Marie's ChemLab: a Mobile Augmented Reality Game to Teach Basic Chemistry to Children

Michaela Arztmann[1], Jessica Lizeth Domínguez Alfaro[2], Jonas Blattgerste[3], Johan Jeuring[4], Peter van Puyvelde[2]

[1] Department of Education, Faculty of Social and Behavioral Sciences, Utrecht University, Utrecht, The Netherlands

[2] Department of Chemical Engineering, Faculty of Engineering, KU Leuven, Leuven, Belgium

[3] Mixed Reality Lab, Electrical Engineering and Information Technology, University of Applied Sciences Emden/Leer, Emden, Germany

[4] Department of Information and Computing Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands

m.arztmann@uu.nl
jessicalizeth.dominguezalfaro@kuleuven.be
jonas.blattgerste@raumtaenzer.com
J.T.Jeuring@uu.nl
peter.vanpuyvelde@kuleuven.be

Abstract: Digital games are widely used in education to motivate students for science. In parallel, the role of augmented reality (AR) in education is increasing. This paper introduces Marie's ChemLab: a marker-less handheld AR application in which K-12 students learn about science. In Marie's ChemLab, users perform interactions in their physical environment using virtual AR objects displayed on a handheld device. The utilized interaction concept is based on the TrainAR framework (Blattgerste et al, 2021), which originally was envisioned for procedural AR training and layered feedback mechanisms. The original framework has been enhanced to support gamified interactions with virtual objects, specifically for K-12 students. In total, 239 early secondary school students played the final application as part of a larger study. The usability score indicated marginally acceptable usability, and the application was successfully implemented in a classroom setting.

Keywords: mobile augmented reality, simulation games, science education, SUS

1. Introduction

Students' interest to engage in academic content in compulsory education declines with age (Shin et al, 2019). Particularly, STEM (Science, Technology, Engineering, Mathematics) subjects are characterized as the least popular among a broad range of students (Renninger et al, 2015). Digital games have been proven effective as an approach to increase students' motivation (Lamb et al, 2018; Arztmann et al, 2022) through their typical characteristics such as providing a feeling of relatedness, autonomy, and competence (Ryan & Rigby, 2020). Augmented Reality (AR) provides an opportunity to create games that offer embodied interactions and immersive visualizations that elicit learning benefits (Hung et al, 2017; Radu 2014). With the increasing availability of smartphones and tablets that are capable of deploying AR scenarios, it also creates the opportunity to implement AR applications in classrooms easily. However, the implementation of newer technologies into classroom settings is still a challenge due to, e.g., themes not matching the classroom's curricula (Poultsakis et al, 2021), accessibility, and cost factors (Blattgerste et al, 2021).

AR for handheld devices allows for immersive visualizations and simulation of concepts using engaging interaction metaphors, like on-screen touches, smartphone translation, and rotation or spatial proximity triggers, that require more embodied interactions than conventional approaches. However, research on using AR in classroom settings thus far has been scarce and has mostly focused on visualization aspects, with very little or no interaction. It is now clear that AR enables the development of simulation games that help students better understand complex phenomena through not only contextualized 3D visualizations but also the freedom to choose their own viewing angle and focus. But what about the possibility of interaction with and exploration of physical environments enriched with 3D content? This would not only be particularly useful for situations that are hard to study in a traditional classroom setting but might also have an effect on the

immersion and subsequent motivational factors of the students. As thus far, most AR applications have been developed for visualization purposes only (Dominguez Alfaro & Van Puyvelde, 2021), which offer little to no interactivity, and available approaches have hardly been validated for learning (Arztmann et al, 2022); we introduce Marie's ChemLab: a mobile AR game to teach basic chemistry to children.

This paper describes the design of the game, including the adjustments that have been made to make it suitable for a young target group, and studies the game's usability and feasibility of implementing it in a classroom environment.

