Human-Centered Design Science Research Evaluation for Gamified Augmented Reality

Nikolche Vasilevski* and James Birt

Faculty of Society and Design, Bond University, Robina, QLD, Australia

As augmented reality (AR) and gamification design artifacts for education proliferate in the mobile and wearable device market, multiple frameworks have been developed to implement AR and gamification. However, there is currently no explicit guidance on designing and conducting a human-centered evaluation activity beyond suggesting possible methods that could be used for evaluation. This study focuses on human-centered design evaluation pattern for gamified AR using Design Science Research Methodology (DSRM) to support educators and developers in constructing immersive AR games. Specifically, we present an evaluation pattern for a location-based educational indigenous experience that can be used as a case study to support the design of augmented (or mixed) reality interfaces, gamification implementations, and location-based services. This is achieved through the evaluation of three design iterations obtained in the development cycle of the solution. The holistic analysis of all iterations showed that the evaluation process could be reused, evolved, and its complexity reduced. Furthermore, the pattern is compatible with formative and summative evaluation and the technical or human-oriented types of evaluation. This approach provides a method to inform the evaluation of gamified AR apps. At the same time, it will enable a more approachable evaluation process to support educators, designers, and developers.

Keywords: augmented reality, gamification, educational games, design science research methodology, design evaluation, design methodology, micro-location

INTRODUCTION

Currently, there is fragmentation in how educators and designers analyze and evaluate immersive gaming experiences. Most educational game studies focus solely on the applied use of the game (e.g., usability or motivation) in the classroom and not on the design methodology and application evaluation (Sommerauer, 2021). Therefore, most educators and developers are left to start from scratch in the design journey citing a lack of reflective research and published methodology. However, as Nelson and Ko (2018) recommend, "the community should wholeheartedly commit to focusing on design and not on refining general theories of learning."

The purpose of design is a translation of existing situations into preferred ones (Simon, 2019). Moreover, "design science . . . creates and evaluates IT artifacts intended to solve identified organizational problems" (Hevner et al., 2004). Emerging from design thinking, Design Science Research Methodology (DSRM) is an iterative methodology aimed at rigorous development of solutions to problems, mainly in the Information Systems (IS) discipline. DSRM solution results in an artifact or multiple artifacts (Peffers et al., 2007). An artifact is commonly known as something

created by human beings for a particular purpose (Geerts, 2011). There are four different types of artifacts: concepts, models, methods, and instantiations. DSRM holds that the artifact must be able to solve an important problem.

DSRM has been used as a viable design method for implementing AR solutions (Vasilevski and Birt, 2019a; Vasilevski and Birt, 2019b), where it was used in designing and developing an educational AR gamified experience to bring people closer to the indigenous community. DSRM has also been proposed as the best-suited framework for implementing gamification as an enhancement service (Morschheuser et al., 2018). DSRM has also been used in the literature on education in designing gameful educational platforms useful to educators and students (El-Masri et al., 2015). Differentiating DSRM from the regular design, Hevner et al. (2004) hint that design science should address an unsolved relevant problem in a unique and novel way or provide a more effective or efficient solution to an already solved problem. "DSRM is intended as a methodology for research; however, one might wonder whether it might also be used as a methodology for design practice" (Peffers et al., 2007).

To design and create compelling and engaging learning, human-centred design is crucial. The benefits of integrating students, teachers, experts in education and technology, and the designers and developers in a collaborative creation process cannot be ignored (Bacca et al., 2015). As education steadily moves from lecture-based to more experiential learning approaches, games can be beneficial, providing hands-on experiences and real-world environments. However, achieving this and measuring its success presents a practical problem to the academics (El-Masri et al., 2015).

DSRM is a systematic process of developing a solution to a known problem, a model that consists of a nominal sequence of six iterative activities (phases) (Peffers et al., 2007). In simple terms, these are 1) Problem identification and motivation phase, which defines the problem and justifies the solution. The definition of the problem should be as concise and straightforward as possible. Next, the problem is split into small solvable chunks that can carry the complexity of the solution in the form of an artifact (s). 2) Define the objectives for a solution phase, infers the objectives of the solution from the problem, and inquires what is possible and feasible. 3) Design and development phase uses the design paradigm to establish the functional and structural requirements for the artifact (s), followed by the actual creation of the artifact (s) that was specified in the past phases. 4) Demonstration phase uses techniques as simulation, experiment, case study, proof, or any appropriate technique to demonstrate the capability of artifact (s) to solve the problem(s). 5) Evaluation phase, through observation and measurement, evaluates "how well the artifact supports a solution to the problem" (Peffers et al., 2007). During this activity, the objectives of the solution are compared against the observed results from the artifact's use during the demonstration. Therefore, measuring success, or the ability of the artifact to solve the problem, is paramount. 6) Communication phase involves disseminating the inquired knowledge about the

artifact and its design, effectiveness, novelty, and utility to researchers and relevant audiences.

