Bootstrapping serious games to assess learning through analytics

Manuel Freire
Dept. of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
0000-0003-4596-3823

Antonio Calvo-Morata
Dept. of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
0000-0001-8701-7582

Iván Martínez-Ortiz
Dept. of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
0000-0001-6595-5690

Baltasar Fernández-Manjón
Dept. of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
0000-0002-8200-6216

***Abstract*—Serious games have several advantages for learning over more traditional activities. One of those advantages is that being highly interactive, they provide a first-hand simulated experience to the users. Serious games can also generate interaction data that can be used to assess learning, both for evaluation and to provide timely feedback for learners or even adapt aspects of the game on the fly. However, this potential is currently far from realized because doing so requires significant investment and expertise as compared to more traditional educational activities. Additionally, game learning analytics continue to be a costly and fragile process. We propose a combination of tools that greatly reduces the associated costs of building serious games with meaningful analytics simplifying the validation of those games and their deployment in real settings. We consider that this can be a key step to building predictive assessment avoiding the need for disruptive external assessment thus bootstrapping serious games to recognize learning through analytics.**

Keywords—Serious games, stealth assessment, data science, learning analytics, xAPI

I. INTRODUCTION

Serious games (SGs) have proven to be useful for learning [1], and are in use in fields as varied as engineering, medicine, the military, and not only for learning or training purposes but also for changing attitudes on complex social issues such as cyberbullying [2]. Serious games for education offer immersive experiences where players obtain experiential learning as they can explore the effects of their choices on simulated environments in a safe but situated experience, also referred to by some authors as authentic learning [3].

Serious games are also an ideal environment to better understand what student know and how they learn: the high degree of interactivity found in SGs is not only useful to keep players engaged in a first-hand experiential learning situation, but also allows collection and analysis of those numerous player interactions to improve the game design, and even to assess learning as and when it happens. Assessing learning without explicit external tests has been termed stealth assessment [4], and is highly desirable because it is seamless and integrated in contrast to traditional approaches which require stopping a fun game to take an immersion-breaking test. However, this user assessment process requires combining the game with a data engineering environment that allows the capture and analysis of interaction data, a process that is termed Game Learning Analytics (GLA) [5].

Despite the relative increase in the popularity of SGs in the last years, they are still far from commonplace in education; and SGs with stealth assessment are even more rare. Taking data from a December 2022 search of IEEE Xplore's Digital Library, while nearly 40% of results for "serious games" were published in the last 5 years, only 10 of those 2.1k recent results mentioned "stealth assessment". This paper describes the problems faced by stakeholders that may want to use SGs both for learning and assessment, and proposes an approach to address those problems. Section II describes the problems and typical approaches to solve them, while Section III describes our proposed approach.

II. OBSTACLES TO USING SERIOUS GAMES FOR ASSESSMENT

We find four main obstacles for a wider adoption of serious games with built-in stealth assessment, which we will describe in greater detail in this section:

1. Game development itself, and the relationship between fun and learning
2. Testing and analytics, both in design and in required infrastructure
3. Game deployment, especially if analytics and assessment are involved
4. Assessment itself, which requires validated games and machine learning

Building a compelling and attractive SG is a costly endeavor; we therefore argue that game design and development constitute the first obstacle to overcome. To build a compelling game, interlocking game mechanics and narrative must be designed, developed, tested, and integrated with suitable artistic resources such as images, animations and sounds. To ensure that a wide audience can play it, the game must be able to run on those platforms available to its intended users; this often requires targeting more than one platform, such as a desktop PCs or mobile devices. Fortunately, there is a growing number of game authoring platforms, such as Unity or Unreal, that support multi-platform development; and of developers that master them, driven by the strength and penetration of commercial video games. However, game developers are not typically educators, and imbuing the learning into the game requires both to collaborate in design and testing.

Partially funded by the Regional Government of Madrid (eMadrid S2018/TCS-4307, co-funded by the European Structural Funds FSE and FEDER), by the Ministry of Education (PID2020-119620RB-I00) and by the Telefónica-Complutense Chair on Digital Education and Serious Games.

