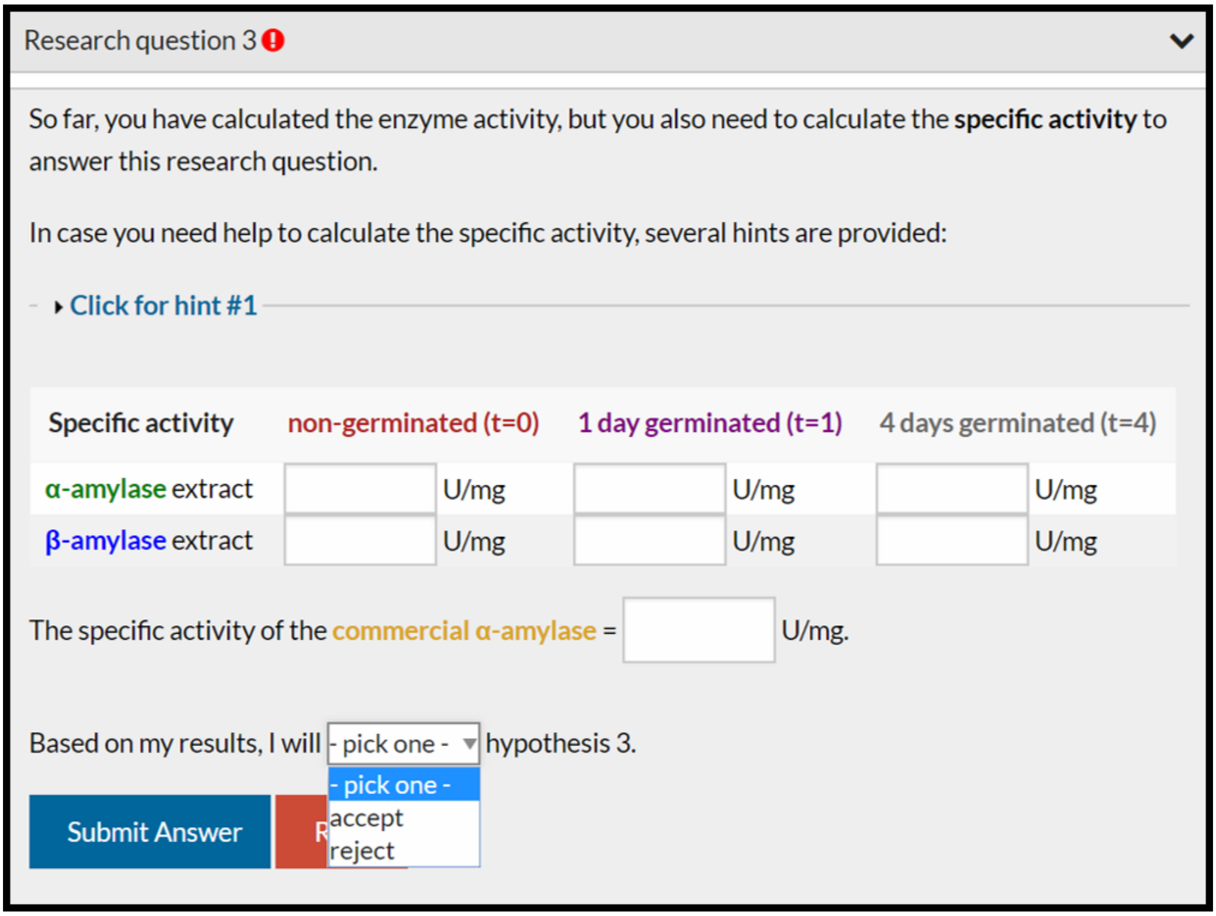

Figure 4.11: LabBuddy VEE screenshot. Example of a research question for which users have to integrate calculated results of two experiments to do further calculations.

4.3 Evaluation

The LabBuddy VEE has been evaluated extensively within its educational context in two successive years, using i) a questionnaire to evaluate design choices, and to study the user's (N = 178) experiences, and ii) pre- and post-knowledge tests. On a 5-point Likert scale, 90% of the users indicated that they learned a lot in the VEE; 92% of the users experienced the VEE as user-friendly; and 84% of the users had the feeling that they were well prepared for the laboratory classes after completing the VEE. From the knowledge tests, it was concluded that the VEE contributed to achievement of the ILOs. The outcomes of the complete evaluation of the user's experiences, as well as the evaluation of the design choices made during the design process are the subject of chapter 3. The VEE was also used in a study on self-regulated learning, in which the relation between students' perceived levels of self-regulation, learning behaviour, and learning outcomes was explored (chapter 6). Overall, we conclude that a VEE such as the one presented in this paper contributes to student preparation for laboratory classes and we hope that the showcase of this VEE provides guidance to teachers and educational designers in the process of designing a VEE.

Chapter 5

Design of interactive protocols that help students to prepare for laboratory work

This chapter was submitted for publication as:

S. Verstege, W. Lamot, J.-P. Vincken, and J. Diederen. "Design of interactive protocols that help students to prepare for laboratory work".

Contents

Abstract

Laboratory education makes up an extensive part of natural science education at universities. To support meaningful learning during laboratory work, students should prepare by carefully studying the protocols. The aim of this study was to design and evaluate a preparative learning material that focuses on protocol steps with the design goal to help students to i) understand the ‘why’ (reason or theory) of protocol steps, ii) make practical choices and iii) do troubleshooting. To reach this goal, protocol steps have been enriched with theoretical, practical, troubleshooting, and/or calculation questions to form interactive protocols. This design process resulted in a list of design choices (i.e. when to include a question, and how to design a question) and a showcase of questions in the interactive protocols. These interactive protocols were implemented and evaluated in a real educational setting. From the evaluation results it was concluded that the interactive protocols were successful in preparing students for the laboratory work. After the laboratory work, students reported a more diverse and slightly less positive contribution of the interactive protocols to their understanding of the protocol steps and their ability to make practical choices. A significant difference was found in the perceived usefulness of the troubleshooting questions before versus after the laboratory work, which is suggested as a topic for further investigation.

5.1 Learning during laboratory work

Imagine you are supervising a laboratory class, and one of your students is about to make a mistake by discarding the supernatant, containing the desired component, instead of the pellet. Despite your exhaustive efforts to make the protocol as clear as possible, there are always students who do not seem to think while executing a protocol. You wonder: how can this phenomenon be explained?

Teaching students how to 'do' science in the laboratory can be split into multiple stages (Seery et al., 2019). In the first stage, students develop experimental skills, and become comfortable and competent in the laboratory environment. This includes practical learning outcomes such as: 'be able to correctly use a pipet'. In the second stage, students learn to predict the effect of changes in protocols, and learn to explain their observations. In this stage, theoretical knowledge is added to the already existing practical knowledge obtained in the first stage. The next stages focus on experimental design, initially with familiar outcomes, and ultimately with open-ended and unfamiliar topics.

In this study we focused on student learning in the first two stages, in which students are typically provided with ready-made protocols. Protocols contain the information required to execute each practical step. But no matter how detailed a protocol is, students might come up with practical questions. For example, the protocol step "Fill three beakers with 2 mL of demi water", can lead to questions such as: What size should the beakers be? Which pipet should I use? Which pipet tip is suitable? Should I change the pipet tip in between? Where do I find all the materials that I need? Since protocols generally do not contain answers to all such questions, students will have to make practical choices. In relation to the second stage, students learn i) to understand the 'why' (reason or theory) of protocol steps, and ii) to combine this theoretical understanding with their experience in making practical choices to do troubleshooting in the case they encounter a problem or unexpected situation while executing a protocol.

The success with which students learn during laboratory work, is among others influenced by their working experience in laboratories, their prior knowledge, the cognitively demanding environment of a laboratory (for example caused by having to work around fellow students, locating materials, and using equipment), and perhaps most importantly, their own goals for laboratory education. Students were found to be primarily guided by affective goals, as opposed to faculty, who tend to primarily focus on cognitive and psychomotor learning (as also reflected in the first two stages) (Bretz et al., 2013). This means that students are primarily focused on completing the laboratory work as soon as possible, which results in very little meaningful learning during the laboratory work (DeKorver and Towns, 2015).

Careful student preparation has proven to increase meaningful learning during the laboratory work (Gregory and Di Trapani, 2012; Jones and Edwards, 2010; Rollnick et al., 2001). Students can for example prepare with learning materials such as videos, lectures, exercises, and computer simulations (Agustian and Seery, 2017; Carnduff and Reid, 2003; Jolley et al., 2016; Jones and Edwards, 2010; Reid and Shah, 2007; Spagnoli et al., 2019; Winberg and Berg, 2007). To the best of our knowledge, such learning materials focus on general understanding of a protocol, without discussing the level of detail: the protocol steps. Especially in terms of the second stage, it is valuable for students to understand the theory behind a protocol step, and understand the reason for this protocol step. Without this understanding, students will not be able to make changes to protocols, or to do troubleshooting in the case they encounter a problem or unexpected situation.

The aim of this study was to design and evaluate a preparative e-learning material that focuses on protocol steps. The design goal was threefold: to help students to i) understand the 'why' (reason or theory) of protocol steps, ii) make practical choices and iii) do troubleshooting. In the following sections, we discuss the design of the preparative e-learning material (section 2), provide a showcase (section 3), and present the evaluation results (section 4).

5.2 Design of interactive protocols

5.2.1 Context of the design

The preparative e-learning material that focuses on protocol steps was designed for the second-year bachelor course 'Food Chemistry', taught at Wageningen University, the Netherlands. The course is an introduction to the chemistry of compounds present in food, and is attended by approximately 200 students every year. During the laboratory classes of this course, students learn about and execute methods that are relevant to analyse food compounds.

Since the learning material should be accessible for many students at the same time, it was required to be online and interactive. With these requirements, the learning material provides students with specific feedback, and a large group of students will be able to complete the assignment at the same time, with minimal supervision.

5.2.2 Design choices

Following the design goal and the requirements, interactive protocols were designed. This was done by enriching existing protocols with closed-ended questions. All closed-ended questions are interactive, which means that students can be provided with answer-specific feedback. Four types of questions were designed: theoretical questions, practical questions, troubleshooting questions, and data processing questions. An overview of the question types, with corresponding aim and example of a learning outcome is provided in Table 5.1.

Table 5.1: Overview of the question types, with corresponding aim and example of a learning outcome.

Question category	Aim	Example of learning outcome
Theoretical	To improve students' understanding of the 'why' (reason or theory) of protocol steps	Understand which components of a sample end up in the pellet, and which components end up in the supernatant upon centrifugation
Practical	To increase students' awareness of practical choices to be made in protocol steps, and to practice making such choices	Be able to choose the appropriate glasswork
Troubleshooting	To help students combine and apply theoretical and practical knowledge to identify and solve (potential) mistakes	Be able to come up with an approach when the pH of a protein solution was increased too much
Data processing	- To improve students' ability to process raw data into results. - To support students' understanding of the reason of protocol steps in which data is gathered (e.g. a weighing step).	Be able to calculate the recovery of protein after isolation

In the following sections, we will first discuss when to include each type of question, after which we elaborate on how to design such questions.

When to include theoretical and practical questions in protocol steps?

The choice to enrich a protocol step with interactive theoretical and/or practical questions is based on the often implicit learning outcomes related to protocol steps, and on motivational aspects. In terms of the learning outcomes, not all protocol steps can contribute to improving students' understanding of the 'why' (reason or theory) of protocol steps or require practical choices to be made. For example, the protocol step: "Place the Dumas aluminium sample cups in the sample tray of the Dumas apparatus", did in our case not contribute to the learning outcomes, so we did not include any question for this step. Consequently, such protocol steps are also irrelevant to focus on in troubleshooting or data processing. In terms of motivational aspects, a selection of the interactive questions is prone to repetition. For example, a practical question on how to use a centrifuge may be applicable multiple times in one protocol, and may be applicable to several protocols within the learning material. While we acknowledge repetition as a key learning aid, we argue that repetitions should be limited to prevent frustration. For this reason, the maximum of two questions on the same topic were included, for the entire learning material.

When to include a troubleshooting scenario?

To help students combine and apply the theoretical and practical knowledge to identify and solve (potential) mistakes, troubleshooting scenarios were introduced. In a troubleshooting scenario, students are presented with a text describing a real-life scenario, followed by pictures of (intermediate) results and/or data resulting from the protocol. At some point in the scenario, an error has been incorporated. In corresponding questions, students are asked to identify the mistake and/or to indicate if the mistake can be fixed, and if so, how. Depending on the context, there were multiple follow-up questions that together form a complete troubleshooting scenario. The decision to include troubleshooting scenarios was based on teacher interviews, which yielded a list of potential mistakes with a significant effect on the outcome, and frequently asked questions by students. Based on this list, the troubleshooting scenarios were designed. In practice it turned out that some protocols are not enhanced with a troubleshooting scenario, while other protocols have multiple troubleshooting scenarios.

When to include data processing questions?

In case executing a protocol leads to the collection of raw data, students should also understand and be able to process this raw data into results (i.e. by doing calculations). For example, when students know they must do mass balance calculations, and understand how these calculations are done, they may better realize why it is important to accurately weigh their samples multiple times while executing the protocol. For this reason, and when appropriate, practice calculations with exemplary raw data were included in the interactive protocols.

How to design interactive closed-ended questions?

So far, it has been discussed when to enhance protocol steps with questions. In this section, the focus is on how to design such questions. To design the interactive closed-ended questions, several design principles from literature were applied (Clark and Mayer, 2016; Diederen, 2005; Van Merriënboer and Kirschner, 2018), as shown in Table 5.2. Note that these design principles are generally applicable to the design of any interactive closed-ended question.

Table 5.2: Design principles (DP) for interactive closed-ended questions, including explanation and application.

Code	Design principle	Explanation and application
DP1	Include only one new concept per question	To minimize unnecessary cognitive load, the content should be broken down into small segments (Clark and Mayer, 2016), which should be presented one by one (Bannert, 2002). Each closed-ended question may contain multiple concepts, but only one of them should be new.
DP2	Avoid including material that does not support the instructional goal	"People learn better when extraneous material is excluded rather than included" (Mayer, 2009a, p. 89). All text and graphics that do not support the instructional goal, e.g. background information, or unneeded variables, should be avoided (Clark and Mayer, 2016).
DP3	Force students to act	In order to trigger students to actively process the information, the interactive protocol requires students to answer all questions correctly before they can proceed.
DP4	Provide feedback	"Providing feedback is an ongoing process in which teachers communicate information to students that helps them better understand what they are to learn, what high-quality performance looks like, and what changes are necessary to improve their learning" (Dean et al., 2012, p. 3). Answer-specific feedback should be given in all closed-ended questions, which should help students to complete the questions independently.
DP5	Provide calculation hints	To guide students' thinking process when doing multi-step calculations, they can be provided by hints. The hints can be accessed one by one, so that students can choose to use only part of the support. The hints, together with the feedback (DP4), will take away many questions that would otherwise have to be answered by a teacher.
DP6	Use different question types	As a motivational element, the interactive protocols use a range of question types such as multiple choice, drag and drop, and fill in the blank questions.

5.3 Showcase of the interactive protocols

The interactive protocols were built in a platform called LabBuddy (Kryt b.v., Wageningen, the Netherlands), and are illustrated in this section. Figure 5.1a shows the first part of a protocol, including the interactive closed-ended questions. Initially, all questions are collapsed to provide a good overview of the protocol. When a student clicks on 'answer question', the question will open within the protocol step. Figure 5.1b shows an opened question containing hints that must be opened one by one, and shows an example of the feedback provided when a correct answer is submitted.

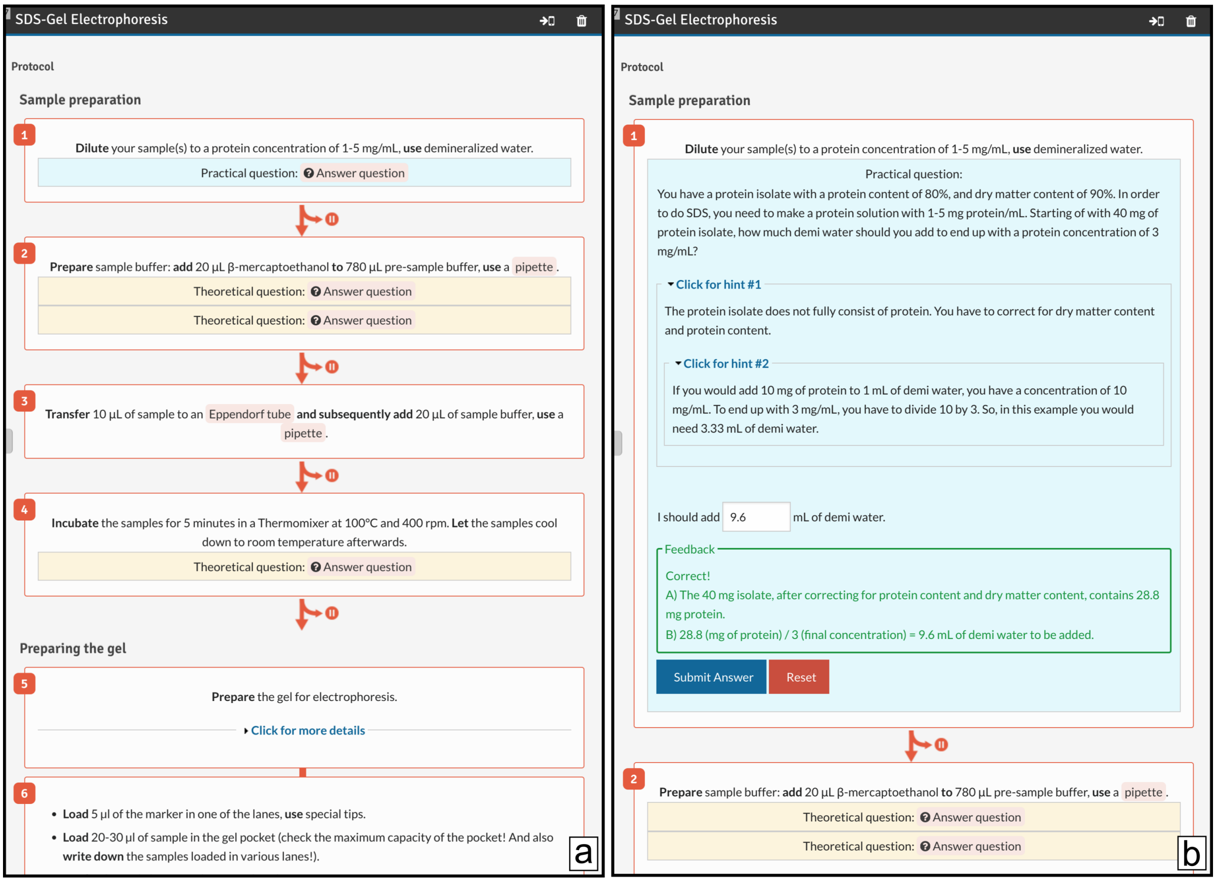

Figure 5.1: a) The first part of an interactive protocol with links that will open the question once clicked on it; b) Opened question in the first protocol step, with visible hints and feedback.

5.3.1 Examples of theoretical and practical questions

In Figure 5.2, examples of theoretical questions (Figures 5.2a, b, and c) and practical questions (Figures 5.2d, e, and f) are shown.

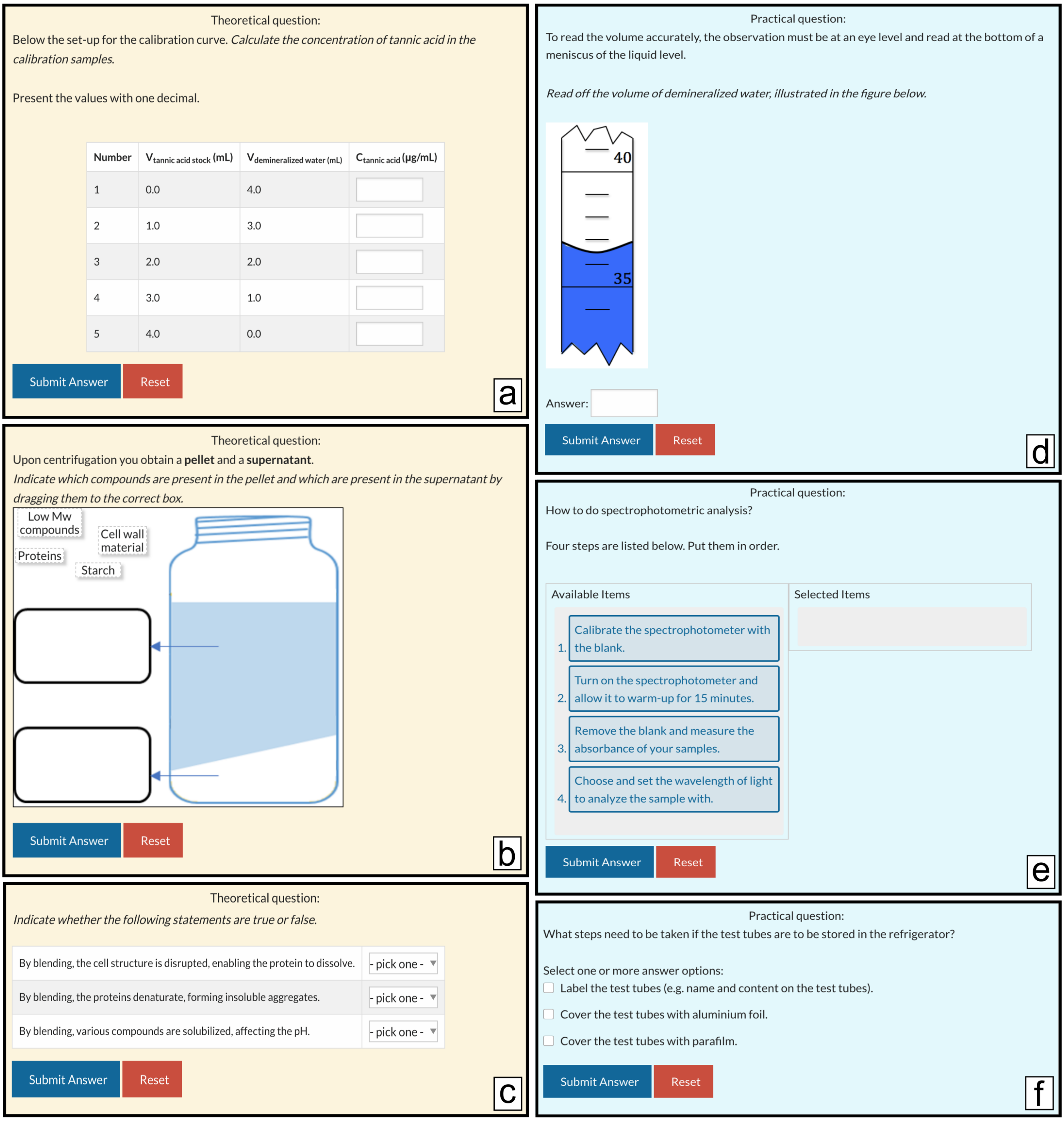

Figure 5.2: Examples of theoretical questions (a, b, and c) and practical questions (d, e, and f).

5.3.2 Examples of questions in troubleshooting scenarios

In Figure 5.3, two troubleshooting questions are shown. The first question describes a situation in which a fellow student loses his/her patience and adds too much NaOH. Then, through this question, advice should be given on how to solve the situation (Figure 5.3a). The second question is part of a scenario in which a student made a poor calibration curve, asking whether he/she will have to redo the calibration curve alone, or also the samples (Figure 5.3b).