Design science addresses the research by creating and evaluating artifacts designed to solve the identified problems in an organization. Evaluation is crucial in providing feedback information and a more in-depth understanding of the problem. This is especially important in education, where feedback and artifact design are core activities. Subsequently, the evaluation would improve the quality of the design process as well as the product. "Evaluation provides evidence that a new technology developed in DSR 'works' or achieves the purpose for which it was designed" (Venable et al., 2012). We think that all solutions in education should be design outcomes that follow the best practices and apply rigor in the process of design. Therefore, using design thinking and testing the solution capability to solve the problem should be paramount.

The reasoning and strategies behind the evaluation can be distinguished in terms of why, when, and how to evaluate. The equally important question is what to evaluate in regard to which properties of the evaluand should be investigated in the evaluation process (Stufflebeam, 2000). When considering the approaches, the evaluation can be approached quantitatively or qualitatively (or both). In terms of techniques can be objective or subjective (Remenyi and Sherwood-Smith, 2012). Regarding the timing, the evaluation can be *ex-ante* or *ex-post* (Irani and Love, 2002; Klecun and Cornford, 2005), or before a candidate system is conceptualized, designed or built, or after, respectively. Considering the functional purpose of the evaluation, there are two ways to evaluate: formative and summative (Remenyi and Sherwood-Smith, 2012; Venable et al., 2012). The primary use of formative evaluation is to provide empirically-based interpretations as a basis for improving the evaluand. At the same time, summative focuses on creating shared meanings of the evaluand considering different contexts. In other words, "when formative functions are paramount, meanings are validated by their consequences, and when summative functions are paramount, consequences are validated by meanings" (Wiliam and Black, 1996). The evaluation can also be sorted by its settings, where there are artificial and naturalistic evaluations as two poles of a continuum. The purpose and settings classification of the evaluation is combined in the Framework for Evaluation in Design Science Research (FEDS) (Venable et al., 2016). FEDS extended revision of the extant work by Pries-Heje et al. (2008), Venable et al. (2012).

The aim of this study is to test the application of DSRM to support the production of human-centered design approaches for AR games, thus addressing the research gap, which is the lack of design methodology and application evaluation for the purpose of immersive games. Our study provides a robust, published evaluation approach available to educators and design researchers, particularly novice ones, which can simplify the research design and reporting. This supports designers and researchers to decide how they can (and perhaps should) conduct the evaluation activities of gamified augmented reality applications.

METHODOLOGY

Below, we highlight the iterative DSR methodology (Peffers et al., 2007) process that was used to produce the evaluation pattern by using FEDS (Venable et al., 2016). We conceptualized, developed, and evaluated a solution in the form of an indigenous artworks tour guide mobile app (**Figure 1**). The app incorporated three major components, AR component, gamification component, and micro-location component from which the evaluation pattern was derived. The app used the components together to replicate an existing indigenous traditional tour on a BYOD mobile device (see).

Through DSRM, we demonstrated and evaluated three iterations (I1, I2, and I3) of the solution capability to solve a problem. The first DSRM iteration (I1) is presented in Vasilevski and Birt (2019b), where we performed a comprehensive analysis of previously published relevant applications to better understand the problem. We focused on the initial development and testing of the indigenous educational experience and usability focusing on AR and user interface. The solution only partially met the predefined objectives resulting in a second iteration. The second iteration (I2) is presented in Vasilevski and Birt (2019a), where the focus was on optimizing the implementation of the AR component and learning experience. The solution partially met the objectives resulting in a third iteration, yet to be published. The third iteration focused on the implementation and evaluation of th gamification component and its interplay with AR. The data are presented in the supplementary materials and online repository (DOI 10.17605/OSF.IO/CJX3D). Each iteration uses a specific methodological approach which is highlighted in the demonstrations section below. All phases of the study were conducted under ethical guidelines in accordance with institutional ethics.

Demonstrations

The demonstration of I1 took place in artificial settings, using qualitative data collection with a small population sample of five experts ($n = 5$). The experts' group consisted of an indigenous culture expert, user-experience expert, service-marketing expert, sense of place expert, and exhibition organization expert. The AR component was tested by following objective measurements, qualitative analysis, and usability evaluation techniques (Billinghurst et al., 2015). The usability testing activity was conducted following the guidelines for usability testing by Pernice (2018) from the Nielsen-Norman group (www.nngroup.com). The data was collected *via* observation and semi-structured interviews. The questionnaire from Hoehle and Venkatesh (2015) was adapted to generate ten focal points for the data collection regarding the artifact's interface. It incorporated ten user interface concepts as objective measurements: Aesthetic graphics, Color, Control obviousness, Entry point, Fingertip-size controls, Font, Gestalt, Hierarchy, Subtle animation, and Transition. Venkatesh and Davis (2000) TAM2 Technology Acceptance Model was used to derive data focal points concerning the level of performance of the solution and how helpful, useful, and effective it was. We analyzed the data by using thematic analysis (Braun and Clarke, 2006).

