Reflecting on Challenges of Conducting Design-Based Research in Large University Courses

Naemi Luckner, Peter Purgathofer, Geraldine Fitzpatrick
TU Wien (Vienna University of Technology), Argentinierstr. 8/193-5, Vienna
Austria
{naemi, purg}@igw.tuwien.ac.at, geraldine.fitzpatrick@tuwien.ac.at

Abstract: Design-based research is an approach frequently applied in research projects that examine technology-enhanced learning environments in real-world settings. This is important if we are to arrive at solutions that support authentic learning contexts. However, as researchers in such a project, there are challenges. While design-based research supports the way we think about design as process and as artefacts, we often struggle with issues not always directly connected, but definitely linked, to our design-based research approach. In this paper, we retrospectively reflect on emergent lessons from our 12-year experience in a design-based research project to develop a technology-enhanced learning environment for large university classes. We discuss some of the issues we experienced and still have to deal with repeatedly while working towards implementing and improving our learning design in large university courses. The discussion and insights in this paper are designed as signposts for researchers in the area that might provide direction, support and, foremost, a basis for further discussion.

Introduction

Teaching large classes at a university level is a challenge (Mulryan-Kyne 2010) to which technology is often seen as a possible solution (e.g., Alexander 2006, Francis 2012, Salter & Purgathofer 2010). However, how to design good technology that works in practice is a non-trivial matter and trying to innovate on designing and deploying technology into authentic learning experiences is challenging.

We are teaching large courses of 600+ students at University. This is a class size that is known to be detrimental to the quality of teaching and learning in many ways. Monks & Schmidt (2010) show that class size as well as overall student load have a negative impact on the assessment of courses and instructors. They also note that large classes influence instructors to alter courses in ways deleterious to students. These findings are echoed in many other studies on class size and quality of teaching or learning (e.g., Kokkelenberg et al. 2008). Unfortunately, class size is not a variable in our context. Due to specifics of the Austrian university system, we have to accept the high number of students and try our best to maintain the overall quality. As this is unavoidable, it has become one of the motivators for our research in teaching and learning about how we can improve the quality of the learning experience for students even with large classes.

Due to the class size, many factors that might support students in a smaller teaching and learning context are lacking, such as personal attention, individual feedback, discourse about lecture topics or feeling connected to one's peers. Additionally, we were not content with the established, traditional format of holding frontal lectures during the semester and administering one test at the end of each semester to test a student's knowledge. Therefore, we decided to create a technology-enhanced learning environment that provides us with the means to do things differently.

While we have a clear learning design in mind, creating a technology-enhanced learning environment, that is based upon ideas and concepts from constructivism (Piaget, 2013) and connectivism (Siemens 2005, Driscoll 2000) to engage and motivate our students, proves to be an ever-changing challenge. Over the past 12 years, we have continuously changed and evolved the system to better support our underlying ideas and theories on how to create a fruitful learning environment. This is a "wicked problem" as originally described by Rittel and Webber (1973), where we treat this as an open-ended process marked by trial and error, trying out ideas and in doing so, learning more about the challenges themselves, which in turn changes our approach on how to deal with them.

To accomplish this task of continuously designing and iterating new features that embody theories of learning, we adhere to the concepts of design-based research - designing in a real-world context while collaborating with practitioners and students, and testing and iterating design artefacts. However, while the ideals and structures of design-based research have always accompanied us, we repeatedly ran into problems we wished we had thought of earlier.

In this paper, we want to familiarise the readers with our design, research and development context and its accompanying challenges. Along with the context we aim at listing tips and remarks concerning emergent issues we found ourselves having to address. While many of those tips can be thought of in hindsight as common sense, they might be helpful as a priori points for researchers setting out to start a design-based research project, or ease the conscience of researchers already entangled in such an endeavour to know that others are dealing with similar issues.

Related Work

Many researchers in the area of technology-enabled learning currently follow a design-based research approach. The origin of design-based research is often traced back to Brown's article on design experiments (Brown 1992) where she discusses the trade-off of giving up the control of doing experiments in a lab by doing design experiments in a real-life setting and the need for a new methodology to conduct research in such an environment, a setting also discussed by Rogers and Marshall (2017) under the term Research in the wild. Ever since, design-based research has been widely discussed and refined.

The goals of design-based research in education are to generally improve learning and teaching (Barb 2014) by identifying significant educational problems (Reeves et al. 2005), guiding technological development in education with a concern for values and principles transported by said technology (Amiel & Reeves 2008), refine innovations so that they matter in real life (Dede 2004, Anderson & Shattuck 2012) and explaining the mechanisms behind these interventions (Barab 2014). Design-based research aims to link practical issues with a theoretical research base (Design-based Research Collective 2003) and advance a common theoretical agenda (Barab & Squire 2004) by continuously developing theories across multiple design investigations (Amiel & Reeves 2008). While developed in a very specific local context, design-based research should aspire to create knowledge that is also of value to other researchers (McKenney & Reeves 2013, Anderson & Shattuck 2012) and commit to theory construction (Reeves et al. 2005). As Amiel and Reeves (2008) point out, there should be a clear focus on educational principles and foundations that might demand a technical solution, rather than designing for the sake of using the newest cutting-edge technology. Another declared goal of design-based research is for the research to be context sensitive rather than pressing for a generalisable outcome (Design-based Research Collective 2003).

One of the main characteristics that distinguish design-based research from other approaches is the *real-life setting* with all its advantages and disadvantages. Such a context has a multitude of collaborators, from participants, teachers and administrators, to developers and researchers (Cobb et al. 2015, Brown 1992), which often leads to a multidisciplinary approach (Reeves et al. 2005, Scanlon et al. 2015). Research is always focussed on a very specific local context, which encompasses complex situations (Barab & Squire, 2004) and wicked problems (Obrenović 2011), but leads to deeper understandings (Design-based Research Collective, 2003) and learning about real-life phenomena (Collins et al. 2004). In these contexts, there a number of differences compared to lab settings: variables cannot be controlled (Collins et al. 2004); the research design needs to be adaptive and flexible (Plomp 2013) in that initial research designs may be insufficiently detailed (Wang & Hannafin 2005); wishes of other parties may influence the study (Kennedy-Clark 2013); it might need to be adapted to changes in external policies or organisational conditions (van den Akker 1999); and methods may need to evolve to better fit the local context (Brown 1992). The real-life context also influences the evaluation of an intervention, since the intervention cannot be frozen at a certain point for evaluation but needs to continue to evolve (Design-based Research Collective 2003). As summarised by Barab (2014), design-based research happens and transacts with naturalistic settings.