The second obstacle is related to game validation (testing) and analytics. Videogames require extensive beta testing and formative assessment to improve the final product, a requirement that grows when the game's primary goal is not simply having fun, but rather learning [6], or any other SG goal, such as increasing awareness about a specific topic. A streamlined approach to test learning in games is to enable analytics that report on how users play, relating actions to learning using Game Learning Analytics (GLA). The main consumers of insights from GLA at the design and development stage should be those in charge of the game's educational design; therefore, they would be ideally able to come together with game developers and design a learning analytics model (LAM; see [7]), describing how in-game interactions map to learning. But this is seldom the case - it is rare to find educators with experience in designing and interpreting game learning analytics. Additionally, use of GLA requires a certain amount of technical infrastructure: the game must include a component that can keep track of in-game actions (which we will term a "tracker"), and this component must communicate with a collector component where all player activity can be stored and linked in order to be analyzed. Finally, player activity must be not only analyzed, but also presented in an understandable way to stakeholders, which, during the testing phase, are mostly developers and educators.

The third hurdle is related to deployment of SGs as educational activities. Besides providing an up-to-date game to the students, the use of GLA introduces additional challenges for deployment, as games must also be configured to report player interactions for analysis, so they must also be able to authenticate their players with the analytics system to ensure that incoming interactions are from actual game-playing students, and to ensure that interactions from the same student can be analyzed as such. While students can be provided with usernames and passwords for this purpose, this introduces significant friction into the process: users have to be enrolled and credentials distributed, and both enrollment and results would be managed in different systems to those that teachers normally use. Ideally, teachers interested in deploying a game in a course would be able to do so with very similar steps to those they follow when adding any other kind of activity to their institutional Learning Management System (LMS). The LMS would then provide details on who is playing the game, and the game would report results back to the LMS for teacher analysis.

Finally, the fourth obstacle relates to game validation and building of stealth assessment models from validated games. Formal validation of learning in a SG is usually performed through a pre-post test [8]: player knowledge is measured using the pre-test before playing, and compared to results on the post-test after playing, with the serious game considered as successfully validated if the increase in knowledge is deemed to be sufficient. Once a game is validated, it is possible to use data science techniques to build predictive models that, given interactions of students and their results in the tests, can associate interactions with learning outcomes. It is certainly possible to conduct stealth assessment without pre-post validations, but then there is no guarantee that assessment is measuring its intended constructs. This can be remedied by conducting a formal validation to test the extent to which both are correlated, as done by [4].

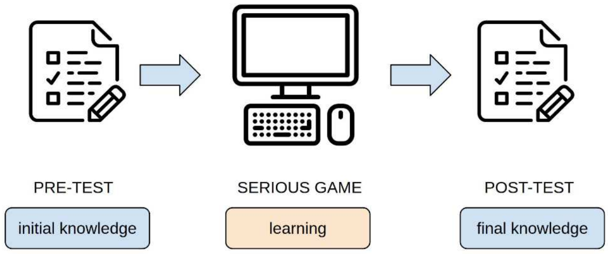

Fig. 1. Steps in traditional pre-post evaluation of a serious game.

Conducting paper-based pre-post evaluations in classrooms requires distributing a questionnaire per student, collecting responses, a game-play session, and then again distributing and collecting the post-gameplay questionnaire (see Figure 1). For each of the students, their two sets of responses and their gameplay interactions need to be linked together. This requires non-trivial logistics, although it can be partially automated through use of online questionnaires. More complex experimental designs, for example with control groups or different game versions, can yield additional information but require expanded logistics to carry out, and are more difficult to justify to participating students.

III. Recognizing Learning Through Analytics

Swimming pools often have shallow ends that are intended to be easily approachable, and deep ends that only committed or experienced users will want to enter. This can be used as a metaphor for potentially complex systems such as those found in educational technology. Bootstrapping a greater adoption of serious games that include assessment can greatly benefit from "shallow ends" for each of the four hurdles identified in Section II, while still allowing advanced users to, if they choose, dive deeper.

A. Game authoring in the shallow end

It is possible to greatly reduce the cost and effort associated with developing a serious game by making it accessible to educators, without the need to involve game developers. This is the approach taken by uAdventure (see Figure 2), a serious game authoring tool built on Unity [9]. Using uAdventure, technically inclined domain experts can create serious games of the point-and-click genre, and package them for the most popular platforms, such as Windows, Linux or MacOS PCs, or mobile devices running iOS or Android. From the point of view of increasing access to authoring, uAdventure takes the approach of making game development easier for domain experts; the opposite approach of making domain expertise easier for developers would fail, as there are too many possible domains of expertise to find a one-size-fits-all solution. uAdventure is available online as open-source, and includes documentation targeted at non-technical authors.

