*Article*

# Does Scientific Evaluation Matter? Improving Digital Simulation Games by Design-Based Research

**Sven Ivens *** and **Monika Oberle**

Department of Political Science, University of Goettingen, 37073 Goettingen, Germany;
monika.oberle@sowi.uni-goettingen.de
* Correspondence: sven.ivens@uni-goettingen.de

check for updates

**Abstract:** Grounded in a design-based research approach, the aim of this article is to determine whether scientific evaluations help to (a) identify and fix problems in educational interventions and (b) eventually foster a more effective and positive evaluated intervention. Therefore, data from a longer-term evaluation of short digital simulation games about the European Parliament for civic education in schools were used. The data included three cycles of interventions with pre- and post-evaluations starting with the first prototype in 2015/2016 ($n = 209$), the second cycle in 2017/18 ($n = 97$), and the last one in 2019/20 ($n = 222$). After each evaluation, major problems and critiques regarding the simulation game were discussed with the developers, and changes were implemented in the game design. The four most important problems, the processes by which they were improved and the reactions of the participants in the following evaluations are pointed out in the article. A comparison of the last and first evaluation cycle showed an overall improvement of the simulation game regarding its effectiveness in transferring EU knowledge and the participants' general satisfaction with the simulation game. This study underlines the value of the design-based research approach for developing educational interventions and can be useful for further work on civic education measures and the implementation of digital simulation games.

**Keywords:** digital simulation game; design-based research; empirical research; civic education; European Union

## 1. Introduction

The effectiveness of active citizenship education programs, tools, and interventions are often measured and evaluated by scientists in many different ways, reaching from qualitative interviews or self-reflections to (quasi-) experimental studies and large n survey studies. But what happens after the evaluation is done and the article is published? Is measuring and evaluating civic education programs and teaching tools an end in itself, or does doing so help to actually increase the effectiveness of these programs? This article investigates the usefulness of scientific evaluation in civic education according to the design-based research approach. This quite new research approach goes a step further, as it focuses not only on evaluating educational interventions but also on enhancing them according to the collected evaluation data and then evaluating them again to see if the changes were effective until the educational intervention leads to a satisfactory result (Anderson and Shattuck 2012). It has recently been used for research on subject didactics in school (Peters and Roviró 2017), on higher education (Ford et al. 2017), blended learning (Ustun and Tracey 2020), and simulation games (Koivisto et al. 2018) but has not yet been used for improving digital simulation games in civic education.

To break new ground, the design-based research approach presented here uses data from an ongoing evaluation study about short digital simulation games in civic education (Oberle et al. 2017). Analog simulation games are a widely used method in civic education (Massing 1997), and simulation

game producers are trying to digitalize them in order to use the positive effects of digitalization for their products (Kaiser et al. 2017). Since 2015, a research team at the chair of Political Science/Didactics of Politics of Göttingen University has been evaluating the usage of a newly developed short digital simulation game about the European Parliament in school classes, developed and implemented by the Berlin-based German company planpolitik (www.planpolitik.de). Three evaluation cycles were conducted in 2015/2016, 2017/2018, and 2019/2020 (for results of the first cycle of evaluated game implementation, see Oberle et al. 2017). Between the iterations, the results were presented to the game developers, and possible dysfunctional elements were pointed out. Based on these data and on the overarching question "Does Scientific Evaluation Matter?" this article will investigate the following research questions:

(1) Did the developers fix the problems that the evaluations revealed, and if so, how did they fix it?
(2) Did the changes made by the game developers lead to a more effective and more positively assessed intervention?

The questions will be answered using data from all three evaluation cycles, mainly focussing on the first and last evaluation as they are similar in their sample composition and are showing a direct contrast between the first prototype and the final product. Additionally, an interview with planpolitik's head developer was made to identify and explain the improvements.

The article is structured as follows: after this introduction the use and literature of digital simulation games in civic education as well as the design-based research approach will be elucidated, followed by a detailed description of the research design. In the results section, the shortcomings of the simulation game, the improvements needed to address them and the effectiveness of the improvements will be explained one after another, ending with a direct comparison of the first and last evaluations of the game. A conclusion will sum up the most important findings and place them in context with the overarching question of the article.

## 2. Digital Simulation Games in Civic Education

Face-to-face simulation games are a widely used method in civic education (Massing 2010), and their value for civic education in schools (Oberle et al. 2018; Oberle et al. 2020) as well as in political science education in universities (Fink 2015; Duchatelet 2019; Lohmann 2019) has been extensively discussed. On the contrary, there is hardly any research, let alone empirical research, about digital simulation games in civic education (Bachen et al. 2015; Oberle et al. 2017). Empirical studies about simulation-based learning in education, in general, show that simulations have a more positive effect when used in combination with modern technology, like VR or digitalization (Chernikova et al. 2020).

There are many expectations and theoretical considerations regarding the benefits of digital simulation games: e.g., they are supposed to improve intrinsic motivation (Le et al. 2013) and strengthen negotiation skills, teamwork, and empathy (Gabriel 2012). The digitalization of simulation games also has the potential to multiply the reach of the didactical method of simulation games, as they can be played online in a fully prepared gaming environment. Thus, the presence of a professional instructor to administer them in the classroom is no longer needed (Kaiser et al. 2017). Furthermore, they can be played independently from location and time. This can make them accessible for a broader range of pupils and, given the technical equipment, relatively easy to implement into regular teaching.

Alongside theoretical articles, most studies regarding the use of digital simulation games are case studies reported by teachers in practitioner journals (DiCamillo and Gradwell 2012), qualitative studies taking a student's perspective (Schnurr et al. 2013), or are based on computer/video games (López and Cáceres 2010; Motyka 2018). Overall, the results of these qualitative studies show that participants are rather satisfied with the digital simulation games but also voice criticism, especially regarding a lack of face-to-face communication and a high workload. However, there is a scarcity of quantitative, quasi-experimental studies regarding the effects of digital simulation games in civic education as well as their development over time. An example of a good quasi-experimental study has been conducted

by Bachen and colleagues (Bachen et al. 2015) who used a pretest-posttest design to analyze the effect of a digital simulation game on 301 American high school students. Their results showed an increase in political interest, especially by uninformed and "low-performing" students. They also criticized the lack of studies of digital simulations in civic and social science education.