Another characteristic of design-based research is the *iterative and progressive* refinement of a technological *design artefact*. Such artefacts embody claims about teaching and learning (Design-based Research Collective 2003), help to demonstrate knowledge that would have been hard to explain otherwise (Schön 1983) and should outlast a study or research setting (Kelly 2004). Collins et al. (2004) talk about the necessity to put such an artefact into a real-world context to see how it works and if it actually addresses the underlying theoretical questions. The Design-based Research Collective (2003) refers to microanalysis of student interaction with an artefact to enable redesign and iterations towards a plausible design solution (Reeves et al. 2005). Kennedy-Clark (2013)

defines iterations as micro-cycles, each of which focuses on a particular, stand-alone aspect of the educational environment in question. Amiel and Reeves (2008) understand the creation of technology as a process rather than an artefact and underpin the need to be aware of values being transported by the designed outcome. This type of iterative research demands a *long-term commitment* of researchers (Reeves et al. 2005) and is developed over years, not months (Scanlon et al. 2015), as single studies would lead to very limited insights (Amyl & Reeves 2008). It also asks for a deep understanding how learning occurs (Barab & Squire 2004) and is geared towards improving, not proving educational theories (Herrington et al. 2007).

Outcomes of design-based research can be divided into three areas as specified by Herrington et al. (2007): scientific, design artefacts and societal. *Scientific outcomes* can be design principles (Anderson & Shattuck 2012), design procedures (Obrenović 2011), and design guidelines to make explicit the implicit decisions in the process (Plomp 2013) as well as a deeper understanding of the problems and issues themselves that are addressed by the research (Obrenović 2011). *Design artefacts* are another outcome of design-based research, ideally artefacts which are reusable elsewhere (Kelly 2004, Herrington et al. 2007), which however is not always the case due to the very context sensitive nature of design-based research. The distinction between scientific and design artefacts as outcomes, however, is not always very clear, as the boundaries between these areas become blurred. While Amyl and Reeves (2008) see the creation of artefacts as a scientific process, Kelly (2004) describes the process of creation itself as an artefact of design-based research. Plomp (2013) adds that outcomes of a design-based research approach should strive to be relevant (also mentioned by McKenney & Reeves 2013), consistent with the addressed challenge, effective by resulting in a desired outcome (as also stated by van den Akker 1999) and practical, also mentioned by van den Akker (1999) as a user perspective on the artefacts. Lastly, the professional development of people involved is mentioned as *societal outcome* (Herrington et al. 2007).

Due to the complex setting, researchers have to be aware of and deal with a plethora of *issues* when attempting design-based research. Similar to the outcomes, these issues can be divided into theoretical, practical and human challenges. Theoretical issues deal with the theoretical and methodological grounding, practical issues come up during the research process itself, and human issues are dealing with the people involved.

Concerning practical research challenges there is a variety of aspects mentioned in related literature. Working on a design-based research project can spread over years (Design-based Research Collective 2003) although financial support is often lacking (Anderson & Shattuck 2012). This long-lasting research produces enormous amounts of data with not enough time to analyse (Collins et al. 2004), while it should be immediately, continuously as well as retrospectively analysed (Wang and Hannifin 2005). Much of the data collected is inconclusive or incomplete (Kennedy-Clark 2015). Researchers need to be clear about what they are evaluating the data for, which outputs they are measuring and how they measure for these outputs to not badly misjudge the outcomes (Collins et al. 2004). Also, Brown (1992) points to the systemic whole, which means that changes in one aspect of an educational environment have implications for all other aspects. So, as van den Akker (1999) confirms, it is impossible to isolate separate variables in the same study due to the dynamic context of the system. This affects studies comparing across designs (Collins et al. 2004) as not only the system changes, but so do courses, assignments and the greater context (Gehringer 2000). A change in the context and its dynamics also makes it difficult or impossible to reproduce studies (Barab & Squire 2004) as there is no experimental control in a real-life setting and the effectiveness of an intervention might change depending on the study context (Collins et al. 2004).

Human challenges can be connected to different people in the research. Evaluation of any collected data depends on the data being trustworthy. An issue mentioned by Dede (2004), as well as by Wang and Hannafin (2005), is the possibility of introducing a Hawthorne effect, a change of subject's behaviour due to the close collaboration between participants and researchers. Another possible problematic involving participants can occur when doing studies in class rooms with an insufficient number of students (Kennedy-Clark 2015). Kennedy-Clark (2013) suggests conducting pre-studies with different groups of participants before conducting the actual study. Van den Akker (1999) mentions that the success of an intervention depends on its implementation, which introduces *developers* as one more party influencing studies. Additionally, due to the long-lasting characteristic of design based research, there is usually a core team of *researchers* while others come and go (Reeves et al. 2005). Any kind of human participation introduces an element of bias that needs to be accounted for. This bias shows itself in different ways, for example as an effect of multiple roles a researcher takes on in a design-based research project, e.g., as researcher, designer and teacher (Barab & Squire 2004), in the choosing of data to evaluate and which data to highlight (Dede 2004, Dominguez et al. 2014), by projecting unconscious mental structures while analysing data (Brown 1992), or by choosing what is meaningful and what not to report (Kelly 2004). However, as Anderson and Shattuck (2012) state, the close relationship of researcher and context adds as much as it detracts. And bias can be

lessened, e.g., by introducing inter-rater reliability (Kelly 2004), actively addressing contrary instances in the evaluation or observe during studies to better understand what is relevant and what is not (Kennedy-Clark 2013).

The human challenges can also be defined to include ethical responsibilities of the researchers. Reeves et al. (Reeves et al. 2005) summarise the ethical responsibilities of design-based research as defining outcomes for pedagogically important educational problems, with emphasis on the content rather than the technology; supporting human interactions and learning; working reflectively and re-iterating until the outcome is reached; and putting the educational outcome before the research outcome (as also mentioned by Hutchings 2003). Similarly, Dominguez et al. (2014) as well as Chang and Gray (2013) advocate research integrity, referring to not only reporting on favourable results and to not base results solely on students with positive grades, or ignore dropouts and thereby distort the outcome. Choosing ethically appropriate methods is often found challenging (e.g., (Chang et al. 2005, Hutchings 2003, Kanuka & Anderson 2007)) for example when having to decide to intervene in a study in order to support participants (Barab & Squire 2011). Concerning research methods, there are on-going debates on using control group studies vs. comparing data to a previous run-through of courses. While direct comparison between classes is perceived as difficult (Gehringer 2000), control groups are also often seen as inappropriate (Hutchings 2003, Chang et al. 2005) for a teaching and learning setting. Hutchings (2003) argues that he cannot ethically subject students to out-dated approaches that he is no longer confident in.