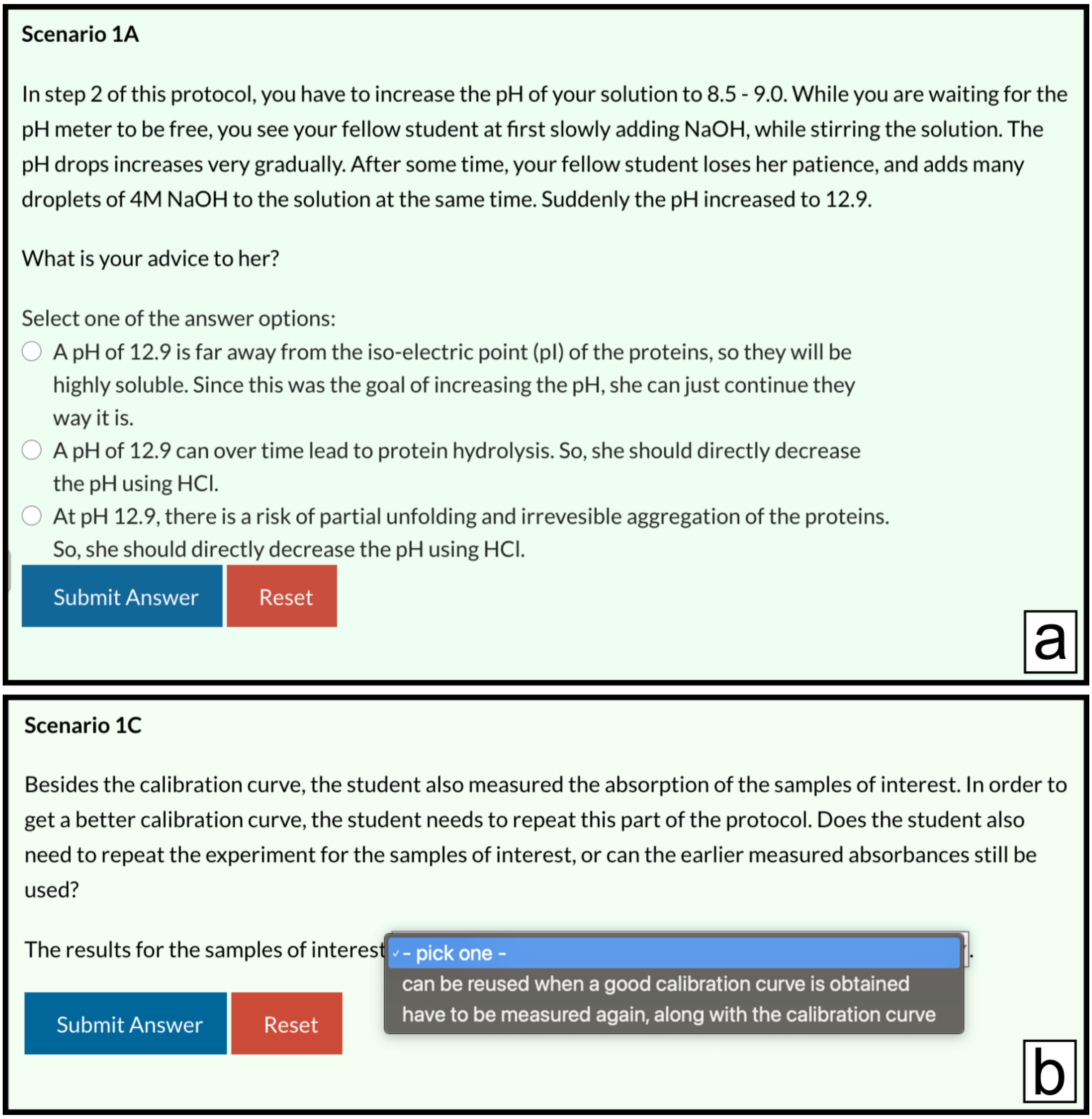

Figure 5.3: Two examples of troubleshooting questions.

5.3.3 Example of a data processing question

In this example, a seven-step data processing question in which students calculate protein recovery for two samples is shown (Figure 5.4a). All intermediate and final calculations can be submitted to the system, which will subsequently indicate for each answer whether it is correct. Furthermore, a worked example, including a visual, is provided as a hint (Figure 5.4b).

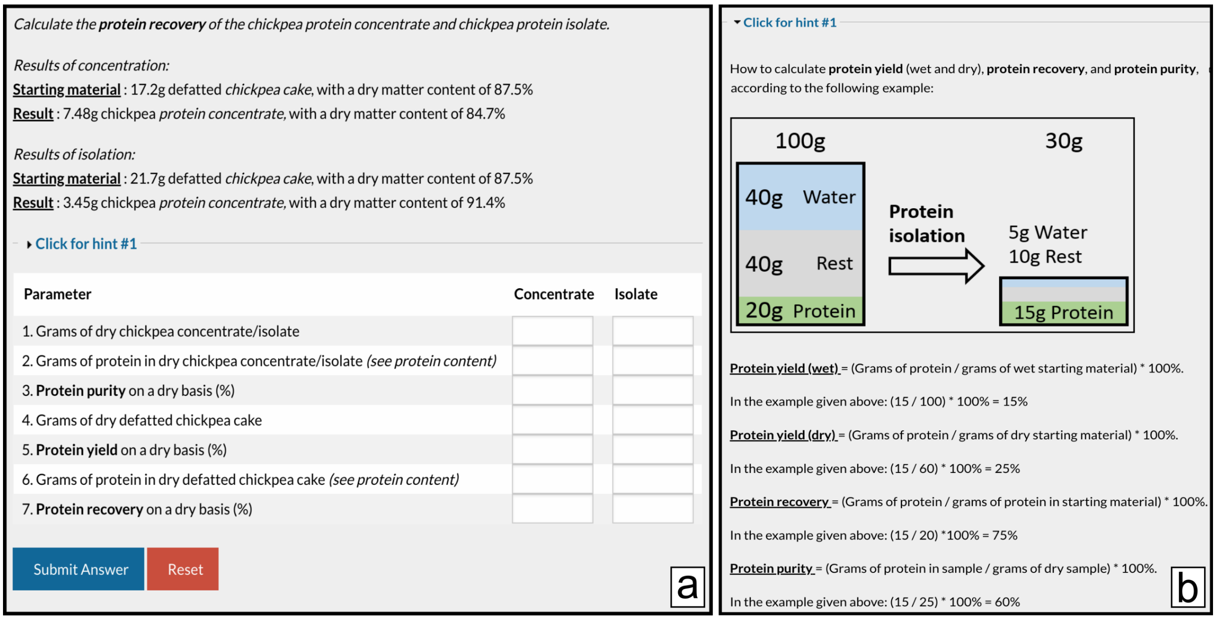

Figure 5.4: a) Example of a data processing question; b) The opened hint, being a worked example with visual.

5.4 Evaluation of the interactive protocols

The designed learning material consisted of seven interactive protocols, and were evaluated in terms of the design goal: to help students to i) understand the 'why' (reason or theory) of protocol steps, ii) make practical choices and iii) do troubleshooting. The seven evaluated protocols consist out of a total of 55 protocol steps, which were enriched with 27 theoretical and 11 practical questions. Five troubleshooting scenarios were introduced, which comprised a total of 12 questions. In four protocols, students obtain raw data that must be processed into results by doing calculations. In those protocols, students can enter a total of 36 (intermediate) values in the corresponding calculation questions.

5.4.1 Participants

All students (N = 200) were enrolled in the bachelor level course Food Chemistry (168 study hours). The course was attended by students who were enrolled in the bachelor study Food Technology (66%), Biotechnology (10%), or another bachelor programme (24%). Students were on average 20.2 (*SD* = 1.8) years old, 59% of the participants was female, while the others were male, and 82% of the participants were Dutch, while the rest had other nationalities.

5.4.2 Procedure

All student activities related to the evaluation procedure are shown in Table 5.3.

Table 5.3: Student activities related to the evaluation of the interactive protocols.

Code	Student activity	Measurement	Day*	Duration
A1	Preparative assignment	N/A	1	2 h
A2	Interactive protocols	Student behaviour**	2	4 h
A3	First questionnaire	Evaluation of design goal	2	5 min
A4	Laboratory classes	N/A	5 - 9	20 h
A5	Second questionnaire	Evaluation of design goal	9	5 min

* Running days relative to the start of the experiment.
** This has been elaborated in chapter 6.

Before students started working in the interactive protocols, they first engaged in a preparative assignment (A1). In this assignment, students designed the experiments for the laboratory work that was scheduled for the week after. The preparative assignment and the interactive protocols (A2) were scheduled as compulsory activities prior to the start of the laboratory classes. Upon completion of the interactive protocols, students were asked to complete the first questionnaire (A3), which led to N = 185 responses. The aim of this questionnaire was to evaluate each of the four question types (theoretical, practical, troubleshooting, and calculation questions) before students started the laboratory work. All questions had a 5-point Likert scale (1 = disagree, to 5 = agree). On the last day of the laboratory classes (A4), students were asked to complete the second questionnaire (A5), which led to N = 190 responses. In this questionnaire, students were once again asked to reflect on the usefulness of the theoretical, practical, troubleshooting, and data processing questions. Only the results from students who filled in both questionnaires (N = 172) were included in the analysis.

5.4.3 Evaluation: was the design goal achieved?

The questionnaire results that were used to evaluate whether the design goal was achieved are shown in Table 5.4. The design goal was threefold: to help students to i) understand the 'why' (reason or theory) of protocol steps, ii) make practical choices and iii) do troubleshooting.

Table 5.4: Combined results of the first (before lab work) and second (after lab work) questionnaires. The values represent the percentage of students ($N = 172$) who selected the corresponding answer option. Shading was used to visualize the distribution of the results. All questions (Q) had a 5-point Likert scale (1 = disagree, to 5 = agree).

Code	**Short description**	**Timing of question**	**1**	**2**	**3**	**4**	**5**	**4+5**
Q1.1	Theoretical questions - understand protocol steps	Before lab work	0	0	9	77	14	91
Q2.1	Level of understanding of theory behind experiments	After lab work	1	8	18	44	29	73
Q1.2	Practical questions - aware of practical choices	Before lab work	1	6	14	53	26	79
Q2.2	Practical questions - helped to make practical choices	After lab work	2	14	16	48	20	68
Q1.3	Troubleshooting - will make fewer mistakes	Before lab work	1	3	26	52	18	70
Q2.3	Troubleshooting - aware of (potential) mistakes	After lab work	6	53	26	11	4	15
Q1.4	Data processing questions - Confidence in processing data	Before lab work	0	2	15	69	14	83
Q2.4	Data processing questions - Was useful to process the data	After lab work	1	3	20	48	28	76

In terms of the first design subgoal (help students to understand the 'why' (reason or theory) of protocol steps), 91% of the students indicated that the theoretical questions in the interactive protocols helped them to understand the reason behind individual protocol steps (Q1.1). After the laboratory classes, a more varying, and overall slightly lower level of understanding was reported (Q2.1).

In terms of the second design subgoal (to help students make practical choices), the results show that 79% of the students indicated that the practical questions helped them become aware of the practical choices that must be made during the laboratory work (Q1.2). More than two-thirds (68%) of the students indicated that the awareness they gained by answering the practical questions in the interactive protocols helped them to make practical choices during the laboratory work (Q2.2).

In terms of the third design subgoal (help students to do troubleshooting), 70% of the students indicated that the knowledge they gained by answering the troubleshooting questions would contribute to making fewer mistakes during the laboratory work (Q1.3). However, after the laboratory work, only 15% of the students reported that the troubleshooting questions made them more aware of (potential) mistakes during the laboratory work (Q2.3).

Being aware of (potential) mistakes during laboratory work might be a bridge too far for many second year BSc students. For students to be aware of (potential) mistakes, they should first have a full understanding of the protocol steps. Second, they should be aware of the level of their own skills. Third, they should be able to reflect on their skills while executing a protocol in a cognitively demanding laboratory setting. Last, students should also be willing to think about potential mistakes, while it has been reported that students' focus is typically on completion of the task in the laboratory as quickly as possible (DeKorver and Towns, 2015). Since many students seem not yet able to or willing to do troubleshooting in the laboratory, it adds extra value to include troubleshooting scenarios to the interactive protocols, so that students will be able to practice troubleshooting.

In terms of data processing, 83% of the students indicated that they were confident that they would be able to process the raw data they would obtain during the laboratory work (Q1.4). This confidence remained after students had completed the actual calculations during the laboratory classes (Q2.4). In line with the results corresponding to the theoretical questions (Q1.1 and Q2.1), the high level of confidence suggests that students also had a good understanding of how to process the raw data obtained by executing the protocol.

5.5 Conclusions

The aim of this study was to design and evaluate a preparative learning material that focuses on protocol steps with the design goal to help students to i) understand the 'why' (reason or theory) of protocol steps, ii) make practical choices and iii) do troubleshooting. To reach this goal, protocol steps have been enriched with theoretical, practical, troubleshooting, and/or calculation questions to form interactive protocols. This design process resulted in a list of design choices (i.e. when to include a question, and how to design a question) and a showcase of questions in the interactive protocols. These interactive protocols were implemented and evaluated in a real educational setting. From the evaluation results it was concluded that the interactive protocols were successful in preparing students for the laboratory work. After the laboratory work, students reported a more diverse and slightly less positive contribution of the interactive protocols to their understanding of the protocol steps and their ability to make practical choices. A significant difference was found in the perceived usefulness of the troubleshooting questions before versus after the laboratory work, which is suggested as a topic for further investigation.

Chapter 6

Relations between students' perceived levels of self-regulation and their corresponding learning behaviour and outcomes in a virtual experiment environment

This chapter was published as:

S. Verstege, H. J. Pijeira-Díaz, O. Noroozi, H. Biemans, and J. Diederen (2019). "Relations between students' perceived levels of self-regulation and their corresponding learning behavior and outcomes in a virtual experiment environment". *Computers in Human Behavior*. DOI: https://doi.org/10.1016/j.chb.2019.02.020

Contents

Abstract

Virtual Experiment Environments (VEEs) have been shown as effective preparation steps for laboratory classes in natural science education. Given the self-directed nature of VEEs, students need adequate Self-Regulated Learning (SRL) skills. This study explores the relation between students' perceived SRL level and their behavior and outcomes in a VEE in the field of enzymology. Ninety-seven higher education students were divided into three groups of perceived SRL level (high, medium, and low). The VEE learning behavior (e.g. number of attempts and hints accessed) and VEE outcomes of these groups were compared while keeping prior knowledge as a covariate. While low self-regulated learners showed the least level of engagement with the VEE, high self-regulated learners showed the most optimum learning activity. Medium self-regulated learners engaged more in gaming the system behavior, and consequently learned the least. These results suggest that there is a nonlinear relationship between perceived SRL level and outcomes, since the intermediate level seems to be detrimental to learning, as explained through behavior. The intermediate level was characterized by an increase in agency, but a lack of goal-directed and planning behavior. Implications for self-regulated learning theory and the design of VEEs in the best interest of students are discussed.

6.1 Introduction and theoretical framework

6.1.1 Laboratory education

Laboratory education is a commonly used practice in natural science education and is generally regarded as a vital component of a chemistry curriculum (Bruck and Towns, 2013). Teachers have several goals for laboratory classes, including hands-on research experience, group work, broader communication skills, data collection and (error) analysis, planning, and laboratory report writing (Bruck and Towns, 2013; Kirschner and Meester, 1988). However, many of these goals have been argued to be better accomplished outside the cognitively demanding learning environment of the laboratory (Hartog et al., 2009).

While it is acknowledged that laboratory practicals play an indispensable role in the conveyance of tacit knowledge, concerns have been raised that not all intended learning outcomes may be achieved by the students, and that some laboratory practicals may even confuse or demotivate students (Kirschner and Huisman, 1998). Doubts on the quality of knowledge construction and learning gains from laboratory classes (Abdulwahed and Nagy, 2009; Hawkes, 2004; Hawkins and Phelps, 2013; Kirschner and Huisman, 1998; Kirschner and Meester, 1988) have been explained by their cognitively demanding nature (Buntine et al., 2007). During laboratory classes students need to plan their time, use (unfamiliar) equipment, find their way around the lab, remember all the safety rules, translate a protocol into action, etc. When performing experiments, students use most of their cognitive abilities for the act itself, the motor skills and the planning skills, acquiring mainly tacit knowledge about the experiments, while underexposing higher cognitive levels of scientific inquiry (Domin, 1999).

In addition, a number of practical challenges in laboratory education have been reported, including, but not limited to, ill-prepared students, difficulties with teaching assistants, high enrollments, diverse majors, a decline in resources and an increase in responsibilities (Bruck and Towns, 2013). The use of laboratories is increasingly costly (Hawkins and Phelps, 2013), and the effective and efficient use of time spent in the laboratory is a necessity for all educational institutions (Kirschner and Meester, 1988).

Careful preparation for laboratory classes, such as practicing with the chemical concepts and data analysis skills, has proven to increase the quality of knowledge construction during laboratory classes (Bannert, 2002; Crandall et al., 2015; Winberg and Berg, 2007). Virtual and simulated laboratories are valuable assets to prepare students for laboratory classes (Abdulwahed and Nagy, 2009), and are also a convenient and effective replacement of laboratory classes, especially in contexts where the laboratory is not an option such as distance education (Hawkins and Phelps, 2013).

6.1.2 Simulations and virtual experiment environments

In order to prepare students for laboratory classes, many different learning activities have been developed over the past decades, including pre-laboratory videos (Spagnoli et al., 2017), animations (Limniou and Whitehead, 2010), pre-laboratory quizzes, tutorials, (digital) assignments (Diederen et al., 2006a; Koehler and Orvis, 2003; Spagnoli et al., 2017; Van der Kolk et al., 2013), and simulations (Crandall et al., 2015; Limniou and Whitehead, 2010). Simulations, virtual laboratories, online laboratories or remote laboratories, have been referred to as non-traditional labs. An extensive review (Brinson, 2015, p. 218) on the difference in achieved learning outcomes between students in a non-traditional lab compared to traditional labs reported that "learning outcome achievement is equal or higher in a non-traditional lab versus traditional labs across all learning outcome categories (knowledge and understanding, inquiry skills, practical skills, perception, analytical skills, and social and scientific communication)". Another review of 17 studies on the effects of computer simulations as a supplement or alternative to traditional laboratory activities in science education concluded that traditional instruction can be successfully enhanced by using computer simulations (Rutten et al., 2012). In most of the 17 reviewed studies, improved learning outcomes were shown. Also, when simulations were used as a preparatory activity, positive effects were found for the comprehension of the lab task as well as for practical laboratory skills during the real lab activity (Rutten et al., 2012).

A simulation environment that can be used to prepare students for laboratory classes, or even (partly) replace them, is referred to as a Virtual Experiment Environment (VEE). A VEE is "any educational resource that enables students to design and/or carry out virtual experiments and/or to process data, and to analyze and interpret results" (Hartog et al., 2009, p. 376). A VEE can be a rich interactive learning environment in which students work on self-directed learning tasks. Students can, for example, design the experiments themselves, answer closed questions, ask for feedback or hints, seek additional information about the chemical methods including visual information such as videos, photos and animations, and process and make sense of data. In order to complete a VEE, students must adopt an active learning attitude and, accordingly, students need self-regulated learning skills.

6.1.3 Self-regulated learning

As educational psychology transitioned from behaviorism to cognitivism and constructivism, learners were no longer considered as passive, receptive subjects, but rather as subjects with agency (Bandura, 1986) responsible for their learning (Hoadley, 2018). The behavioural, motivational, and metacognitive active participation of students in their own learning processes is referred to as self-regulated learning (SRL) (Zimmerman, 1989). As students set their own learning goals, plan learning activities, execute tasks, adopt, enact, develop and refine strategies, monitor and control their learning process, and adapt it to meet their goals, they engage in SRL (Hadwin et al., 2011). SRL is inherent in goal-directed activity (Winne, 1997) and essentially a cyclical process (Zimmerman, 2015). Given SRL's wide scope, different models of SRL have been proposed in literature (for reviews, see Panadero (2017) and Winne (2015)). Although there is overlap between them, the models focus on different aspects of SRL. This article adopts Zimmerman's cyclical phases model, for its balance in cognition, emotion and motivation as facets of learning, and because it has been the most widely adopted according to its number of citations (Panadero, 2017). Zimmerman's model has been predominant as it offers a robust and complete vision of different types of subprocesses that articulate the interaction of students' personal and behavioural characteristics (Moos and Ringdal, 2012).

The cyclical phases model (Zimmerman, 2000; Zimmerman and Moylan, 2009) consists of three phases, namely forethought, performance and self-reflection. Each phase, in turn, is divided in two categories. The forethought phase comprises the categories of task analysis and self-motivation beliefs. During the task analysis, the students set their goals and strategically do their planning, which depends on motivational beliefs in relation to the task, such as self-efficacy (i.e. to what extent they feel capable of accomplishing the task), expectations, interest, and value. The categories of self-control and self-observation determine the performance phase. As part of self-control, students choose the strategies to approach the task, gather information, seek help, and manage the available time, among other activities. Meanwhile, self-observation essentially refers to monitoring their performance. Monitoring is central to the third phase, self-reflection, where students evaluate their performance and react to the evaluation by making adaptations to the forethought phase (e.g. adjusting the goals and modifying the plan), thus initiating another cycle of SRL. In sum, SRL is a cyclical process through which students take command of their own learning, stemming from task identification, planning learning activities, monitoring progress, diagnosing problems, testing their learning outcomes, adjusting, and reflecting (Schunk and Zimmerman, 2011; Zimmerman, 2000). SRL is thus expected to increase efficiency when the learner addresses analogous tasks in the future (Winne and Perry, 1994).

According to Vermunt and Donche (2017), the way students approach a learning activity such as a VEE depends, among others, on their learning patterns. A learning pattern is an interrelationship between the conception of learning, learning motivation, processing strategies, and regulation strategies. These learning patterns influence the learning outcomes of the students and are in turn influenced by contextual factors and personal factors. Regulation strategies are a central element of the learning patterns. For example, students who have a meaning-directed learning pattern, learn in a self-regulated way, not limiting themselves to prescribed materials through consulting literature and other sources.

6.1.4 SRL in relation to outcomes and behavior

Students who set superior goals proactively, monitor their learning intentionally, use strategies effectively, and respond to personal feedback adaptively not only attain mastery more quickly, but also are more motivated to sustain their efforts to learn (Zimmerman, 2013). Students' beliefs about their efficacy to regulate their academic learning activities have been found to be highly predictive of their academic achievement (Zimmerman and Schunk, 2003). The personal capabilities that enable students to develop as independent, self-regulated learners are highly related to their achievement (Chen, 2002; Zimmerman and Pons, 1986).

Several SRL strategies have been listed, of which metacognition, time management, effort regulation, and critical thinking were found to have higher associations with academic achievement (Broadbent and Poon, 2015; Richardson et al., 2012; Zimmerman and Pons, 1986). Most prior research has focused on the relation between regulation and academic achievement in traditional classrooms (Boekaerts and Corno, 2005; Moos and Ringdal, 2012; Vermunt and Donche, 2017).

More recently, the increasing availability and use of online learning environments (e.g. massive open online courses (Kay et al., 2013) and learning management systems (Coates et al., 2005)) are leading to a growing interest in the exploration of the interrelation among students' SRL level and behavior and outcomes in these environments, given the increased autonomy they confer upon students (Kizilcec et al., 2017; Tempelaar et al., 2015). Students' behavior in online learning environments has been measured through a variety of indicators based on trace data, seamlessly recorded by these environments. The behaviors chosen for analysis have customarily been determined by the characteristics and features of the environments (Lowes et al., 2015). Commonly explored students' behaviors in online learning environments include number of logins, time on task, number of discussion forum posts viewed and authored, number of assignments completed, and number of accesses to different course pages (Gašević et al., 2016; Kovanović et al., 2015; Tempelaar et al., 2015). Although, in general, higher levels of engagement in online learning environments have been associated with better learning outcomes (Lowes et al., 2015), the findings in terms of the predictive power of such behavioural variables on students' learning outcomes have been mixed, spanning all the spectrum from weak to strong (Gašević et al., 2016; Tempelaar et al., 2015). The study of these relations using massive open online courses has enabled big sample sizes, in the order of hundreds (e.g. Tempelaar et al., 2015) or even thousands (e.g. Gašević et al., 2016) of participants. However, other types of online learning environments have different characteristics than massive open online courses and require specific research, since the characteristics and purpose of the environment strongly determine the students' behavior (Gašević et al., 2016; Winne and Hadwin, 1998). This paper focuses on a particular type of online learning environment (a VEE) designed to enhance students' learning outcomes from their laboratory practices. To date, little is known about the relation between students' SRL skills with their learning behavior and outcomes in the context of a VEE, which is precisely what this research addresses.

6.1.5 Aim of the study

The aim of this study was to explore differences in learning behavior, in terms of how often students access multiple forms of information included in a VEE, and learning outcomes between groups of students with different perceived (i.e. self-reported) levels of self-regulation. The following research questions were formulated:

1. What is the relationship between students' perceived SRL level and their VEE behavior?
2. What is the relationship between students' perceived SRL level and their VEE learning outcomes?
3. What is the relationship between students' VEE behavior and VEE learning outcomes according to their perceived SRL level?

6.2 Method

6.2.1 Participants

This study was conducted at a university in the Netherlands. All participants (N = 97), a group of students with heterogeneous educational backgrounds, were enrolled in a 168-h MSc level course related to enzymology. All students had previously obtained a BSc degree in natural sciences: 39% from the Netherlands, 56% from a university outside the Netherlands, and 5% was unknown. All students were enrolled in an MSc study program, being Food technology (60%), Biotechnology (33%), or other (7%). Each MSc study program is subdivided in several specializations. The prior knowledge of students was diverse because of the aforementioned differences in prior education, MSc study program, and specialization. The students were 23.3 years old (SD = 2.4) on average. About 63% of the students was female and the rest were males.

6.2.2 Independent and dependent variables

The independent variable in this study was the self-reported or, in other words, perceived level of SRL. The dependent variables were students' VEE behavior (i.e. attempts to complete preparation questions, number of clicks, feedback on experimental design, background information, attempts for simulation questions, answers requested, and hints accessed) and VEE learning outcomes in terms of content-specific knowledge.

Students' prior knowledge was included as a covariate in the data analyses to be able to control for differences in this regard.

6.2.3 The LabBuddy VEE

The VEE was purposefully developed for the course as a preparation for the laboratory classes. The VEE is an extended version of a previously developed web-based experiment designer and web laboratory manual (Van der Kolk et al., 2012), nowadays known as LabBuddy (Kryt b.v., Wageningen, the Netherlands) (Van der Kolk et al., 2013). Figure 6.1 depicts four screenshots of the LabBuddy VEE, including some annotated functionality.