Lastly, there was a pilot study by Oberle and colleagues (Oberle et al. 2017), which analyzed the effects of a short digital simulation game from the German company planpolitik about decision-making in the European Parliament. This research will be elaborated upon further in the research design portion as the study presented here is based on the continuing evaluation of this digital simulation game. The pilot study suggested that there is not enough research on the topic and that the theoretical expectations in digital simulation games might be exaggerated as the study shows that an analog simulation game worked better in comparison. Furthermore, the digital simulation game had hardly any effect on any of the student dispositions measured; it even had a negative effect on objective EU knowledge. Additionally, the students assessed the analog simulation game more positively than the digital one. However, the authors also note that the digital simulation game was in an early stage, and digital simulation games in civic education, in general, are a very new development, which has not yet been studied sufficiently (Kaiser et al. 2017). To tackle this research deficit, the Göttingen research team continued to evaluate the digital simulation, shared the research with the developers, and used a design-based research approach to further improve the digital simulation game as a useful educational measure for teaching about the European Union (EU).

## 3. Design-Based Research in Educational Science

Design-based research is an increasingly popular practical research method that helps to bring research and practice together in order to generate effective educational interventions and methods (Van den Akker et al. 2006). Educational researchers started using it at the beginning of the 21st century to "increase the impact, transfer and translation of education research into improved practice" (Anderson and Shattuck 2012, p. 2) and simultaneously develop their theories further. It is mainly used to study innovative learning environments, especially those using new technologies and/or complex interventions in a classroom setting (Sandoval and Bell 2004) and should be used in new fields where pedagogical content knowledge, e.g., regarding instructional strategies, is poor (Ford et al. 2017). Thus, it is a promising method to study digital simulation games about (European) politics.

Design-based research can complement experimental and quasi-experimental research but differs from it in some important aspects (Hoadley 2004; Jen et al. 2015). However, it goes further than just evaluating the intervention and its effects: the method at hand also focuses on optimizing the interventions. Therefore, the improvement of the intervention itself is an outcome (Hoadley 2004; Design-Based Research Collective 2003). To accomplish this, the research methodology is quite open, and researchers can "select and use differing methods, selecting them as they see need" (Maxcy 2003, p. 59), which includes but is not limited to (quasi-) experimental pre-/post designs, developmental evaluations (Patton 2011), interviews, observations, and questionnaires.

Design-based research also needs more than one evaluation cycle and often multiple iterations of testing, evaluating, and improving the interventions until the educational intervention is suitable for its purpose (Lewis et al. 2020), which makes it "difficult to know when (or if ever) the research program is completed" (Anderson and Shattuck 2012, p. 2). On that score, design-based researchers argue that the implementation and evaluation of the intervention have to be taken out of the laboratory and have to be studied in a real-world setting (Barab and Squire 2004).

Lastly, design-based research builds on close cooperation between practitioners/developers and the researchers themselves (Kuhn and Quigley 1997; Štemberger and Cencic 2016), as the practitioners usually lack the necessary skills for scientific research, and researchers lack the technical knowledge and practical skills to develop and implement the intervention.

This close relationship between practitioners and researchers as well as the in-depth involvement of researchers in the development process of the intervention are two main points of criticism against

design-based research (Barab and Squire 2004). Similar to many qualitative methods, there is a "narrow line between objectivity and bias" (Anderson and Shattuck 2012, p. 5) for researchers using the method, and similar to qualitative methods, there are different ways to minimize this bias (Onwuegbuzie and Leech 2007). One would be to triangulate the data drawn from the evaluations with a researcher who is not involved in the project.

Another problem of design-based research is the long duration of design-based research projects, as they usually require at least two iterations, often more, to be complete (Anderson and Shattuck 2012). To tackle this, Herrington and colleagues (Herrington et al. 2007) recommend using design-based research in doctoral dissertations over a four-year period or in multi-year research agendas.

Many examples of the use of design-based research have been published in edited volumes in recent years (see, for example, Kay and Luckin 2018). One Ph.D. thesis, in particular, was focused on the study and development of analog simulation games for civic education, the cumulative dissertation of Knogler (2014). In his dissertation, he developed a school course over three weeks, including an analog simulation game about the future of a community's energy supply (Knogler and Lewalter 2014). Following the design-based research approach, he tested the prototype with 112 pupils in German high schools, evaluated the test with a pretest-posttest design, and used the results to further develop the intervention. The re-developed intervention was used on another sample of 156 German pupils and tested again with the same pretest-posttest design. With that design, he could show that only the re-designed intervention had a positive effect on the pupils' appreciation of the value of science. All in all, the prototype had hardly any effect, whereas the re-design intervention affected all measured dimensions (Knogler and Lewalter 2014). The research design in the article is similar to the design by Knogler but with multiple cycles of testing, evaluation, developing, and testing again, and will be described in detail in the next chapter.

## 4. Research Design

Following the design-based research approach, the design of this research was focused on an intervention, a 90-min-long synchronous digital simulation game. The simulation game was designed for secondary level high school students in Germany and is about the European Parliament and its political decision-making process regarding the topics of asylum and data protection policy. The game was developed by the Berlin-based company planpolitik and implemented with an accompanying scientific evaluation by the chair of political science/civic education at the University of Göttingen in 2015/2016. The results of the scientific evaluation were then discussed with the developers from planpolitik, and changes were made accordingly. A second iteration with a smaller sample (see Table 1) followed in 2017/2018, and a third iteration with a larger sample in 2019/2020. For this article, we mainly focus on the differences between the first and the last evaluation cycles as both samples are quite similar, and the sample of the second iteration is quite small (see Table 1).