Since uAdventure runs on top of Unity, game developers can add advanced features easily, and can include third-party resources for Unity into uAdventure games such as self-contained mini-games – this provides a deep end which allows games of arbitrary technical complexity to be built. An example of this feature can be found in [10]. These features make it possible to extend the functionality of uAdventure to make more complete games with a wider range of game mechanics, adaptable to very different educational objectives and contexts.

B. Integrated authoring and analytics

Fig. 2. uAdventure's main interface. Analytics is automatically integrated into games, and deployed games can send interaction traces to any configured xAPI-compliant server, such as Simva.

An important insight is that adding analytics to a finished game is not as effective as building analytics into the game from the beginning. First, many important design decisions are difficult to change once the game is near completion; having analytics available during early testing can ensure that changes occur earlier rather than later, and this is important in development as failing early (knowing that it is not working educationally) reduces costs. Second, games built with analytics from the start will tend to be designed to be easier to test than those where analytics is added at the very end, resulting in better data to work with. Finally, an argument can be made as to stakeholder buy-in: if both educators and developers that build the game have collaborated early to set up analytics, it is much more likely that they will have reached agreement and feel ownership of the resulting analytics dashboards built to display those analytics to teachers who deploy the game in their classrooms [11]. In our experience, dashboards built by developers late in the process with little input from educators are likely to be perceived as opaque by teachers.

Our shallow-end approach to integrated authoring and analytics is again embodied in uAdventure, which includes built-in analytics. Any game built with uAdventure will have analytics enabled, and a built-in tracker component will record and store player interactions for potential analysis. A server-side tool to collect these interactions for analysis is also available, in the form of Simva, which is described more fully in the next subsection. Of course, unless customized, analytics will be basic, and report only on events that can be automatically identified as learning-related by linking the game's structure to suitable xAPI Serious Games (xAPI-SG) concepts [9]. However, this again provides useful scaffolding on which more nuanced analytics can be built. Once potential authors can explore, at minimal cost, how their games are being played, they can make much more informed choices on how to improve both the games and their associated analytics and reporting.

Both uAdventure, by including xAPI-SG standards-compliant trackers in games; and Simva, by collecting traces from xAPI-SG trackers, provide critical pieces of analytics infrastructure for analytics that would otherwise need to be sourced or developed by stakeholders. From the point of view of infrastructure, only an analytics dashboard is missing. To fill this gap, we have developed TxMon [12] (depicted in Figure 3), which can directly access Simva traces to build

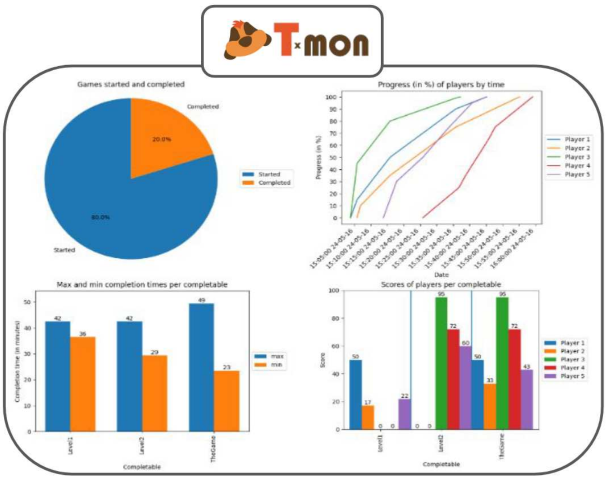

Fig. 3. TxMon provides default dashboards for xAPI-SG compliant analytics, and can retrieve them directly from Simva, where u-Adventure can place them with minimal setup.

generic dashboards. Again, both Simva and TxMon are open-source and available online.

While we provide a user-friendly and consistent experience when using our tools ecosystem tools (e.g. uAdventure, Simva, TxMon), through compatibility with the xAPI standard in uAdventure, it is possible to use a generic Learning Record Store (LRS) to store the analytics generated by any uAdventure game. Authors are therefore free to choose where to store analytics, and always retain control over their data. This also facilitates compliance with data stewardship requirements. Moreover, by using the xAPI-SG application profile, we facilitate the interoperability and portability of the data collected, and encourage the creation and reuse of tools, as the xAPI-SG profile provides a well-defined set of common semantics for data collected from games.