We performed the demonstration of the I2 part *in-situ* and part in artificial settings. For the *in-situ* demonstration, we approached the same group of five experts ($n = 5$) from I1. We selected six participants ($n = 6$) from the student population at an Australian university campus for the simulated environment demonstration. We used the same methodology from I1 (Vasilevski and Birt, 2019b) and built upon it in terms of the settings and the recorded details during observation and the interviews. As a result, there was more data gathered, and the quality of data improved.

The I3 demonstration took place *in situ*, and the number of participants was significantly higher than the previous iterations. Forty-two participants ($n = 42$) used the provided smartphone devices to experience the educational artwork tour on their own. The data collection methods were also updated. We collected the quantitative data *via* pre and post-activity questionnaires (**Supplementary Data Sheet**). The questionnaire consisted of adapted questions from Hoehle and Venkatesh (2015) original questionnaire with seven-item Likert scales. These were the same concepts from I1 and I2. Relating to AR specifically, the extended reality (XR) user experience questions were also added to the instrument as an adapted version of Birt and Cowling (2018) instrument, validated in Birt and Vasilevski (2021), measuring the constructs of utility, engagement, and experience in XR. Concerning the gamification service, we measured several other constructs such as social dimension, attitude towards the app, ease of use, usefulness, playfulness, and enjoyment, adapted from Koivisto and Hamari (2014). Finally, the quantitative data were subjected to parametric descriptive and inferential analysis in the SPSS software package.

The collected qualitative data was in the form of reflective essays that the participants submitted within 2 weeks after the activity and reflective comments embedded in the post-questionnaire. We used Nvivo software to analyze the qualitative data following the thematic analysis methods (Braun and Clarke, 2006). Observation was also part of the data collection. The above was in line with Billinghurst et al. (2015) AR evaluation guidelines, looking at Qualitative analysis, Usability evaluation techniques, and Informal evaluations.

To choose the appropriate evaluation approach specific to the project, Venable et al. (2016) developed a four-step process: 1) explication of the evaluation goals, 2) choosing the strategy or strategies of evaluation, 3) determination of the properties to be evaluated and 4) designing the subsequent individual evaluation episode or episodes.

We used the *Hierarchy of evaluation criteria* Prat et al. (2014) to select the properties to be evaluated (see **Figure 2**). All demonstration activities were evaluated concerning these criteria.

We used the FEDS framework (Venable et al., 2016) to evaluate the artifacts through the strategy of why, when, how, and what to evaluate. FEDS is two-dimensional in nature. The first dimension can be formative to summative and concerns the functional purpose of the evaluation. The second dimension can be artificial to naturalistic and concerns the paradigm of the evaluation. FEDS design process of evaluation follows four steps: 1) explicating the goals of the evaluation; 2) choosing the evaluation strategy(s); 3) determining the properties for
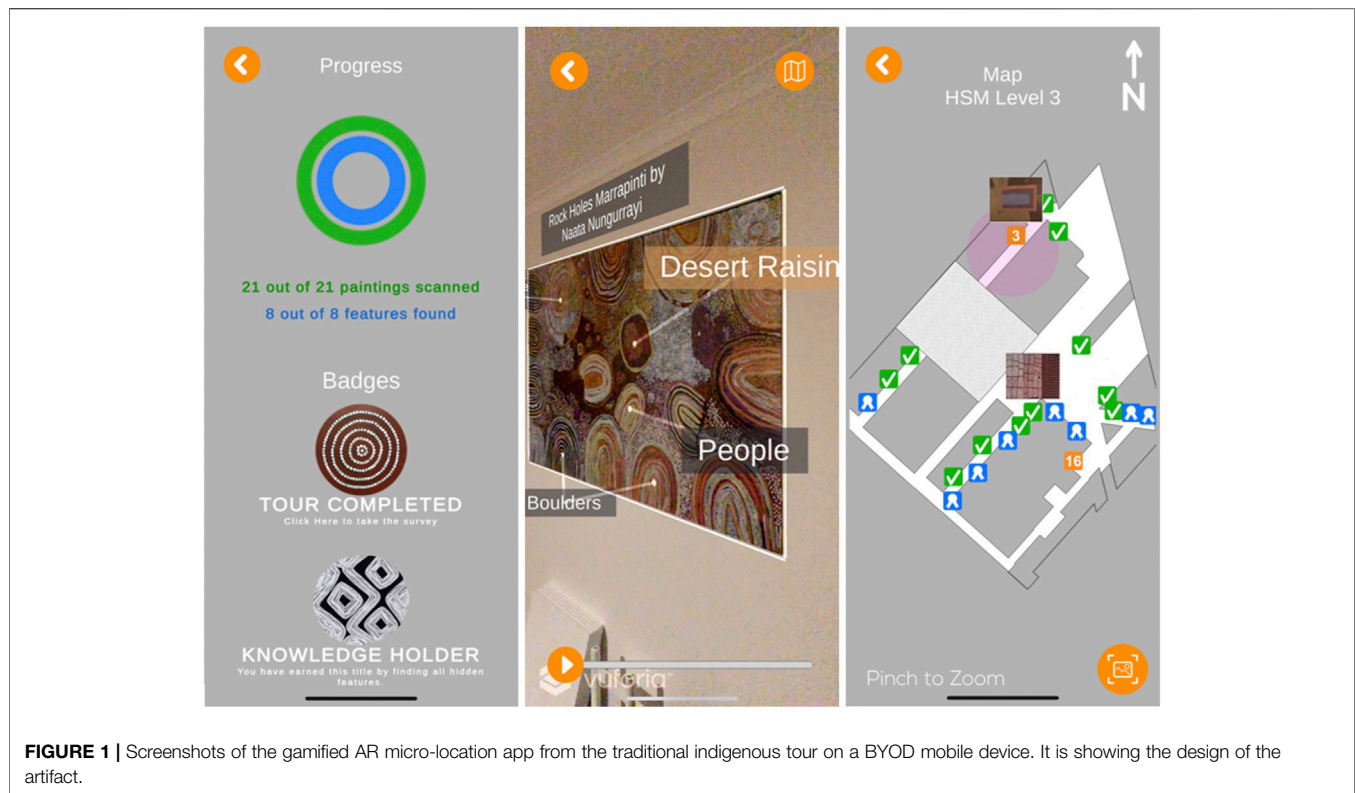
FIGURE 1 | Screenshots of the gamified AR micro-location app from the traditional indigenous tour on a BYOD mobile device. It is showing the design of the artifact.