2. Marie's ChemLab

Marie's ChemLab is a marker-less handheld AR, single-player game designed for tablets. The game consists of three sequential levels, situated in different scenarios (kitchen, forest, and lab). Each level introduces a student to a different chemical concept, and they are asked to complete various tasks to advance in the levels. To help guide a student, an intelligent agent (Marie, "the expert") provides messages on essential things to look at, try or remember (Figure 1a). At the end of each level, the game shows post-it notes with essential concepts to internalize (Figure 1b).
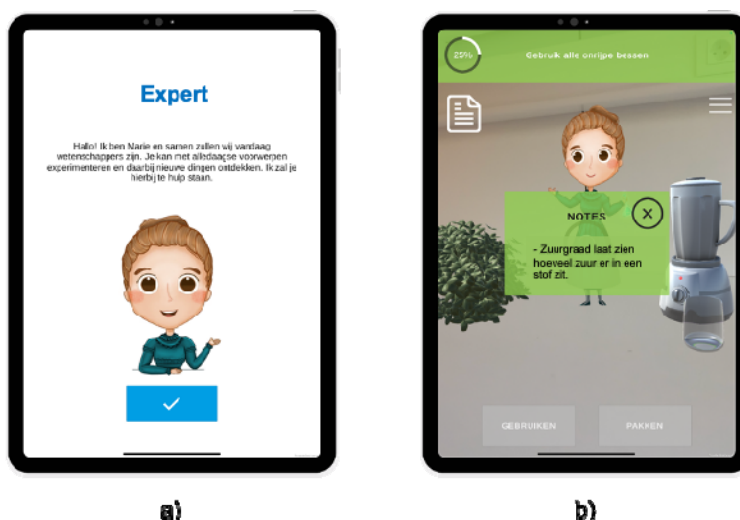
Figure 1: a) Introductory scene of Marie "the expert"; b) example end post-it note from level 2

The game starts on level one ("Fruit Salad") (Figure 2a). The scenario is located in a kitchen with different fruits to be tested. The main goal is to compare the acidity of three fruits and associate their "sourness" with the level of acidity. A student starts by ranking the fruits in three bowls: "sour," "less sour," and "sweet." Afterward, a toolbox appears with different tools such as a hammer, a bubble pump, and a box with pH paper. All of these tools can be explored. If the students are successful, they discover that with the help of the hammer, they can cut the ranked fruits and measure the pH with the pH paper. Finally, the player can observe the discoloration of the paper and rank again according to the sourness of each fruit in the three bowls.

In level two ("Smoothie party"), the student is in a forest (Figure 2b). The scenario displays a bush full of blackberries of different ripeness. Marie is also there. The level starts when the user feeds Marie with each of the berries. Depending on the berry's ripeness, she reacts with a sour, happy (sweet), or neutral face. Marie's reaction aims to clarify that each berry has different acid and sugar content. Afterward, the student prepares some juice of unripe berries and compares the difference when lime is added to this juice and juice made of sweet berries. Finally, the student learns that blackberry juice changes color when lemon/lime is added; thus, it can be used as an indicator.

The game ends with level 3 ("Kitchen Detectives") (Figure 2c). The scenario is located in a laboratory, where the pH of three different kitchen products needs to be measured and ranked accordingly. The student can measure the pH using the tools they learned about in levels one (pH paper) and two (blackberry juice).

| a) Fruit Salad | b) Smoothie Party | c) Kitchen Detectives |

Figure 2: Marie's ChemLab Levels

2.1 Development and Design

Marie's ChemLab was designed and developed using the User-Centered Design (UCD) approach and the Design Implementation Framework (DIF) (Stone et al, 2018). Thus, usability indicators and feedback from the students are integrated iteratively to improve the application.

Additionally, Marie's ChemLab was designed with the aim of triggering students' interest in chemistry by providing a playful environment with relatable content. By adding game mechanics intended to satisfy students' need for autonomy, relatedness, and competence (e.g., cue messages at the top of the screen; Ryan & Rigby, 2020) we intended to create a game environment that sparks students' interest for chemistry, which would be based on the first phase of the four-phase-interest development model by Hidi & Renninger (2006). As such, interest may develop from brief situational interest to maintained-situational and emerging-individual interest, given that students are provided with additional opportunities to engage with this topic by their teachers or their individual environment.