C. Game deployment

From a low entry-barrier point of view, deploying serious games as classroom activities should be as easy as deploying quizzes or questionnaires. Typically, teachers in secondary schools or higher education would use their institution's Learning Management System (LMS), such as Moodle or Canvas, to propose these activities to the corresponding group of students, and would be able to immediately check who has completed the activity and who has not. Those teachers would then later have easy access to results for all participants. Since serious games can generate additional analytics, we would want the corresponding analytics to be accessible also from within the LMS, instead of only reporting on game completion and overall score – in addition to making a global activity score directly accessible, just as expected from any other activity.

To support this vision, we have added LMS integration features into Simva [13]. Simva can act as a mediator tool between the LMS and the game, allowing one-click launch of games from the LMS by using links that, when clicked, provide user information to the game, which is then launched in the student's computer by Simva. The game reports analytics for the launching user to Simva, and Simva reports back completion and scores to the LMS. This Simva-mediated deployment greatly reduces the potential burdens of traditional deployment of serious games in settings where LMSs are already used, as no additional credentials need to be created or distributed, and there is no need to manually move results from one system to another. For teachers that are often not technically inclined, these are very important barriers to remove.

The integration process between the LMS and Simva is based on the IMS LTI 1.3 specification, that is the de-facto standard to integrate new external tools into a LMS. This not only allows us to cover a wide range of LMS, but also leverages the existing and growing familiarity by educators of this type of tool integration.

Additionally, Simva also gives teachers greater control over access to deployed uAdventure games, for example allowing them to prevent users from launching games outside the of a specific class activity; or allowing them to subdivide a classroom into subgroups so that each would play a different game from the others. Finally, it also provides minimal analytics feedback to teachers, for example displaying which students have completed the game and which have not yet started. We have not yet integrated dashboard-links to Simva-mediated LMS deployments; as of this writing, teachers would still need to log into Simva to access analytics dashboards.

D. *Validation and stealth assessment*

Formal validation is a straightforward way to measure the extent to which a serious game promotes learning. However, it presents several logistical challenges: ensuring that each student first answers the pre-test, then plays the game, and finally answers the post-test requires good organization, where any error distributing tests, collecting results, or in linking together the 3 activities for each student can lead to unusable data. To lower this friction, Simva includes support for this (and many other) common experimental designs [9]. In this role, Simva is therefore a "shallow end of the pool" for launching experiments and collecting their data. One critical component for this is built-in support for questionnaires, achieved through integration of LimeSurvey, an open-source questionnaire management software.

Formal validation with full playthroughs of players paves the way to using machine learning to build predictive models of learning from those playthroughs. This in turn can be used to provide stealth assessment to players while they play: once the model is trained, it can be used on playthroughs without pre or post-tests, and to the extent that the players are similar to those that the model was trained with, similar degrees of agreement between model predictions and actual post-test scores can be expected. (Alonso-Fernández 2021) proposes a set of mappings from xAPI-SG to machine learning features to predict post-test improvement in serious games. These can be extended to include additional parts of the game's learning analytics model. However, building a LAM and choosing parts of it to include when building prediction models may be too much to ask of stakeholders when attempting to keep entry barriers as low as possible. Therefore, we propose fully automated building of (very basic) prediction models directly from the generic, non-customized xAPI-SG traces that are automatically generated when authoring any uAdventure game. This last part is currently being implemented and relies on connecting a new module to Simva to first build and then apply predictive models to allow minimal cost, but also possibly noisy, stealth assessment to any validated uAdventure game.

IV. Games as Assessment Instruments

Serious games are already used as effective educational content in certain domains such as health, business or the military. However, in mainstream education, and particularly in schools, SGs, when used at all, are mainly present as complementary or motivational content. We blame the obstacles identified in Section II, and in particular the complexity of integrating SGs in the educational flow and of extracting reliable student assessments from SGs. There is a long history of attempting to fill this gap: in his seminal 1970 book *Serious Games*, Clark Abt proposed serious games not only as a way to improve education, but also as an instrument to improve the evaluation of complex skills, including not only cognitive, but also negotiation or organizational skills [12, p. 25]. However, this circumstance is gradually beginning to change as games or gamified environments are appearing that can assess skills in a rigorous and validated way. Perhaps the most significant example is Duolingo, a gamified environment that emerged as a motivating way to learn and practice foreign languages, which currently offers a validated system to evaluate the language level of a student – with the resulting assessments accepted by thousands of universities in the world [13].