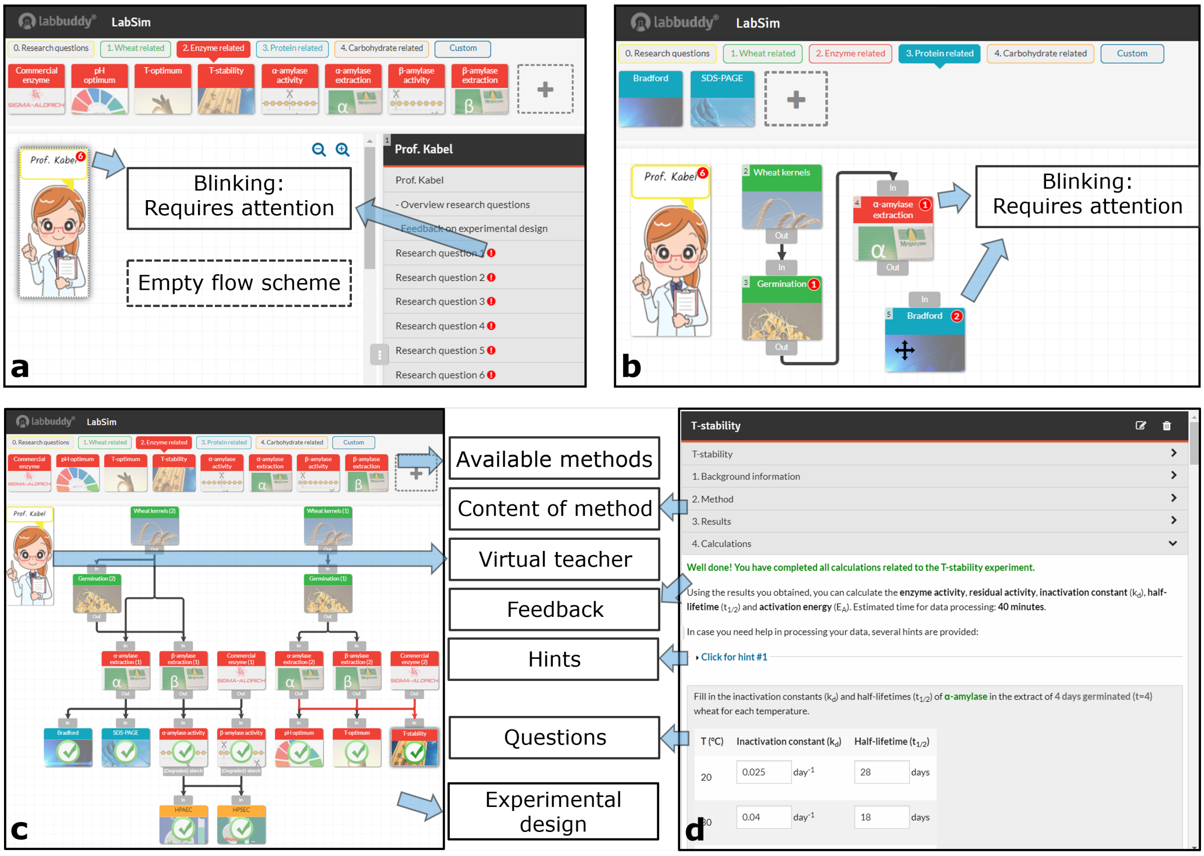

Figure 6.1: LabBuddy VEE screenshots. a) Starting situation. b) Situation where three chemical methods have been added to the flow scheme, and the fourth ("Bradford") is currently being dragged into the flow scheme. c) Example of a completed flow scheme. d) Example of the content of a method.

Figure 6.2 depicts a schematic overview of the LabBuddy VEE procedure including labels of the corresponding students' VEE behavior (i.e. B1 through B7), which are described later on in Table 6.2. Students are provided with preparation questions (B1)

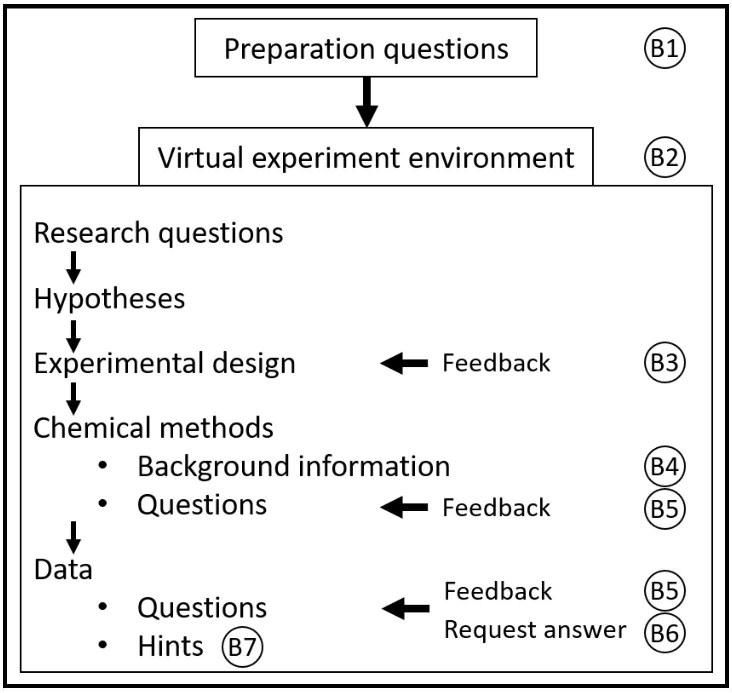

Figure 6.2: Flow scheme of student activities: preparation questions and virtual experiment environment. Labels B1 – B7 indicate where student behavior was measured, as further explained in section 6.2.4.

before entering the VEE, to ensure that they have the required basic knowledge needed to understand the topic of the VEE.

This VEE (B2) contains a self-directed learning task through which students can actively learn and apply concepts related to the laboratory class. The task consists of answering six research questions, which are provided by a virtual teacher (Figure 6.1a). To start off, students define a hypothesis on each research question. Next, they make an experimental design in the form of a flow scheme (Figures 6.1b and 6.1c), where each block corresponds to an available chemical method. In other words, students must select and connect the chemical methods to each other in a logical sequence. To verify the correctness of the flow scheme, students can request feedback from the virtual teacher (B3). In case the flow scheme is incorrect, the virtual teacher guides them towards the correct solution. Each chemical method, which must be further opened by students, contains the following tabs (Figure 6.1d): 1) background information, 2) method (containing the detailed lab protocol), 3) results (containing raw data), and, if applicable, 4) calculations. The background information tab often contains, next to general information about the method, a collapsible that students can open for more detailed information (B4) on the method. To motivate students to go through the information, the background information and method tabs incorporate one or more closed questions (B5). All questions belonging to one method must be answered correctly before any results can be unlocked. The results obtained from the VEE are directly affected by the flow scheme made by the students. For example, in case the flow scheme is incomplete, an incomplete dataset is provided by the VEE. Simulating real laboratories, all virtual chemical methods yield raw data, which students must process and perform calculations on using for example Microsoft Excel. If a student cannot find the correct answer to a calculation, he/she can request it from the VEE (B6), after which the VEE provides him/her with the correct answer. Students can also opt to access hints (B7) that help them while processing the data and/or performing the calculations. Depending on the nature of the question, up to four hints are available per question, and 28 hints in total for the whole simulation. The results of the calculations must be entered in the calculations tab. Once all prerequisites to answer a research question are fulfilled, students can consult the virtual teacher to interpret their results, do further calculations if necessary, and finally accept or reject their hypotheses. Whenever there is something that requires the students' attention (e.g. feedback or an unanswered question), a blinking signal (Figures 6.1a and 6.1b) will appear to guide students in their progress.

6.2.4 Procedure

The course in which the VEE was used, took place over a period of eight weeks. The experiment started at the beginning of the second week. Student activities in relation to the VEE with corresponding measurements and time schedule are shown in Table 6.1. Attendance was compulsory for all activities. Before entering the VEE, students were given 30 minutes to complete a pre-knowledge test (A1) on paper. The pre-knowledge test consisted of 13 open questions, covering all content-specific learning outcomes of the VEE. Each correctly answered question was rewarded with one point, yielding a range of scores for students between zero and 13 points. The assessment of the pre-knowledge test was done by a domain-related expert, who used a detailed key. The key included how many points should be awarded for specific answers. A randomly selected sample ($N = 10$) was checked by another domain-related expert to make sure that the assessment was accurate and reproducible. The scoring by both experts was identical. Most of the students ($N = 93$) completed the pre-knowledge test, while four students were absent that day. The scores of this test ($M = 5.1$, $SD = 2.3$) were used as a measurement of their content-specific prior knowledge.

After the pre-knowledge test, students started working individually on 17 closed questions on the computer, which served as preparation (A2) for the VEE. Once the preparation questions were completed, a link was provided to access the VEE (A3). Students used the VEE for an average of 7.5 hours distributed over the first three days of the start of the experiment. Computer rooms were available and three carefully instructed supervisors were present to answer possible questions of students. However, due to the self-explanatory nature of the VEE, only few questions were raised by the students. Alternatively or additionally, students were able to access the VEE from any

Table 6.1: Student activities with corresponding measurement(s) and time indications.

Code	Student activity	Measurement	Day*	Duration
A1	Pre-knowledge test	content-specific prior knowledge	1	30 min
A2	Preparation questions	student behavior	1	60 min
A3	VEE	student behavior	1-3	7.5 h
A4	Post-knowledge test	content-specific knowledge	9	30 min
A5	Laboratory classes	N/A	9-12, 15-19	27 h
A6	Survey	perceived SRL level	26	15 min

*Running days relative to the start of the experiment. There were no course activities in between the VEE and the post-knowledge test.

device connected to the Internet at any time over the period established for VEE usage (i.e. days one to three). Students took advantage of this possibility, as 19% of the digital traces were logged outside of the scheduled computer room time of days one and two of the experiment. After the scheduled time of day three, all students had completed the VEE.

The VEE was intended to prepare students for the laboratory classes. Accordingly, a post-knowledge test was conducted (A4) just before the start of the laboratory classes to determine the learning outcomes of the VEE. This test consisted of 13 open questions slightly different than the pre-knowledge test. The assessment procedure of the post-knowledge test was the same as for the pre-knowledge test. Four students were absent on the post-knowledge test day, and therefore did not complete it. The scores (M = 7.2, SD = 2.6) were used as a measure of the content-specific knowledge the students (N = 93) had at the time of initiating their laboratory practice.

There were no other course activities in between the pre- and post-knowledge test, other than the VEE. Students were not aware that a post-knowledge test would follow, to prevent targeted studying for it. The validity of these self-made knowledge tests was obtained through three experts including the main teacher of the course. A survey on perceived SRL level (A6) was performed at a later date, independent from the VEE.

Measurement of students' VEE behavior

Students' behavior in the LabBuddy VEE (Figure 6.2, labels B1 – B7) was measured by digital trace data, which is most commonly aggregated by frequency counts in studies linked to SRL (Hadwin et al., 2007; Van Laer and Elen, 2018). Therefore, aligned with the literature and particularizing for the LabBuddy VEE, the seven aspects of students' behavior measured are described in Table 6.2.

The focus was clearly on count measures, while time-on-task measures have been excluded on the twofold basis of measurement challenges and relevancy, which is

Table 6.2: Explanation of the student behavioural aspects analyzed.

Code	Behavioural aspect
B1	Number of extra attempts to complete the preparation questions (= total number of attempts minus minimal attempts possible for completion (17))
B2	Number of meaningful clicks in the simulation
B3	Number of times student accessed the virtual teacher for feedback on the experimental design
B4	Number of unique background information units accessed
B5	Number of extra attempts to complete all questions in the simulation (= total number of attempts minus minimal attempts possible for completion (49))
B6	Number of times a student requested the answer to a calculation
B7	Number of unique hints accessed

elaborated on next. Kovanović et al. (2015) conducted a thorough study of 15 different time-on-task estimation strategies, and how they compare to count measures in explaining differences in students' learning outcomes. They conducted experiments in two different settings: blended learning and fully online learning. Based on their findings, they concluded that "with all the challenges in accurate estimation of time-on-task, given the off-task behaviours, [...] perhaps using time-on-task measures should be reconsidered and counts measures be more promoted" (Kovanović et al., 2015, p. 104). Still, they found time-on-task to be an important variable to consider in fully online learning settings, but not in blended learning, which is the case in our study. This was explained by evidence that "the relative amount of activity per student is much higher in the fully online course" (Kovanović et al., 2015, p. 103).

Further clarification needed in relation to some of the count measures used follows. Meaningful clicks (B2) were considered those necessary to carry out the VEE assignment, including use of the support provided by the environment. For example, when a student clicked to open a hint within an analysis method, the click was added to the count of meaningful clicks, but the click to close it afterwards was not counted. This is because it was optional to close a tab or hint, and after switching to, for example, another analysis method, all previously opened hints were closed automatically. Moreover, access to a specific background information unit (B4) or hint (B7) was only logged the first time it was accessed.

Measurement of perceived SRL level

The self-regulation scale of the Inventory of Learning Styles (ILS) (Vermunt, 1998), a questionnaire for higher education, was used to measure students' self-reported SRL level (A6). ILS was chosen, because the ILS inventory was developed based on qualitative and quantitative empirical studies with university students and has been applied and validated in many studies ever since (Vermunt and Donche, 2017). The scale consists of seven items on self-regulation strategies, such as "When I have difficulty grasping a particular piece of subject matter, I try to analyze why it is difficult for me" and "I do more than I am expected to do in a course". Students rated each item on a five-point Likert scale ranging from "I do this seldom or never" to "I do this almost always". The reliability and validity of these scales have been reported as adequate in various contexts (Roth et al., 2016; Vermunt and Vermetten, 2004). In this study, the reliability coefficient for the self-regulation scale of the ILS was satisfactory (*Cronbach's* α = .80). A small number of participants (N = 8) was not considered for the analysis since five students were absent and three failed to complete the questionnaire. Based on the scale scores of those who completed it, students were subdivided into three groups of approximately equal size: low self-reported SRL level (N = 31, M = 2.0, SD = 0.4), medium self-reported SRL level (N = 30, M = 2.8, SD = 0.1), and high self-reported SRL level (N = 28, M = 3.4, SD = 0.3). This categorization resulted in statistically highly significantly different groups (F (2, 86) = 128.16, $p < .001$, η^2 = .75).

6.2.5 Analysis

Univariate analyses of variance were used to find out whether the three self-regulation groups were significantly different in terms of their VEE behavior, and significantly different in terms of their VEE learning outcomes as indexed by the post-knowledge test results. The results of the post-knowledge test were corrected for variance in prior knowledge by considering prior knowledge as a covariate. This method was justified and chosen over a repeated measurement analysis, since the pre- and post-knowledge questions tested the same concepts but were not identical. If a significant overall difference between the three groups was found, repeated contrast analyses were carried out to examine statistical differences among the various groups.

A correlation analysis for each perceived SRL level was performed to investigate the relationship of the VEE learning outcomes to the students' prior knowledge and their VEE behavior.

6.3 Results

6.3.1 Perceived level of SRL in relation to VEE behavior

The results for the behavior of students with high, medium, and low level of perceived self-regulation in the preparation questions and the VEE are shown in Table 6.3.

Table 6.3: Quantitative data on each behavior variable, grouped by perceived self-regulating ability.

Code	Variable	SRL	M	SD	F	Sig	η^2
B1	Attempts preparation questions	Low	15.67[a]	9.25	3.13*	.05	.07
		Medium	24.68[b]	16.02			
		High	21.62	11.23			
		Total	20.51	12.89			
B2	Number of meaningful clicks	Low	1154.20[a]	375.66	4.10*	.02	.09
		Medium	630.21[b]	744.77			
		High	482.00	605.49			
		Total	414.33	617.89			
B3	Feedback on experimental design	Low	7.73	5.92	0.13	.88	.00
		Medium	8.11	5.19			
		High	8.54	7.08			
		Total	8.11	6.02			
B4	Background information	Low	5.20	3.27	0.80	.45	.02
		Medium	5.71	2.62			
		High	5.81	3.29			
		Total	5.56	3.05			
B5	Attempts simulation questions	Low	74.17	31.36	0.76	.47	.02
		Medium	87.29	26.99			
		High	81.04	37.40			
		Total	80.67	32.12			
B6	Answer requested	Low	23.90[a]	20.32	5.27*	.01	.12
		Medium	47.50[b]	29.83			
		High	28.77[a]	19.95			
		Total	33.27	25.69			
B7	Hints accessed	Low	18.60	5.30	0.95	.39	.02
		Medium	20.96	4.13			
		High	19.92	5.84			
		Total	19.80	5.16			

* Significant at the .05 level.
[a] Self-regulation group is significantly different from [b] in terms of the behavior variable.

The attempts to complete the preparation questions (B1) were found to be significantly lower (F (3, 80) = 3.13, p = .05, η^2 = .07) for students with a low level of SRL (M = 15, SD = 9) compared to students with a medium level of SRL (M = 25, SD = 16). The total number of meaningful clicks in the simulation (B2) was found to be significantly lower (F (3, 80) = 4.10, p = .02, η^2 = .09) for students with a low level of SRL (M = 1154, SD = 376) compared to students with a medium level of SRL (M = 1630, SD = 744). The total number of answers requested (B6) was found to be significantly higher (F (3, 80) = 5.27,

p = .01, η^2 = .12) for students with a medium level of SRL (M = 48, SD = 30), compared to students with a low level of SRL (M = 24, SD = 20) and students with a high level of SRL (M = 29, SD = 20). No significant differences were found between the three levels of SRL and the other behavioral aspects including the number of times students accessed the virtual teacher for feedback on the experimental design (B3), the number of unique background information units accessed (B4), the number of extra attempts to complete the VEE (B5), and the number of unique hints accessed (B7).

6.3.2 Perceived level of SRL in relation to VEE learning outcomes

The descriptive statistics of the pre- and post-knowledge tests are shown in Table 6.4. A univariate analysis of variance showed that the three perceived self-regulation groups were significantly different in terms of students' VEE learning outcomes as measured by the post-knowledge test (F (2, 77) = 5.36, $p < .01$, η^2 = .12). The follow-up repeated contrast analysis showed a statistically significant difference (p = .02) in VEE outcomes between students with low level of self-regulation and medium level of self-regulation. There was also a statistically significant difference ($p < .01$) in the post-knowledge test between students with medium and high level of self-regulation. The sample size for this statistical test was 81, which was the number of students who completed both the survey and the post-knowledge test.

Table 6.4: Descriptive statistics of knowledge tests by perceived SRL level.

Perceived SRL level	Pre-knowledge test		Post-knowledge test	
	M	SD	M	SD
Low	5.4	2.4	7.8	2.4
Medium	4.3	1.7	5.8	2.4
High	5.8	2.3	8.3	2.0

6.3.3 VEE learning outcomes in relation to VEE behavior, prior knowledge, and perceived SRL level

The results of the correlation analyses for each perceived SRL level are shown in Table 6.5. Many of the correlations were statistically significant, and moderate to strong. The VEE learning outcomes showed positive strong correlation to the prior knowledge for the low and medium groups of perceived SRL level, and moderate positive correlation with the high group. In general, the VEE behavior correlated negatively with the learning outcomes.

Table 6.5: Correlation of VEE learning outcomes to prior knowledge and VEE behavior by perceived SRL level.

		VEE behavioral variables						
Perceived SRL level	Prior knowledge	B1	B2	B3	B4	B5	B6	B7
Low	.74**	-.29	-.60**	.03	-.45	-.48**	-.46*	-.42*
Medium	.69**	-.41*	-.28	.16	-.20	-.25	-.53**	.01
High	.41*	-.62**	-.46	-.07	-.18	-.46**	-.43*	-.46*

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

6.4 Discussion

6.4.1 Perceived level of SRL in relation to VEE behavior

A clear pattern emerged for the three VEE behavioral variables with significant differences in terms of perceived SRL levels, namely the total number of attempts to complete the preparation questions (B1), the number of meaningful clicks (B2), and the number of times the answer was requested (B6). The average values for the three variables were ascending for low, high, and medium SRL levels, in that order (i.e. low $<$ high $<$ medium). Low self-regulated learners showed the lowest level of activity, engaging the least with the VEE and its resources. Aligned with the SRL theory, low self-regulated learners thus proved to be the least agentic and least active group, with students taking minimum control of their learning process (De Bruin and Van Merriënboer, 2017). At the following level in the SRL hierarchy, the VEE activity of medium self-regulated learners peaked. It is known that SRL skills develop with time and practice (Panadero, 2017). Our findings suggest that, as students develop their SRL skills from low to medium, they go from the lowest to the highest measured level of activity and use of learning resources at their disposal. Their agency and willingness to own their learning process increase, but their use of strategies is not mature, and their behavior shows low level of planning (too much unnecessary activity). Finally, as the students evolve from medium to high levels of SRL, their activity decreases to a more optimum level in between that of low and medium self-regulated learners. The findings are thus consistent with the current body of SRL knowledge (cf. Bjork et al., 2013; Zimmerman, 1998), which characterizes the behavior of highly self-regulated learners as more goal-directed, relevant to their learning process, involving a more carefully considered and purposeful use of strategies, and a more rational use of resources.

The preparation for the VEE (B1) consisted of closed questions designed to compensate for the differences in prior knowledge by providing content-specific feedback to incorrect answers. Therefore, although the number of attempts to complete the preparation questions (B1) follows the pattern described above, it is affected also by prior knowledge differences, still important at this point of the course. This might explain why students with low SRL level required the least number of extra attempts. It is important to note that least number of attempts does not necessarily mean greater learning at this stage. Errors and mistakes are typically viewed as something to avoid during the learning process, but research suggests, by contrast, that making errors creates opportunities that are often an essential component of efficient learning (Bjork et al., 2013). Feedback on errors is especially effective for learning if the errors were made with high confidence (Butterfield and Metcalfe, 2001). Thus, the content-specific feedback may lead the students to learning gains if properly used. Again, the same pattern was found when it comes to the number of answers requested. Medium self-regulated learners went beyond having significant differences with both low and high ones, to account, on average, for twice as many requested answers as low self-regulated learners (48 vs. 24). This conspicuous result clearly shows that medium self-regulated learners followed a minimum effort approach and abused the VEE's functionality of providing the answers to the calculations, rather than doing the calculations themselves. Attempting to succeed in an interactive learning environment (a VEE in this case) by exploiting properties of the system rather than by learning the material is a strategy known as "gaming the system" (Baker et al., 2010). Previous work has shown that one of the common ways in which learners become careless and disengaged, is precisely gaming the system (Azevedo, 2015). Because learners intentionally decide how to regulate their engagement in (or disengagement from) learning, gaming the system, as procrastination and other self-handicapping techniques, is a form of SRL (Winne, 2015). In other words, SRL is not always beneficial and we should be talking of particular forms of SRL that lead to better learning, rather than SRL in general. Our results highlight medium self-regulated learners as clearly engaging in gaming the system, significantly more than low and high self-regulated learners. However, it is worth noting that independently of the SRL levels, on average, all students engaged in gaming the system, 24, 48, and 29 times for the low, medium, and high SRL levels respectively. Lack of motivation and interest have been found among the leading causes associated with this behavior (Baker et al., 2008). User interface features designed to increase interest have been reported to decrease gaming (Graesser et al., 2018). Nonetheless, apart from designing to enhance interest, it is important to explore various forms of SRL instruction to prepare students to learn in relatively new contexts (Cutumisu et al., 2015) such as VEEs.

The other VEE behavioral variables, namely the feedback requested on the experiment scheme (B3), the additional background information requested (B4), the number of attempts to answer the questions (B5), and the number of hints used (B7), showed no statistically significant difference among the three levels of SRL. The feedback on the experiment scheme (B3) and the use of hints (B7) can be associated to the SRL strategy of help seeking, during the performance phase (Zimmerman, 2013). When students do not ask for help when they need it, they run the risk of undermining their learning and achievement (Ryan et al., 2001). Normally, one would expect that students with different SRL level vary in their help-seeking behavior. However, the specificity of the feedback on the tasks usually available in online learning environments (e.g. VEEs), may not give enough room for differences to appear, as opposed to more open learning situations where knowing when and how to seek and apply help is an important part of SRL (Greene and Azevedo, 2010; Roll et al., 2014).

Similarly, the specificity and format of additional information (B4) available to the students, which can be associated with the performance phase SRL strategy of seeking information (Zimmerman, 2013), together with the other support features available (e.g. feedback on scheme, hints, and answers), might explain the lack of significant differences in this variable.