**Table 1.** Pretest-Posttest Sample Description for All Three Evaluation Cycles.

|  | Sample 2015/2016 $n = 209$ | Sample 2017/2018 $n = 97$ | Sample 2019/2020 $n = 222$ |
|---|---|---|---|
| Sex Female | 54.1% | 57.3% | 61.1% |
| Grammar School | 71.3% | 100% | 79.8% |
| Age | 15.89 (SD 0.72) Range 15–18 Years | 15.59 (SD 4.57) Range 15–18 Years | 16.86 (SD 2.52) Range 14–29 Years |
| Migration Background * | not enquired | 36.2% | 23% |
| Cultural Capital ** | 4.73 (SD 1.29) | 4.57 (SD 1.25) | 4.74 (SD 1.35) |
| Classes | 13 | 6 | 13 |

* = measured by place of birth and parents' place of birth; ** = measured by amount of books in parents' household (Sieben and Lechner 2019).

The main defaults of the simulation game, which the first evaluation pointed out, will be presented. The changes planpolitik made according to this evaluation will be disclosed using an interview with Konstantin Kaiser from planpolitik, who is in charge of digital simulation games and their development. The results of the last evaluation will clarify whether the changes had a positive effect on the result and evaluation of the simulation game in 2019/2020. The results and evaluation of the 2015/2016 and 2019/2020 iterations will be compared in the crucial points to determine whether the designed-based research approach has led to more effective evaluation.

### 4.1. Sample Description

All classes that played the simulation game were from schools in the German state of Lower Saxony since the European Information Center of Lower Saxony is funding the implementation of the games at school. The European Information Center is also in charge of promoting the simulation game and recruiting classes to play it. Therefore, the sample selection was not random; teachers signed in with their classes voluntarily. That also led to the different sizes of the respective samples (see Table 1), as there were fewer classes signing in for the game in 2017/2018, as well as fewer teachers willing to facilitate the scientific evaluation. The samples not only differ in size, but also in school type, as in 2015/2016 pupils from grammar schools and comprehensive schools participated, in 2017/2018 only pupils from grammar schools, and lastly, in 2019/2020 classes from grammar schools and vocational schools signed up to take part in the simulation game. This resulted in an age difference between the last and the first sample, which is a critical point of this study and will be further addressed in the discussion of the limitations of this article.

Nevertheless, the focus of this article is on the first and third evaluation cycles, as these samples are quite similar in size, cultural capital, and the percentage of pupils attending grammar schools. A possible migration background was not assessed in the first study. All three samples consist of pupils who participated both in the pretest and in the posttest.

### 4.2. Research Instruments

The accompanying scientific evaluations were made in a pre-, post-, and follow-up design including mainly closed yet some open questions as well as an EU knowledge test in the first and third samples, modeled based on the EU knowledge test by Monika Oberle (Oberle 2012). The survey measures different concepts including subjective EU knowledge, internal efficacy, attitudes towards the EU, and political interest (derived from Deutsche 2010; Gille et al. 2006; Kerr et al. 2010; Oberle and Forstmann 2015; Vetter 2013; Westle 2006). The evaluation of the simulation game is carried out in the posttest with 19 four-scaled Likert items with 1 for "not at all" to 4 for "fully agree" which measure the simulation game in the dimensions of general satisfaction, subjective learning effect, and motivational effect (see Table 2). All statistical analyses conducted for this paper were completed with SPSS 25. The posttest survey also includes open questions regarding what participants liked, disliked, and would improve in the simulation game, which were analyzed and categorized with the qualitative content analysis after (Mayring 2010) using MAXQDA 2018 (VERBI GmbH, Berlin Germany).

**Table 2.** Measurement Model of Simulation Game Assessment.

| Dimensions Simulation Game Assessment | Number of Items | Sample 2015/2016 *n*209 Alpha Value | Sample 2017/2018 *n*97 Alpha Value | Sample 2019/2020 *n*222 Alpha Value | Item Examples |
|---|---|---|---|---|---|
| General Satisfaction | 9 | 0.80 | 0.81 | 0.85 | "Altogether, how satisfied are you with the simulation game?" |
| Subjective Learning Effect | 6 | 0.84 | 0.83 | 0.81 | "Through the simulation game… I all in all better understand how the EU works." |
| Motivational Effect | 4 | 0.90 | 0.82 | 0.91 | "The simulation game has motivated me to further occupy myself with the EU." |

Lastly, a guided expert interview was conducted with Konstantin Kaiser from planpolitik to document the improvements made in the simulation game and to find out whether the implemented improvements followed the suggestions from the scientific evaluation.

## 5. Results

To answer the research questions properly, the result section is structured as follows. First, the main dysfunctions of the simulation game, which were revealed in the first evaluation, will be displayed one after another. The improvements planpolitik made will be displayed, and the third evaluation will be checked for changes in the evaluation regarding the dysfunctional area. After all dysfunctions have been explained, discussed, and the effects of the improvements have been verified, the effectiveness and assessment of the 2019/2020 implementations will be compared to the first one to answer the second research question.

### 5.1. The Chat Function

One of the most criticized functions in the first application of the simulation game was the chat function. The chat function was on an extra page in the simulation and participants had to actively click on the page to see if they had new messages as opposed to modern social media pages, where you get a notification that you have new messages. Therefore 29.4% of the participants in 2015/2016 rated the technical features of the chat system as rather bad and 11.8% as very bad, similar to the rates of the communication via chat in general with 25.6% as rather bad and 15.9% as very bad. In the open questions, 26.5% of all comments complained about technical problems with the chat system, and only 8.2% of comments mentioned the chat system in a positive way (see Appendix A).

Since this was a major dysfunction of the simulation game made clear by the evaluation, changes in this area were recommended, which planpolitik also took seriously, as can be seen from this interview quote:

*Interviewer*: "It can be seen from the results of 2015/2016 both in the open questions and in the evaluation of the simulation game that the chat and the chat function were seen very critically by the participants. Did you (planpolitik) react to this result of the evaluation?"