The third block of challenges has to do with the scientific outcome of design-based research. This block addresses problems with objectivity, generalizability, reliability and validity. Previously discussed challenges with an ever-changing research team, close involvement of researchers and practitioners and the accompanying bias, which is unavoidable, make it challenging to produce data that can be considered objective. The setting of design-based research is often too complex to make sure that the outcomes are reliable. Too few iterations might not provide enough data to make reliable statements about the underlying theories (Herrington et al. 2007, Zheng 2015). It is also not always clear if the observed effects are caused by the study's intervention, or happen as a result of an interplay between other changes in the design (Reimann 2011) as not all aspects can be controlled (Collins et al. 2004). Wang and Hannafin (2005) advise setting detailed practical goals for the evaluation since researchers cannot study everything at once. The validity of design-based research can be strengthened by collaborating with various experts in the field (Kennedy-Clark 2013), increasing the alignment of theory, design and practice (Design-based Research Collective 2003) and consistently linking the design with state-of-the-art components (van den Akker 1999). After having dealt with problems of objectivity, reliability and validity of the data, researchers face the challenge of generalizability of the results. The very localised context makes it hard to draw generalisable solutions that would also apply to a different context (Herrington et al. 2007). Kelly (2004) criticises drawing general conclusions from a very narrow set of subjects and over studies that are often adapted to serve local needs. McKenney and Reeves (2013) address the issue of current studies mostly presenting the potential of interventions rather than actual impact on educational research. Barab (2014) discusses the validity of so called storied truths and the need to transcend specific cases when building knowledge. Edelson (2002) proposes concentrating on domain theories so theories about the world rather than the design per se.

There is also a theoretical discussion about calling design-based research a methodology. Dede (2004) and Kelly (2004) point toward missing standardisation in the research process. Wang and Hannafin (2005) agree by pointing out an issue of standardisation of when to abandon a design or declare it successful. Walker (2011) criticizes a lack of discussion of epistemology, especially in connection to the mixed-methods approach.

Finally, there are lists of guidelines towards how to conduct design-based research (Hevner & Chatterjee 2010, Collins et al. 2004, Barab 2014, Plomp 2013) and recommendations for researchers in the field (Scanlon et al. 2013, Reeves et al. 2005). We want to add to this second type of guidelines by discussing our local experiences and reflections as we see existing gaps in understanding more about the tensions in design-based research projects. We argue that most problems that arise are unique to each context, while at the same time they are caused by common root causes. Describing the problems and learnings from individual projects then helps identify those root causes as it allows us to abstract from the concrete examples to general principles. A multitude of case studies such as this one could lead to a more systematic analysis across a corpus of studies, which in turn could address the issues of design-based research as a methodology missing standardisation as raised by Dede (2004) and Kelly (2004).

Setting the Context

Like many design-based research projects, we are working in the real-life context of a university where we develop a technology-enhanced learning environment for students of informatics/computer science catering to our local

needs. Having explored different options using existing technology such as Wikis, Blogs or Twitter to enhance our web-based learning setup, we quickly encountered limits as to what was possible and what was not. Furthermore, we discovered issues with each of these technologies that we could not overlook in our goal to create an emancipated learning environment to empower students to take responsibility of their own learning. Issues included: off-topic conversations and discussions that would take over the systems; unmonitored discussions where good input would get lost among a vast number of comments; and students hiding behind anonymity, being rude and aggressive to peers or derailing discussions purposefully. Many of these issues have ethical repercussions, such as causing harm to participating students that we felt a desperate need to deal with.

Hence, we took the extra step of designing and maintaining a system that lets us explore teaching and learning interventions freely and iteratively, with the advantage of stepping in and changing problematic modules of the environment if necessary. Even though we do not have funding for the project, it is running in its 12th year now and we are continuing to iterate for each course based on our experiences with previous iterations.

Generally, the lectures we teach deal with issues of technology assessment and design. Typical instances were *Basics of Human-Computer Interaction*, *Introduction to Informatics and Society*, *Ethics of Technology Design*, or *Introduction to the Philosophy of Science*. Due to regular changes in the curriculum, one design cycle lasts 6 or 12 months, depending on when and how often we are holding lectures that use our technology-enhanced learning environment. For each lecture, the content usually stays similar from one semester to another, however, we have had 3 major lecture changes over the last few years, which led to considerable adaptations in the lecture content and changes about for whom the lectures are compulsory, which in turn significantly impacts the number of students attending each lecture. Additionally, even when the lectures are relatively stable, we continuously update content as well as our general learning design. The impact of this for research is that we very rarely end up with data that can be used for comparative studies between semesters.

Each design iteration of our web-based learning platform is deployed during one semester, which lasts 4 months, during which we are teaching courses of around 600+ students per course and up to 2 courses at once. The general set-up during each semester needs to stay the same and the students need to be informed about the setting at the beginning of the semester. Therefore, we are also constrained by the practical/ethical requirement that there should not be any significant changes to the platform design during the semester. This is a fact that we have struggled with often in the past, since the real life context of teaching gives us very direct feedback on which intervention works, which does not and which might have potential but needs to be adapted to better respond to the challenge it is meant to solve.

Method and Research Question

This paper reflects on our learning experiences in engaging in a design-based research project, focusing on our research approach and process rather than the outcomes of our research, which is discussed in other papers such as Luckner and Purgathofer (2014) or Luckner and Purgathofer (2015). To do so, we are using an autoethnographic stance to present a retrospective study interlaced with an extensive research into related work in the area. We use this retrospective approach, since our emphasis during the actual iteration phases is on how to best design our interactive learning platform to support and help students in their studies, and only thereafter, shifts to analysing the challenges in doing so.

Taking an autoethnographic lens (Ellis et al. 2010), we are not looking to provide an exhaustive account of all our experiences but we are trying to reflect on a subset of challenges that might be useful to other researchers and yield insights or solutions for the process of design-based research. Following Brookfield (1998), we exercise a continual reflection on our practice of and experiences during design-based research, and have repeatedly learned from conversations with colleagues and other researchers in the field that doing so can be a valuable contribution to other, similar research projects. In this learning process, we are relying on the reflective practice described by Schön's (1983) reflection-in-action argument.

For this retrospective analysis, we use data collected continuously throughout our design and research process. We are working with a multitude of data sources, such as: extensive log data, notes from discussions with students - usually taken by at least 2 researchers in parallel, and compared thereafter; systematic feedback collected from students, tutors and other practitioners through surveys and online discussions, expert reviews, and reflections on this data; as well as on our own experiences of the process. For the reflection we worked with thematic clustering

of topics, an iterative process informed by thematic analysis (Braun & Clarke 2006). We used an extensive body of related work to position and reflect these insights within the greater context of design-based research.