6.4.2 Perceived level of SRL in relation to VEE learning outcomes

SRL is regarded as "constitutive of success in learning, problem solving, transfer, and academic success in general" (Winne, 1997, p. 397). Based on previous findings (Broadbent and Fuller-Tyszkiewicz, 2018; Schunk, 2012), students with a high level of self-regulation skills are expected to have higher learning outcomes, as was the case in our study. However, by showing medium self-regulated learners to have the worst results, our findings indicate that there is no linear relationship between the level of (perceived) SRL and learning outcomes. Rather, as learners progress in their SRL skills in between low and high, there seems to be an intermediate phase that is detrimental for learning. Apparently, this phase (i.e. medium SRL) is characterized on the one hand by an increase in the students' perceived agency (as reflected in higher scores in the SRL questionnaire used), and on the other hand, by an abuse or random use of resources available, showing lack of both planning and goal-directed activity, as clearly evidenced in the VEE behavior discussed in section 6.4.1. The agency in these learners is not yet matched by a strategic control of their learning process (Järvelä et al., 2018). Effective learners seek deeper help (Luckin and Hammerton, 2002), while excessive help is associated with shallower learning (Mathews and Mitrović, 2008). Particularly, of the help and support available to the students in the VEE used, medium SRL learners abused the request of answers, thus avoiding obtaining the results by themselves and learning to do so. In other words, as previously discussed, they engaged in gaming the system behavior. Their lower learning outcomes come then as no surprise, since gaming the system has been repeatedly reported to be conducive to poorer learning (Azevedo, 2015; Baker et al., 2008; Cutumisu et al., 2015; Roll et al., 2014).

Low self-regulated learners outperformed medium, and actually, they were not that far from students with high SRL, which seems to be explained by the fact that they had a higher prior knowledge compared to medium self-regulated learners, and similar prior knowledge compared to high self-regulated learners (Table 6.4).

6.4.3 VEE learning outcomes in relation to VEE behavior, prior knowledge, and perceived SRL level

Two main ideas clearly arise from the results. First, prior knowledge as measured by the pre-knowledge test showed statistically significant and moderate to strong correlations with the VEE learning outcomes. This is an expected result since, historically, prior grades tend to be the best predictors of subsequent academic success (Zimmerman, 2013). As the perceived level of SRL increased from low to high, the correlations of outcomes to prior knowledge decreased from strong to moderate.

Prior knowledge is undeniably an advantage, and highly self-regulated learners activate this automatically (Pintrich, 2000). But, in the case of insufficient prior knowledge, highly self-regulated learners have the skills to cover the knowledge gap using effective strategies. This can also explain the reduced correlation of prior knowledge of highly self-regulated learners to their learning outcomes, as compared to students at the low and medium levels of SLR. Aligned with the previous discussions, prior knowledge accounted for more than half of the variance (R^2 = .55) of the learning outcomes of low SRL students, and almost half of the variance of (R^2 = .48) of medium, while 17% of the variance (R^2 = .17) of high SRL students, the latter having obtained the best results in the post-knowledge test.

Second, in general, VEE behavior negatively correlated with the outcomes. Previous studies indicate that there is indeed a link between student activity in online learning environments and outcomes as measured by final grades (Chen, 2002; Cho and Kim, 2013; Colthorpe et al., 2015). Higher levels of activity are often associated with better outcomes and greater student satisfaction (Lowes et al., 2015). Our results support that there is a relationship between VEE behavior and outcome, but a negative one in this case. The explanation is found in the number of answers requested (B6). Table 5 shows that, apart from prior knowledge, the number of answers requested is the only variable with statistically significant (and at least) moderate correlations. The highest negative correlation of this variable (B6) to VEE outcomes is found in the medium SRL group, which had the poorest outcomes. As discussed in previous sections, this gaming the system behavior (Baker et al., 2008) evidences disengagement (Azevedo, 2015) and impairs learning. The generally negative correlations of VEE behavior to outcomes suggest that it would have been more efficient for the medium self-regulated learners to not interact that much with the VEE, mostly because they tried to "abuse" its resources. This result should be taken into account for the design of online learning environments.

6.5 Conclusions

In many disciplines, laboratory education is an essential part of the curriculum, as it gives students the possibility to put the learned theories and procedures to the test, and gain practical, hands-on experience on the subject matter. However, laboratory education is expensive, and it has been shown that, often, students do not come to the laboratory sufficiently prepared, which prevents them from obtaining the proper skills from this practical learning opportunity. To mitigate this issue, VEEs such as the LabBuddy VEE used in our experiment, have been developed to take advantage of online learning environments for supporting students in familiarizing with the laboratory environment, its methods, content-specific knowledge, and resources. Nonetheless, it is important that students take ownership of their learning processes. Students differ in their ability to set goals, plan how to meet their goals, use strategies to address tasks, monitor their progress towards their goals, evaluate their performance, and adapt their learning processes cyclically based on monitoring and control loops. In other words, students vary in their level of SRL, but SRL can be taught and supported so that learners become better at it (Winne, 1997), leading them to better learning outcomes (Bjork et al., 2013).

In this study, students' perceived SRL level (i.e. low, medium, high) was explored in relation to their VEE behavior (e.g. number of attempts and hints, number of answers requested) and their VEE learning outcomes, while controlling for their prior knowledge. In addition, the relationships between VEE behavior and VEE learning outcomes depending on their SRL level were investigated. The results showed that, on the one hand, as expected, students with a high SRL level obtained better outcomes, but on the other hand, the relationship between these variables was not linear, since medium self-regulated learners obtained the worst results. The medium groups' VEE behavior suggested lack of goal-directed activity and planning, and showed the highest level of activity with the VEE but overusing resources that impair learning, such as number of attempts at answering the questions. This suggested random behavior, and excessive request of answers (statistically significantly more than both their low and high counterparts, and twice as many times as the students with low SRL). Such behavior, known as gaming the system, has been repeatedly reported to be detrimental to learning (Roll et al., 2014). The findings thus suggest the existence of an intermediate SRL level between low and high, characterized by an increase in perceived agency, ownership over the learning process and use of strategies and resources, but still lacking goal-directed activity and appropriate planning and execution to meet the goals. The medium SRL level seems to be a developing but not yet mature SRL level, a transitional state between the low and high levels. The random and detrimental to learning behavior of this group, indicates that special attention should be paid in SRL interventions to this group, and to support students to move to a higher SRL level as soon as possible, where they will benefit from their increase in agency, but use it in a purposeful way in the accomplishment of their goals, and improving their learning outcomes. The phases of the SRL process have been largely studied and reflected in a variety of models (Panadero, 2017). However, to the best of our knowledge, the phases of how the SRL skills develop over time have received less attention.

In this study, it was also shown that prior knowledge correlated with the learning outcomes, aligned with a historical tendency, and showed an inverse relation to the SRL level. The more self-regulated students perceived they were, the less they relied on their prior knowledge for their learning outcomes, since their ability to acquire new knowledge was likely to be superior.

In general, and independently of SRL level, VEE behavior correlated negatively with VEE learning outcomes. This has important implications for the design of VEEs, given their role of supporting students' laboratory practice. It is paramount to iteratively design VEEs to maximize students' engagement and minimize the harmful gaming the system behavior (Graesser et al., 2018). Students might benefit from a limited access to ready-made answers, and only been provided with task solutions when there is evidence in the VEE that they have been attempting the solution. Instead, the use of hints and seeking information should be encouraged and seen by the students as the standard procedure. Challenges are recognized as triggers of SRL (Perry, 1998). The VEE should be designed for information and help-seeking strategies, giving answers only exceptionally. In the real laboratory, this support is not available, and they need to learn how to perform the calculations. The VEE acts as a scaffold to support learning that is removed in the real laboratory practice. SRL is paramount to optimizing learning and performance, but formal training is scarce (Bjork et al., 2013). Accordingly, it might be important to embed SRL support in the VEE, for example, through prompts (Hadwin et al., 2005).

The use of a questionnaire in this study has its limitations (Winne, 2015). It is known that students are not always accurate when reporting their own use of strategies (Boekaerts and Corno, 2005), and that questionnaires are static tools (Greene and Azevedo, 2010). However, self-reported data is still the most frequent measure of SRL (Panadero, 2017), and the field has heavily relied on them (Hadwin et al., 2011). Self-reported data is still regarded as a valid tool when it is sufficiently tailored to the specific context in which the study is being conducted (Panadero et al., 2016).

Future research could focus on the development phases of SRL, on how to prevent students from staying at a medium SRL level, and on the effects of VEEs designed to avoid gaming the system behavior and to promote SRL.

Chapter 7

Using sequential pattern mining to understand how students use guidance while doing calculations

S. Verstege, Y. Zhang, P.A. Wierenga, L. Paquette, and J. Diederen.

Contents

Abstract

In natural science education, many experiments lead to the collection of raw data, which needs to be processed into results by doing calculations. Teaching students how to approach such calculations can be done using digital learning materials that provide guidance. The goal of this study was to investigate students' behaviour regarding the use of guidance while doing calculations, and to relate this behaviour to learning. Sequential pattern mining was used to i) identify students' behaviour patterns while doing calculations in an online learning environment, ii) study the relation between use of guidance and early success on first attempt at submitting a calculated value, iii) study the relation between students' use of guidance and learning gain, and iv) study the relation between students' use of guidance and prior knowledge. Students made a lot of use of the guidance provided in the learning task. Students who used the option to check their intermediate calculations and students who studied worked examples were more likely to make a successful first attempt to complete the calculation when compared to students who did not use this guidance. Guidance in the form of hints was used a lot, although the usefulness of the hints alone was not quantified. We did not find a relation between learning gain and use of guidance, while we did find a trend that students with a low prior knowledge used more guidance compared to students with a higher prior knowledge. The results of this study are a first indication that learning materials designed based on the design principles that students should be provided with hints and with the opportunity to check their intermediate calculations, contribute to enabling students to independently complete calculations in digital learning materials.

7.1 Introduction

7.1.1 Doing calculations in natural science education

In natural science education, many experiments lead to the collection of raw data, which needs to be processed into results by doing calculations. As such, teaching students how to approach calculations, or in more general terms, teaching students how to solve problems, is an important aspect of natural science education. A problem occurs "whenever there is a gap between where you are now and where you want to be, and you don't know how to find a way to cross that gap" (Hayes, 2013), and problem-solving is "What you do, when you don't know what to do" (Wheatley, 1984). Problem solving is a complex skill, requiring a multi-step approach to get from the problem to the solution (Belland, 2011). To train students from being a novice (i.e. experiencing procedural difficulties during problem solving) towards being an expert in solving calculation problems (i.e. approaching a problem guided by knowledge and experience), they should be provided with a general problem-solving approach, and opportunities to practice using this approach. Yuriev et al. (2017) introduced a problem-solving approach based on problem solving in chemistry and related fields, consisting of five steps: i) define and deconstruct the problem, ii) analyse the problem and its relevant context, iii) create a plan based on the relationship between the data and the unknowns, iv) implement the planned steps, and v) evaluate the results, which can lead to either troubleshooting or to the solution to the problem.

While doing multi-step calculations, students should ideally go through all of these five steps and will have to overcome the challenges that each of the five steps may present to them. To help students face the challenges when doing multi-step calculations, they can be provided with a learning material in which they can practice applying a problem-solving approach. This learning material should provide guidance such as hints and feedback. An advantage of the use of online learning materials over traditional instruction is this support can be given to many students simultaneously and just-in-time (Diederen et al., 2003). In traditional classrooms, the teacher experiences how their personal guidance is received by students, while in an online learning material, this feedback by students is less visible. The shift in guidance from teachers to the online learning material raises questions such as: How do students use the guidance, and does the guidance help to do the calculations?

One way to answer these questions, is by applying learning analytics to evaluate and improve the design of learning materials (e.g. Law and Liang, 2020; Lockyer et al., 2013; Mangaroska and Giannakos, 2019; Wiley et al., 2020). Learning analytics involves the collection, analysis, and interpretation of (log) data of students' interactions with the environment, for purposes of understanding and optimizing learning in the environment

(Gasevic et al., 2015; Greller and Drachsler, 2012; Hernández-Leo et al., 2019; Siemens and Long, 2011).

7.1.2 Sequential pattern mining and its application in learning process data

Sequential pattern mining (SPM) is an exploratory data analysis method defined as the problem of finding interesting sequences in a dataset containing multiple sequences of items (Agrawal and Srikant, 1995). In the context of educational data, an item may be an action that a student executes in a computer-based educational system. For example, in a learning management system a sequence of student actions might indicate how students navigate the course content by watching lectures, downloading lecture notes, and submitting an assignment solution. SPM differs from the other learning analytics methods by focusing on the sequential relationships between items (Baker, 2010).

SPM has been applied to data from a wide range of educational settings–including Learning Management Systems, Massive Open Online Courses, Intelligent Tutoring Systems, computer-supported collaborative learning environments, educational games, and course enrolment systems. Within those contexts, SPM has showed how it can be used to highlight similarities and differences in engagement and learning across students with regards to self-regulated learning (Kinnebrew et al., 2014; Taub and Azevedo, 2018; Taub et al., 2018), inquiry-based learning (Chen and Wang, 2020; Kang et al., 2017), and collaborative learning (Zheng et al., 2019; Zhu et al., 2019).

Researchers have applied SPM to educational data for a variety of purposes such as mining learning behaviours, enriching educational theories, and evaluating the efficacy of interventions. Frequent sequential patterns generated by SPM from learning process data such as event logs can be used to identify common behaviour patterns across students (Zhou et al., 2010). These behaviour patterns may reveal how students navigate their activities within a learning environment and inform the learning design update (Mirzaei and Sahebi, 2019). For example, Kang et al. (2017) applied SPM to gameplay logs in Alien Rescue, a serious game for teaching middle school students scientific problem-solving skills, to investigate how students' playing behaviours might vary on different days.

SPM has been used to investigate how interpretation of mined sequential patterns relate to educational theories. For instance, Taub and Azevedo (2018) applied SPM to investigate how self-regulated learning behaviours and emotions influenced learning and gameplay within Crystal Island, an educational game that teaches scientific inquiry skills and microbiology. They linked gameplay behaviours, such as hypothesis testing, to metacognitive monitoring strategies aligned with self-regulated learning theories. In the context of Betty's Brain, an open-ended learning environment where students learn about complex scientific phenomena, Kinnebrew et al. (2017) showed how sequential patterns discovered using SPM may not always align to predefined theoretical problem-solving models. They argued that these patterns may represent new learning strategies that can be integrated in existing models to extend them.

Studies have also explored whether sequential learning behaviour patterns could capture the effect of interventions. For example, Wong et al. (2019) designed weekly prompt videos to facilitate students to think about their plan, monitoring, and reflection on learning in a Coursera course. They compared students who watched at least one prompt video (prompt viewers) and those not watching any prompt video (non-viewers). The group of prompt viewers shared more sequential behaviour patterns than non-viewers. Besides, prompt viewers tended to watch videos in the order that instructors planned.

7.1.3 Learning task

The learning task that was used as the context for this study resulted from an educational design research project, and was embedded within a larger virtual experiment environment in which students i) make an experimental design of chemical methods to answer the given research questions, ii) study the chemical methods' background information, iii) obtain raw data corresponding to the chemical methods, and iv) process the raw data into results by means of interactive calculation questions (chapter 4). The last part of the learning task requires students to complete three types of interactive calculation questions. Each type of calculation question requires students to use a different approach and must be solved multiple times for different datasets (Figure 7.1).

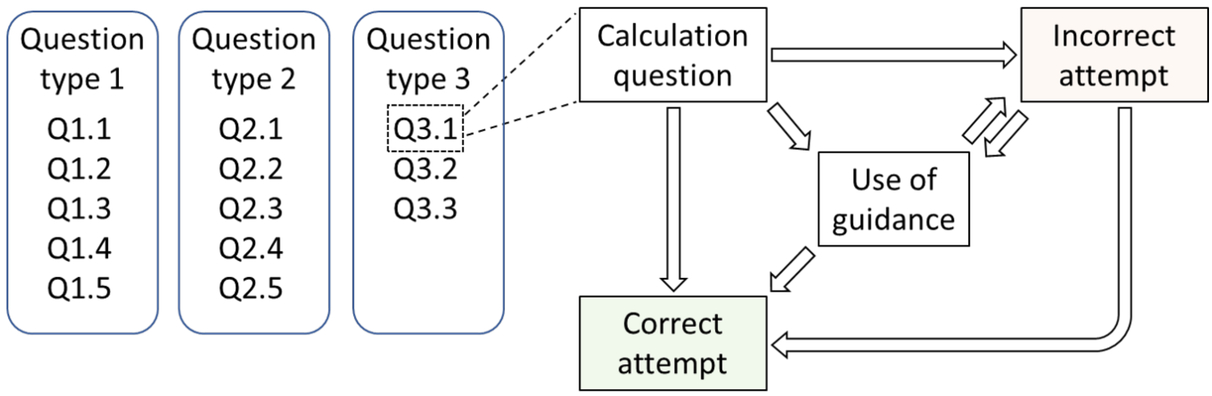

Figure 7.1: Schematic overview of the three types of calculation questions, including the number of corresponding questions per type, followed by a flow diagram on how students can get to a correct attempt for a question.

All interactive calculation questions provide students with guidance, so that students should be able to complete the calculations without supervision. First, students can choose to access procedural hints (Figure 7.2a). Hints can be opened one at the time, while each hint provides a link to the next hint (if available). Each hint provides detailed information on the (next) calculation step. All hints together provide the procedural information required to solve the problem, guiding students with the second, third and fourth steps of the problem-solving approach (Yuriev et al., 2017). Second, when an answer is submitted, the system will indicate with a green check or a red cross which values are correct, and which values are incorrect (Figure 7.2b), which supports the fifth step of Yuriev's problem-solving approach. Third, students have the option to check their intermediate calculations, which decompose the calculation into smaller parts (Figure 7.2c). These intermediate calculations will be checked by the system, which will indicate whether they are correct or incorrect (which respectively are the first and the fifth step of Yuriev's problem-solving approach. This functionality can be used to check if the procedure (possibly acquired through the hints) was carried out correctly. When an intermediate calculation is incorrect, students can choose to request the correct intermediate value from the system. An overview of an interactive calculation question is given in Figure 7.2d.

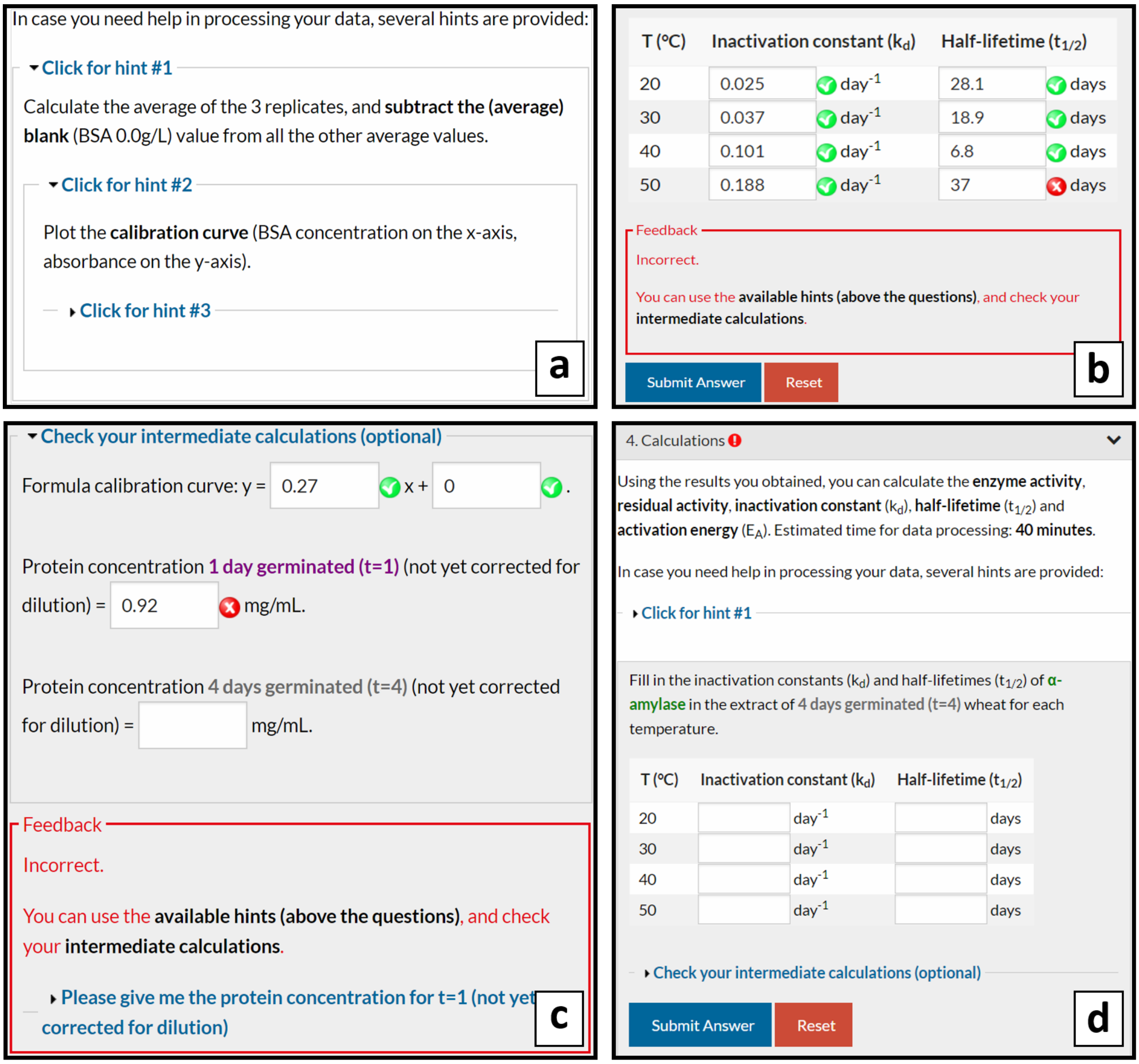

Figure 7.2: Screenshots of interactive calculation questions. a) Procedural hints; b) Feedback on submitted answers; c) Example of checking intermediate calculations, with feedback; d) Overview of an interactive calculation question.

7.1.4 Goal and research questions

The goal of this study was to investigate students' behaviour regarding the use of guidance while doing calculations, and to relate this behaviour to learning. The following research questions were formulated:

1. Which behaviour patterns can be observed while students are doing calculations in an online environment?
2. Is there a relation between use of guidance and early success on first attempt at submitting a calculated value?

3. Is there a relation between students' use of guidance and learning gain?
4. Is there a relation between students' use of guidance and prior knowledge?

7.2 Method

7.2.1 Participants

All students (N = 81) were enrolled in a master level course in enzymology (168 study hours) at a university in the Netherlands. All students had previously obtained a bachelor level degree in natural sciences: 53% from the Netherlands, 37% from a university outside the Netherlands, and 10% was unknown. All students were enrolled in a MSc programme, either Food Technology (81%) or Biotechnology (19%). The students were 22.9 years old (SD = 1.9) on average. About 64% of the students was female, while the rest were males.

7.2.2 Study design

All student activities related to the study are shown in Table 7.1.

Table 7.1: Student activities related to the study, with corresponding measurements, measurement output, timing, and duration.

Code	Student activity	Measurement	Measurement output	Day*	Duration
A1	Pre-test	Ability to approach calculations	Test score	1	30 min
A2	Learning task	Student behaviour while doing calculations	User log data	1 and 2	4 h
A3	Post-test	Ability to approach calculations	Test score	8	30 min

* Running days relative to the start of the experiment. There were no course activities in between the learning task and the post-test.

The calculations were embedded in a virtual experiment environment that was implemented in a master level course. Given the pre-requisites for this course, students were expected to have already developed problem-solving skills to some degree. To measure students' ability to approach calculations, they were asked to complete the pre-test (A1), for which they got a maximum of 30 minutes. The pre-test consisted of three open questions related to approaching calculations, on which students could score a maximum of 13 points. An example of an open question is: "Describe as detailed as possible all calculation steps you need to do to calculate...". Most of the students (N =

77) completed the pre-test. The scores (M = 5.2, SD = 2.7) were used as a measure of students' ability to approach calculations.

After the pre-test, students started working in the virtual experiment environment, which included calculations (A2) for which they needed an average of 4 hours to solve. Most students completed the whole VEE, including the calculations, in 8 hours, at day 2 of the experiment. User log data was collected, which is explained in more detail in section 7.2.3. All students (N = 81) completed the assignment in the virtual experiment environment (including the calculations) as this was a compulsory course activity. One week after the pre-test, just before the start of the next course activity, students were asked to complete the post-test (A3), for which they got a maximum of 30 minutes. The post-test was identical to the pre-test, and was completed by N = 79 students. Students were not informed about the post-test, to prevent targeted studying for it. There were no course activities in between doing the calculations and the post-test. Students voluntarily completed the pre- and post-test (A1 and A3), while they were aware that the anonymized results could be used for research purposes. The scores (M = 8.5, SD = 2.0) were used as a measure of students' ability to approach calculations.

7.2.3 Log data pre-processing

User log data was collected for the complete learning task in the virtual experiment environment. As we were mainly interested in students' behaviour patterns while doing calculations, we only kept the user log data that was related to completing calculations for any of the three question types. Based on the context of the user log data, the logs were labelled with action codes, which are described in Table 7.2. To distinguish between code representing a single action and code representing consecutive actions, consecutive actions of the same type were condensed and labelled as one action code by adding a suffix (Kinnebrew et al., 2013). The used suffixes are: '+', which means two or more consecutive actions, '+2', means: exactly two actions, '+3+' means: three or more. For hints, '+all' means: all available hints. The suffixes '+2', and '+3+' were only used for the GI and AI actions, to aid the interpretation of those actions. The suffix '+all' was only used for opening hints, as there is a maximum number of hints that can be opened.