*Planpolitik*: "Yes, for sure! That was a little bit like a slap in the face for us. In technical developments, especially in this phase, there are always problems. The implementations that we made there at the beginning with the schools that you also evaluated 2015, 2016 were simply super important for us to find all these problems and to see them as well, and of course we responded to that and then made improvements in many ways."

The mentioned improvements include a notification for new messages so that participants can react swiftly to messages, and the flow of the game could be enhanced. Furthermore, a like and dislike function for contributions in debates and in the group forum was added. The positive effect of the changes can already be seen in the 2017/2018 evaluation and are still present in the 2019/2020 evaluation (see Table 3). The average ratings for the technical features of the chat and the general communication by chat are initially increasing between the first and second cycles towards a more positive view and remain that way in the later evaluation (see Table 3). Therefore, participants are more satisfied with the improved chat in the later versions.

A similar development can be seen in the open questions: in 2019/2020, 13.3% of all comments actually mention the chat system as something they explicitly liked about the game. The negative comments about the chat systems decrease to 8.4% in the 2017/2018 evaluation and to 12.4% in the 2019/2020 evaluation, which is a decrease by more than half compared to 26.5% in 2015/2016.

**Table 3.** Mean Value Comparison of Chat Function Assessment Using Cohen's d (Cohen 1988).

| Category | Mean Values (SD) | | | Cohen's d Values | | |
|---|---|---|---|---|---|---|
| | 2015/2016 | 2017/2018 | 2019/2020 | Cohen's d 2015/2016 to 2017/2018 | Cohen's 2015/2016 to 2019/2020 | Cohen's d 2017/2018 to 2019/2020 |
| Technical features of the chat | 2.75 (0.89) | 3.02 (0.70) | 2.95 (0.76) | 0.31 | 0.23 | −0.10 |
| Communication by chat | 2.67 (0.99) | 3.06 (0.72) | 3.03 (0.81) | 0.45 | 0.39 | −0.04 |

Cohen's d: No effect >0.2, small effect 0.2–0.5, medium effect 0.5–0.8, large effect <0.8; mean values: 4 high approval, 1 disapproval.

*5.2. The Final Vote/The End of the Game*

The simulation game ends with a final vote on the legislation draft the students were required to produce during the simulation of the European Parliament. In 2015/2016, the draft was a text written by the participants themselves. It took some time to agree on single formulations, leading to many votes on the draft in the parliament to end the game, and the draft was not satisfying for many participants. This already was a problem in the 2015/2016 evaluation, as the result of the simulation was that the category rated the worst in the evaluation with 49% of the participants having assessed it as rather bad or very bad. In the small 2017/2018 evaluation, after fixing the chat problem, the discontent was even bigger, with nearly 58% of the participants ticking rather bad or very bad in the evaluation. Furthermore, 16.9% of the complaints in the open questions focused on the issue of the result of the game and the final voting process (see Appendix A). The two evaluations led to a change in planpolitik's voting system, as you can see in this part of the interview:

*Interviewer*: "In 2017/2018, the end of the simulation game specifically the final vote was especially criticized. Did you change or customize the voting regarding the final vote of the game after the critical results of the evaluation?"

*Planpolitik*: "No, we did not. That is also confusing for me. No idea. We did not change anything."

Interviewer: "So absolutely nothing? Since 2015 is was always the same?"

*Planpolitik*: "Although no, yes. Right. In 2015 it was still free text. And so everyone could formulate each word individually [ . . . ] we changed it so that we use prewritten sentence building blocks as drop-downs, that's what you call them, so you can select them by clicking on such a field."

The 2017/2018 implementation was still with the old version of the game, and the evaluation of this cycle made the dysfunction of the final voting system even more evident. The improvement towards a drag and drop version was made by planpolitik, and the participants can now build their legislation draft by choosing out of different text blocks going in one political direction or another, which can be changed in order to achieve a majority approval for the draft. The positive effect of this change is clearly visible in the newest evaluation. Compared to the average assessment of the end result of the simulation game in 2017/2018 (M = 2.33; SD = 0.98), the 2019/2020 evaluation had a medium positive increase of the assessment with a Cohen's d value of 0.67 (M = 2.93; SD = 0.81), thus moving the end result and the final vote from the worst-assessed category to a rather highly assessed category in the newest evaluation. Furthermore, there was only one complaint in the open questions about the final voting system (see Appendix A).

*5.3. Role Description/Role Identification*

A smaller but noteworthy change in the simulation game is the role description. In the first evaluation, the participants were quite satisfied with their role description, but the evaluation indicated a lack of role identification, and as for the open questions, only 8.2% of the comments were mentioning positive role identification. Planpolitik reacted and changed the role description after 2015/2016:

*Interviewer*: "Since 2015/2016 role identification is increasing; in 2015/2016 it was relatively low with 9% and nearly doubled after that. Did you do something with the role description over time?"

*Planpolitik*: "Yes, we have revised the role description. The page used to be just statistical information, like a book, just plain text. We have divided these (pages) into three modules, let's say. And then small questions in between. So we have included several interactive elements."

The role description changed from a predetermined role description to an interactive role design process that should increase the role identification as the participants have a greater influence on their own role. The later evaluations can support this claim: in 2017/2018, 17.1% of all comments on the game mention a positive role identification, similar to 2019/2020 with 19%, doubling the percentage of positive comments compared to 2015/2016 (see Appendix A).