2.2 Learning goals

Each level introduces a different learning goal with the provided content. In level one, students should develop a basic understanding of sourness and that sourness can be measured by the use of pH-strips. Level two builds upon these concepts and introduces a natural indicator as an alternative to the pH strips. Finally, level three is aimed at transferring the learned content of measuring other materials with the various tools introduced to different tools that can be measured, showing that also fluids that are found in the kitchen can be acidic.

The iterative process began with several content ideation sessions with secondary and primary school teachers. The idea was to create levels that are based on young students' experiences and, therefore, can easily be connected to basic concepts of daily life chemistry. As such, the topic of acidity was chosen. It is important to create a basic understanding of the term acidity, which forms the basis of understanding pH values and, in further education, the basis of titration experiments. A three-level structure was chosen for the game to divide the concept of acidity into sub-concepts. This structure and each of the levels were based on a real classroom intervention for the topic, thus the game follows the format of an existing lesson plan which was provided by a teacher. This ensures comparable results of the intervention and allows teachers to integrate the application into their curriculum more easily.

2.3 Iteration and prototyping

A first digital prototype was developed using the TrainAR framework (Blattgerste et al, 2021). TrainAR is an interaction framework with an authoring tool, in which interactive procedural task training applications can be

developed. The user of the prototype can interact with objects that elicit feedback to the user's actions. The user can grab, release, interact and combine virtual objects to advance in each level.

The original TrainAR framework was developed with procedural tasks in mind. It can also be used to create games that follow a rule-based approach without enforcing a particular order on the rules. Marie's ChemLab is a rule-based game where the user must accomplish a set of defined goals at each level. Therefore, the user can explore the environment and freely decide on the sequence of actions. They are guided on each level with messages given by the expert (Figure 3a), pointers at the top of the screen (Figure 3b), and layered feedback (sound and outlines) embedded in the system (Figure 3c).
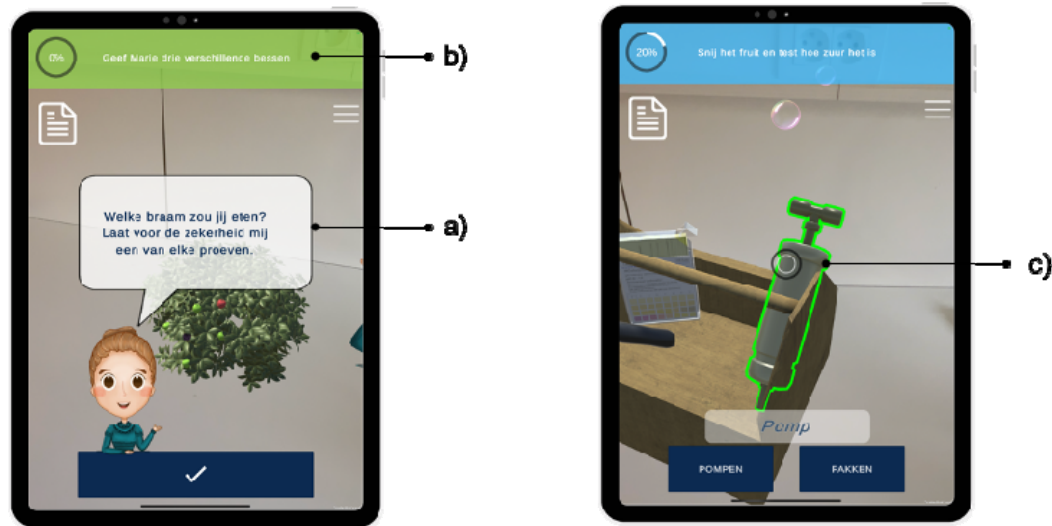
Figure 3: Design elements: a) message given by Marie in level 2 (the intelligent agent); b) on-screen pointers; c) example of object outline