evaluation; 4) designing the individual evaluation episode(s). While incorporating the above features, FEDS provides comprehensive guidance on conducting the evaluation.

The evaluation trajectory depends on the circumstances of the DSRM project. As mapped on the two dimensions of FEDS (see **Figure 3**), Venable et al. (2016) propose four trajectories for evaluating with a strategically different approach. These are Quick and Simple, Human Risk and Effectiveness evaluation, Technical Risk and Efficacy evaluation, and Purely Technical Artefact. All four strategies rely on the balance concerning speed, quality, cost, and environment.
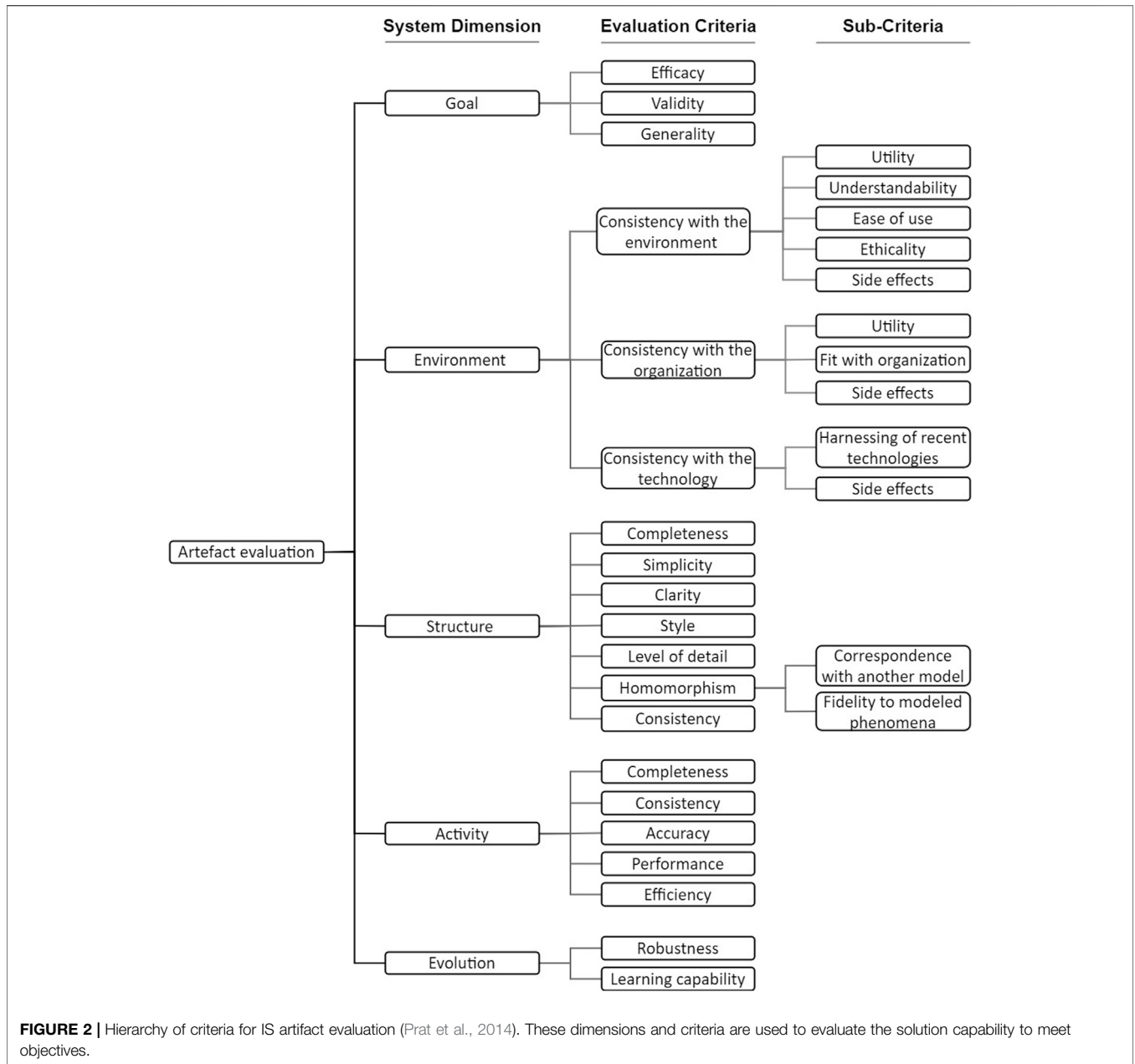
Evaluations