### 5.4. The Learning Effect

The last, yet very important factor that the design-based research impacted is the effect of the simulation game on objective EU knowledge measured by the knowledge test. In the first evaluation of the simulation game, the results of the knowledge test showed a small decrease of EU knowledge; thus, the participation in the simulation game had a negative effect on the objective EU Knowledge of the participants (Cohen's d −0.21). This was, of course, not intended and the opposite effect of what the game should actually do. The evaluation pointed this out, and planpolitik reacted and changed the setting of the digital simulation game:

*Planpolitik*: "We changed nothing in the game itself but we added a presentation for the teachers, so in the introduction before the game starts, there is now a presentation for the teacher which the teacher will show the pupils. Now we are doing an online simulation game. This and that will happen and that helps to structure the expectations of the pupils and then they know better what is coming for them and what they will be learning about and that there is a committee and they are simulating a law-making process and just the classification where they are and what they are doing right now. Before that we just threw them into cold water. Thereby, I think, it is clearer to the people what they are doing and then they can better learn and save the information."

In educational science, there are clear indicators that combining and structure in educational interventions help the learner to better understand and receive knowledge (Jang et al. 2010; Sierens et al. 2009). The positive impact of the restructuring can also be seen in the 2019/2020 evaluation as the game now has a positive effect on objective knowledge (see Table 4).

**Table 4.** Mean Value Comparison of Objective Knowledge Using Cohen's d and *t*-test.

|  | Pretest M (SD) | Posttest M (SD) | Cohen's d | *t*-test |
|---|---|---|---|---|
| Objective Knowledge 2015/2016 | 11.11 (2.64) | 10.41 (3.84) | −0.21 | −2.59 ** |
| Objective Knowledge 2019/2020 | 12.18 (2.16) | 13.25 (3.34) | 0.34 | 8.06 ** |

Cohen's d: No effect >0.2, small effect 0.2–0.5, medium effect 0.5–0.8, large effect <0.8; *t*-test Significance Level: ** = $p < 0.01$.

So the effect on objective knowledge turned from a small significant decrease to a small significant increase, which can be expected for a 90-minute intervention.

### 5.5. General Comparison

For the second research question about the passage to a more effective and appreciated digital simulation game through the use of design-based research and scientific evaluation, in this section, a comparison between the first and the most recent evaluation will be done. Therefore, the participants' evaluations of the games and the effect on the participants will be highlighted and compared, the

first one being the general assessment of the three dimensions of the simulation game assessment (see Table 5).

**Table 5.** Mean Value Comparison of All Three Dimensions of Simulation Game Assessment in the First and Last Evaluation Cycles.

| Dimensions Simulation Game Assessment | Sample 2015/2016 *n*209 M (SD) | Sample 2019/2020 *n*222 M (SD) | Cohen's d 2015/2016 to 2019/2020 |
|---|---|---|---|
| General Satisfaction | 2.87 (0.52) | 3.02 (0.43) | 0.31 |
| Subjective Learning Effect | 2.67 (0.59) | 2.77 (0.49) | 0.18 |
| Motivational Effect | 2.22 (0.64) | 2.30 (0.68) | 0.12 |

The last evaluation was the most positive one in all dimensions, especially in the dimension of general satisfaction Not only does the general assessment of the simulation games show positive development, but the direct comparison of the adjectives the participants associate with the games also makes the advancements of the latest version clearer (see Figure 1).
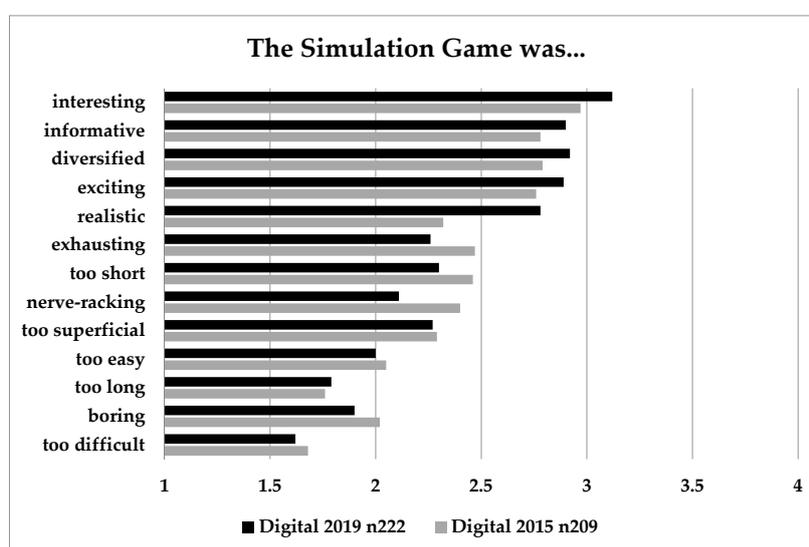


**Figure 1.** Comparison Simulation Game Adjectives Assessed by Participants in 2015/2016 and 2019/2020; 1 Indicates Strong Rejection and 4 Strong Agreement.

Overall, the participants rate the 2019/2020 simulation game, on the one hand, in all positive aspects higher, like the diversification of the game, find it is more interesting and exciting, and especially perceive it as more realistic, and on the other hand, in all negative aspects rate lower, particularly the nerve-racking and exhausting part, than the first version.

Looking at the effects of the simulation games, the 2015/2016 version had nearly no effect on all measured constructs (attitudes towards the EU, willingness of political participation, political interests, internal efficacy) except a negative effect on EU knowledge and a small positive effect on the willingness to engage in illegal political participation (Cohen's d 0.24), like political vandalism and civil disobedience. The newest version of the simulation game had a positive effect on objective EU knowledge and still had a small positive effect on the willingness to participate in politics illegally (Cohen's d 0.20). Additionally, there is a significant positive effect on internal efficacy (*t*-test 3.69 ***) that has only a very small effect size according to the Cohen's d value (0.17). Lastly, in the third evaluation cycle, pupils' general attitudes towards the European Union became slightly less positive (−0.21) after having played the game, but the general attitudes towards the EU were very positive in this sample from the beginning. Comparing the mean values of general EU attitudes in the posttests, the values of the third cycle are still higher than those of the first cycle.