A first prototype was evaluated in two non-representative preliminary small-scale pilots with students in 5th grade. Based on observations and their feedback, the prototype was adjusted to the final version used for the evaluation of this paper. Particularly, changes concerning challenges with the position and size of the interaction button (Figure 4) and the structure of the tutorial were addressed.
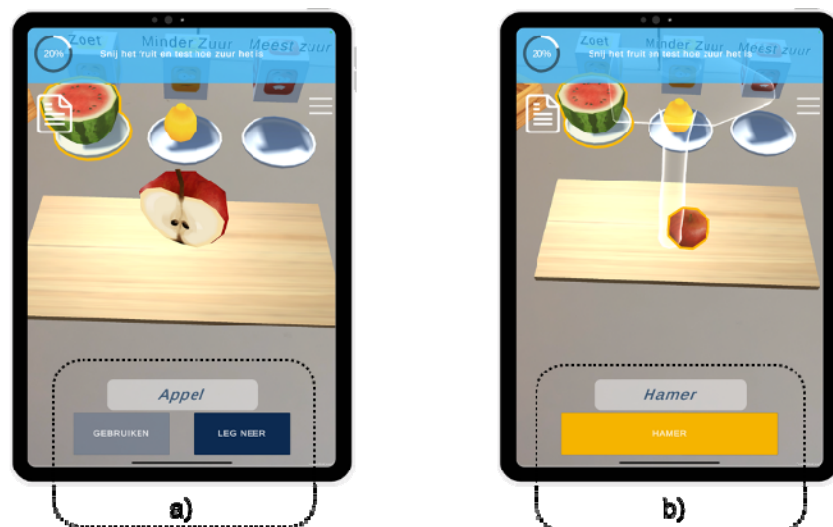
Figure 4: a) two interaction buttons showing the option to "use" or "release" the object (apple); b) when an interaction between two objects is possible, the combine button appears as one single yellow button.

2.4 Playtesting

As part of a larger study, 239 year 1 and 2 students (M_{age} = 12.9, SD = 0.8) from Dutch secondary schools played the game in one of their science lessons. The game was played on iOS tablets (iPad 9) that were provided by the researchers. The students had 30 minutes to play the game; after that the game automatically moved them to the post-test. In case students finished early, they were provided with riddles that they could solve until their classmates completed the game. To measure usability, the adapted version of the System Usability Scales (SUS) for children introduced by Putnam et al (2020) was used. To create a Dutch version, the back-translation method was applied. The Dutch items can be found in the Appendix. Ethical approval for this study has been granted by the local Institutional Review Board.

3. Evaluation results

3.1 Classroom implementation

Even though the students had 30 minutes to play the game before they were moved to the post-test, the majority needed less time and were able to finish with the last level within 20 minutes. None of the students communicated complaints about the device being too heavy, and no apparent problems with properly holding tablets while playing the applications could be observed. All students were able to play the game simultaneously. The researchers were mainly available to answer questions regarding possible technical issues. Additionally, they would intervene when the scene was placed too far away (i.e., on the ground instead of the table) or when a glitch of the application was experienced.