Table 7.2: Action codes, including condensed actions, with description and corresponding category.

Action code	Description	Category
AC	Student submits a calculated value, and is provided with feedback that the value is correct	Attempt
AC+	Multiple consecutive AC	Multiple attempts
AI	Student submits a calculated value, and is provided with feedback that the value is correct	Attempt
AI+2	Two consecutive AI	Multiple attempts
AI+3+	Three or more consecutive AI	Multiple attempts
GH	Student opens a hint	Guidance
GH+	At least two hints opened consecutively	Multiple guidance
GH+all	All hints opened consecutively	Multiple guidance
GW	Student accesses a worked example (only available after mistake in GI, in T-stability)	Guidance
NI	Navigate to the section in which intermediate values can be submitted (required before GI is possible)	Guidance
GI	Student submits a value of an intermediate calculation, and is provided with feedback whether the value is correct	Guidance
GI+2	Two consecutive GI	Multiple guidance
GI+3+	Three or more consecutive GI	Multiple guidance
GA	After submitting an incorrect intermediate value, the student the student requests and receives the correct intermediate value	Guidance
GA+2	Two consecutive GA	Multiple guidance
GA+3+	Three or more consecutive GA	Multiple guidance
GC	After submitting a correct value (AC), or after correctly submitting all intermediate values (GI) and submitting the final calculated value incorrectly (AI), students click a link in the feedback which provides a pdf with the full calculation	Information
NB	Navigate to background information	Information
AD	Student acquires an excel file with the raw data required to answer a calculation question	Not available*
NC	Navigate to an interactive calculation question	Not available*

* These actions are required by the system.

In addition, all possible actions codes were assigned a category based on their purpose within the learning environment. All codes action in which students submit a value in the VEE were categorized as 'attempt'. All actions in which students make use of guidance were categorized as 'guidance'. All actions providing students with general information were categorized as 'information', for example when students clicked a link to access a pdf document providing details about the required calculations (GC). In the case of condensed actions, 'multiple' was added to the categories. The actions AD and NC were not categorized, as these actions are required by the system in specific situations. They were included in the table and in some results as they do provide meaning. For example, the sequence AD => NC => AC indicates that students acquired the dataset, navigated to an interactive calculation question, and submitted a correct value, without using support in the meantime. After the pre-processing, a student's log data contained 224.37 actions on average (SD = 116.07).

7.2.4 Analyses

To find behaviour patterns of students while doing calculations in the online environment (research question 1), the cSPADE algorithm was applied (Zaki, 2000). This is an efficient SPM algorithm that has been used in other educational studies to find frequent sequential patterns (Jiang et al., 2015; Kang et al., 2017; Wong et al., 2019). We implemented the cSPADE algorithm via the *arulesSequences* package in R (Buchta and Hahsler, 2020). The algorithm computed the support value for each sequential pattern, which represents the proportion of students who used the sequential pattern. We wanted to cover as many sequential patterns as possible, as such we set the minimum support to a small value, 0.2, which meant that only sequential patterns with support no less than 0.2 were counted as frequent sequential patterns. We tried other values (including 0.1, 0.3, 0.4, and 0.5), but higher minimum support ignored many interesting patterns, and lower minimum support (e.g. 0.1) generated excessive meaningless patterns. We specified the maximum gap as 1 so that the algorithm would only consider action sequences that matched a sequential pattern exactly as an instance of the sequential pattern. For example, AI followed by GI further followed by AC was an instance of AI => GI => AC but not an instance of AI => AC. For each sequential pattern returned by the cSPADE algorithm, its average occurrence in a student's log data was computed. The cSPADE algorithm generated 324 sequential patterns whose support was greater than 0.2. These patterns were inspected, after which all meaningless patterns were removed. Meaningless patterns occurred for example when students first have to navigate to a calculation question, before they can make an attempt to answer it (e.g. NC => AC). Such a pattern does not contain more information than its sub pattern without the first element (e.g. AC). After removing the meaningless patterns, 88 sequential patterns remained.

We investigated if there was a relation between use of guidance and early success by comparing differences in use of guidance between students whose first attempt is correct versus students whose first attempt is incorrect (research question 2). The learning task that is subject of this study involves three types of calculation questions (as described in section 7.1.3. Each type of calculation question requires students to use a different approach and has to be done multiple times for different datasets. Since it is likely that the amount of guidance that students need to solve multiple calculation questions of the same question type will decrease over time, we chose to examine the difference in use of guidance before students' first attempt for each calculation type. Students' log data was split into three subsets with each containing the actions corresponding to each of the question types. For each subset of log data, a student's first attempt to answer a calculation question and all actions after this first attempt were removed. The average length of the remaining sequences was relatively short, being 15.53 (SD = 13.25) for a correct first attempt, and 8.32 (SD = 6.96) for an incorrect first attempt at question type 1. The average length of the remaining sequences at question type 2 were 7.68 (SD = 7.23) for a correct first attempt, and 5.59 (SD = 7.39) for an incorrect first attempt. The average length of the remaining sequences at question type 3 were 17.35 (SD = 10.37) for a correct first attempt, and 14.25 (SD = 20.22) for an incorrect first attempt.

Differential SPM (Kinnebrew et al., 2013) was used to identify sequential patterns used differentially by students who submit the correct answer at their first attempt versus students who find an incorrect answer at their first attempt across the three subsets of log data. Differential SPM typically achieves its goal through the following steps:

1. Discover frequent sequential pattern candidates within each group through traditional SPM algorithms, such as the cSPADE algorithm.
2. Compute the occurrences per sequential pattern per student for all candidate patterns identified in the first step, no matter in which group a pattern is frequent.
3. Apply statistical tests, such as t-tests, to identify candidate patterns that are statistically significantly different in the occurrences between groups.

Because the investigated sequences were short (9.33 actions on average), it is unlikely that a student will execute the same pattern more than once. As such, instead of treating the frequency of occurrences of sequential patterns as continuous, we treated them as binary variables: whether a sequential pattern occurred in a student's action sequence or not. The Fisher exact test was used to examine the differences in the binary occurrence between the groups whose first attempts were correct versus incorrect (He et al., 2019). As such, the support values between groups were compared. The p-values were adjusted by the Benjamini and Yekutieli correction to control the false discovery rate due to multiple comparison (Benjamini and Yekutieli, 2001). The magnitude of the difference in a sequential pattern was characterized by the odds ratio = ((n11+0.5) * (n22+0.5)) / ((n12+0.5) * (n21+0.5)). n11 is the number of students whose first attempt was correct and executed the pattern, n12 is the number of students whose first attempt was incorrect and executed the pattern, n21 was the number of students whose first attempt was correct and did not execute the pattern, and n22 was the number of students whose first attempt was incorrect and did not execute the pattern. Addition of the +0.5s in the formula for the odds ration makes the odds ratio less biased (Gart and Zweifel, 1967). Rosenthal (1996) proposed a guideline for interpreting odds ratios: an odds ratio > 1.5 (or < 0.67), > 2.5 (or < 0.4) and > 4 (or < 0.25) represents a small, medium, and large effect, respectively.

To investigate the potential relation between students' use of guidance and learning gain (research question 3), we applied differential SPM to identify sequential patterns used differentially by students with high, medium, and low learning gains. Interpreting learning gain values is a challenge since they are dependent on students' prior knowledge. In other words, students who get a high score on the pre-test have limited potential to improve towards the post-test. To provide as much contextual meaning to the learning gains as possible, the following approach was taken. K-means clustering was used to create two sets of three clusters of students, based on both the pre- and post-test scores. Through both sets of clusters, students were grouped based on whether their result on each test was low, medium or high. The average pre- and post-test scores for each cluster are shown in Table 7.3. We also calculated the average learning gain for students in each cluster to illustrate that students with lower pre-test scores were more likely to show larger learning gains.

Table 7.3: Clusters with corresponding average pre- and post-test scores and average learning gain.

Cluster *pre*-test	**# students**	**Average score (SD)**	**Average learning gain (SD)**
Low	19	1.66 (1.13)	4.84 (2.66)
Medium	37	5.18 (0.97)	3.34 (2.11)
High	21	8.50 (1.13)	1.31 (1.51)
Cluster *post*-test	**# students**	**Average score (SD)**	**Average learning gain (SD)**
Low	20	5.78 (0.79)	2.38 (2.01)
Medium	24	8.19 (0.79)	3.52 (2.95)
High	35	10.24 (0.71)	3.94 (2.60)

Based on the pre-test score, students were assigned to cluster low, medium, or high, and based on the post-test score students were once again assigned to cluster low, medium, or high. A student could for example be in cluster medium based on the pre-test score, learn a lot during the learning task, and end up in cluster high based on the post test score. This led to a total of nine possibilities, which were grouped and labelled as low learning, medium learning, or high learning. Students (N = 16) were assigned to low learning if they dropped one or two clusters (e.g. cluster medium -> low). Students (N = 35) were assigned to medium learning if they remained in the same cluster (e.g. cluster medium -> medium). Students (N = 25) were assigned to high learning if they increased one or two clusters (e.g. cluster medium -> high) (Table 7.4).

Table 7.4: The pre- and post-test clusters including the number of students assigned to each of the options, with corresponding label of learning and average learning gain.

Cluster pre-test	**Cluster post-test**	# students	Learning	**Average learning gain (SD)**
High	High	15	Medium learning	1.73 (1.40)
High	Medium	6	Low learning	0.25 (1.33)
High	Low	0	N/A	N/A
Medium	High	16	High learning	4.94 (1.42)
Medium	Medium	11	Medium learning	3.14 (1.53)
Medium	Low	10	Low learning	1.00 (1.11)
Low	High	3	High learning	7.50 (0.87)
Low	Medium	6	High learning	6.58 (1.74)
Low	Low	9	Medium learning	3.44 (1.57)

The Kruskal Wallis test was used to examine whether the three learning groups had statistically significant differences in sequential patterns. The effect size measure, η^2 was reported. A rule of thumb for interpret this measure is that η^2 greater than 0.01, 0.06, and 0.14 represents small, medium, and large effects respectively (Cohen, 1988). If the Kruskal Wallis test indicates that a sequential pattern was used differentially by the three groups, Mann-Whitney U tests were conducted on each pair of the three groups to identify between which two there was a difference. The Benjamini and Yekutieli correction was conducted to control the false discovery rate due to multiple comparison. An effect size, Cohen's r, was reported for the Mann-Whitney U test, which ranges between -1 and 1 (Fritz et al., 2012). Cohen's guideline for r is that the absolute value of r greater than 0.5, 0.3, and 0.1 represents a larger, medium, and small effect, respectively (cited in Fritz et al. (2012)). The same analyses were repeated to examine the three pre-test score groups' differences in sequential patterns (research question 4). A summary of the analyses per research question is shown in Table 7.5.

Table 7.5: Summary of the analyses for each research question.

Research question	SPM Algorithm	Input data	Statistical test	Independent variable	Dependent variable	Effect size
1	cSPADE	All log data	N/A	N/A	N/A	N/A
2	Differential SPM	Log data before the first attempt to questions	Fisher exact test	Correct first attempt versus incorrect first attempt	Whether a sequential pattern existed in one's log data (binary)	Odds ratio
3	Differential SPM	All log data	Kruskal Wallis test & Mann-Whitney U test	High, medium, and low learning	The occurrences of a sequential pattern in one's log data (continuous)	η^2 & r
4	Differential SPM	All log data	Kruskal Wallis test & Mann-Whitney U test	High, medium, and low prior knowledge	The occurrences of a sequential pattern in one's log data (continuous)	η^2 & r

7.3 Results and discussion

7.3.1 Frequent sequential patterns throughout the learning task

After analysis of all log data and filtering of meaningless patterns, 88 frequent sequential patterns remained. These remaining patterns were categorized, using the categories enumerated in Table 7.2, as either (multiple) attempt, (multiple) support, information, and the combination of these codes. The 18 sequential patterns with the highest support are shown in Table 7.6.

Table 7.6: The eighteen sequential patterns (SP) with the highest support, average occurrences and corresponding category.

Code	Sequential pattern	Support	Average occurrences	Category
SP1	AC	1.00	15.52	Attempt
SP2	AI	0.99	6.04	Attempt
SP3	NI	0.96	8.37	Support
SP4	GH+	0.90	2.86	Multiple support
SP5	AI+2	0.89	2.11	Multiple attempt
SP6	GH+all	0.87	2.80	Multiple support
SP7	GH	0.86	3.13	Support
SP8	AI+3+	0.84	2.61	Multiple attempt
SP9	GI	0.81	5.91	Support
SP10	AI =>AC	0.80	1.81	Multiple attempt
SP11	AD =>NC =>AC	0.76	2.30	Attempt, no support
SP12	AI =>NI	0.75	1.63	Attempt =>support
SP13	GI+2	0.73	2.14	Multiple support
SP14	GI+3+	0.70	2.87	Multiple support
SP15	GA	0.67	2.30	Support
SP16	AC+	0.65	1.70	Multiple attempt
SP17	AD =>NC =>AI	0.61	1.03	Attempt, no support
SP18	NB =>NC =>NI	0.58	1.22	Information =>support

The results indicate that students use a lot of support while working on the calculation questions. Almost all students navigated to the section in which intermediate values can be submitted (SP3). Students submitted many intermediate calculations (SP9, 13, and 14), but most of those were not submitted directly after navigating towards this section. Taking all 88 frequent sequential patterns in consideration, the only match that includes submitting an intermediate value directly after navigating to the just mentioned section, is the sequential pattern NI => GI+3+, with a support of 0.48 and an average occurrence of 0.94. When students navigate to the section in which intermediate values can be submitted, but do not proceed in submitting one or more values, this can be considered as using a hint (i.e. which intermediate values can be calculated).

From SP4, 6, and 7 it is clear that students use a lot of hints. What stands out is that students were more likely to open two or more hints consecutively (SP4 and 6) than opening one hint at a time (SP7). Since most calculations are done outside the system (e.g. in Microsoft Excel), and the time factor is not included in this analysis, it is not possible to conclude whether students quickly opened all hints, which could indicate students who are "gaming the system" and trying to have the learning environment provide them with the answer (Baker et al., 2008), or whether they have been following the hint, after which they needed more support, and opened the next hint.

More or less the same holds for multiple consecutive incorrect attempts (SP5 and 8). Based on this analysis we cannot tell whether students were guessing the correct value, or whether they really tried to (re)do their calculations after they received feedback on their mistake. SP5, 8, 10, and 16 all show multiple consecutive attempts. This means that students sometimes decide not to use support in between attempts. Another situation in which students do not seem to use support, is when they make an attempt directly after acquiring the raw data required to do the calculation (SP11 and 17). However, students may have used support before acquiring the raw data. It should be noted that given the learning task, students needed to apply a different calculation approach for each question type, but they had to repeat this approach multiple times for different datasets. Given the results, it is possible that students initially used support to find the correct calculation approach, during which they learned, and subsequently required less support when solving similar problems.

7.3.2 Use of guidance before first attempt

Table 7.7 shows all frequent sequential patterns (i.e. the support for the sequential pattern was higher than 0.2) found before students' first attempt. For each frequent sequential pattern, the support was given for students whose first attempt was correct and for students whose first attempt was incorrect.

Table 7.7: Frequent sequential patterns (i.e. when the support is higher than 0.2) before first attempt, being either correct or incorrect, for different question types.

Code	**Sequential pattern**	**Question type 1**			**Question type 2**			**Question type 3**		
		Support		**Odds ratio**	**Support**		**Odds ratio**	**Support**		**Odds ratio**
		Correct (N = 30)	Incorrect (N = 50)		Correct (N = 38)	Incorrect (N = 41)		Correct (N = 26)	Incorrect (N = 53)	
SP3	NI	0.63	0.30	3.88	0.47	0.20	3.56	0.62	0.30	3.57
SP4	GH+	0.40	0.46	0.79	0.26	0.27	0.98	0.42	0.32	1.55
SP6	GH+all	0.37	0.24	1.82	0.26	0.20	1.45	0.42	0.45	0.89
SP7	GH	0.60	0.32	3.09	0.34	0.15	2.89	0.39	0.28	1.58
SP9	GI	0.27	0.00	38.16**	0.34	0.00	43.94***	0.46	0.06	12.44**
SP13	GI+2				0.26	0.00	30.58**	0.42	0.04	15.28**
SP14	GI+3+	0.20	0.00	26.8*				0.39	0.04	13.11**
SP15	GA	0.32	0.00	39.15***						
SP19	GI =>GW							0.39	0.04	13.11**
SP20	GI+2 =>GW							0.23	0.00	33.93*
SP21	GW =>GI							0.31	0.04	9.46*
SP22	GW =>GI+2							0.23	0.00	33.93*
SP23	GW =>GI+3+							0.23	0.04	6.53
SP24	NI =>NB	0.27	0.08	3.90						
SP25	GH =>NB	0.23	0.16	1.60				0.23	0.23	1.05
SP26	GH+ =>NB	0.33	0.42	0.70				0.31	0.28	1.14
SP27	GH+all =>NI							0.27	0.08	4.23
SP28	GH+all =>NB	0.27	0.14	2.19				0.12	0.26	0.41
SP29	NB =>NC =>GH	0.23	0.18	1.39				0.23	0.21	1.17
SP30	NB =>NC =>NI	0.23	0.08	3.30						

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. p was adjusted by the Benjamini and Yekutieli correction.

The support for frequent sequential patterns was in general higher for students whose first attempt was correct versus students whose first attempt was incorrect. In other words, students who use more support in the system are more likely to have their first attempt correct. The support for almost all frequent sequential patterns that contain intermediate calculations (i.e. GI, GI+2 or GI+3+) was significantly higher for students whose first attempt was correct versus students whose first attempt was incorrect (SP9, 13, 14, 19 – 23). Moreover, the support for frequent sequential patterns that involve opening a worked example, which was only available for question type 3 (GW, sequential patterns 19 – 23), was significantly higher for students whose first attempt was correct versus students whose first attempt was incorrect. The use of GW was only possible after submitting an incorrect intermediate value, so this result is dependent on the use of GI. These results suggest that decomposing the calculation into smaller parts and providing students with feedback on their intermediate calculations and/or subsequent worked examples, contribute to students' ability to successfully calculate the correct answer on their first attempt. The support for sequential patterns that contain hints (SP4, 6, 7, 25 – 29) was usually frequent for both groups while no statistically significantly differences in support between students whose first attempt was correct versus students whose first attempt was incorrect. This suggests that the use of hints alone did not contribute to students' ability to directly calculate the correct answer.

7.3.3 Use of guidance in relation to learning gain

After splitting the participants in three groups of high, medium, and low learning, 127 frequent sequential patterns (support > 0.2) were found in one or more of the three groups.

No pattern showed a statistically significant difference between the different learning groups after the Benjamini and Yekutieli correction, but seven were significant before correction (Table 7.8). Those seven patterns were investigated to see if they might reveal interesting insights about which behaviour patterns may be most likely to lead to higher learning gains, based on their effect size, support and average occurrences. In particular, despite not showing a statistically significant differences after controlling for false discovery, patterns such as SP31 and SP 32 showed large effect sizes (η^2 of 0.24 and 0.19 respectively). However, upon further investigation, the interpretation of both patterns was considered as not meaningful due to the rarity (based on overall support and average occurrences) of both patterns.

Table 7.8: Sequential patterns that were used differentially by high, medium, and low learning students (SP31 – 37) and corresponding effect sizes.

Code	**Sequential pattern**	**Average occurrences** High learning group (N = 25)	Medium learning group (N = 35)	Low learning group (N = 16)	**Overall support (N = 76)**	η^2	**H**	**r** low-medium	low-high	medium-high
SP31	AI =>NI =>GI =>AC	0	0	0.31	<0.20	0.24	19.81 *	0.48	0.46	N/A
SP32	AD =>NB =>NC =>GH+all	0	0	0.25	<0.20	0.19	15.63 *	0.43	0.41	N/A
SP33	GC =>NC =>AC	0.04	0.09	0.38	<0.20	0.10	9.01 *	0.33	0.37	0.04
SP34	GI =>GW =>GI+2	0.24	0.06	0	<0.20	0.07	7.46 *	0.14	0.33	0.26
SP35	GI+2 =>GA	0.28	0.26	0.56	0.24	0.07	6.83 *	0.33	0.32	0.01
SP36	AD =>NC =>AI+3+	0.12	0.69	0.50	0.30	0.07	6.99 *	0.1	0.25	0.34
SP37	GH+all =>NB =>NB =>NC =>AC	0.08	0.03	0.25	<0.20	0.06	6.42 *	0.34	0.23	0.12

* $p < 0.05$ *before* the Benjamini and Yekutieli correction.

In fact, all of the seven sequential patterns presented in Table 7.8 show relatively low average number of occurrences and overall support when compared to sequential patterns identified in previous analyses (see Table 7.6). In general, there were no clear trends differentiating the occurrence of meaningful patterns between the different learning groups, and we were not able to draw any meaningful conclusions about how specific patterns related to how students use guidance might relate to improved learning gains.

The results suggest that when students are categorized based on their learning gain, no conclusions can be drawn about differential use of guidance when doing calculations. In other words, students with a high learning gain did not use the guidance in a different way than students with a low learning gain. A possible explanation for this result is that the three groups of learning are heterogeneous in terms of students' prior knowledge. For example, the medium learning group is a cluster of students who initially were categorized in the high, medium, or low prior knowledge cluster (see Table 7.4). Hence, prior knowledge may be a better indicator to explain differences in the students' behaviour when doing calculations. Therefore, we also investigated whether there are significant different patterns in relation to students' prior knowledge.

7.3.4 Use of guidance in relation to prior knowledge

After splitting the participants in three groups of high, medium, and low prior knowledge, 142 sequential patterns were found with a support larger than 0.2 in at least one of the three groups. No statistically significant differences were found after the Benjamini and Yekutieli correction, but 29 patterns were significant before the correction and had medium to high effect sizes (η^2 greater or equal to 0.06). Of these 29 patterns, we show the 11 patterns that had an overall support of greater than 0.2, since patterns with a low overall support (smaller than 0.2) were interpreted as not meaningful (Table 7.9). These 11 patterns were investigated to identify insights into possible relationships between prior knowledge and use of guidance when performing calculations within the learning environment.

What stands out is that the differentially used sequential patterns include all types of guidance that could potentially be used. In terms of the prior knowledge groups, the results suggest that there is no difference between use of guidance for students with high or medium prior knowledge. However, the average occurrences of all sequential patterns for students in the low prior knowledge group is higher compared to the other two groups. The trend suggests that students with a low prior knowledge used more guidance compared to students with a higher prior knowledge. However, additional data should be collected to make sure that these trends can be observed in other datasets and are not due to the random chance of false discovery caused by the high number of discovered patterns.

Table 7.9: Sequential patterns that were used differentially by students with high, medium, and low prior knowledge, and corresponding effect sizes.

Code	**Sequential pattern**	**Average occurrences**			**Overall support (N = 77)**	η^2	**H**	**r**		
		High prior knowledge group (N = 21)	Medium prior knowledge group (N = 37)	Low prior knowledge group (N = 19)				low-medium	low-high	medium-high
SP4	GH+	2.48	2.51	4.00	0.90	0.08	7.68*	0.36	0.36	0.03
SP6	GH+all	2.38	2.24	4.33	0.87	0.06	6.41*	0.33	0.32	0.02
SP7	GH	2.71	2.73	4.67	0.86	0.08	7.58*	0.32	0.40	0.08
SP14	GI+3+	2.76	1.84	5.44	0.70	0.07	7.19*	0.35	0.25	0.15
SP18	NB =>NC =>NI	0.67	1.03	2.39	0.58	0.12	10.8*	0.32	0.49	0.2
SP38	AC =>GC	0.48	1.05	2.17	0.51	0.11	10.2*	0.31	0.49	0.19
SP39	NB =>NC =>AI =>AC	0.19	0.65	0.78	0.39	0.06	6.42*	0.09	0.40	0.27
SP40	NB =>NC =>GI	0.29	0.51	1.11	0.34	0.08	8.05*	0.27	0.43	0.15
SP41	NB =>NC =>GI+2	0.33	0.08	0.61	0.23	0.12	11.0*	0.44	0.17	0.32
SP42	NB =>NC =>NI =>GI	0.10	0.19	0.72	0.22	0.09	8.41*	0.31	0.40	0.10
SP43	NI =>GH	0.14	0.19	0.50	0.22	0.06	6.58*	0.30	0.34	0.03

* $p < 0.05$ *before* the Benjamini and Yekutieli correction.

7.4 Limitations

SPM algorithms require researchers to predetermine some parameters, such as the minimum support. The predetermined parameters are similar to the hyperparameters of machine learning models, but, currently, there is no way to tune the predetermined parameters in SPM as tuning the hyperparameters in machine learning. Thus, this study chose the minimum support on the basis of the trade-off between including more interesting learning behaviour patterns and excluding less meaningless patterns. We acknowledge that the decision might be a little subjective. There needs to be a guideline for the parameter setting of SPM.