With this approach, we focused and singled out certain experiences retrospectively that had an impact on our work from a very personal and individual perspective and reflected those in the context of our personal values and backgrounds, as well as the literature. While other researchers might single out different experiences and come to different conclusions, if put in the same context, we see these experiences as one case study and a grounding to continue discussions and reflections on the topic from different points of view. Additionally, as we present the particularities of our context, we hope to inspire others to do so as well and add to an ever-growing corpus of studies, which in turn can then be used to draw out similarities and contradictions across different contexts. This might lead to a more exhaustive, grounded, contextualised understanding of the setting and hence to a generalisable documentation of design-based research as a method.

The aims of this paper therefore are (1) to provide the reader with a graspable set of challenges and (2) to report on our reflections on the research process and practices of a design-based research project. Additionally, we aim (3) to initiate academic exchange and find a way to communicate about qualitative, design-based research. Over all, we are led by the research question: "*What were the challenges and issues in taking a design-based research approach to develop an enhanced learning environment for large university classes?*".

Challenges

In the following section we reflect on 12+ years working with a design-based research approach in the field of technology-based learning and learning design. While we wholeheartedly subscribe to design-based research as a method to conducting research in our technology-enhanced learning environment, we continuously reflect on the issues we face in the process. In doing so, we regularly stumble across insights that in retrospect appear common sense. Many underlying issues are already discussed in related research, however, based on our specific context and long-term perspective, we believe that we can expand on this knowledge.

In this Section, we try to identify common issues and themes that researchers face when conducting design-based research. We use a classification of issues as research challenges, practical a.k.a. development challenges, and people challenges as already established in the Related Work Section earlier. An additional issue, context, precedes the discussion of these three challenges, since the context of a design-based research project plays a vital role for the approach. For each issue, we briefly describe our specifics. We then list and discuss specific predictions that we deduced from our experience. We use this typology of challenges - context, research, development and people - to help us better reflect on their impact on different areas of our research.

Many of these challenges touch on more than one category and could also be reflected on from a different lens. Figure 2 shows a map of the interrelated nature of the challenges discussed in this paper and gives an overview of all challenges we will refer to.

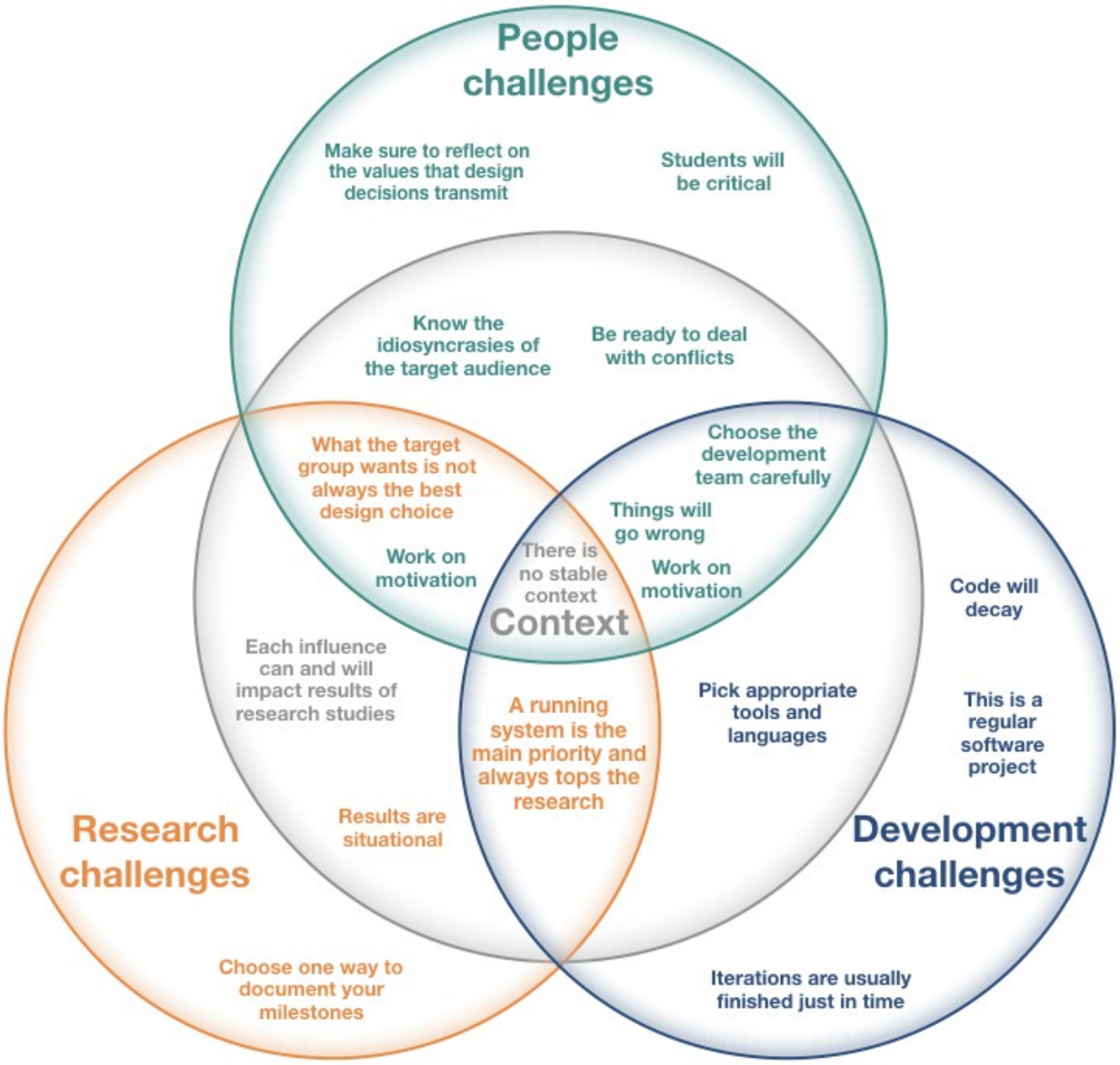

Figure 2. Mapping the challenges

Context

We previously gave an overview of our context in the Section "Setting the context". From this, we can draw out a number of lessons:

There is no stable context. While most related literature talks about design-based research taking place in specific local contexts (e.g., Barab & Squire 2004, Design-based Research Collective 2003), the authors do not often underline the fact, that these contexts themselves are not stable and prune to context changes when it comes to policies, technology and so on. Even though we have been working in the same organisation and offering courses on the same general topic, our project context changed repeatedly and significantly. The list of potential changes in the context is endless, for example we had regular changes in the teaching staff with variable degrees of involvement of practitioners, lecturers as well as tutors; or changes in the curriculum which moved our lecture from a later to an earlier semester, hence influencing the maturity of students and many more. We can directly influence some changes in the context, others, however, will affect the project, no matter what we do or plan for.