3.2 Usability

When analyzing the perceived usability through SUS scores using the SUS Analysis Toolkit by Blattgerste et al (2022), the overall SUS score of Maries' ChemLab was 54.7 (SD = 15.19). This result would indicate "OK" usability and be contextualized in the first quartile of SUS scores gathered in meta-analysis datasets, according to Bangor et al (2009). According to Bangor et al (2008), this usability score is still marginally acceptable. When grouping the SUS scores by age, these results stay persistent. As visualized in Figure 5, the n = 77 students aged 11 to 12 reported an average SUS score of 54.16 (SD = 15.19), the n = 110 students aged 13 reported an average SUS score of 54.34 (SD = 17.09), and the remaining 52 students that were aged 14-15 reported an average score of 56.25 (SD = 13.53). All scores would still indicate "OK" usability, be contextualized in the first quartile of SUS scores, and be marginally acceptable. While there is a slight upward trend observable, where usability scores increase with the age group, a one-way ANOVA revealed that there was no statistically significant difference in reported perceived usability through the SUS scores based on the participants' age groups ($F(161.638, 58707.472) = [0.324]$, p = 0.723).
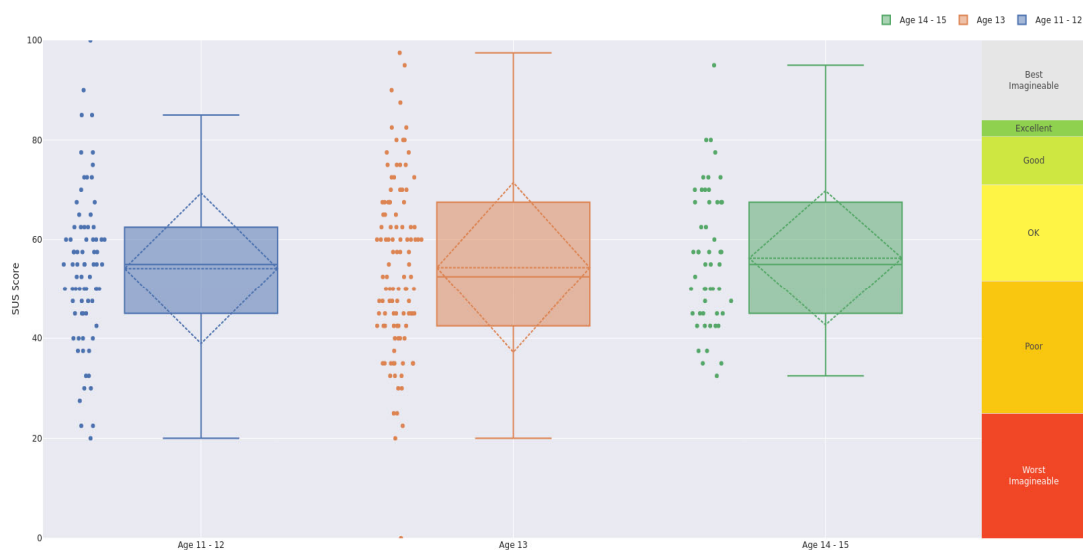
Figure 5: The SUS scores of the evaluation of Marie's ChemLab, grouped by age, contextualized on the adjective scale by Bangor et al (2009). Dotted lines show the mean and SD, box plots show the median, quartiles and all data points besides the plot.

For the original SUS questionnaire by Brooke (1996), 100% conclusive results can be reached with around 14 SUS scores per SUS study score. While we used a translation of a variant of the questionnaire by Putnam et al (2020), the sample size of n = 236 should yield conclusive results, according to Tullis et al (2004). Nonetheless,the adjusted SUS scale showed a rather low internal reliability, with a Cronbach's alpha of .446. Since the proposed variation of Putnam et al (2020) failed to establish full construct validity, the results need to be interpreted with caution.

4. Discussion

AR provides the opportunity to create game environments that offer embodied interactions and immersive visualizations that can increase the learning and motivation of students. As AR applications thus far have mainly been developed for visualization purposes only and have not added interactivity, Marie's ChemLab was designed to be easily used in a school environment by combining both visualization and interactivity and using accessible hardware for implementation. Additionally, it aimed to provide a first impression of daily life chemistry that should trigger interest in students.

This paper shows that AR technology can be successfully implemented in a classroom. The perceived usability of the game with a SUS study score of 54.7 (SD = 15.19), however, can be improved. The observed results of the adapted SUS need to be interpreted with caution, since the scale failed to reach a sufficient reliability score. Different from the adjusted version of Putnam (2020), the reliability was calculated for the overall scale and not in different dimensions. Additionally, our target group was older than the original age group used to validate the children's version. Nevertheless, our results could be in agreement with Radu's (2016) findings, which showed that children of different ages encounter different usability issues with handheld AR such as inaccuracy in interacting with objects in the virtual environment or difficulties in remembering the spatial locations of virtual objects. A possible explanation for this might be that in Marie's ChemLab, the students are required to aim the crosshair at a specific object. This interaction may be problematic for some students since it requires an understanding of the spatial relationship between the two-dimensional projection of the device and the real environment. (Radu et al, 2016).