In conclusion, the new version is now expanding the knowledge of its participants and slightly enhancing their internal political efficacy, which is what the educational intervention was meant to do. It still inspires them a bit towards illegal political participation and decreases the initially very high level of positive attitudes towards the EU. On the other EU-related political dispositions captured in the evaluation, the simulation game has no effect; given that it only takes 90 min, strong effects on attitudes were not expected.

## 6. Discussion

The results support certain aspects of the design-based approach, for instance, the importance of multiple iterations and evaluation cycles (Lewis et al. 2020; Anderson and Shattuck 2012; Herrington et al. 2007). The participants in the first iteration were very focused on the problems with the chat, drawing attention to this point of criticism. Therefore, problems regarding the "final vote/the end of the game", which were already present but not the main point of criticism in the first iteration, were neglected but became more visible during the second iteration. The results also underline the usefulness of mixed-method approaches in design-based research (Maxcy 2003). For example, while statistical analysis uncovered the problem with the chat system, it was the qualitative content analysis (Mayring 2010) of the open questions that showed what the problem actually was, thus making it easier for the developers to fix it.

So did scientific evaluation matter in the presented case? In the example of these short digital simulation games, a design-based research approach designed after Anderson's and Shattuck's (Anderson and Shattuck 2012) was implemented. The implementation followed the recommendations of design-based research literature and was leaning on the work of (Knogler and Lewalter 2014), including a close cooperation between researchers and developers (Kuhn and Quigley 1997), as well as multiple iterations over time. This led to a more effective and positively rated version of the educational intervention that is now expanding the knowledge of its participants and slightly enhancing their internal political efficacy, which is what the digital simulation game was meant to do (Kaiser et al. 2017; Oberle et al. 2017). Even though the effects of the short simulation games are still small, they now do help students to learn about the European Parliament; indeed, the size of the effects is considerable in view of the limited length of the intervention (90 min). Therefore, the scientific evaluation did matter in developing a better-functioning education measure, which is, according to the design-based research approach, a goal in itself (Hoadley 2004; Design-Based Research Collective 2003). The positive results are also in line with similar studies from the fields of blended-learning (Ustun and Tracey 2020) and face-to-face simulation games (Knogler and Lewalter 2014; Koivisto et al. 2018), showing the design-based-research approach as a suitable way to enhance digital simulation games in civic education, too.

Of course, limitations of the study need to be taken into account for an appropriate interpretation of its results. The main limitation of the study is its quasi-experimental setting; as opposed to purely experimental studies, it is not guaranteed that the shifts in the effects and assessments between the three evaluations are coming from the improvements planpolitik made or from other influences. In experimental settings, other potential factors of influence can be controlled, whereas in quasi-experimental settings this can only be done to a limited extent as there is no control of the Wi-Fi connection and technical equipment of the school, the behavior of the pupils, or other possible influences. Another important limitation is the sample composition and size: the 2017/2018 sample is very small and all samples having different compositions of school types leading to differences in age and other potentially influencing factors. As already underlined in the description of the sample, another point of limitation is that the sample selection was not random as the classes were signed up for participation by their teachers on a voluntary basis. Therefore, the results should be interpreted with care. Overall, the sample size is still too small and arbitrary for general assumptions but the results can point in a direction for further large n studies in this field.

## 7. Conclusions

Four major flaws of the first version of the digital simulation game—respectively, the dysfunctional chat, the end result/final vote of the game, the role description, and the negative learning effect—were pointed out in an attempt to answer the first research question. Following the design-based research approach (Anderson and Shattuck 2012), dysfunctions were processed and could be improved, leading to a more positive assessment by the participants in the later evaluations. Thus, the final evaluation of the simulation game was more positive than the evaluation before, and the game's effects regarding participants' objective EU knowledge and internal political efficacy were enhanced as compared to the earlier evaluations (Oberle et al. 2017).

Similar to the research of (Knogler and Lewalter 2014), the simulation game was enhanced through design-based research and is now more useful for practical applications. The study has limitations and its results cannot be generalized, but they do support the request that educational interventions like simulation games, digital and analog ones, should be accompanied by empirical evaluations to enhance their potential and to point out possible design/application failures. Close cooperation between evaluators and developers can lead to a more effective intervention (Štemberger and Cencic 2016); new ideas and prototypes should be tested and investigated closely, and this effort should not stop after one cycle of evaluation.

Scientific evaluation matters and can improve programs, tools, and methods in the important field of civic education. Further long term studies, including large n studies as well as qualitative studies, should accompany new civic education programs inspired by the design-based research approach to critically analyze the effectiveness of such measures and to make sure learners can benefit from the best civic education possible.

## Appendix A

**Table A1.** Extract of the Results of the Qualitative Content Analysis Focusing on the Problems Discussed in the Article.

| What Did You Like about the Simulation Game? | | | | | |
|---|---|---|---|---|---|
| | 2015/16 | | 2017/18 | | 2019/20 | |
| Category | Mention | Percent | Mention | Percent | Mention | Percent |
| Role identification | 12 | 8.2% | 14 | 17.1% | 43 | 19% |
| Chat functions | 12 | 8.2% | – | – | 30 | 13.3% |
| (End-)Voting in the game | 3 | 2.1% | – | – | 5 | 2.2% |
| Overall | 146 | 100% | 82 | 100% | 226 | 100% |
| **What Did You <u>Not</u> Like about the Simulation Game?** | | | | | |
| | 2015/16 | | 2017/18 | | 2019/20 | |
| Category | Mention | Percent | Mention | Percent | Mention | Percent |
| End of the game/final vote | 23 | 14.2% | 14 | 16.9% | 1 | 0.5% |
| Technical problems with chat system | 43 | 26.5% | 7 | 8.4% | 25 | 12.4% |
| Overall | 162 | 100% | 83 | 100% | 202 | 100% |

## References

Anderson, Terry, and Julie Shattuck. 2012. Design-based research: A decade of progress in education research? *Educational Researcher* 41: 16–25. [CrossRef]

Bachen, Christine M., Pedro F. Hernández-Ramos, Chad Raphael, and Amanda Waldron. 2015. Civic play and civic gaps: Can life simulation games advance educational equity? *Journal of Information Technology & Politics* 12: 378–95. [CrossRef]

Barab, Sasha, and Kurt Squire. 2004. Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences* 13: 1–14. [CrossRef]

Chernikova, Olga, Nicole Heitzmann, Matthias Stadler, Doris Holzberger, Tina Seidel, and Frank Fischer. 2020. Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research* 90: 499–541. [CrossRef]

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale: Erlbaum. [CrossRef]

Design-Based Research Collective. 2003. Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher* 32: 5–8. [CrossRef]

Deutsche, Shell. 2010. *16. Shell Jugendstudie. Jugend 2010. Eine Pragmatische Generation Behauptet Sich*. Frankfurt: Fischer.