Each influence can and will impact results of research studies. Each of the aforementioned changes directly influenced our study results, as they are part of the 'systemic whole' (Brown 1992). We get different results when a lecture is mandatory or voluntary for students. We get different results when we change the type of exercises students have to solve. We get different results depending on the content of a lecture, which in turn impacts how students interact with the system. A lecture about design for example creates a more design critical environment, whereas a lecture about the societal impact of technologies will yield more context critical students. Thus, students in the first course will more likely critique design, whereas students in the second course will more likely comment on privacy, security, or vulnerability of the system. We get different results when we have a bigger lecture staff to support students and so on. While all of these influences will quite obviously impact results, we regularly discover influences that we are not aware of at the time of a study, for example, seeing a steep drop in the quality of peer reviews in the middle of a study, only to find out by chance that it collided with a big test in another course mandatory for many students at the time. These influences are harder to deal with as they yield unexpected and unforeseeable outcomes.

Research Challenges

Most practical research challenges previously mentioned in the Related Research Section, also apply to our context: Each iteration takes at least 6 months in which we cannot change the system, apart from fixing existing bugs in the implementation. The learning context does not easily lend itself to *controlled studies* from an ethical point of view, since we have to guarantee equal chances for all students and we want to avoid using out-dated designs on sections of our students who would form the control group, when we believe that another solution might work better (see also Hutchings 2003). *Comparative studies* between semesters are mostly impossible due to changes in designs and curricula as mentioned above. Changes in the design are usually not restricted to just one area, hence, any outcome cannot be definitely ascribed to a certain design decision. Additionally, internal database structures typically change from one semester to the next, which complicates rerunning evaluations since they cannot simply be repeated the same way as before and which is very time intensive. Changes in lecture content and curriculum can impact the mind-set students are in when using the learning environment and hence the quality and content of collected data. For example, if a lecture is compulsory they might work towards a "tick in the box", if a lecture is chosen voluntarily, it impacts the motivation to succeed. Additionally, even if the design stays the same from one semester to the next, the group of students we work with might in part have previous experience with the system or the lecture and will hence react differently than students with no previous experience with the setting, which might distort comparative data as well.

We deal with this situation by largely focussing on qualitative data collection on different levels. For example, we conduct voluntary surveys with the students throughout and after the semester; we have an issue tracker where students can participate in gathering issues with the technology; we have so called 'student café' sessions, where we invite groups of students in an informal setting to talk to us about their grievances and approval of the system and the general learning design; we regularly meet tutors and lecturers who provide feedback to the system from their points of view; and we also rely on 'reflective practitioner' methods to include our own experiences of the research and design process. Furthermore, we do regularly evaluate our vast amount of log data and student-generated context, to better understand how our interventions impact said data.

Some lessons we learned from this research process over the years are:

What the target group wants is not always the best design choice. We come from a Human-Computer Interaction context and hence adopt many user-centred design methods such as focus groups or workshops with students and tutors to better understand their needs and issues with the system. These approaches introduce a whole other set of challenges, such as weighing priorities of users with our own strategic goals for the learning environment. For example, students might not always be a trustworthy source when it comes to their opinion about the learning design of the system, since their interests often lean towards less course work or easier completion of a course, while our goal is a better learning experience and learning outcome. This is not necessarily reconcilable.

A running system is the main priority and always tops the research. Due to the large number of students we end up with huge amounts of mostly qualitative data after each semester that has to be evaluated swiftly to spark the next design/technical iteration of the learning environment. This iteration cycle becomes especially hard when using the system in 2 consecutive semesters with less than 4 weeks in between. Even when we try to continuously evaluate data throughout the semester, the focus is on keeping the system running smoothly and evaluation is seldom prioritised. Another issue in this category is deciding which interventions to prioritise in

upcoming design iterations. Often we find ourselves with a long list of ideas and fixes to implement, but since we usually work under tight time constraints, we have to choose which can be finished and which will have to wait for another year. In such situations, we often find ourselves again choosing the practical outcome before the scientific outcome, since the system still first and foremost has to be functional for the next semester and cater to students, a very appropriate criticism made by Walker (2011) on the design-based research approach.

Results are situational. We collect different types of data, such as for example the quality and frequency of student's work, interactions with the system and so on. Data that we collect at the beginning of a semester usually differs from data collected toward the end, in that students have a learning curve, not only content-wise but also in the use of the learning environment and the learning design. They become increasingly more familiar with the system, which makes for different evaluation results in different phases of the semester. Additionally, when using the system simultaneously in two different lectures, we often saw diverging outcomes of our studies, finding our hypotheses about how an intervention will impact the students' progress supported in one lecture, but not seeing any impact in another. This led us to understand, that not only the context in which the lecture is held and by whom has an impact, but also the type of content, however similar, and hence the students' approach to using the system greatly influences success or failure of an intervention. As a result we can usually only improve our understanding of a studied problem rather than finding definite yes or no answers to given hypothesis and often have to step away from our goal of trying to find generalisable outcomes that would add to a common research agenda as advocated for by Barab and Squire (2004).

Choose one way to document your milestones. A very practical research challenge when doing long-term research is to keep track of changes made or new design ideas. We did not understand the consequences of this challenge for a long time, as early on we had not anticipated for the project to become such a long undertaking. For example, over the years we used numerous different services for internal communication, some of which were discontinued, some of which we dropped due to usability issues. Therefore, a lot of documentation of our research and project process got lost, which left us with a fragmented and incomplete version and it is very hard to recreate actions, design decisions and reasoning after the fact.

Development Challenges

As in every software project, we face constant issues regarding the development of the system. As already mentioned by related work, a system is only as good as its developers. Being an unfunded project, our developers are mainly drawn from students who are generally interested in the topic or who want to use the work towards their Bachelors or Masters project. This means that the quality of the code as well as motivation of the developers varies, as does meeting deadlines and finishing features in time to test and deploy them before the start of a new semester and the total time of involvement in the project.

Pick appropriate tools and languages. In our project, we had 2 major resets of the code base where we restarted from scratch. The first time, we felt like we had learned so much about the project that it made more sense to start anew rather than letting a new team of developers work with and clean a decaying codebase. Because of the new team's preferences, we switched from Python-based Django to Java. The specific dynamics and flexibility necessitated by our research by design approach collided with this decision, as the new environment was more geared towards a slow-and-rigorous approach that we could not accommodate. Thus, two years later the next development team switched back to Django, which felt like a more suitable development environment for the kind of backend/frontend separation and independence we needed. While both switches had made sense at the time, redoing the whole project took a lot of time and effort, which felt like reinventing the wheel. In a similar vein, we learned to avoid the 'hip' framework or toolkit of the day, instead relying on components that show a promise of continued and mature further development.