5. Conclusions

This paper makes three contributions. First, Marie's ChemLabs design rationale and the didactic considerations during its development were described. Secondly, details of the application with its three levels and the changes that were made during development were reported. Finally, Marie's ChemLab was evaluated with 239 students from Dutch secondary schools spanning year 1 and 2, with a focus on the perceived usability and potential of interactive handheld AR games in this context. The results showed that although the usability of Marie's ChemLab can be improved, the game can successfully be used in a real classroom setting. This can show interested practitioners and researchers that students are able to play through the levels of Marie's ChemLab independently and can interact with the application at their own pace. Moreover, in case of implementation, the teacher would only need to be available to answer possible questions.

Further improvements of Marie's ChemLab can be focused on adding an adaptive system that provides instructions based on the student's individual abilities. In spite of its limitations, the study certainly adds to our understanding of the use of interactive games in AR.

6. Funding

7. Acknowledgments

8. References

Arztmann, M., Hornstra, L., Jeuring, J. and Kester, L. (2022) "Effects of games in STEM education: a meta-analysis on the moderating role of student background characteristics", *Studies in Science Education*.

Bangor, A., Kortum, P. T. and Miller, J. T. (2008) "An empirical evaluation of the system usability scale", *International Journal of Human–Computer Interaction*, Vol 24, No 6, pp 574-594. https://doi.org/10.1080/03057267.2022.2057732

Bangor, A., Kortum, P. and Miller, J. (2009) "Determining what individual SUS scores mean: Adding an adjective rating scale", *Journal of Usability Studies*, Vol 4, No 3, pp 114-123.

Blattgerste, J., Luksch, K., Lewa, C. and Pfeiffer, T. (2021) "TrainAR: A Scalable Interaction Concept and Didactic Framework for Procedural Trainings Using Handheld Augmented Reality", *Multimodal Technologies and Interaction*, Vol 5, No 7, p 30. https://doi.org/10.3390/mti5070030

Blattgerste, J., Behrends, J. and Pfeiffer, T. (2022) "A Web-Based Analysis Toolkit for the System Usability Scale", in *The 15th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '22)*. Corfu, Greece.

Brooke, J. (1996) "Sus: a 'quick and dirty' usability scale", *Usability Evaluation in Industry*, Vol 189, No 3.

Dominguez Alfaro, J.L. and Puyvelde, P. Van (2021) "Mobile Augmented Reality Apps in Education: Exploring the User Experience Through Large-Scale Public Reviews", in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pp. 428–450. https://doi.org/10.1007/978-3-030-87595-4_32

Domínguez Alfaro, J.L., Gantois, S., Blattgerste, J., De Croon, R., Verbert, K., Pfeiffer, T. and Van Puyvelde, P. (2022) "Mobile Augmented Reality Laboratory for Learning Acid–Base Titration", *Journal of Chemical Education*, Vol 99, No 2, pp. 531–537. https://doi.org/10.1021/acs.jchemed.1c00894

Hidi, S. and Renninger, K. A. (2006), "The four-phase model of interest development", *Educational Psychologist*, Vol 41, pp. 111–127. http://dx.doi.org/10.1207/s15326985ep4102_4

Hung, Y. H., Chen, C. H. and Huang, S. W. (2017) "Applying augmented reality to enhance learning: a study of different teaching materials", *Journal of Computer Assisted Learning*, Vol 33, No 3, pp. 252-266. https://doi.org/10.1111/jcal.12173

Lamb, R.L., Annetta, L., Firestone, J. and Etopio, E. (2018) "A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations", *Computers in Human Behavior*, Vol. 80, pp. 158-167. https://doi.org/10.1016/j.chb.2017.10.040

Poultsakis, S., Papadakis, S., Kalogiannakis, M., and Psycharis, S. (2021) "The management of Digital Learning Objects of Natural Sciences and Digital Experiment Simulation Tools by teachers", *Advances in Mobile Learning Educational Research*, Vol 1, No 2, pp. 58-71.