Studies that applied SPM into educational data have considered the time factor (Dermy and Brun, 2020; Emara et al., 2018). For instance, Emara et al. (2018) added long or short suffixes to reading actions based on whether a read is longer than 3 seconds. However, the LabBuddy system recorded the time of action log at the minute level, and thus, we were unable to compute the duration of each action. As a result, it could be that similar sequential patterns should have been interpreted differently. For example, students who make three consecutive incorrect attempts could just be guessing (gaming the system), or they could have tried to do the calculations, and turned out to have made a mistake each time.

The studied learning task was part of a larger virtual experiment environment. Since the virtual experiment environment was designed with the purpose to provide students with the best possible learning experience, rather than with the purpose to investigate how students use support while doing calculations (which is only a part of the assignment in the VEE), the context for this study was not ideal. In the ideal situation, the learning task would be extracted from the larger virtual experiment environment, and it would have been more structured in terms of the order in which students have to progress through the learning task. In the current situation, students are free to determine the order in which they do the calculations. Since there are three question types, with each multiple questions, it is hard to control for what students might or might not have previously learned in the other question types.

7.5 Conclusions

The goal of this study was to investigate students' behaviour regarding the use of guidance while doing calculations, and to relate this behaviour to learning. Using sequential pattern mining, we were able to find meaningful patterns of student behaviour. Students made a lot of use of the guidance provided in the learning task. Students who used the option to check their intermediate calculations were more likely to make a successful first attempt to complete the calculation when compared to students who did not use this guidance. This effect was not found for the use of guidance in the form of hints. Although we were not able to quantify the usefulness of the hints in this study, the hints were used a lot, so it seems that many students find them a useful source of guidance. Guidance in the form of worked examples was found to have a positive contribution to students' success at first attempt, which was also expected based on literature (Clark and Mayer, 2016). To improve the guidance, worked examples should also be included for the other question types. Despite the large amount of guidance used by students and its established contribution to students' success when they first submit the answer to a calculation, we were unable to find a relation between learning gain and students' use of guidance. When exploring the relation between students' use of guidance and their prior knowledge, we found a trend that students with a low prior knowledge used more guidance compared to students with a higher prior knowledge. However, additional research should be done to confirm this result.

When designing the learning task that was the subject of this study, we used among others the design principles "provide access to hints (one by one) that guide the students' thinking process" and "Provide the opportunity for students to check their intermediate calculations" (chapter 3). These design principles were based on our experience with designing digital learning materials (Diederen et al., 2003; Diederen et al., 2006b; Van der Kolk et al., 2012; Van der Kolk et al., 2013; this thesis). Despite the potential and relevance of such guidance in digital learning materials, the use of hints and intermediate calculations are not specifically investigated in educational theories. The results of this study are a first indication that learning materials designed based on these two specific design principles indeed contribute to enabling students to independently complete calculations in a digital learning task.

Chapter 8

General discussion

Contents

In this thesis an educational design research (EDR) approach was used to answer the central research question: How to design a virtual experiment environment (VEE) that helps students to prepare for laboratory classes? The EDR approach resulted in both *theoretical knowledge* to answer the central research question, as well as *practical solutions* to solve the complex educational problem. The theoretical knowledge was captured in what was defined as a 'blueprint to design VEEs' (chapters 3 and 5). The practical solutions are the VEEs that were designed and improved while cycling through the EDR approach.

Throughout the design process, I have attempted (and sometimes succeeded) to establish collaborations between university teachers, educational designers, and educational scientists. Let me first define these three types of people.

When I look around in my own department, the laboratory of food chemistry, the vast majority of the 'research and teaching staff' has obtained a PhD in the field of food chemistry or related. Besides doing food chemistry research, they are also responsible for coordinating and teaching bachelor and master courses, a job for which they have not been educated. This is what I refer to as *university teachers*: scientists who also teach in their field of research.

An *educational designer* is someone who designs solutions for complex educational problems. They have typically studied the field of educational design research, so they know how to approach complex educational problems and have experience in designing solutions. Educational designers are rarely found in universities.

An *educational scientist* is someone who does research in which they seek to describe and understand how people learn. Given the context of my dissertation, this could be done by investigating student behaviour and student learning in a learning material that was designed by an educational designer, who in turn did that to solve a university teachers' complex educational problem.

Over the years, I learned that these three types of people approach problems from different perspectives. These different perspectives can result in challenges and long discussions, which, as long as we keep respecting each other's perspectives, may lead to new insights. In the following sections I will discuss my findings from those different perspectives.

8.1 The university teacher perspective

Collaboration is they key to better education

One of the challenges of university level teaching, as opposed to teaching in primary and secondary education, is that there are no ready-made methods and materials that university teachers can use. To make the life of university teachers even more difficult, the technological innovations in the last decades have changed their role from distributors of knowledge (e.g. through lectures), to being designers of learning experiences (Mor and Craft, 2012). Despite university teachers' efforts to provide the best education they can, they are typically not trained to design learning experiences. As a result, many learning experiences are designed by intuition (Cen et al., 2006; Koedinger et al., 2013), which does not necessarily result in high quality learning materials. As a supplement for using intuition, and to improve the quality of (new) learning materials, university teachers should engage in the field of educational design research. But knowing that university teachers often do not have the time to do so, they should collaborate with educational designers.

Laboratory classes can (partially) be replaced

The trigger question of my research project was: To what degree can laboratory classes be replaced by other learning activities? In the first chapter, I wrote that it is not doable to replace laboratory classes, since almost all university teachers indicated one of the goals of their laboratory education to be 'obtaining hands-on experience'. However, due to the COVID-19 pandemic, we learned that it is possible to (re-)design courses in such a way that laboratory work can (partially) be replaced. On top of that, it seems that many of the newly designed course materials (including some VEEs) in my department are there to stay even after the pandemic. The Journal of Chemical Education dedicated a special issue to the insights gained while teaching chemistry in the time of COVID-19, which contains many articles involving the (re-)design of laboratory education (Holme, 2020).

The blueprint, although targeted to design VEEs that *prepare* students, was used multiple times to design VEEs that were intended to (partially) *replace* laboratory classes. I would suggest to create an extended version of the current blueprint for the purpose of (partially) replacing laboratory classes. By design, the current blueprint focuses on the pre- and post-experiment phase, while mostly ignoring the experiment phase itself (see Figure 3.1, p. 28). For an extended version of the blueprint, additional visualization of the (chemical) experiments should be considered, for example by adding interactive video materials (Mitrovic et al., 2017; Vural, 2013).

So, although university teachers generally indicate their laboratory classes cannot be (partially) replaced due to the need for students to obtain hands-on experience, it turns out this issue was smaller than expected. These findings could (or should) be the start of a new era. An era in which university teachers and programme directors question the need for laboratory classes, and consider alternatives such as VEEs instead. Ultimately, this should result in a better balance between the investment in and the outcomes of laboratory classes.

8.2 The educational designer perspective

Design principles

In the process of designing a new learning material, a large number of design choices have to be made. The educational design research approach can provide structure to this process, and will ultimately and among others lead to a list of design principles (DPs) related to the design goal. Educational designers (with a related design goal) can use such a list of DPs as advice in their decision making process. However, following up all DPs is no guarantee for the design of a successful learning material. For example, my DP13: "Provide access to hints (one by one) that guide the students' thinking process", leaves a lot of room for interpretation. In which situations do you provide hints? How many hints do you provide? What should be the content of the hints? The answers to these questions are context-specific, and are hence not part of the blueprint. In the end, the educational designer makes design choices based on relevant advice (such as DPs), experience, and intuition. Since the design choices are only partially informed by DPs, the relation between the DPs and the final design remains vague. It is up to educational designers to put a little more effort in communicating the rationale behind the design choices they make. The recently established journals: Educational Design Research and the Journal of Formative Design in Learning, can hopefully provide a platform for an increasing number of publications that introduce and discuss the rationales behind design choices.

Based on the design of multiple VEEs, fourteen DPs were described in the blueprint. I am convinced that I forgot to add several 'hidden DPs'. These hidden DPs could for example follow from the design choices that I made subconsciously, or were induced by the platform in which I built my designs (which was based on design principles as well). For example, the appearance of many elements and the location of most feedback in the VEEs were predetermined by the platform. The best way to identify these hidden DPs, would be to have the blueprint applied by other educational designers, who use a different platform to build the VEE. Based on the findings of these educational designers, the blueprint can be fine-tuned, so that it can provide even better advice in the future.

Support system

To enable students to complete the assignment in a VEE independently (design requirement 3), they should be provided with support such as procedural information, formative feedback, hints, and the opportunity to check intermediate calculations (design principles 11 – 14), which are the kinds of support that would traditionally be provided by a teacher. Advantages of designing a system that provides such support are: i) students are provided with just-in-time support by the system, ii) students can access support at any time they need it (i.e. they do not have to wait for a teacher), and iii) the quality of the information and support is the same for all students (which is often not the case during tutorials in which student assistants are helping the university teacher). Disadvantages are that i) students may need unforeseen support (which a teacher would be able to provide), ii) the system does not 'know' and cannot ask where exactly a student gets stuck (i.e. the system provides a lot of general support, rather than specific for what a student might need), and iii) students can engage in gaming the system behaviour (i.e. (ab)using the available support to complete the assignment as soon as possible, rather than using it to learn) (Baker et al., 2008). Before the COVID-19 pandemic, students would work in the VEEs in a computer room on campus, for which we established a workable teacher to student ratio of 1:50. Depending on the quality of the designed learning material, teachers can be replaced to a large degree. Nevertheless, it remains important to always have some supervision, since besides answering students' questions that remain unanswered by the system, this provides the opportunity for teachers to learn in which places many students get stuck, which provides opportunities for fine-tuning of the designed learning materials.

Student responsibility

Another implication of design requirement 3, to enable students to complete the assignment in the VEE independently, is that students are given the responsibility to decide how much support they want to use. This responsibility can be handled by students who are academically committed, interested, and want to do well (student type 1). But students who are not curious about a particular subject, have no ambition to excel, and are just at university to obtain a qualification for a job (student type 2) (Biggs, 1999), will handle this responsibility differently. Over the last decades, the student intake at universities has strongly increased, which resulted in an increase of type 2 students (Biggs, 1999). When it turned out many students were engaging in gaming the system behaviour (chapter 6), I adapted the support system, to make it less easy to request answers from the system. Less easy means: I removed all references to the possibility to request answers, and introduced the option for students to check their intermediate calculations. Only after an incorrect intermediate value is submitted, the student can ask the support system to provide the correct intermediate value (which requires extra clicks and more effort). Only if students submit all correct intermediate values (which they could have requested) as well as the incorrect final answer, the system can be asked to provide the final answer. All this was done to make the life of student type 2 a little harder, hoping that they would engage in active learning rather than gaming the system, while at the same time introducing more and helpful support for student type 1. In terms of student type 2, I think this approach was successful to some degree, as the number of answers requested from the system sharply decreased: Before the adaptation, students requested an average of 35 final answers from the system, whereas after the adaptation this was on average 8 intermediate or final answers per student.

I believe that teachers and educational designers should focus on students who are actually willing to learn, so should target for student type 1. Nevertheless, I have three design suggestions that could potentially help to get student type 2 motivated to actively engage with the VEE: i) at the start of the assignment in the VEE, extra information could be provided on what students will gain by completing the assignment in a 'good' way, ii) the VEE could provide just-in-time information such as: 'for your learning process it would be best to try yourself before opening the hints', and iii) game mechanics such as storytelling, badges, points could be introduced (Kalogiannakis et al., 2021). Such additions or changes to the design could potentially make a big difference for many students.

8.3 The educational scientist perspective

In chapters 6 and 7, I used a VEE as the context for investigation in the field of learning sciences, looking at self-regulated learning (SRL) and student behaviour in relation to learning using sequential pattern mining. With these studies I thoroughly evaluated the designed and implemented VEEs, based on which their design, and in extension the blueprint, was improved (e.g. based on the evaluation described in chapter 6, the option to check intermediate calculations was incorporated in the design). However, I have some doubts regarding the implications of the outcomes of these studies in different contexts.

Learning materials as context

The learning materials that are used as the context for investigating a construct such as SRL are often poorly described in literature: the provided details of the learning material are often insufficient to truly understand the content, depth, nature and level of the learning material. For example, in my own chapters 6 and 7, I only provide a very general overview of the learning material, which means that the reader will not truly understand how the system works. I believe that it is very hard for any reader to interpret the outcomes of such a study, and translate this knowledge to a different situation, without having a proper understanding of the learning material itself. This is why I suggest for any future publications that use a learning material as the context for their study, that a detailed showcase of the learning material should be provided, either be in the same paper, or as a separate piece of work. This is also why I published a showcase of the VEE (chapter 4).

Measuring students' level of self-regulated learning

After considering several instruments that measure students' level of self-regulation (chapter 6), we chose to use the self-regulation scale of the 'inventory of learning styles' questionnaire (Vermunt, 1998), since it was developed in a university education context, and it was validated in many studies (Vermunt and Donche, 2017). This questionnaire is over 20 years old, which means that the statements in the questionnaire were made at a time when computers did not play a large role in education. In Table 8.1, the relevance of each item in the self-regulation scale in the context of VEEs is discussed. Out of the seven items, five are not relevant or unlikely to happen in the context of VEEs. So, the behaviour based on which a student scores high on this questionnaire, is related to general study behaviour, and not to behaviour that could be expected for completing tasks such as in the VEEs. Given that to the best of my knowledge there is no better way of measuring SRL in the context of VEEs, perhaps it would have been better to look at another construct that is more closely related to student behaviour in VEEs. Despite all the previous, we were able to draw meaningful conclusions and improve the design of the VEE. One year later, I engaged in the next cycle of the EDR approach, and repeated the exact same procedure with a new group of students and with the improved VEE. We were unable to replicate the results and corresponding conclusions of the previous year. This could potentially be explained by one or a combination of the following factors: i) a different group of students, ii) the improvements made to the VEE, and/or iii) the relevance of using SRL in relation to the type of learning material (VEE).

Table 8.1: All items in the self-regulation scale of the inventory of learning styles questionnaire, with corresponding relevance in the context of VEEs.

Item in the self-regulation scale	Relevance in the context of VEEs
To test my learning progress when I have studied a textbook, I try to formulate the main points in my own words.	This is not relevant for VEEs, since the VEEs contain only short pieces of text, usually presented near corresponding interactive questions.
When I start reading a new chapter or article, I first think about the best way to study it.	This is not relevant for VEEs, since there is no reading of chapters or articles to be done.
When I have difficulty grasping a particular piece of subject matter, I try to analyse why it is difficult for me.	This could happen in the context of a VEE.
To test my learning progress, I try to answer questions about the subject matter which I make up myself.	This is unlikely to happen in a VEE, since the VEE contains many interactive questions to test students' learning process. Moreover, the student is working on a task, not studying for a test.
To test whether I have mastered the subject matter, I try to think up other examples and problems besides the ones given in the study materials or by the teacher.	This could happen in the context of a VEE. Although, again, the student is working on a task, not studying for a test.
To test my own progress, I try to describe the content of a paragraph in my own words.	This is not relevant for VEEs. They contain very few paragraphs of text, as opposed to for example textbooks.
When I am studying, I also pursue learning goals that have not been set by the teacher but by myself.	This is unlikely to happen in a VEE, since the learning goals are predefined, and all information that is not required to successfully complete the learning task is left out (design principle 4).

Learning gain

In general, educational scientists are fond of the metric 'learning gain'. Besides the discussion on how to group students, the true meaning of learning gain should also be a point of discussion. Imagine student A, who scored 10% on the pre-test and 20% on the post-test, and student B, who scored 85% on the pre-test, and 90% on the post-test. Looking at absolute values, the conclusion would be that student A learned a little more than student B, and when splitting the population into three groups, both students would probably end up in the category low learning. It is obvious that student B had much less potential to learn, and that when looking at the absolute values, student A did not perform better than student B. In my university context, I repeatedly found the differences in students' prior knowledge to vary between approximately 0% and 85% of the maximum possible score. If the differences in students' prior knowledge are as large as in my context, one should wonder if it is valuable to look at learning gain in the first place.

In chapter 7 we used a different approach to correct for differences in prior knowledge. We created two sets of three clusters of students through k-means clustering, one set of three clusters based on pre-test scores and one set of three clusters based on post-test scores. Students were assigned to the category low, medium, or high learning, based on whether they dropped one cluster between the pre- and post-test scores (e.g. they were categorized in the medium cluster based on the pre-test, and they were categorized in the low cluster based on the post-test) remained in the same cluster, or increased a cluster (See chapter 7). The trouble that remains is that students who start in the highest cluster, can never improve to a higher cluster, and vice versa for students who start in the lowest cluster. In conclusion, the use of the metric learning gain can only be meaningful if the variation in students' scores on the pre-test is limited. But if this is not the case, I consider the metric learning gain of limited or no value.

8.4 Concluding remark

After several years of educational design research, I developed a blueprint that can provide university teachers and educational designers with advice on how to design virtual experiment environments that help students to prepare for laboratory classes. I hope that an increasing number of collaborations between university teachers, educational designers, and educational scientists will be established, leading to the best possible education for future generations.

References

Abdulwahed, M. and Z. K. Nagy (2009). "Applying Kolb's experiential learning cycle for laboratory education". *Journal of Engineering Education* 98, 283–294. DOI: https://doi.org/10.1002/j.2168-9830.2009.tb01025.x.

Agrawal, R. and R. Srikant (1995). "Mining sequential patterns". In: *Proceedings of the eleventh international conference on data engineering*, 3–14. DOI: https://doi.org/10.1109/ICDE.1995.380415.

Agustian, H. Y. and M. K. Seery (2017). "Reasserting the role of pre-laboratory activities in chemistry education: A proposed framework for their design". *Chemistry Education Research and Practice* 18, 518–532. DOI: https://doi.org/10.1039/C7RP00140A.

Agustian, H. Y. (2020). "Students' learning experience in the chemistry laboratory and their views of science: In defence of pedagogical and philosophical validation of undergraduate chemistry laboratory education". Thesis. DOI: http://dx.doi.org/10.7488/era/107.

Aldowah, H., H. Al-Samarraie, and W. M. Fauzy (2019). "Educational data mining and learning analytics for 21st century higher education: A review and synthesis". *Telematics and Informatics* 37, 13–49. DOI: https://doi.org/10.1016/j.tele.2019.01.007.

Alkhaldi, T., I. Pranata, and R. I. Athauda (2016). "A review of contemporary virtual and remote laboratory implementations: observations and findings". *Journal of Computers in Education* 3, 329–351. DOI: https://doi.org/10.1007/s40692-016-0068-z.

Anderson, T. and J. Shattuck (2012). "Design-based research: A decade of progress in education research?" *Educational Researcher* 41, 16–25. DOI: https://doi.org/10.3102/0013189x11428813.

Ayres, P. (2006). "Impact of reducing intrinsic cognitive load on learning in a mathematical domain". *Applied Cognitive Psychology* 20, 287–298. DOI: https://doi.org/10.1002/acp.1245.

Azevedo, R. (2015). "Defining and measuring engagement and learning in science: conceptual, theoretical, methodological, and analytical issues". *Educational Psychologist* 50, 84–94. DOI: https://doi.org/10.1080/00461520.2015.1004069.

Baker, R. (2010). "Data mining for education". In: *International encyclopedia of education*. Ed. by B. McGaw, P. Peterson, and E. Baker. Vol. 7. Elsevier Ltd., 112–118.

Baker, R., S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser (2010). "Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments". *International Journal of Human-Computer Studies* 68, 223–241. DOI: https://doi.org/10.1016/j.ijhcs.2009.12.003.

Baker, R., J. Walonoski, N. Heffernan, I. Roll, A. Corbett, and K. Koedinger (2008). "Why students engage in "gaming the system" behavior in interactive learning environments". *Journal of Interactive Learning Research* 19, 185–224.

Bakker, A. (2018). *Design research in education: a practical guide for early carreer researchers*. New York, NY: Routledge.

Bandura, A. (1986). "Social foundation of thought and action: a social-cognitive view", Englewood Cliffs, NY: Prentice–Hall.

Bannert, M. (2002). "Managing cognitive load—recent trends in cognitive load theory". *Learning and Instruction* 12, 139–146. DOI: https://doi.org/10.1016/S0959-4752(01)00021-4.

Belland, B. R. (2011). "Distributed cognition as a lens to understand the effects of scaffolds: The role of transfer of responsibility". *Educational Psychology Review* 23, 577–600. DOI: https://doi.org/10.1007/s10648-011-9176-5.

Benjamini, Y. and D. Yekutieli (2001). "The control of the dalse discovery rate in multiple testing under dependency". *The Annals of Statistics* 29, 1165–1188. DOI: https://www.jstor.org/stable/2674075.

Bereiter, C. (2014). "Principled practical knowledge: Not a bridge but a ladder". *Journal of the Learning Sciences* 23, 4–17. DOI: https://doi.org/10.1080/10508406.2013.812533.

Biggs, J. (1999). "What the student does: Teaching for enhanced learning". *Higher Education Research & Development* 18, 57–75. DOI: https://doi.org/10.1080/0729436990180105.

Bjork, R. A., J. Dunlosky, and N. Kornell (2013). "Self-regulated learning: beliefs, techniques, and illusions". *Annual Review of Psychology* 64, 417–444. DOI: https://doi.org/10.1146/annurev-psych-113011-143823.

Boekaerts, M. and L. Corno (2005). "Self-regulation in the classroom: a perspective on assessment and intervention". *Applied Psychology* 54, 199–231. DOI: https://doi.org/10.1111/j.1464-0597.2005.00205.x.

Bowers, W. G. (1924). "The advantages of laboratory work in the study of elementary chemistry". *Journal of Chemical Education* 1, 205. DOI: https://doi.org/10.1021/ed001p205.

Brahm, T. (2017). "Design-based research in the context of transitioning to VET: Developing interventions through research-practice collaboration". *Educational Design Research* 1. DOI: https://doi.org/10.15460/eder.1.2.1163.

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. New York, NY: Springer Science & Business Media.

Bretz, S. L. (2019). "Evidence for the importance of laboratory courses". *Journal of Chemical Education* 96, 193–195. DOI: https://doi.org/10.1021/acs.jchemed.8b00874.

Bretz, S. L., M. Fay, L. B. Bruck, and M. H. Towns (2013). "What faculty interviews reveal about meaningful learning in the undergraduate chemistry laboratory". *Journal of Chemical Education* 90, 281–288. DOI: https://doi.org/10.1021/ed300384r.

Brinson, J. R. (2015). "Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research". *Computers & Education* 87, 218–237. DOI: https://doi.org/10.1016/j.compedu.2015.07.003.

Broadbent, J. and M. Fuller-Tyszkiewicz (2018). "Profiles in self-regulated learning and their correlates for online and blended learning students". *Educational Technology Research and Development*, 1–21. DOI: https://doi.org/10.1007/s11423-018-9595-9.

Broadbent, J. and W. Poon (2015). "Self-regulated learning strategies & academic achievement in online higher education learning environments: a systematic review". *Internet and Higher Education* 27, 1–13. DOI: https://doi.org/10.1016/j.iheduc.2015.04.007.

Bruck, A. D. and M. Towns (2013). "Development, implementation, and analysis of a national survey of faculty goals for undergraduate chemistry laboratory". *Journal of Chemical Education* 90, 685–693. DOI: https://doi.org/10.1021/ed300371n.

Brydges, R., J. Manzone, D. Shanks, R. Hatala, S. J. Hamstra, B. Zendejas, and D. A. Cook (2015). "Self-regulated learning in simulation-based training: A systematic review and meta-analysis". *Medical Education* 49, 368–378. DOI: https://doi.org/10.1111/medu.12649.

Buchta, C. and M. Hahsler (2020). "Package arulesSequences". DOI: https://CRAN.R-project.org/package=arulesSequences.

Budai, T. and M. Kuczmann (2018). "Towards a modern, integrated virtual laboratory system". *Acta Polytechnica Hungarica* 15, 191–204.

Buntine, M. A., J. R. Read, S. C. Barrie, R. B. Bucat, G. T. Crisp, A. V. George, I. M. Jamie, and S. H. Kable (2007). "Advancing chemistry by enhancing learning in the laboratory (ACELL): a model for providing professional and personal development and facilitating improved student laboratory learning outcomes". *Chemistry Education Research and Practice* 8, 232–254. DOI: https://doi.org/10.1039/B6RP90033J.

Butterfield, B. and J. Metcalfe (2001). "Errors committed with high confidence are hypercorrected". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27, 1491–1494. DOI: https://doi.org/10.1037/0278-7393.27.6.1491.

Carnduff, J. and N. Reid (2003). *Enhancing undergraduate chemistry laboratories: pre-laboratory and post-laboratory exercises*. London, UK: Royal Society of Chemistry.

Cen, H., K. Koedinger, and B. Junker (2006). "Learning factors analysis – a general method for cognitive model evaluation and improvement". In: *International Conference on Intelligent Tutoring Systems*. Intelligent Tutoring Systems. Springer Berlin Heidelberg, 164–175. DOI: https://doi.org/10.1007/11774303_17.