The results presented below are concerning the strategies and the techniques that we developed and implemented to evaluate the solution in three cycles. The empirical evaluation results from the demonstration activity are outside the scope of this study.

All iterations included development, refinement, feature addition, and upgrades of the artifacts. The nature of the solution required the inclusion of the human aspect since inception and followed throughout the process. Regarding the map of the evaluation, the process took a path closest to the Human Risk and Effectiveness evaluation strategy. This trajectory is illustrated in **Figure 3**, highlighted with green color. At the early stage, the strategy relied more on the formative evaluation and was more artificial in nature. As the artifacts evolved, the strategy path progressed toward a balanced mixture of the two

dimensions. After the process was past two-thirds, the evaluation transitioned into almost entirely summative and naturalistic, allowing for more rigor in the evaluation.

As we evaluated the instantiation artifact following the Human Risk and Effectiveness (see **Figure 3**), the other two model artifacts were evaluated *via* the instantiation evaluation. In essence, the evaluation compares the objectives and the inquired results from the Demonstration activity. The properties for evaluation were selected from the *Hierarchy of evaluation criteria* developed by Prat et al. (2014), which were goal, environment, structure, and activity dimensions as relevant for this project. We evaluated the *Environment dimension via* all the sub-dimensions, consistency with the environment, consistency with the organization, and consistency with the technology. For the *Structure dimension*, we evaluated the completeness, simplicity, clarity, style, level of detail, and consistency criteria. For the *Activity dimension*, we evaluated the completeness, consistency, accuracy, performance, and efficiency criteria. Evaluation of all dimensions and respective criteria included the methods explicated in the methodology section. The corresponding methods and the respective criteria that we used for all three iterations, as well as the proposed ones for a fourth future iteration (i4), are presented in **Figure 4**. In **Figure 4** we present the AR component evaluation, however, this can be generalized to other application components, such as VR, gamification and location-based services.

The evaluation process for all iterations is summarized in the five points below:

FIGURE 2 | Hierarchy of criteria for IS artifact evaluation (Prat et al., 2014). These dimensions and criteria are used to evaluate the solution capability to meet objectives.

1. We conducted interviews to investigate if the design meets the requirements and the expectations of the users and the stakeholders. These interviews were timed before, during, and after the artifact development and were vital in collecting feedback from experts and the target users.
2. To evaluate the components and technologies of the artifact, we conducted experiments and simulations throughout all iterations.
3. At the beginning of the development process, we tested the artifact's performance and the ability to meet the requirements in a closed simulated environment. As the process evolved, we refined the artifacts and introduced new features with every iteration, which depended on the previous evaluation episode

and the objectives of the solution. In addition, we gathered observation and interview data before, during, and after every activity and used it to debug, refine and upgrade the artifacts.
4. When the artifact had matured, providing sufficient performance, and implemented the key functionalities, we migrated the whole tested testing process in a real environment and with real users. We collected quantitative, observation, and qualitative data, shifting to more summative settings.
5. The evaluation showed that to provide a complete solution and meet the objectives in this project, a fourth iteration was required, where the artifact was to be deployed for the parent study experimental intervention.
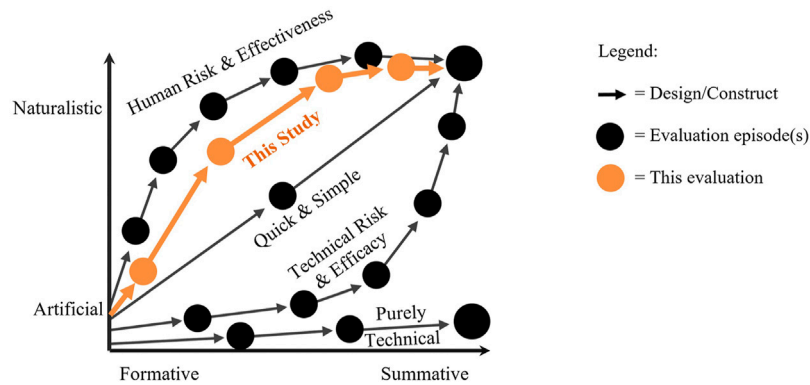
FIGURE 3 | Map of the dimensions, trajectories, and aspects of evaluation (Venable et al., 2016). The panel shows the possible paths that the evaluation process could take regarding sample size and the settings.