Putnam, C., Puthenmadom, M., Cuerdo, M. A., Wang, W. and Paul, N. (2020) "Adaptation of the System Usability Scale for User Testing with Children", Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.25082/amler.2021.02.002

Radu, I. (2014) "Augmented reality in education: a meta-review and cross-media analysis", *Personal and ubiquitous computing*, Vol. 18, No. (6), pp. 1533-1543. https://doi.org/10.1007/s00779-013-0747-y

Radu, I., Macintyre, B. and Lourenco, S. (2016) "Comparing children's crosshair and finger interactions in handheld augmented reality: Relationships between usability and child development", *Proceedings of IDC 2016 - The 15th International Conference on Interaction Design and Children*, pp. 288–298. https://doi.org/10.1145/2930674.2930726

Renninger, K.A., Costello Kensey, C.N., Stevens, S.J. and Lehman, D.L. (2015) "Perceptions of Science and Their Role in the Development of Interest", in *Interest in Mathematics and Science Learning*, K. Renninger, M. Nieswandt and S. Hidi Eds. American Educational Research Association, Washington, pp. 93-110. www.jstor.org./stable/j.ctt1s474j0

Ryan, R.M. and Rigby, C.S. (2020) "Motivational Foundations of Game-Based Learning", in *Handbook for Game-Based Learning,* J.L. Plass, R.E. Mayer and B.D. Homer Eds. The MIT Press, Cambridge, pp.153-176.

Shin, D.D., Lee, M., Ha, J.E., Park, J.H., Ahn, H.S., et al (2019) "Science for all: Boosting the science motivation of elementary school students with utility value intervention", *Learning and Instruction*, Vol. 60, pp. 104-116. https://doi.org/10.1016/j.learninstruc.2018.12.003

Stone, M. L. et al (2018) "The Design Implementation Framework", in R.D. Roscoe, S.D. Craig and I. Douglas Eds. *End-User Considerations in Educational Technology Design*. IGI Global, pp. 76–98.

Tullis, T. S. and Stetson, J. N. (2004) „A comparison of questionnaires for assessing website usability", *Usability professional association conference,* Vol 1, pp 1-12.

9. Appendix

Table 1: Overview including the original SUS items, the adapted children's version and the Dutch translation.

Original SUS by Brooke (1996)	Adapted SUS for children by Putnam et al (2020)	Dutch translation of the adapted SUS for children
I think that I would like to use this system frequently.	If I had this app on my iPad, I think that I would like to play it a lot.	Als ik deze app op mijn iPad had, zou ik het vaak spelen.
I found the system unnecessarily complex.	I was confused many times when I was playing this app.	Tijdens het spelen van de app was ik vaak in de war.
I thought this system was easy to use.	I thought this app was easy to use.	Ik denk dat de app makkelijk te gebruiken is.
I think that I would need the support of a technical person to be able to use this system.	I would need help from an adult to continue to play this app.	Ik zou hulp van een volwassene nodig hebben om deze app te spelen.
I found the various functions in this system were well integrated.	I always felt like I knew what to do next when I played this app.	Tijdens het spelen van de app wist ik hoe ik verder moest.
I thought there was too much inconsistency in this system.	Some of the things I had to do when playing this app did not make sense.	Sommige dingen die ik in de app moest doen, begreep ik niet.
I would imagine that most people would learn to use this system very quickly.	I think most of my friends could learn to play this app very quickly.	Ik denk dat de meeste van mijn vrienden snel zouden leren hoe zij de app moeten spelen.
I found this system very cumbersome to use.	Some of the things I had to do to play this app were kind of weird.	Sommige dingen die ik tijdens het spelen van de app moest doen, waren raar.
I felt very confident using this system.	I was confident when I was playing this app.	Ik had zelfvertrouwen tijdens het spelen van de app.

| I needed to learn a lot of things before I could get going with this system. | I had to learn a lot of things before playing this app well. | Ik moest veel leren voor ik de app goed kon spelen. |