DiCamillo, Lorrei, and Jill M. Gradwell. 2012. Using simulations to teach middle grades US history in an age of accountability. *RMLE Online* 35: 1–16. [CrossRef]

Duchatelet, Dorothy. 2019. "Yes I Can! Can I?" At the Heart of Self Efficacy for Negotiating in Role Play Simulations of Political Decision-Making. Ph.D. thesis, University of Antwerp, Antwerpen, Belgium.

Fink, Simon. 2015. 36 different chocolate directives: How does the setting influence negotiation outcomes in an EU simulation? *European Political Science* 14: 241–53. [CrossRef]

Ford, Cristi, Darragh McNally, and Kate Ford. 2017. Using Design-Based Research in Higher Education Innovation. *Online Learning* 21: 50–67. [CrossRef]

Gabriel, Sonja. 2012. Spielend lernen?! Einsatz von digitalen Spielen im Unterricht. In *Zukunft des Lernens. Wie digitale Medien, Aus- und Weiterbildung verändern*. Edited by Edith Blaschitz, Gerhard Brandhofer, Christian Nosko and Gerhard Schwed. Glückstadt: Verlag Werner Hülsbüsch, pp. 275–90.

Gille, Martina, Sabine Sardei-Biermann, Wolfgang Gaiser, and Johann de Rijke. 2006. Jugendliche und junge Erwachsene in Deutschland. Lebensverhältnisse, Werte und gesellschaftliche Beteiligung 12- bis 29-Jähriger. Wiesbaden: Springer VS.

Herrington, Jan, Susan McKenney, Thomas Reeves, and Ron Oliver. 2007. Design-based research and doctoral students: Guidelines for preparing a dissertation proposal. In *Proceedings of ED-MEDIA 200—World Conference on Educational Multimedia, Hypermedia & Telecommunications*. Edited by Craig Montgomerie and Jane Seale. Vancouver: Association for the Advancement of Computing in Education (AACE), pp. 4089–97.

Hoadley, Christopher M. 2004. Methodological alignment in design-based research. *Educational Psychologist* 39: 203–12. [CrossRef]

Jang, Hyungshim, Johnmarshall Reeve, and Edward L. Deci. 2010. Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology* 102: 588. [CrossRef]

Jen, Enyi, Sidney Moon, and Ala Samarapungavan. 2015. Using design-based research in gifted education. *Gifted Child Quarterly* 59: 190–200. [CrossRef]

Kaiser, Konstantin, Simon Raiser, and Björn Warkalla. 2017. Einfach digital? Onlineplanspiele in der politischen Bildung. In *Handbuch Planspiele in der politischen Bildung*. Edited by Andreas Petrik and Stefan Rappenglück. Frankfurt am Main: Wochenschau, pp. 267–75.

Kay, Judy, and Rosemary Luckin. 2018. Rethinking Learning in the Digital Age: Making the Learning Sciences Count. In Proceedings of the 13th International Conference of the Learning Sciences (ICLS), London, UK, June 23–27.

Kerr, David, Linda Sturman, Wolfram Schulz, and Bethan Burge. 2010. *ICCS 2009 European Report: Civic Knowledge, Attitudes, and Engagement among Lower-Secondary Students in 24 European Countries*. Amsterdam: IEA.

Knogler, Maximilian. 2014. Investigating Student Interest in the Context of Problem-based Learning A Design-based Research Study. Ph.D. thesis, Technical University of Munich, Munich, Germany.

Knogler, Maximilian, and Doris Lewalter. 2014. What Makes Simulation Games Motivating? Design–Based Research on Learners' Motivation in Simulation Gaming. In *The Shift from Teaching to Learning: Individual, Collective and Organizational Learning through Gaming Simulation*. Edited by Willy C. Kriz. Bielefeld: Bertelsmann Verlag, pp. 150–62.

Koivisto, Jaana M., Elina Haavisto, Hannele Niemi, Päivi Haho, Saara-Maria Nylund, and Jari Multisilta. 2018. Design principles for simulation games for learning clinical reasoning: A design-based research approach. *Nurse Education Today* 60: 114–20. [CrossRef] [PubMed]

Kuhn, Gary W., and B. Allen Quigley. 1997. Understanding and using action research in practice settings. In *Creating Practical Knowledge through Action Research*. Edited by B. Allen Quigley and Gary W. Kuhne. San Fransisco: Jossey-Bass, pp. 23–40.

Le, Son, Peter Weber, and Martin Ebner. 2013. Game-Based Learning. Spielend Lernen? In *Lehrbuch für Lernen und Lehren mit Technologien*, 2nd ed. Edited by Martin Ebner and Sandra Schön. Berlin: Epubli, pp. 267–76.

Lewis, Daniel Rees, Spencer E. Carlson, Christopher K. Riesbeck, Kristine J. Lu, Elizabeth M. Gerber, and Matthew W. Easterday. 2020. The Logic of Effective Iteration in Design-Based Research. In *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020*. Edited by Melissa Gresalfi and Ilana S. Horn. Nashville: International Society of Learning Sciences, vol. 2, pp. 1149–56. [CrossRef]

Lohmann, Jan Robert. 2019. Simulations Matter—Wirkungsweisen und Mehrwert von Politiksimulationen. Ph.D. thesis, University of Passau, Passau, Germany.