DISCUSSION

Regarding the human-centered first evaluation dimension in most occurrences, formative evaluation should be conducted during the beginning of the design and development process. The summative evaluation should be introduced after an artifact is mature enough and passed the basic evaluation. However, this does not exclude summative during the early evaluation and vice versa. As per Venable et al. (2016), the strength of formative evaluation is the reduced risks when the design uncertainties are significant, which in most cases is at the beginning of the process. On the other hand, summative evaluation provides the highest rigor and, consequently, the reliability of the inquired knowledge. Artificial evaluation can include processes such as laboratory experiments, mathematical proofs, and simulations and can be empirical (Hevner et al., 2004), or as in this case, preferably non-empirical.

The benefits of artificial evaluation could be stronger reliability, better replicability, and falsifiability. Moreover, it is inherently simpler and less expensive. However, it has limitations such as reductionists abstraction and unrealism that can produce misleading results regarding the users, systems, or problems. On the other hand, naturalistic evaluation probes the solution capabilities in natural, authentic settings, including real people and real systems. As the naturalistic evaluation is empirical in nature, it relies on case studies, field studies, field experiments, action research, and surveys. The benefits range from stronger internal validity to rigorous assessment of the artifact. The major limitations present the difficulty and cost of conducting the demonstration and evaluation. This could lead to the exclusion of some of the variables, which might negatively impact the realistic artifact efficacy.

The maturity of the artifact allows for more rigorous evaluation by moving to summative and naturalistic evaluations that would include larger sample sizes and realistic and *in-situ* evaluation environments, in line with Venable et al. (2016). This is an opportunity that should be used to introduce as much diversity as possible in terms of technology as well as the human aspect. As Hevner et al. (2004) suggest, the artifact should be implemented in its "natural" settings, in the organizational environment, surrounded with the impediments of the individual

and social battle for its acceptance and use. All this would provide more objective and detailed insight on the performance and the capability of the artifact to solve the problem.
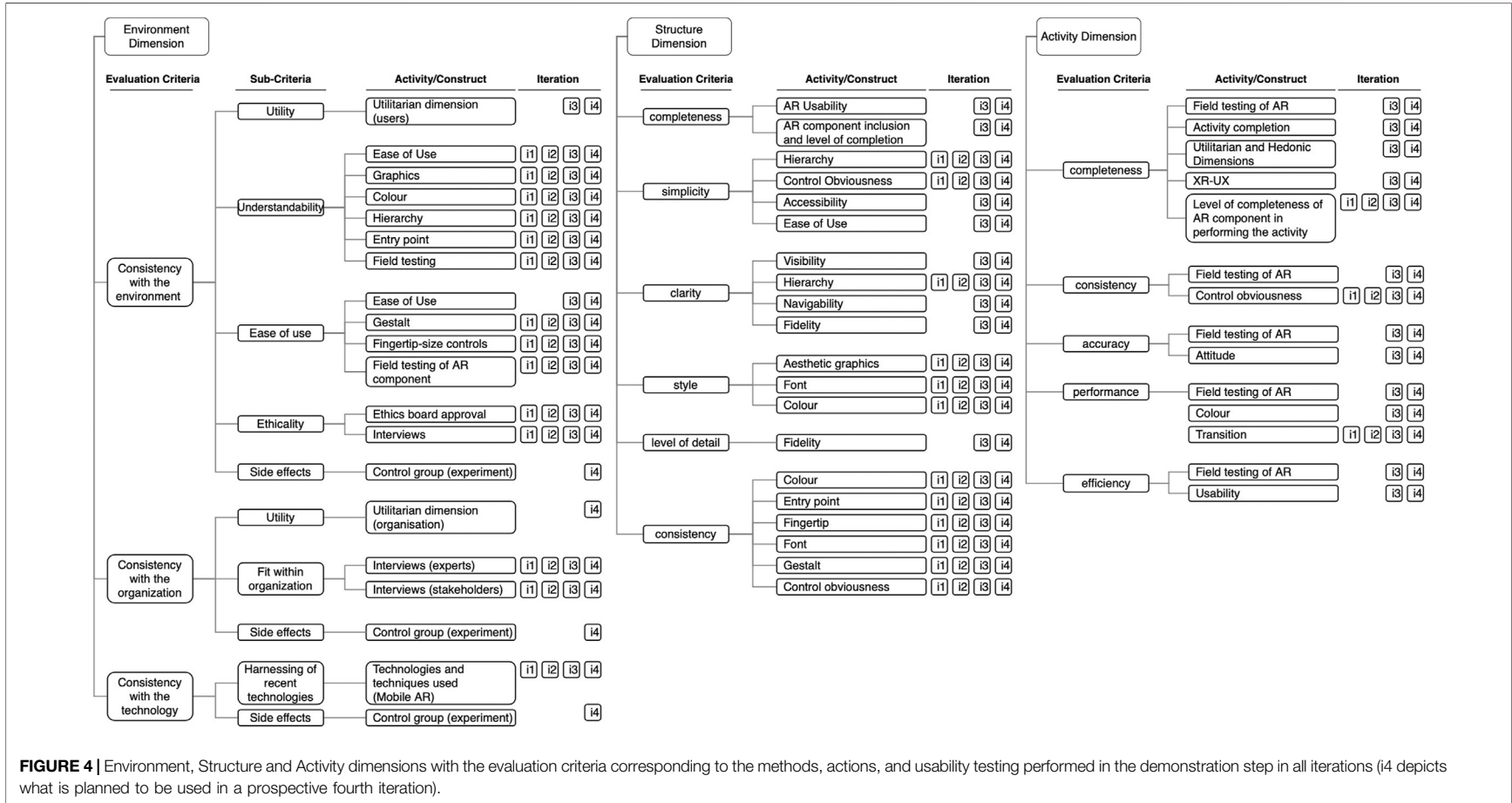
Continuous evaluation of the experiments and simulations is crucial to determine the optimums and the limits of the implementations and the symbiotic fit of the components. The mixed-type data used in the evaluation showed a holistic view of the state of the system and provided a pragmatic base for refinements and upgrades to the artifact as per Hevner et al. (2004). Furthermore, the evaluation of each iteration enabled and informed the next cycle. Thus, the process was cyclic in as many iterations as required to meet the required performance and objectives and ready to provide a solution to the problem in line with Peffers et al. (2007).