Chen, C.-M. and W.-F. Wang (2020). "Mining effective learning behaviors in a web-based inquiry science environment". *Journal of Science Education and Technology* 29, 519. DOI: https://doi.org/10.1007/s10956-020-09833-9.

Chen, C. (2002). “Self-regulated learning strategies and regulated learning strategies and achievement in an introduction to information systems course”. *Information Technology, Learning, and Performance Journal* 20, 11–25.

Chen, W. and T. C. Reeves (2019). “Twelve tips for conducting educational design research in medical education”. *Medical Teacher*, 1–7. DOI: https://doi.org/10.1080/0142159X.2019.1657231.

Cho, M.-H. and B. J. Kim (2013). “Students’ self-regulation for interaction with others in online learning environments”. *Internet and Higher Education* 17, 69–75. DOI: https://doi.org/10.1016/j.iheduc.2012.11.001.

Clark, R. C. and R. E. Mayer (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. Hoboken, NJ: John Wiley & Sons.

Coates, H., R. James, and G. Baldwin (2005). “A critical examination of the effects of learning management systems on university teaching and learning”. *Tertiary Education and Management* 11, 19–36. DOI: https://doi.org/10.1080/13583883.2005.9967137.

Cohen, J. (1988). “The effect size index: d. Statistical power analysis for the behavioral sciences”. *Abingdon-on-Thames: Routledge Academic*.

Colthorpe, K., K. Zimbardi, L. Ainscough, and S. Anderson (2015). “Know thy student! Combining learning analytics and critical reflections to develop a targeted intervention for promoting self-regulated learning”. *Journal of Learning Analytics* 2, 134–155. DOI: https://doi.org/10.18608/jla.2015.21.7.

Crandall, P. G., R. K. Engler III, D. E. Beck, S. A. Killian, C. A. O’Bryan, N. Jarvis, and E. Clausen (2015). “Development of an augmented reality game to teach abstract concepts in food chemistry”. *Journal of Food Science Education* 14, 18–23. DOI: https://doi.org/10.1111/1541-4329.12048.

Cronje, J. (2013). “What is this thing called “design” in instructional design research?—the ABC instant research question generator”. In: *Media in education*. Ed. by A. Moreira, O. Benavides, and A. J. Mendes. New York, NY: Springer New York, 15–28. DOI: 10.1007/978-1-4614-3175-6_2.

Cutumisu, M., K. P. Blair, D. B. Chin, and D. L. Schwartz (2015). “Posterlet: a game-based assessment of children’s choices to seek feedback and to revise”. *Journal of Learning Analytics* 2, 49–71. DOI: https://doi.org/10.18608/jla.2015.21.4.

De Bruin, A. B. H. and J. J. G. Van Merriënboer (2017). “Bridging Cognitive Load and Self-Regulated Learning Research: A complementary approach to contemporary issues in educational research”. *Learning and Instruction*. DOI: http://dx.doi.org/10.1016/j.learninstruc.2017.06.001.

De Jong, T. and W. R. Van Joolingen (1998). “Scientific discovery learning with computer simulations of conceptual domains”. *Review of educational research* 68, 179–201. DOI: https://doi.org/10.3102/00346543068002179.

Dean, C. B., E. R. Hubbell, H. Pitler, and B. Stone (2012). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

DeKorver, B. K. and M. H. Towns (2015). “General chemistry students’ goals for chemistry laboratory coursework”. *Journal of Chemical Education* 92, 2031–2037. DOI: https://doi.org/10.1021/acs.jchemed.5b00463.

Dermy, O. and A. Brun (2020). “Can we take advantage of time-interval pattern mining to model students activity?” In: *The 13th International Conference on Educational Data Mining*. EDM 2020. Ifrane, Marocco: Association for Computing Machinery.

Diederen, J. (2005). “Design and evaluation of digital activating learning materials for Food Chemistry education”. Thesis.

Diederen, J., H. Gruppen, R. J. Hartog, and A. G. Voragen (2006a). “Design and evaluation of digital assignments on research experiments within food chemistry”. *Journal of Science Education and Technology* 15, 227. DOI: https://doi.org/10.1007/s10956-006-9010-x.

Diederen, J., H. Gruppen, R. Hartog, G. Moerland, and A. G. Voragen (2003). “Design of activating digital learning material for food chemistry education”. *Chemistry Education Research and Practice* 4, 353–371. DOI: https://doi.org/10.1039/B3RP90020G.

Diederen, J., H. Gruppen, R. Hartog, and A. G. Voragen (2006b). “Design and evaluation of digital assignments on research experiments within food chemistry”. *Journal of Science Education and Technology* 15.3-4, 227–246.

Domin, D. S. (1999). “A review of laboratory instruction styles”. *Journal of Chemical Education* 76, 543–547. DOI: https://doi.org/10.1021/ed076p543.

Dunn, R., J. Hattie, and T. Bowles (2019). “Exploring the experiences of teachers undertaking Educational Design Research (EDR) as a form of teacher professional learning”. *Professional Development in Education* 45, 151–167. DOI: https://doi.org/10.1080/19415257.2018.1500389.

Dybå, T. and T. Dingsøyr (2008). “Empirical studies of agile software development: A systematic review”. *Information and software technology* 50, 833–859. DOI: https://doi.org/10.1016/j.infsof.2008.01.006.

Dym, C. L., P. Little, E. J. Orwin, and E. Spjut (2014). *Engineering design: A project-based introduction*. Hoboken, NJ: John Wiley & Sons.

Emara, M., R. Rajendran, G. Biswas, M. Okasha, and A. A. Elbanna (2018). “Do students’ learning behaviors differ when they collaborate in open-ended learning environments?” 2.CSCW. DOI: https://doi.org/10.1145/3274318.

Euler, D. and M. Collenberg (2018). “Design-based research in economic education”. *Educational Design Research* 2. DOI: https://doi.org/10.15460/eder.2.2.1271.

Faulconer, E. K. and A. B. Gruss (2018). “A review to weigh the pros and cons of online, remote, and distance science laboratory experiences”. *International Review of Research in Open and Distributed Learning* 19, 155–168. DOI: https://doi.org/10.19173/irrodl.v19i2.3386.

Fritz, C. O., P. E. Morris, and J. J. Richler (2012). “Effect size estimates: Current use, calculations, and interpretation”. *Journal of Experimental Psychology: General* 141, 2. DOI: https://doi.org/10.1037/a0024338.

Gart, J. J. and J. R. Zweifel (1967). “On the bias of various estimators of the logit and Its variance with application to quantal bioassay”. *Biometrika* 54, 181–187. DOI: https://doi.org/10.2307/2333861.

Gašević, D., S. Dawson, T. Rogers, and D. Gasevic (2016). “Learning analytics should not promote one size fits all: the effects of instructional conditions in predicting academic success”. *Internet and Higher Education* 28, 68–84. DOI: https://doi.org/10.1016/j.iheduc.2015.10.002.

Gasevic, D., S. Dawson, and G. Siemens (2015). “Let’s not forget: Learning analytics are about learning”. *TechTrends* 59, 64–71. DOI: https://doi.org/10.1007/s11528-014-0822-x.

Graesser, A. C., X. Hu, and R. Sottilare (2018). “Intelligent tutoring systems”. In: *International Handbook of the Learning Sciences*. Ed. by F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, and P. Reimann. New York, NY: Routledge, 246–255.

Greene, J. A. and R. Azevedo (2010). “The measurement of learners’ self-regulated cognitive and metacognitive processes while using computer-based learning environments”. *Educational Psychologist* 45, 203–209. DOI: https://doi.org/10.1080/00461520.2010.515935.

Gregory, S.-J. and G. Di Trapani (2012). “A blended learning approach to laboratory preparation”. *International Journal of Innovation in Science and Mathematics Education* 20.

Greller, W. and H. Drachsler (2012). “Translating learning into numbers: A generic framework for learning analytics”. *Journal of Educational Technology & Society* 15, 42–57. DOI: https://www.jstor.org/stable/pdf/jeductechsoci.15.3.42.pdf.

Hadwin, A. F., S. Järvelä, and M. Miller (2011). “Self-regulated, co-regulated, and socially shared regulation of learning”. In: *Handbook of Self-regulation of Learning and Performance*. Ed. by B. J. Zimmerman and D. H. Schunk. Vol. 30. New York, NY: Taylor & Francis, 65–84.

Hadwin, A. F., J. C. Nesbit, D. Jamieson-Noel, J. Code, and P. H. Winne (2007). “Examining trace data to explore self-regulated learning”. *Metacognition and Learning* 2, 107–124. DOI: https://doi.org/10.1007/s11409-007-9016-7.

Hadwin, A. F., L. Wozney, and O. Pontin (2005). “Scaffolding the appropriation of self-regulatory activity: a socio-cultural analysis of changes in teacher–student discourse about a graduate research portfolio". *Instructional Science* 33, 413–450. DOI: https://doi.org/10.1007/s11251-005-1274-7.

Hartog, R. J., A. J. Beulens, and J. Tramper (2012). "Digital learning resources in higher education: Designing for large-scale use". *Journal of Universal Computer Science* 18, 2274–2300.

Hartog, R. J., H. Van der Schaaf, A. J. Beulens, and J. Tramper (2009). "Virtual experiments in university education". In: *Looking toward the future of technology-enhanced education: Ubiquitous learning and the digital native*. Ed. by M. Ebner and M. Schiefner. Hershey, PA: IGI Global, 373–393. DOI: https://doi.org/10.4018/978-1-61520-678-0.ch021.

Hartog, R., H. Scholten, and A. Beulens (2013). "Output-Classes for Faculty-Based Design-Oriented Research on Digital Learning Resources in Higher Education". In: *Proceedings of the Informing Science and Information Technology Education Conference*. Informing Science Institute, 69–98.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Hattie, J. and H. Timperley (2007). "The power of feedback". *Review of Educational Research* 77, 81–112. DOI: https://doi.org/10.3102/003465430298487.

Hawkes, S. J. (2004). "Chemistry is not a laboratory science". *Journal of Chemical Education* 81, 1257.

Hawkins, I. and A. J. Phelps (2013). "Virtual laboratory vs. traditional laboratory: which is more effective for teaching electrochemistry?" *Chemistry Education Research and Practice* 14, 516–523. DOI: https://doi.org/10.1039/C3RP00070B.

Hayes, J. R. (2013). *The complete problem solver*. Hillsdale, NJ: Routledge.

He, Z., S. Zhang, F. Gu, and J. Wu (2019). "Mining conditional discriminative sequential patterns". *Information Sciences* 478, 524–539. DOI: https://doi.org/10.1016/j.ins.2018.11.043.

Heradio, R., L. de la Torre, D. Galan, F. J. Cabrerizo, E. Herrera-Viedma, and S. Dormido (2016). "Virtual and remote labs in education: A bibliometric analysis". *Computers & Education* 98, 14–38. DOI: https://doi.org/10.1016/j.compedu.2016.03.010.

Hernández-Leo, D., R. Martinez-Maldonado, A. Pardo, J. A. Muñoz-Cristóbal, and M. J. Rodríguez-Triana (2019). "Analytics for learning design: A layered framework and tools". *British Journal of Educational Technology* 50, 139–152. DOI: https://doi.org/10.1111/bjet.12645.

Hoadley, C. (2018). "A short history of the learning sciences". In: *International Handbook of the Learning Sciences*. Ed. by F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, and P. Reimann. Abingdon: Routledge, 11–23.

Hofstein, A. and V. N. Lunetta (2004). “The laboratory in science education: Foundations for the twenty-first century”. *Science Education* 88, 28–54. DOI: https://doi.org/10.1002/sce.10106.

Holme, T. A. (2020). “Introduction to the journal of chemical education special issue on insights gained while teaching chemistry in the time of COVID-19”. *Journal of Chemical Education* 97, 2375–2377. DOI: https://doi.org/10.1021/acs.jchemed.0c01087.

Hung, J.-L. and K. Zhang (2008). “Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching”. *MERLOT Journal of Online Learning and Teaching* 4, 426–437.

Järvelä, S., A. F. Hadwin, J. Malmberg, and M. Miller (2018). “Contemporary perspectives of regulated learning in collaboration”. In: *Handbook of the Learning Sciences*. Ed. by F. Fischer, C. E. Hmelo-Silver, P. Reimann, and S. R. Goldman. New York: NY: Routledge, 127–136.

Jiang, Y., L. Paquette, R. Baker, and J. Clarke-Midura (2015). “Comparing novice and experienced students within virtual performance assessments”. *International Educational Data Mining Society*.

Johnstone, A. (1997). “And some fell on good ground”. *University Chemistry Education* 1, 8–13.

Johnstone, A. and A. Al-Shuaili (2001). “Learning in the laboratory; some thoughts from the literature”. *University Chemistry Education* 5, 42–51.

Jolley, D. F., S. R. Wilson, C. Kelso, G. O’Brien, and C. E. Mason (2016). “Analytical thinking, analytical action: Using prelab video demonstrations and e-quizzes to improve undergraduate preparedness for analytical chemistry practical classes”. *Journal of Chemical Education* 93, 1855–1862. DOI: https://doi.org/10.1021/acs.jchemed.6b00266.

Jones, S. M. and A. Edwards (2010). “Online pre-laboratory exercises enhance student preparedness for first year biology practical classes”. *International Journal of Innovation in Science and Mathematics Education* 18.

Kalogiannakis, M., S. Papadakis, and A.-I. Zourmpakis (2021). “Gamification in science education. A systematic review of the literature”. *Education Sciences* 11.1. DOI: https://doi.org/10.3390/educsci11010022.

Kang, J., M. Liu, and W. Qu (2017). “Using gameplay data to examine learning behavior patterns in a serious game”. *Computers in Human Behavior* 72, 757–770. DOI: https://doi.org/10.1016/j.chb.2016.09.062.

Kay, J., P. Reimann, E. Diebold, and B. Kummerfeld (2013). “MOOCs: so many learners, so much potential ...” *IEEE Intelligent Systems* 28, 70–77. DOI: https://doi.org/10.1109/MIS.2013.66.

Keller, J. M. (1983). "Motivational design of instruction". In: *Instructional design theories and models: An overview of their current status*. Ed. by C. M. Reigeluth. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. Publishers, 383–434.

– (1987). "Development and use of the ARCS model of instructional design". *Journal of Instructional Development* 10, 2–10. DOI: https://doi.org/10.1007/bf02905780.

Kidron, A. and Y. Kali (2017). "Extending the applicability of design-based research through research-practice partnerships". *Educational Design Research* 1. DOI: https://doi.org/10.15460/eder.1.2.1157.

Kinnebrew, J. S., S. S. Killingsworth, D. B. Clark, G. Biswas, P. Sengupta, J. Minstrell, M. Martinez-Garza, and K. Krinks (2017). "Contextual markup and mining in digital games for science learning: Connecting player behaviors to learning goals". *IEEE Transactions on Learning Technologies* 10, 93–103. DOI: https://doi.org/10.1109/TLT.2016.2521372.

Kinnebrew, J. S., J. R. Segedy, and G. Biswas (2014). "Analyzing the temporal evolution of students' behaviors in open-ended learning environments". *Metacognition and Learning* 9, 187–215. DOI: https://doi.org/10.1007/s11409-014-9112-4.

Kinnebrew, J. S., K. M. Loretz, and G. Biswas (2013). "A contextualized, differential sequence mining method to derive students' learning behavior patterns". *Journal of Educational Data Mining* 5, 190–219.

Kirschner, P. and W. Huisman (1998). "'Dry laboratories' in science education; computer-based practical work". *International Journal of Science Education* 20, 665–682. DOI: https://doi.org/10.1080/0950069980200605.

Kirschner, P. and M. Meester (1988). "The laboratory in higher science education: problems, premises and objectives". *Higher Education* 17, 81–98. DOI: https://doi.org/10.1007/BF00130901.

Kirstein, J. and V. Nordmeier (2007). "Multimedia representation of experiments in physics". *European Journal of Physics* 28, S115. DOI: https://doi.org/10.1088/0143-0807/28/3/S11.

Kizilcec, R. F., M. Pérez-Sanagustín, and J. J. Maldonado (2017). "Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses". *Computers & Education* 104, 18–33. DOI: https://doi.org/10.1016/j.compedu.2016.10.001.

Koedinger, K. R., J. C. Stamper, E. A. McLaughlin, and T. Nixon (2013). "Using data-driven discovery of better student models to improve student learning". In: *International Conference on Artificial Intelligence in Education*. Artificial Intelligence in Education. Springer Berlin Heidelberg, 421–430. DOI: https://doi.org/10.1007/978-3-642-39112-5_43.

Koehler, B. P. and J. N. Orvis (2003). “Internet-based prelaboratory tutorials and computer-based probes in general chemistry”. *Journal of Chemical Education* 80, 606–608. DOI: https://doi.org/10.1021/ed080p606.

Koivisto, J. M., E. Haavisto, H. Niemi, P. Haho, S. Nylund, and J. Multisilta (2018). “Design principles for simulation games for learning clinical reasoning: A design-based research approach”. *Nurse Education Today* 60, 114–120. DOI: https://doi.org/10.1016/j.nedt.2017.10.002.

Kovanović, V., D. Gašević, S. Dawson, S. Joksimović, R. Baker, and M. Hatala (2015). “Does time-on-task estimation matter? Implications for the validity of learning analytics findings”. *Journal of Learning Analytics* 2, 81–110. DOI: https://doi.org/10.18608/jla.2015.23.6.

Lambert, D. and M. Jacobsen (2019). “Implementing an intervention into a grade six learning environment: A design-based research framework”. *Educational Design Research* 3. DOI: https://doi.org/10.15460/eder.3.1.1388.

Lambourne, R. (2007). “Laboratory-based teaching and the physics innovations centre for excellence in teaching and learning”. *European journal of physics* 28, S29. DOI: https://doi.org/10.1088/0143-0807/28/3/S03.

Law, N. and L. Liang (2020). “A multilevel framework and method for learning analytics integrated learning design”. *Journal of Learning Analytics* 7, 98–117. DOI: https://doi.org/10.18608/jla.2020.73.8.

Limniou, M., N. Papadopoulos, A. Giannakoudakis, D. Roberts, and O. Otto (2007). “The integration of a viscosity simulator in a chemistry laboratory”. *Chemistry Education Research and Practice* 8, 220–231. DOI: https://doi.org/10.1039/B6RP90032A.

Limniou, M. and C. Whitehead (2010). “Online general pre-laboratory training course for facilitating first year chemical laboratory use”. *Cypriot Journal of Educational Sciences* 5, 39–55.

Lockyer, L., E. Heathcote, and S. Dawson (2013). “Informing pedagogical action: aligning learning analytics with learning design”. *American Behavioral Scientist* 57, 1439–1459. DOI: https://doi.org/10.1177/0002764213479367.

Lowes, S., P. Lin, and B. Kinghorn (2015). “Exploring the link between online behaviours and course performance in asynchronous online high school courses”. *Journal of Learning Analytics* 2, 169–194. DOI: https://doi.org/10.18608/jla.2015.22.13.

Luckin, R. and L. Hammerton (2002). “Getting to know me: helping learners understand their own learning needs through metacognitive scaffolding”. In: *International Conference on Intelligent Tutoring Systems, Berlin, Heidelberg*. Intelligent Tutoring Systems. Springer Berlin Heidelberg, 759–771. DOI: https://doi.org/10.1007/3-540-47987-2_76.

Lyle, K. S. and W. R. Robinson (2002). "An action research report: Improving pre-laboratory preparation of first-year university chemistry students". *Journal of Chemical Education* 79, 663. DOI: https://doi.org/10.1021/ed079p663.

Ma, J. and J. V. Nickerson (2006). "Hands-on, simulated, and remote laboratories: A comparative literature review". *ACM Computing Surveys* 38, 7–30. DOI: https://doi.org/10.1145/1132960.1132961.

Mangaroska, K. and M. Giannakos (2019). "Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning". *IEEE Transactions on Learning Technologies* 12, 516–534. DOI: https://doi.org/10.1109/TLT.2018.2868673.

Mathews, M. and T. Mitrović (2008). "How does students' help-seeking behaviour affect learning?" In: *International Conference on Intelligent Tutoring Systems*. Ed. by B. P. Woolf, E. Aïmeur, R. Nkambou, and S. Lajoie. Intelligent Tutoring Systems. Springer, 363–372.

Mayer, R. E. (2009a). *Multimedia learning*. New York, NY: Cambridge University Press.

– (2009b). "Segmenting principle". In: *Multimedia learning*. Ed. by R. E. Mayer. New York, NY: Cambridge University Press, 175–188. DOI: DOI:10.1017/CBO9780511811678.004.

– (2009c). "The science of instruction: Determining what works in multimedia learning". In: *Multimedia learning*. Ed. by R. E. Mayer. New York, NY: Cambridge University Press, 28–56. DOI: DOI:10.1017/CBO9780511811678.004.

– (2014). "Incorporating motivation into multimedia learning". *Learning and Instruction* 29, 171–173. DOI: https://doi.org/10.1016/j.learninstruc.2013.04.003.

Mayer, R. E. and C. Pilegard (2014). "Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles". In: *The Cambridge handbook of multimedia learning*. Ed. by R. E. Mayer. New York: NY: Cambride University Press, 316–344.

McKenney, S. and T. C. Reeves (2019). *Conducting educational design research*. New York, NY: Routledge.

Mercer-Chalmers, J. D., C. L. Goodfellow, and G. J. Price (2004). "Using a VLE to enhance a foundation chemistry laboratory module". *CAL-laborate* 12, 14–18.

Mirzaei, M. and S. Sahebi (2019). "Modeling students' behavior using sequential patterns to predict their performance". In: Artificial Intelligence in Education. Springer International Publishing, 350–353. DOI: https://doi.org/10.1007/978-3-030-23207-8_64.

Mitrovic, A., V. Dimitrova, L. Lau, A. Weerasinghe, and M. Mathews (2017). "Supporting constructive video-based learning: Requirements elicitation from exploratory studies". In: *International Conference on Artificial Intelligence in Education*. Artificial Intelligence in Education. Springer International Publishing, 224–237. DOI: https://doi.org/10.1007/978-3-319-61425-0_19.

Moos, D. C. and A. Ringdal (2012). "Self-regulated learning in the classroom: a literature review on the teacher's role". *Education Research International* 2012, 1–15. DOI: https://doi.org/10.1155/2012/423284.

Mor, Y. and B. Craft (2012). "Learning design: Reflections upon the current landscape". *Research in Learning Technology* 20. DOI: http://dx.doi.org/doi:10.3402/rlt.v20i0.19196.

Neelen, M. and P. A. Kirschner (2020). *Evidence-informed learning design: Creating training to improve performance*. London, UK: Kogan Page.

Panadero, E. (2017). "A review of self-regulated learning: six models and four directions for research". *Frontiers in psychology* 8, 1–28. DOI: https://doi.org/10.3389/fpsyg.2017.00422.

Panadero, E., J. Klug, and S. Järvelä (2016). "Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand". *Scandinavian Journal of Educational Research* 60, 723–735. DOI: https://doi.org/10.1080/00313831.2015.1066436.

Perry, N. E. (1998). "Young children's self-regulated learning and contexts that support it". *Journal of Educational Psychology* 90, 715–729. DOI: https://doi.org/10.1037/0022-0663.90.4.715.

Pintrich, P. R. (2000). "The Role of Goal Orientation in Self-Regulated Learning". In: *Handbook of Self-Regulation*. Ed. by M. Boekaerts, P. R. Pintrich, and M. Zeidner. San Diego, CA: Academic Press, 451–502. DOI: https://doi.org/10.1016/B978-012109890-2/50043-3.

Pirker, J. and C. Gütl (2015). "Educational gamified science simulations". In: *Gamification in Education and Business*. Ed. by T. Reiners and L. Wood. Cham, Germany: Springer, 253–275. DOI: https://doi.org/10.1007/978-3-319-10208-5_13.

Plomp, T. (2013). "Educational design research: An introduction". In: *Educational design research*. Ed. by T. Plomp and N. Nieveen. Enschede, The Netherlands: Netherlands Institute for Curriculum Development, 11–50.

Potkonjak, V., M. Gardner, V. Callaghan, P. Mattila, C. Guetl, V. M. Petrović, and K. Jovanović (2016). "Virtual laboratories for education in science, technology, and engineering: A review". *Computers & Education* 95, 309–327. DOI: https://doi.org/10.1016/j.compedu.2016.02.002.

Reid, N. and I. Shah (2007). "The role of laboratory work in university chemistry". *Chemistry Education Research and Practice* 8, 172–185. DOI: https://doi.org/10.1039/B5RP90026C.

Richardson, M., C. Abraham, and R. Bond (2012). "Psychological correlates of university students' academic performance: a systematic review and meta-analysis". *Psychological Bulletin* 138, 353–387. DOI: https://dx.doi.org/10.1037/a0026838.

Roll, I., R. Baker, V. Aleven, and K. R. Koedinger (2014). "On the benefits of seeking (and avoiding) help in online problem-solving environments". *Journal of the Learning Sciences* 23, 537–560. DOI: https://doi.org/10.1080/10508406.2014.883977.