López, José M. C., and Myriam J. M. Cáceres. 2010. Virtual games in social science education. *Computers & Education* 55: 1336–45. [CrossRef]

Massing, Peter. 1997. *Handlungsorientierter Politikunterricht—Ausgewählte Methoden*. Schwalbach/Ts.: Wochenschau.

Massing, Peter. 2010. Planspiele und Entscheidungsspiele. In *Methodentraining für den Politikunterricht I. Mikromethoden—Makromethoden*, 4th ed. Edited by Siegfried Frech, Hans-Werner Kuhn and Peter Massing. Schwalbach: Wochenschau Verlag, pp. 163–94.

Maxcy, Spencer J. 2003. Pragmatic threads in mixed methods research in the social sciences: The search for multiple modes of inquiry and the end of the philosophy of formalism. In *Handbook of Mixed Methods in Social and Behavioral Research*. Edited by Abbas Tashakkori and Charles Teddlie. Thousand Oaks: Sage, pp. 51–89.

Mayring, Philipp. 2010. Qualitative Inhaltsanalyse. In *Handbuch Qualitative Forschung in der Psychologie*. Edited by Günter Mey and Katja Mruck. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 601–13. [CrossRef]

Motyka, Marc. 2018. *Digitales, Spielbasiertes Lernen im Politikunterricht*. Wiesbaden: Springer VS. [CrossRef]

Oberle, Monika. 2012. *Politisches Wissen über die Europäische Union. Subjektive und Objektive Politikkenntnisse von Jugendlichen*. Wiesbaden: Springer VS. [CrossRef]

Oberle, Monika, and Johanna Forstmann. 2015. Förderung EU-bezogener Kompetenzen bei Schüler/innen—Zum Einfluss des politischen Fachunterrichts. In *Die Europäische Union Erfolgreich Vermitteln. Perspektiven der Politischen EU-Bildung Heute*. Edited by Monika Oberle. Wiesbaden: Springer VS, pp. 81–98. [CrossRef]

Oberle, Monika, Simon Raiser, Björn Warkalla, Konstantin Kaiser, and Johanna Leunig. 2017. Online-Planspiele in der politischen Bildung—Ergebnisse einer Pilotstudie. In *Medienkompetenz—Herausforderung für Politik, Politische Bildung und Medienbildung*. Edited by Harald Gapski, Monika Oberle and Walter Staufer. Bonn: Bundeszentrale für politische Bildung, pp. 243–55.

Oberle, Monika, Sven Ivens, and Johanna Leunig. 2018. Effects of EU Simulation Games on Secondary School Students' Political Motivations, Attitudes and Knowledge: Results of an Intervention Study. In *Simulations of Decision-Making as Active Learning Tools*. Edited by Peter Bursens, Vincent Donche, David Gijbels and Pieter Spooren. Cham: Springer, pp. 145–64. [CrossRef]

Oberle, Monika, Johanna Leunig, and Sven Ivens. 2020. What do students learn from political simulation games? A mixed-method approach exploring the relation between conceptual and attitudinal changes. *European Political Science*, 1–20. [CrossRef]

Onwuegbuzie, Anthony, and Nancy L. Leech. 2007. Validity and Qualitative Research: An Oxymoron? *Quality & Quantity* 41: 233–49. [CrossRef]

Patton, Michael Quinn. 2011. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press.

Peters, Maria, and Bàrbara Roviró. 2017. Fachdidaktischer Forschungsverbund FaBiT: Erforschung von Wandel im Fachunterricht mit dem Bremer Modell des Design-Based Research. In *Making Change Happen*. Edited by Sabine Doff and Regine Komoss. Wiesbaden: Springer VS, pp. 19–32. [CrossRef]

Sandoval, William A., and Philip Bell. 2004. Design-based research methods for studying learning in context: Introduction. *Educational Psychologist* 39: 199–201. [CrossRef]

Schnurr, Matthew A., Elizabeth De Santo, and Rachael Craig. 2013. Using a Blended-Learning approach to simulate the negotiation of a multilateral environmental agreement. *International Studies Perspectives* 14: 109–20. [CrossRef]

Sieben, Swen, and Clemens M. Lechner. 2019. Measuring cultural capital through the number of books in the household. *Measurement Instruments for the Social Sciences* 1: 1–6. [CrossRef]

Sierens, Eline, Maarten Vansteenkiste, Luc Goossens, Bart Soenens, and Filip Dochy. 2009. The synergistic relationship of perceived autonomy support and structure in the prediction of self-regulated learning. *British Journal of Educational Psychology* 79: 57–68. [CrossRef]

Štemberger, Tina, and Majda Cencic. 2016. Design based research: The way of developing and implementing. *World Journal on Educational Technology: Current Issues* 8: 180–89. [CrossRef]

Ustun, Ahmet Berk, and Monica Walch Tracey. 2020. An effective way of designing blended learning: A three phase design-based research approach. *Education and Information Technologies* 25: 1529–52. [CrossRef]

Van den Akker, Jan, Koeno Gravemeijer, Susan McKenney, and Nienke Nieveen. 2006. *Educational Design Research*. Abingdon-on-Thames: Routledge.

Vetter, Angelika. 2013. Zufriedenheit der Bürger mit politischen Institutionen und der Demokratie. In *Politik im Kontext: Ist alle Politik lokale Politik?* Edited by Jan W. van Deth and Markus Tausendpfund. Wiesbaden: Springer VS, pp. 359–80. [CrossRef]

Westle, Bettina. 2006. Politisches Interesse, subjektive politische Kompetenz und politisches Wissen–eine Fallstudie mit Jugendlichen im Nürnberger Raum. In *Jugend und Politik—"Voll normal!* Edited by Edeltraut Roller, Frank Brettschneider and Jan W. van Deth. Wiesbaden: VS Verlag, pp. 209–40. [CrossRef]