To evaluate the evaluation process used here, we look at the three objectives of DSRM. The evaluation in DSRM should: 1) be consistent with prior DS theory and practice, 2) provide a nominal process for conducting DS research in IS, and 3) provide a mental model for the research outputs characteristics. All three objectives are addressed below.

First, the evaluation process is consistent with the extant literature on the subject and best evaluation practices in IS. The evaluation approach was derived from multiple sources that converge on the subject (Pries-Heje et al., 2008; Prat et al., 2014; Venable et al., 2016). It is also consistent with the best practices, using the latest research and practice on usability, technology, and user experience testing (Billinghurst et al., 2015; Hoehle et al., 2016; Pernice, 2018).

Second, the evaluation followed the nominal process of DSRM (Peffers et al., 2007) and the evaluation guidelines. We showed how we evaluated the artifacts throughout the iterations following the process consistent with DSRM. In each iteration, the process worked as intended and was effective for its purposes.

Third, it provides a mental model for the presentation of the Design Science Research output. The evaluation process is explicated to a level that is comprehendible, relevant, and replicable. The steps, activities, and methods were described. The way we designed the evaluation can be used as a design pattern for the evaluation of similar or a variety of design projects.

FIGURE 4 | Environment, Structure and Activity dimensions with the evaluation criteria corresponding to the methods, actions, and usability testing performed in the demonstration step in all iterations (i4 depicts what is planned to be used in a prospective fourth iteration).

CONCLUSION

In conclusion, our study highlights a robust method of evaluation of gamified AR applications that can be used in education, design, and development. This is supported through our project solution case study for the development of a gamified AR micro-location indigenous artworks-tour mobile app. To the best of our knowledge, we are one of the first to provide a way of evaluating human-centered design artifacts through DSRM. We presented the six activities of DSRM throughout three development cycles, focusing on the evaluation. Our recommendations are based on the latest literature and best design practices, as well as the experiences gained throughout the process. DSRM showed to be an irreplaceable toolkit for designing and developing solutions to complex problems that emerge in intricate environments, such as education. Evaluation, as a critical part of the DSRM process, provides the rigor and robustness that can cope with solving high-complexity problems. Our research had some limitations in the form of technological limitations and the ability to test the other trajectories for evaluation and provide a broader picture, which had to be noted and could provide a foundation for future research. The evaluation path we showed is compatible with both formative and summative evaluation, as well as the technical or human risk and effectiveness types of evaluation. In the spirit of DSRM, the biggest strength of this study is the knowledge and experience shared, which is novel and provides support for educators and developers looking to design cutting edge solutions. We hope that this is a step towards a structured use of design patterns and the evaluation of gamified AR apps that should inform the artifact evaluation and the design process in a holistic manner in the fields of gamification and immersive technology.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by NV00009 at Health and Medical Research Council (NHMRC). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2021.713718/full#supplementary-material

REFERENCES

Bacca, J., Baldiris, S., Fabregat, R., Kinshuk, S., and Graf, S. (2015). Mobile augmented reality in vocational education and training. *Proced. Comp. Sci.* 75, 49–58. doi:10.1016/j.procs.2015.12.203

Billinghurst, M., Clark, A., and Lee, G. (2015). A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction* 8 (2-3), 73–272. doi:10.1561/9781601989215

Birt, J., and Cowling, M. (2018). Assessing mobile mixed reality affordances as a comparative visualisation pedagogy for design communication. *Res. Learn. Tech.* 26, 1–25. doi:10.25304/rlt.v26.2128

Birt, J., and Vasilevski, N. (2021). Comparison of single and multiuser immersive mobile virtual reality usability in construction education. *Educ. Tech. Soc.* 24 (2), 93–106.