Code will decay. While it is debatable if it makes sense to sometimes restart from scratch again, an unavoidable issue in software development is code decay and code maintenance. With regular updates or discontinuations of web frameworks and plugins, there is a constant need to maintain and update the existing code base to keep it running. Working with out-dated versions is a valid option for some time, but with rising awareness of security risks posed by old software, this becomes less and less attractive. Also, often enough there is some new feature that requires an update to a newer version in one software module of the project, which ends up cascading into the rest of the project. It is highly frustrating for the team to see things break without it being their fault. A solution to deal with this would be to schedule regular maintenance and refactoring runs, however, lacking the

resources in a non-funded but ever-changing system like ours, code maintenance and sustainability is often rather low on the priority list.

Iterations are usually finished just in time. This is a fact we had to learn the hard way. Over years of development, we kept planning on finishing the development of a new design early to find time for a more thorough testing phase. However, in reality we usually finished just in time for the next iteration to start. There are usually many reasons for this issue to arise, some unpredictable, others however, are structural. The project is conducted as a non-funded pet project. As a consequence, it cannot always be the first priority of all researchers involved, which sometimes leads to a general delay in the schedule. Another accompanying factor of the un-funded nature of the project is the fact that we depend on students to do much of the development work, and that often brings along a certain "just-in-time" mentality when it comes to deadlines. This per se would not be problematic, if it is accounted for in the planning, however, many modules of the system are tightly intertwined, and we have not always managed to put together a working list of milestones to make the development run smoothly without a delay in one module delaying another.

This is a regular software project with all its strengths and weaknesses. The biggest strength of running the project team and building our own environment is the immediate control we have over the systems. Very problematic issues, however, are bugs and security. This is a multi-layered problem since it ties in with technological as well as societal problems. The technological issue with bugs is rather straightforward: we try to avoid them at all costs but can never remove them from the system entirely. Our strategies to find them in a timely manner is to do as much testing of new features as possible, and also invite outsiders for trial runs before the new semester starts. These outsiders, usually a mix of tutors, students and developers, are given random roles in the system and are invited to do structured test runs going through every possible interaction we can think of. While the test runs usually point us towards major issues in the system, we always stumble across additional bugs during the semester, which we try to fix quickly. Most bugs do not affect key features of the system, however, we did occasionally come across major security breaches during the running of our project. One of the worst cases was a port left open by a developer of a sub module, which made all previous work of every student attending the lecture available to everyone and opened the doors for malicious use of data such as for plagiarism. This brings about societal issues concerning development, i.e., ethical or learning consequences of bugs.

People challenges

There are multiple challenges concerning the people involved in the project. This being a pragmatic project that tries to solve problems in teaching in addition to a research project, we work with people from diverse backgrounds and with equally diverse interests and strength; people who sometimes take on more than one role in the project; people who themselves constantly change and adapt to their changing circumstances.

The real-life context not only affects the research, but each *student* using the learning platform deals with real life consequences if they fail because of a failed technology experiment on our side. Students are the main target group; the project would not exist without them. They are not just data points towards a better educational system. In our context, we are using the technology-enhanced learning platform in compulsory courses. Hence it will always affect grades, as well as engagement and motivation in the lecture. In order to avoid negative consequences for students, we have to carefully consider how to be transparent about the inevitable technological failures, as well as how to give students an opportunity to succeed unimpaired by said failure. When designing a new intervention, we need to ask ourselves, whom the intervention is for. Are we trying to support at-risk students, are we trying to further engage top students or are we working towards encouraging the intermediate student segment? These decisions will impact what the new iteration will look like, which features take priority, and also how we want to evaluate data to better understand the intervention.

Tutors are another user group we have to keep in mind when designing our system. While this group is by far smaller in numbers - only about 12 tutors per semester - the system will affect their work ethic, their workload and the quality of their work. However, we usually have a good rapport with the currently active group of tutors and they are happy about every change that makes their life easier, so this aspect of the system can actually be adapted even during the semester as it does not concern students grades immediately, thus we have shorter iteration cycles with more immediate feedback as to how the new version works. This makes for a very interesting segment of the project, as we can quickly see how changes affect motivation and speed regarding the tutors' work, which sometimes also sparks ideas for the next iteration of the student's view of the platform.

We have *lecturers* who are not directly involved in the research aspect of the project, however, they are happy to use the system as long as it does not provide them with additional work and it supports the underlying learning design. While this group is one of the least influential in the redesign process, we do have to make sure to keep the system approachable and easy to use.

We, the authors of this paper, are holding very different roles in the development, maintenance and running of the learning platform. We are teachers, *designers* as well as *researchers* but were also students, tutors and developers in previous iterations of the system. This gives us a unique holistic view on the system as a whole and an intimate understanding of the needs of different groups of people.

As becomes apparent in our setting and is also mentioned in related research (e.g. Kennedy-Clark 2013), people hold very different roles in the project. Figure 1 shows a snapshot of our current setup with former students who are now tutors or developers, tutors who are also developers, lecturers who are also researchers and designers, researchers who also develop and so on. Orange overlaps in the figure depict connections of the recent past - tutors and developers in the system have previously taken courses as students in which the system was used and hence, might have greater feelings of solidarity and empathy towards current students. Purple overlaps show conflicting roles in the project team. This figure illustrates the constant inner struggle when it comes to prioritising and points to why conflicts can arise in the decision-making process.

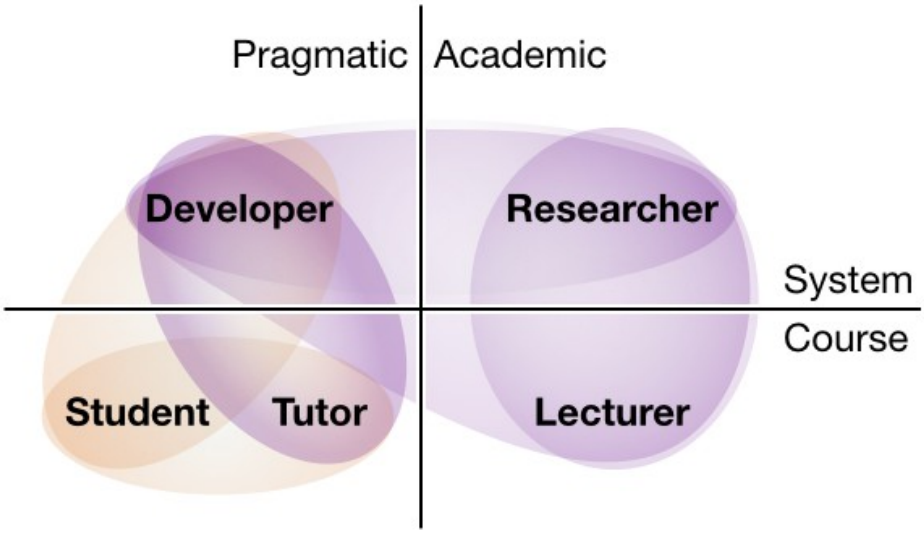

Figure 1. Snapshot of conflicting roles in the project

While we are facing the "usual" issues of design-based research, as described in the Related Work, here are our lessons learned concerning people challenges:

Students will be critical. Many students are generally motivated and appreciate a learning concept that differs from 90% of university courses. However, we cannot create a system that everyone is happy to use. Even if we feel like students will profit and take away more from using our learning design, students' motivations often differ from our goals. Additionally, even if they might generally be pleased with the learning design, students seem to find it easier to publicly voice criticism than to praise. We learned from conducting surveys and directly speaking to groups of students after each semester that the publicly expressed discontent, even if it was supported by much peer approval, is not representative of what the majority of students think of the system and the learning design.

Know the idiosyncrasies of the target audience. For example, we are working with and for students of informatics/computer science. This is an immediate advantage, as we need developers for our system. However, it also makes our work harder because informatics students are less content with custom-made software, thus often complain about shortcomings or bugs in the software from a somewhat qualified perspective, or even take it on themselves to test for flaws in the system. Being critical of software that they have to use is a specific trait of informatics students, as it makes them appear critical and knowledgeable in their field at the same time. But because it is hard to be critical and give productive feedback, this then produces some very well written, unpleasant feedback

addressed to the designers and developers. Also, some students try to take advantage of found flaws rather than report or help fix them, which create additional stress for the development team.

Choose the development team carefully. And if you cannot find suitable replacements for team member in time, fewer might be more. While the workload might sometimes look insurmountable with a smaller team, taking on additional team members might be more of a struggle than it's worth. Each member of the team needs attention, especially at the beginning and needs to feel like being part of the group.

Work on faculty and development team motivation. Design-based projects are long-term projects. They demand constant attention and effort, but don't always provide adequate motivation to keep moving on. While we are the core team that has accompanied the project since the beginning, we choose, supervise and guide an ever-changing team of collaborators who join for various reasons and might not be as invested in the success or failure as we are. It is sometimes hard not to get disheartened when things do not turn out the way we planned, when design interventions are not finished on time, or when we see a lack of engagement in issues we feel very strongly about. There is a constant need to not only keep our own motivation in check, but to radiate it out to your team to keep them engaged, especially if things go wrong. External rewards are sparse - so it is important to be reminded of the internal motivation to drive the project.

Things will go wrong. Not every intervention will yield the desired results. As researchers, designers and teachers we have to be ready to 'kill our darlings' and discontinue interventions when they repeatedly fail. This step is not easy especially since there is usually a large chunk of work attached to even the smallest seeming change in the design. Additionally, we sometimes had to let go of interventions that worked very well but that were made useless by another redesign.

Be ready to deal with conflicts. With a vast number of people involved and their needs not always lining up, there will be disagreements. Even without external people, conflicts can arise due to our different roles in the project. For us, a helpful step to solving conflicts, personal and design-related ones, is to have a strongly connected core team to agree and disagree with, people to bounce ideas off, regular meetings with all different people in the system, workshops with experts, ideally with different backgrounds and points of view, and different channels to vent when frustrated or exhilarated.

Make sure to reflect on the values that design decisions transmit. As mentioned by Amyl and Reeves (2008), each change in the design will change how students can interact, not only with the system, but with the lecture context, with their own work and with other users of the system, be it peers, tutors or teachers which transports certain values. As researchers, designers and developers of the learning platform you should try to understand the possible impact of an intervention before you deploy it.

Conclusion

Design-based research is "research in the wild". Thus, it is often unpredictable, and it has to cope with many issues that more controlled forms of research are spared from. While a failed hypothesis in science advances the knowledge in the field in general, a failed hypothesis in our project has repercussions in real life. It might just produce miserable experiences for students, for tutors, and for us; it can lead to serious stress for either party; or, in the worst case, it causes students to fail the course. We carry a grave responsibility to treat our students equally, and to never let them suffer from bad design decisions on our part. This is a source of constant doubt and countless ethical challenges. Again and again, we face decisions to either compromise our research or our students. In all conscious decisions we of course prioritise students, promising them to always err on their side.

This points to a basic conflict in our research through design approach; on the one hand, we want to design a good, usable and meaningful system. Therefore, we introduce changes designed to fix known flaws, to accommodate specific needs and to adapt to an ever-changing context. On the other hand, constant modifications of the system we design also induce changes in the context. Design-based research thus has to cope with a process that can be described as »hermeneutic«: every modification changes what can happen in an unpredictable way, and what happens in turn makes modifications necessary. We know of no way to avoid this reciprocal dependency in long-term design-based research, and have learned to accept it as the driver of reflection and insight. To give an example of how we try to institutionalise a reflective culture, we implemented a tracker system where students can at any time post and discuss feedback, critique, ideas, and bugs with us.

For this paper, we compiled reflections and insights from 12+ years of design-based research into predictions, structured by a categorisation of issues that come up during our work. These predictions are discussed

so readers can extrapolate their relevance in different contexts. At the same time, these assertions can be understood as communication starters. They reflect our experiences more than anything else. While we are working in a very specific context, we would be interested in a constructive dialogue and in building up a corpus of cases from which we can later derive more generalisable lessons.

References

Alexander, P.M. (2006). Virtual Teamwork in Very Large Undergraduate Classes. Computers & Education, 47(2), 127-147. Retrieved April 9, 2018 from https://www.learntechlib.org/p/67459/

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. Journal of Educational Technology & Society, 11(4), 29–40.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? Educational Researcher, 41(1), 16–25.

Barab, S. (2014). Design-Based Research: A Methodological Toolkit for Engineering Change. In Handbook of Learning Sciences. Cambridge University Press.

Barab, S., & Squire, K. (2004). Design-Based Research: Putting a Stake in the Ground. Journal of the Learning Sciences, 13(1).