Rollnick, M., S. Zwane, M. Staskun, S. Lotz, and G. Green (2001). "Improving pre-laboratory preparation of first year university chemistry students". *International Journal of Science Education* 23, 1053–1071. DOI: https://doi.org/10.1080/09500690110038576.

Rosenthal, J. A. (1996). "Qualitative descriptors of strength of association and effect size". *Journal of Social Service Research* 21, 37–59. DOI: https://doi.org/10.1300/J079v21n04_02.

Roth, A., S. Ogrin, and B. Schmitz (2016). "Assessing self-regulated learning in higher education: a systematic literature review of self-report instruments". *Educational Assessment, Evaluation and Accountability* 28, 225–250. DOI: https://doi.org/10.1007/s11092-015-9229-2.

Rutten, N., W. R. van Joolingen, and J. T. van der Veen (2012). "The learning effects of computer simulations in science education". *Computers & Education* 58, 136–153. DOI: https://doi.org/10.1016/j.compedu.2011.07.017.

Ryan, A. M., P. R. Pintrich, and C. Midgley (2001). "Avoiding seeking help in the classroom: who and why?" *Educational Psychology Review* 13, 93–114. DOI: https://doi.org/10.1023/a:1009013420053.

Schunk, D. H. (2012). *Learning theories an educational perspective (sixth edition)*. Boston, MA: Pearson.

Schunk, D. H. and B. Zimmerman (2011). *Handbook of self-regulation of learning and performance*. New York, NY: Taylor & Francis.

Seery, M. K., H. Y. Agustian, and X. Zhang (2019). "A framework for learning in the chemistry laboratory". *Israel Journal of Chemistry* 59, 546–553. DOI: https://doi.org/10.1002/ijch.201800093.

Shute, V. J. (2008). "Focus on formative feedback". *Review of Educational Research* 78, 153–189. DOI: https://doi.org/10.3102\%2F0034654307313795.

Siemens, G. and P. Long (2011). "Penetrating the fog: Analytics in learning and education". *EDUCAUSE review* 46, 30.

Spagnoli, D., C. Rummey, N. Y. Man, S. S. Wills, and T. D. Clemons (2019). "Designing online pre-laboratory activities for chemistry undergraduate laboratories". In: *Teaching Chemistry in Higher Education: A Festschrift in Honour of Professor Tina Overton*. Creathach Press, 315–322.

Spagnoli, D., L. Wong, S. Maisey, and T. D. Clemons (2017). "Prepare, do, review: a model used to reduce the negative feelings towards laboratory classes in an introductory chemistry undergraduate unit". *Chemistry Education Research and Practice* 18. 10.1039/C6RP00157B, 26–44. DOI: https://doi.org/10.1039/c6rp00157b.

Sweller, J. (1994). “Cognitive load theory, learning difficulty, and instructional design”. *Learning and instruction* 4, 295–312. DOI: https://doi.org/10.1016/0959-4752(94)90003-5.

Tamir, P. (1976). “The role of the laboratory in science teaching. Technical report 10”.

Taub, M. and R. Azevedo (2018). “Using sequence mining to analyze metacognitive monitoring and scientific inquiry based on levels of efficiency and emotions during game-based learning”. *Journal of Educational Data Mining* 10, 1–26. DOI: https://doi.org/10.5281/zenodo.3554711.

Taub, M., R. Azevedo, A. E. Bradbury, G. C. Millar, and J. Lester (2018). “Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment”. *Learning and Instruction* 54, 93–103. DOI: https://doi.org/10.1016/j.learninstruc.2017.08.005.

Tempelaar, D. T., B. Rienties, and B. Giesbers (2015). “In search for the most informative data for feedback generation: Learning analytics in a data-rich context”. *Computers in Human Behavior* 47, 157–167. DOI: https://doi.org/10.1016/j.chb.2014.05.038.

Thalheimer, W. (2016). *Performance-focused smile sheets: A radical rethinking of a dangerous art form*. Minneapolis, MN: Work-Learning Press.

Thomas, M. K., A. Shyjka, S. Kumm, and R. Gjomemo (2019). “Educational design research for the development of a collectible card game for cybersecurity learning”. *Journal of Formative Design in Learning* 3, 27–38. DOI: https://doi.org/10.1007/s41686-019-00027-0.

Van den Akker, J. (2013). “Curricular development research as specimen of educational design research”. In: *Educational design research*. Ed. by T. Plomp and N. Nieveen. Enschede, The Netherlands: Netherlands Institute for Curriculum Development, 53–70.

Van der Kleij, F. M., R. C. W. Feskens, and T. J. H. M. Eggen (2015). “Effects of feedback in a computer-based learning environment on students’ learning outcomes: A meta-analysis”. *Review of Educational Research* 85, 475–511. DOI: https://doi.org/10.3102/0034654314564881.

Van der Kolk, J. (2013). “Design of an electronic performance support system for food chemistry laboratory classes”. Thesis.

Van der Kolk, J., G. Beldman, R. J. Hartog, and H. Gruppen (2012). “Students using a novel web-based laboratory class support system: a case study in food chemistry education”. *Journal of Chemical Education* 89, 103–108. DOI: https://doi.org/10.1021/ed1005294.

Van der Kolk, J., G. Beldman, R. J. Hartog, G. Moerland, and H. Gruppen (2013). “ExperD: web-based support for laboratory class workflow design”. In: *5th International Conference on Computer Supported Education, Aachen, Germany*, 25–32. DOI: https://dx.doi.org/10.5220/0004349800250032.

Van Laer, S. and J. Elen (2018). "Adults' self-regulatory behaviour profiles in blended learning environments and their implications for design". *Technology, Knowledge and Learning*, 1–31. DOI: https://doi.org/10.1007/s10758-017-9351-y.

Van Merriënboer, J. J. G. and P. A. Kirschner (2018). *Ten steps to complex learning: A systematic approach to four-component instructional design*. New York, NY: Routledge.

– (2012). *Ten steps to complex learning: A systematic approach to four-component instructional design*. 2nd ed. New York, NY: Routledge.

Vermunt, J. D. (1998). "The regulation of constructive learning processes". *British Journal of Educational Psychology* 68, 149–171. DOI: https://doi.org/10.1111/j.2044-8279.1998.tb01281.x.

Vermunt, J. D. and V. Donche (2017). "A learning patterns perspective on student learning in higher education: state of the art and moving forward". *Educational Psychology Review* 29, 269–299. DOI: https://doi.org/10.1007/s10648-017-9414-6.

Vermunt, J. D. and Y. J. Vermetten (2004). "Patterns in student learning: relationships between learning strategies, conceptions of learning, and learning orientations". *Educational Psychology Review* 16, 359–384. DOI: https://doi.org/10.1007/s10648-004-0005-y.

Verstege, S. and J. Diederen (2019). "Virtual experiment environment: A showcase of a preparation tool for laboratory classes". In: *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), 1406–1416.

Verstege, S., H. J. Pijeira-Díaz, O. Noroozi, H. Biemans, and J. Diederen (2019). "Relations between students' perceived levels of self-regulation and their corresponding learning behavior and outcomes in a virtual experiment environment". *Computers in Human Behavior*. DOI: https://doi.org/10.1016/j.chb.2019.02.020.

Verstege, S., J. Van der Kolk, J. Diederen, R. J. Hartog, and H. Gruppen (2016). "Exploring the potential of simulations as preparation tools for university laboratory classes". In: *EDULEARN16 Conference*. 8th International Conference on Education and New Learning Technologies. IATED. DOI: https://doi.org/10.21125/edulearn.2016.1429.

Verstege, S., J.-P. Vincken, and J. Diederen (2021). "Blueprint to design virtual experiment environments". *Computers and Education Open* 2. DOI: https://doi.org/10.1016/j.caeo.2021.100039.

Vural, O. F. (2013). "The impact of a question-embedded video-based learning tool on e-learning". *Educational Sciences: Theory and Practice* 13, 1315–1323. DOI: https://files.eric.ed.gov/fulltext/EJ1017292.pdf.

Wheatley, G. (1984). "Problem solving in school mathematics (MEPS Technical Report No. 84.01)". *West Lafayette, IN: Purdue University, School of Mathematics and Science Center*.

Wiley, K. J., Y. Dimitriadis, A. Bradford, and M. C. Linn (2020). "From Theory to Action: Developing and Evaluating Learning Analytics for Learning Design". In: LAK '20. Frankfurt, Germany: Association for Computing Machinery, 569–578. DOI: https://doi.org/10.1145/3375462.3375540.

Winberg, T. M. and C. A. R. Berg (2007). "Students' cognitive focus during a chemistry laboratory exercise: effects of a computer-simulated prelab". *Journal of Research in Science Teaching* 44, 1108–1133. DOI: https://doi.org/10.1002/tea.20217.

Winne, P. H. (1997). "Experimenting to bootstrap self-regulated learning". *Journal of Educational Psychology* 89, 397–410. DOI: https://doi.org/10.1037/0022-0663.89.3.397.

– (2015). "Self-regulated learning". In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Ed. by J. D. Wright. Oxford: Elsevier, 535–540. DOI: https://doi.org/10.1016/B978-0-08-097086-8.25091-5.

Winne, P. H. and A. F. Hadwin (1998). "Studying as self-regulated learning." In: *Metacognition in educational theory and practice*. Ed. by D. J. Hacker, J. Dunlosky, and A. C. Graesser. Mahwah, NJ: Lawrence Erlbaum Associates, 277–304.

Winne, P. H. and N. E. Perry (1994). "Educational psychology". In: *Encyclopedia of Human Behavior*. Ed. by V. S. Ramachandran. San Diego, CA: Academic Press, 213–223.

Wong, J., M. Khalil, M. Baars, B. B. de Koning, and F. Paas (2019). "Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course". *Computers & Education* 140, 103595. DOI: https://doi.org/10.1016/j.compedu.2019.103595.

Yuriev, E., S. Naidu, L. S. Schembri, and J. L. Short (2017). "Scaffolding the development of problem-solving skills in chemistry: Guiding novice students out of dead ends and false starts". *Chemistry Education Research and Practice* 18, 486–504. DOI: https://doi.org/10.1039/C7RP00009J.

Zaki, M. J. (2000). "Sequence mining in categorical domains: Incorporating constraints". In: Proceedings of the ninth international conference on Information and knowledge management. Association for Computing Machinery, 422–429. DOI: https://doi.org/10.1145/354756.354849.

Zheng, J., W. Xing, and G. Zhu (2019). "Examining sequential patterns of self- and socially shared regulation of STEM learning in a CSCL environment". *Computers & Education* 136, 34–48. DOI: https://doi.org/10.1016/j.compedu.2019.03.005.

Zheng, L. (2015). "A systematic literature review of design-based research from 2004 to 2013". *Journal of Computers in Education* 2, 399–420. DOI: https://doi.org/10.1007/s40692-015-0036-z.

Zhou, M., Y. Xu, J. C. Nesbit, and P. H. Winne (2010). "Sequential pattern analysis of learning logs: Methodology and applications". *Handbook of educational data mining* 107, 107–121.

Zhu, G., W. Xing, and V. Popov (2019). "Uncovering the sequential patterns in transformative and non-transformative discourse during collaborative inquiry learning". *The Internet and Higher Education* 41, 51–61. DOI: https://doi.org/10.1016/j.iheduc.2019.02.001.

Zimmerman, B. J. (1989). "A social cognitive view of self-regulated academic learning". *Journal of Educational Psychology* 81, 329–339. DOI: https://doi.org/10.1037/0022-0663.81.3.329.

– (1998). "Comparing skillful and naive self-regulated learners". In: *Self-regulated learning: from teaching to self-reflective practice*. Ed. by D. H. Schunk and B. J. Zimmerman. New York, NY: Guilford, 6–10.

– (2000). "Attaining self-regulation: a social cognitive perspective". In: *Handbook of Self-Regulation*. Ed. by M. Boekaerts, P. R. Pintrich, and M. Zeidner. San Diego: Academic Press, 13–39. DOI: https://doi.org/10.1016/B978-012109890-2/50031-7.

– (2013). "From cognitive modeling to self-regulation: a social cognitive career path". *Educational Psychologist* 48, 135–147. DOI: https://doi.org/10.1080/00461520.2013.794676.

– (2015). "Self-Regulated Learning: Theories, Measures, and Outcomes". In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Ed. by J. D. Wright. Oxford: Elsevier, 541–546. DOI: https://doi.org/10.1016/B978-0-08-097086-8.26060-1.

Zimmerman, B. J. and A. R. Moylan (2009). "Self-regulation: where metacognition and motivation intersect". In: *Handbook of Metacognition in Education*. Ed. by D. J. Hacker, J. Dunlosky, and A. C. Graesser. New York, NY: Routledge, 311–328.

Zimmerman, B. J. and M. M. Pons (1986). "Development of a structured interview for assessing student use of self-regulated learning strategies". *American Educational Research Journal* 23, 614–628. DOI: https://doi.org/10.3102/00028312023004614.

Zimmerman, B. J. and D. H. Schunk (2003). "Albert Bandura: the scholar and his contributions to educational psychology". In: *Educational psychology: a century of contributions.* Ed. by B. J. Zimmerman and D. H. Schunk. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 431–457.

Summary

Laboratory education plays an important role in natural science education. Despite its prevalence in many curricula, it is surprisingly difficult to find evidence that the time, effort, and costs invested in laboratory classes are in balance with students' achievement of intended learning outcomes during laboratory education. Since laboratory classes contribute to teaching students how to 'do' science, many teachers enjoy being personally involved in the teaching, and students generally enjoy laboratory classes, there is little support for replacing the usually inefficient laboratory classes. Instead, there is a general consensus that student *preparation* prior to laboratory classes will increase the quality of knowledge construction during the laboratory classes.

Students can be prepared for laboratory classes by means of non-traditional labs (NTLs) such as virtual experiment environments (VEEs). What stands out in publications about NTLs, is that they hardly provide any details on *how* the NTLs were designed. As an educational designer, I was missing the link between established guidelines for designing instruction and the design of the NTLs. This resulted in my central research question: How to design a VEE that helps students to prepare for laboratory classes?

To answer the central research question, I adopted the educational design research approach. Educational design research can be defined as "a genre of research in which the iterative development of solutions to practical and complex educational problems also provides the context for empirical investigation, which yields theoretical understanding that can inform the work of others". What sets this type of research apart from many other types of research, is that it is committed to develop theoretical understanding and practical solutions simultaneously. The educational design research approach typically consists of three phases: i) analysis & exploration, ii) design & construction, and iii) summative evaluation. These phases are described in detail in chapter 3, and are represented by the different chapters throughout this dissertation.

The analysis & exploration phase is represented by **chapter 2**. In this chapter, the complex educational problem and its context were defined. Stakeholders were identified, and provided with a questionnaire to get their opinion about the current state of laboratory education. Several explorative follow-up interviews were conducted to get a more complete overview. Based on the results of the questionnaire, the interviews, and literature research, it was concluded that there is potential for virtual experiment environments as preparation tools for laboratory classes, and an initial list of design requirements was identified, which was used as the starting point for the design & construction phase.

The design & construction phase is represented by chapters 3, 4, and 5. In **chapter 3**, a blueprint to design virtual experiment environments is provided. Following an extensive description and overview of the educational design research approach, solutions for the complex educational problem (chapter 2) were explored, which finally resulted in the design of four VEEs. Based on recurrent testing and formative evaluations of the VEEs, a blueprint was established that consists of three design requirements, fourteen design principles, and a design architecture. Such information in the context of VEEs is new to literature, and has the potential to help teachers and instructional designers to design high quality VEEs.

In **chapter 4**, a showcase of a VEE is provided. The user interface, the available interactions, and the feedback- and support mechanism are presented and discussed. This contextual information is limited in many publications (including my own, in chapters 6 and 7) that use an existing environment to study a construct such as self-regulated learning. The showcase, when combined with the blueprint (chapter 3), helps to place the studies presented in this thesis in the right context, and gives a complete overview that may as such provide guidance and/or inspiration to teachers and educational designers.

In **chapter 5**, an extension of the initial design (chapters 3 and 4) is introduced and evaluated. The extended design is focused on students' understanding of protocol steps, by enriching existing protocols with theoretical, practical, troubleshooting, and/or calculation questions, to form interactive protocols.

The summative evaluation phase is represented by chapters 3, 5, 6, and 7. In chapters 3 and 5, the designs were mostly evaluated based on questionnaires, so that means based on students' self-reported data. The disadvantage of self-reported data is that the assumption must be made that students are capable of reflecting on their own behaviour, are willing to answer the questions honestly, and that all students have the same interpretation of the questions. To obtain more objective information on student behaviour in the VEEs, logging data was introduced in chapters 6 and 7.

In **chapter 6**, a VEE was used as the context for exploring the relation between students' perceived level of self-regulated learning (SRL), their achieved learning outcomes in the VEE, and their logged behaviour in the VEE.

In **chapter 7**, sequential pattern mining was used to study how students use support while doing calculations in a VEE. The use of support was related to students' learning gain and prior knowledge, with the goal to evaluate the support mechanism that is part of the blueprint (chapter 3).

In **chapter 8**, I discuss my findings from the perspectives of the university teacher, the educational designer, and the educational scientist.

In conclusion, the complex educational problem was solved by the practical solutions: the VEEs. The theoretical understanding acquired in the educational design research process, which was captured in the blueprint, which is the answer to the central research question.

Acknowledgements

First of all, I want to thank my family: Ad, Helma, Renske, & Monique for your unconditional love and support ♡. Without you, I would not have been the person I am today.

Julia, thank you for giving me the opportunity to pursue a PhD under your supervision. I will never forget the day you picked me up and brought me to Jean-Paul's office. After permanently deleting (in my opinion) disastrous course materials, I was absolutely convinced you were going to fire me. As you can imagine, what happened next came as a surprise to me...

The past five years have been a great experience to me. I found great joy in designing my first virtual experiment environment. It required me to come up with creative solutions for the many design challenges that I had to overcome. Then, suddenly, no less than 97 MSc students got to work in my creation. I was happily surprised when the first results showed me that students appreciated the virtual experiment environment far more than I expected. Many design projects followed, and I found myself challenged to write down my findings in such a way that the world may find use in them.

Along the road, I received help and support, and gained inspiration from many people. If you were one of them, you know it. Thank you very much!

Julia, I really enjoyed working with you. You allowed me the time and the freedom to shape my project, which I appreciated very much. You were always approachable for my many questions, and, not irrelevant, provided me with useful answers to them. I have noticed that there are no colleagues who dare to schedule a meeting with you after five. I was definitely blessed that our regular meetings were scheduled until five :) What I appreciate about you as a person is that you see the good in people, drawing a nearly endless patience from that source.

I want to thank all my co-authors, in order of appearance: Koos, Julia, Rob, Harry, Héctor, Omid, Harm, Jean-Paul, Wander, Yingbin, Peter & Luc for adding your expertise to my project.

I want to thank my colleagues at LabBuddy, and especially Koos, Tineke, and Albert, for the (technical) support you have provided me with throughout my project. I appreciate your involvement in design choices that were to be made.

Mirjam, thank you for allowing me to implement my first creations in your course. I am grateful for your patience and for the trust you gave me.

Wander, thank you for your contributions to my work. First as a student, and later as a colleague. You were of great help to me in the process of designing, improving, and evaluating multiple simulations. I was very happy to have you around as a sparring partner during the last phase of my PhD project.

Peter, you are probably the most critical thinker I have ever met, which made the conversations and discussions we had so much fun. I want to thank you for your sincere interest in my work. Your suggestions were at the heart of the support system that was later developed.

Rob, thank you for teaching me how to write. When writing my first manuscript, you spent so much time helping me. Thank you for the elaborate feedback you provided me with over the years.

Renske, Monique & Koos, thank you for reading several of my manuscripts. I was happily surprised by the amount and quality of your feedback.

Maarten, thank you guiding me in my search for pixels. Thanks to your efforts, the world can now see what it is looking for.

Mark, although we haven't shared our office over the last 1.5 years, I enjoyed the time I spent in there. I appreciate your critical view on my project, including all the forthcoming suggestions. Our not-too-much-noise-policy was perfect.

Jolanda, I enjoyed the many conversations we had about ... anything. I was always eager to come up with a question for you :) Thank you for your help in organizing my FCH-life.

Last but not least, I want to thank my thesis students: Lieve, Lindsey, Jamie, Sander, Wander, Vincent, Jord, Soraya, Tom & Oscar, for your contributions to my work, and perhaps more importantly, to my development as a supervisor.

About the author

Sjors Verstege was born on July 21st, 1991 in Nieuwegein, the Netherlands. In 2009 he received his secondary school diploma from Cals College in Nieuwegein. After that, he started the BSc programme Food Technology, with a minor 'Education', at Wageningen University. He proceeded with the MSc programme Food Technology at Wageningen University, specializing in Ingredient Functionality, with the minor 'Chemistry for Education'. He finished his MSc study in 2016 with an internship at LabBuddy, during which he explored the potential of simulations as preparation tools for university laboratory classes. He published the results of this research and presented them at the EDULEARN16 conference in Barcelona, Spain. In August 2016, he continued this project as a PhD candidate at the Laboratory of Food Chemistry at Wageningen University, under the supervision of Julia Diederen and Harry Gruppen, who was later to be replaced by Jean-Paul Vincken. The results of his PhD research are presented in this thesis. Currently, Sjors is working as productmanager at LabBuddy.

List of publications

S. Verstege, J. Van der Kolk, J. Diederen, R. J. Hartog, and H. Gruppen (2016). "Exploring the potential of simulations as preparation tools for university laboratory classes". In: *EDULEARN16 Conference*. 8th International Conference on Education and New Learning Technologies. IATED. doi: https://doi.org/10.21125/edulearn.2016.1429.

S. Verstege and J. Diederen (2019). "Virtual experiment environment: A showcase of a preparation tool for laboratory classes". In: *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), 1406–1416.

S. Verstege, H. J. Pijeira-Díaz, O. Noroozi, H. Biemans, and J. Diederen (2019). "Relations between students' perceived levels of self-regulation and their corresponding learning behavior and outcomes in a virtual experiment environment". *Computers in Human Behavior*. doi: https://doi.org/10.1016/j.chb.2019.02.020.

S. Verstege, J. Diederen, and J.-P. Vincken. "Blueprint to design virtual experiment environments". *Computers and Education Open*. doi: https://doi.org/10.1016/j.caeo.2021.100039.

S. Verstege, W. Lamot, J.-P. Vincken, and J. Diederen. "Design of interactive protocols that help students to prepare for laboratory work". *Submitted for publication*

S. Verstege, Y. Zhang, P.A. Wierenga, L. Paquette, J. Diederen. "Using sequential pattern mining to understand how students use guidance while doing calculations" *Manuscript in preparation*

Overview of completed training activities

With the training and education activities listed below, the PhD candidate has complied with the requirements set by the Graduate School VLAG, which comprises of a minimum total of 30 EC (≈ 22 weeks of activities). The activities were subdivided in discipline specific activities, general courses, and optionals. Furthermore, the PhD candidate was involved in education, by means of supervision of courses and thesis students.

Name of the learning activity	Organization	Year
Discipline specific activities (14.5 EC)		
Games and virtual worlds in education	SURF	2016
'Onderwijsdagen'	SURF	2016
Teacher day**	WUR	2016 - 2020
Smartphone-based food analysis	VLAG	2017
Virtual reality in education	SURF	2017
'Onderwijsfestival'**	VSNU	2018
EdMedia conference*	AACE	2019
Present day practicals conference*	UvA[a]	2019
Educational data mining conference	EDM	2020
Big Data and Education (MOOC)	Penn	2020
General courses (6.5 EC)		
PhD week	VLAG	2016
'Communiceren met kinderen'	WUR	2017
Famelab presentation course	WUR	2017
Brain based teaching	WUR	2017
PhD carousel	WGS	2017 - 2019
Gamified design	WUR	2019
Designing exams	WUR	2019
Supervising group work	WUR	2019
(Re)designing a course	WUR	2020
Optionals (13.3 EC)		
Preparation of research proposal	FCH	2016
Weekly group meetings	FCH	2016 - 2021
PhD trip to Japan	FCH	2016
PhD trip to Italy & Austria***	FCH	2018

* Oral presentation; ** Poster presentation; *** Organising committee; [a] In cooperation with Leiden University and LabBuddy.

Explanation of abbreviations: SURF: 'coöperatieve vereniging van Nederlandse onderwijs- en onderzoeksinstellingen op het gebied van informatie- en communicatietechnologie.'; WUR: Wageningen University & Research; VLAG: Advanced studies in Food Technology, Agrobiotechnoloty, Nutrition and Health Sciences; VNSU: 'De Vereniging van Universiteiten'; AACE: Association for the Advancement of Computing in Education; UvA: 'Universiteit van Amsterdam'; EDM: International Educational Data Mining Society; Penn: University of Pennsylvania; WSG: Water Systems and Global Change; FCH: Laboratory of Food Chemistry.

The work described in this thesis was performed at the Laboratory of Food Chemistry, Wageningen University & Research, The Netherlands.

Financial support from Wageningen University & Research for printing this thesis is gratefully acknowledged.

Cover design by Maarten van der Velde

This thesis was printed by Gildeprint
Edition 75 copies