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3 (2), 77–101. doi:10.1191/1478088706qp063oa

El-Masri, M., Tarhini, A., Hassouna, M., and Elyas, T. (2015). "A Design Science Approach to Gamify Education: From Games to Platforms," in *ECIS* (Germany: Münster).

Geerts, G. L. (2011). A design science research methodology and its application to accounting information systems research. *Int. J. Account. Inf. Syst.* 12 (2), 142–151. doi:10.1016/j.accinf.2011.02.004

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS Q.* 28 (1), 75–105. doi:10.2307/25148625

Hoehle, H., Aljafari, R., and Venkatesh, V. (2016). Leveraging Microsoft's mobile usability guidelines: Conceptualizing and developing scales for mobile application usability. *Int. J. Human-Computer Stud.* 89, 35–53. doi:10.1016/j.ijhcs.2016.02.001

Hoehle, H., Venkatesh, V., and Venkatesh, V. (2015). Mobile application usability: conceptualization and instrument development. *Misq* 39 (2), 435–472. doi:10.25300/misq/2015/39.2.08

Irani, Z., and Love, P. E. D. (2002). Developing a frame of reference for ex-ante IT/IS investment evaluation. *Eur. J. Inf. Syst.* 11 (1), 74–82. doi:10.1057/palgrave.ejis.3000411

Klecun, E., and Cornford, T. (2005). A critical approach to evaluation. *Eur. J. Inf. Syst.* 14 (3), 229–243. doi:10.1057/palgrave.ejis.3000540

Koivisto, J., and Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Comput. Hum. Behav.* 35, 179–188. doi:10.1016/j.chb.2014.03.007

Morschheuser, B., Hassan, L., Werder, K., and Hamari, J. (2018). How to design gamification? A method for engineering gamified software. *Inf. Softw. Tech.* 95, 219–237. doi:10.1016/j.infsof.2017.10.015

Nelson, G. L., and Ko, A. J. (2018). "On use of theory in computing education research," in *Proceedings of the 2018 ACM Conference on International Computing Education Research* (New York: ACM), 31–39.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *J. Manag. Inf. Syst.* 24 (3), 45–77. doi:10.2753/MIS0742-1222240302

Pernice, K. (2018). *User Interviews: How, when, and Why to Conduct Them [Online]. Nielsen Norman Group.* Available: https://www.nngroup.com/articles/user-interviews/(Accessed June 16, 2019).

Prat, N., Comyn-Wattiau, I., and Akoka, J. (2014). Artifact Evaluation in Information Systems Design-Science Research-a Holistic View. *PACIS*, 23.(Year)

Pries-Heje, J., Baskerville, R., and Venable, J. R. (2008). Strategies for Design Science Research Evaluation. *ECIS*, 255–266.

Remenyi, D., and Sherwood-Smith, M. (2012). *IT investment: making a business case*. UK: Routledge.

Simon, H. A. (2019). *The sciences of the artificial*. Massachusetts: MIT press.

Sommerauer, P. (2021). "Augmented Reality in VET: Benefits of a qualitative and quantitative study for workplace training," in *Proceedings of the 54th Hawaii International Conference on System Sciences* (Hawaii: HICSS), 1623.

Stufflebeam, D. L. (2000). "The CIPP model for evaluation," in *Evaluation models* (Berlin: Springer), 279–317.

Vasilevski, N., and Birt, J. (2019a). "Optimizing Augmented Reality Outcomes in a Gamified Place Experience Application through Design Science Research," in *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry* (New York: Association for Computing Machinery (ACM))), 1–2.

Vasilevski, N., and Birt, J. (2019b). "Towards Optimizing Place Experience Using Design Science Research and Augmented Reality Gamification," in *12th Australasian Simulation Congress* (Singapore: Springer), 77–92.

Venable, J., Pries-Heje, J., and Baskerville, R. (2012). "A comprehensive framework for evaluation in design science research," in *International Conference on Design Science Research in Information Systems* (Berlin: Springer), 423–438.

Venable, J., Pries-Heje, J., and Baskerville, R. (2016). FEDS: a framework for evaluation in design science research. *Eur. J. Inf. Syst.* 25 (1), 77–89. doi:10.1057/ejis.2014.36

Venkatesh, V., and Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Manag. Sci.* 46 (2), 186–204. doi:10.1287/mnsc.46.2.186.11926

Wiliam, D., and Black, P. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment?. *Br. Educ. Res. J.* 22 (5), 537–548. doi:10.1080/0141192960220502