Brookfield, Stephen D. (1998). Critically reflective practice. Journal of Continuing Education in the Health Professions. 18 (4): 197–205. doi:10.1002/chp.1340180402

Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. Journal of the Learning Sciences, 2(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chang, R. L., & Gray, K. (2013). Ethics of research into learning and teaching with Web 2.0: reflections on eight case studies. Journal of Computing in Higher Education, 25(3), 147–165. https://doi.org/10.1007/s12528-013-9071-9

Cobb, P., Confrey, J., Lehrer, R., & Schauble, L. (2015). Design Experiments in Educational Research. Educational Researcher, 32(1), 9–13. https://doi.org/10.3102/0013189X032001009

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. Journal of the Learning Sciences, 13(1), 15–42. https://doi.org/10.1207/s15327809jls1301_2

Dede, C. (2004). If Design-Based Research is the Answer, What is the Question? A Commentary on Collins, Joseph, and Bielaczyc; diSessa and Cobb; and Fishman, Marx, Blumenthal, Krajcik, and Soloway in the JLS Special Issue on Design-Based Research. Journal of the Learning Sciences, 13(1), 105–114. https://doi.org/10.1207/s15327809jls1301_5

Design-based Research Collective, T. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. Educational Researcher, 32(1), 5–8. https://doi.org/10.3102/0013189X032001005

Dominguez, C., Nascimento, M. M., Payan-Carreira, R., Cruz, G., Silva, H., Lopes, J., … Morais, E. (2014). Adding value to the learning process by online peer review activities: towards the elaboration of a methodology to promote critical thinking in future engineers. European Journal of Engineering Education.

Driscoll, M. (2000). Psychology of Learning for Instruction. Needham Heights, MA, Allyn & Bacon.

Edelson, D. C. (2002). Design research: What we learn when we engage in design. The Journal of the Learning Sciences, 11(1), 105–121.

Ellis, C., Adams, T. E., & Bochner, A. P. (2010). Autoethnography: An Overview. Forum: Qualitative Social Research, 12(1). Retrieved from http://nbn-resolving.de/urn:nbn:de:0114-fqs1101108.

Francis, R.W. (2012). Engaged: Making Large Classes Feel Small through Blended Learning Instructional Strategies that Promote Increased Student Performance. Journal of College Teaching & Learning, 9(2), 147-152. Retrieved April 9, 2018 from https://www.learntechlib.org/p/49913/

Gehringer, E. F. (2000). Strategies and mechanisms for electronic peer review. In Frontiers in Education Conference, 2000. FIE 2000. 30th Annual (Vol. 1, p. F1B/2–F1B/7 vol.1). https://doi.org/10.1109/FIE.2000.897675

Herrington, J., Mckenney, S., Oliver, R., Herrington, J., & Mckenney, S. (2007). Design-based research and doctoral students : Guidelines for preparing a dissertation proposal. EdMedia World Conference on Educational Media & Technology, (2005).

Hutchings, P. (2003). Competing Goods. Change, (September/October).

Kanuka, H., & Anderson, T. (2007). Ethical issues in qualitative e-learning research. International Journal of Qualitative Methods, 6(2), 20–39.

Kelly, A. (Eamonn). (2004). Design Research in Education: Yes, but is it Methodological? Journal of the Learning Sciences, 13(1), 115–128. https://doi.org/10.1207/s15327809jls1301_6

Kennedy-Clark, S. (2013). Research by design: Design-based research and the higher degree research student. Journal of Learning Design, 6(2).

Kennedy-Clark, S. (2015). Reflection: Research by design: Design-based research and the higher degree research student. Journal of Learning Design, 8(3).

Kokkelenberg, E. C., Dillon, M., & Christy, S. M. (2008). The effects of class size on student grades at a public university. Economics of Education Review, 27(2), 221–233. https://doi.org/10.1016/j.econedurev.2006.09.011

Luckner, N., & Purgathofer, P. (2014). Explorative Design as an Approach to Understanding Social Online Learning Tools. International Journal on Advances in Intelligent Systems, 7(3&4), 493 – 506.

Luckner, N., & Purgathofer, P. (2015). Exploring the Use of Peer Review in Large University Courses. Interaction Design and Architecture(s) Journal, (25), 21–38.

McKenney, S., & Reeves, T. C. (2013). Systematic review of design-based research progress: Is a little knowledge a dangerous thing? Educational Researcher, 42(2), 97–100.

Monks, J. & Schmidt, R. (2010). The impact of class size and number of students on outcomes in higher education. Retrieved May 8, 2018 from Cornell University, School of Industrial and Labor Relations site: http://digitalcommons.ilr.cornell.edu/workingpapers/114/

Mulryan-Kyne, C. (2010). Teaching large classes at college and university level: challenges and opportunities. Teaching in Higher Education, 15(2), 175–185. https://doi.org/10.1080/13562511003620001

Obrenović, Ž. (2011). Design-based research: what we learn when we engage in design of interactive systems. Interactions, 18(5), 56–59.

Piaget, J. (2013). The Construction of Reality in the Child (Vol. 82). Routledge.

Plomp, T. (2013). Educational design research: An introduction. Educational design research.

Reeves, T. C., Herrington, J., & Oliver, R. (2005). Design research: A socially responsible approach to instructional technology research in higher education. Journal of Computing in Higher Education, 16(2), 96–115.

Reimann, P. (2011). Design-Based Research. In L. Markauskaite, P. Freebody, & J. Irwin (Eds.), Methodological Choice and Design. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-8933-5

Rittel, Horst W. J. & Webber, Melvin M. (1973). Dilemmas in a General Theory of Planning. Policy Sciences. 4: 155–169. doi:10.1007/bf01405730.

Rogers, Y. & Marshall, P. (2017). Research in the Wild. *Synthesis Lectures on Human-Centered Informatics*, *10*(3), i-97.

Salter, D., & Purgathofer, P. (2010). Students Use of Laptops in Large Lecture Classes : Distraction , Partial Attention or Productive Use? Proc. ED-MEDIA 2010, 2010(1), 2136–2141.

Scanlon, E., O'Shea, T. M. M., & McAndrew, P. (2015). Technology-Enhanced Learning: Evidence-based Improvement (pp. 229–232). ACM Press. https://doi.org/10.1145/2724660.2728664

Schön, D. A. (1983). The Reflective Practitioner: How Professionals Think in Action. (T. Smith, Ed.) (Vol. 1). Basic Books.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning, 2(1), 3–10.

van den Akker, J. (1999). Principles and Methods of Development Research. In R. M. Branch, K. Gustafson, N. Nieveen, T. Plomp, & J. van den Akker (Eds.), Design Approaches and Tools in Education and Training (pp. 1–14). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-011-4255-7

Walker, R. (2011). Design-Based Research: Reflections on Some Epistemological Issues and Practices. In L. Markauskaite, P. Freebody, & J. Irwin (Eds.), Methodological Choice and Design. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-8933-5

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. Educational Technology Research and Development, 53(4), 5–23. https://doi.org/10.1007/BF02504682

Zheng, L. (2015). A systematic literature review of design-based research from 2004 to 2013. Journal of Computers in Education, 2(4), 399–420. https://doi.org/10.1007/s40692-015-0036-z