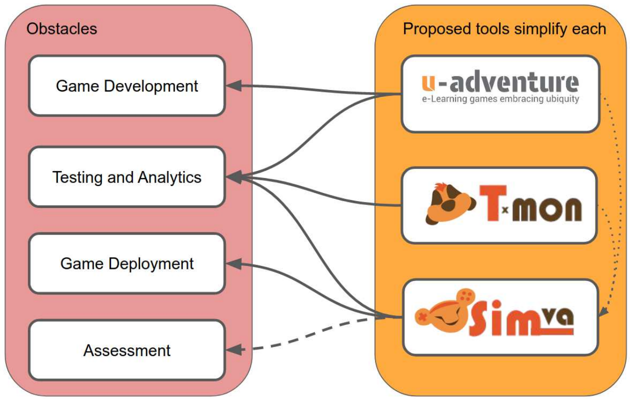

Fig. 4. Obstacles to using serious games with built-in assessment and proposed tools to address them. Dashed lines in the center indicate work-in-progress. Dotted lines in the left indicate inter-tool communication: both u-adventure and TxMon communicate with Simva.

Games are starting to be used to evaluate more complex situations, for example in personnel selection processes. But they are still isolated examples and more experience is needed for this to be generalized and systematized. And this has only just begun, as games through game analytics can provide large and rich data sets that enable the use of AI techniques which will open new opportunities for their applicability. This means that we have to start taking into account not only technical aspects such as, for example, that in order to systematize solutions it is necessary to use standards (e.g. xAPI) and reuse code or applications (e.g. LRS) that simplify the process and make it replicable. Other ethical and legal aspects must also be considered. For example, why implement an analytics system if there are already free game analytics services such as those offered by Unity or gameanalytics.com? In these cases, it is necessary to consider data ownership; since, by using a third-party service, we may be sharing sensitive data and incurring significant hidden costs attempting to analyze it later. It is also necessary to consider whether such services comply with European privacy regulations (GDPR), and whether they can be made compatible with the intended experimental design or with laws and regulations regarding school deployments. Such deployments can be especially stringent when students are legally minors. In any case, we believe that in the medium term it will be common to have skills and abilities that have been evaluated or even certified by a serious game.

V. Conclusion

Although serious games applied to education have proven their usefulness in countless occasions and contexts, their use as tools for stealth evaluation is still scarce. There are several major hurdles that make their use difficult and costly. To address each of those hurdles, we have presented a series of tools (see Figure 4), and improvements to those tools, that could bootstrap wider adoption of serious games for both learning and assessment of learning, powered through analytics. We are currently working on implementing and testing this approach, with most tools already fully functional and available online as open source. There are two main pieces of the puzzle that have yet to be integrated into their tools. One is to provide access to minimal TxMon dashboards when deploying uAdventure games as LMS activities, and the other is to provide minimal but fully automated stealth analytics for validated games. Both would also need to be documented to be usable to the non-technical teachers who we would like to run experiments with to validate the entire approach.

It is important to highlight that the use of well-defined and separate components to solve each of the obstacles found when applying game assessment allows us, or others, to combine these modules with other external solutions. For example, it is not necessary to use u-Adventure when creating games, we can integrate the tracker submodule with Unity to develop games with any other editor that may be built on this engine. Another example is that we can use TxMon with any other solution that can provide data in the xAPI-SG format, such as a standards-compliant LRS. In the same sense, a proper use of standards which allows us to integrate our solutions with external tools is also important to maintain compliance with applicable privacy regulations such as, in the EU area, the GDPR.

Acknowledgment

This work has been partially funded by the Regional Government of Madrid (eMadrid S2018/TCS-4307, co-funded by the European Structural Funds FSE and FEDER), by the Ministry of Education (PID2020-119620RB-I00) and by the Telefónica-Complutense Chair on Digital Education and Serious Games.

References

[1] E. A. Boyle *et al.*, “An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games,” *Comput. Educ.*, vol. 94, pp. 178–192, Mar. 2016.

[2] A. Calvo-Morata, C. Alonso-Fernández, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, “Creating awareness on bullying and cyberbullying among young people: Validating the effectiveness and design of the serious game Conectado,” *Telemat. Informatics*, vol. 60, 2021.

[3] J. Herrington, T. C. Reeves, and R. Oliver, “Authentic Learning Environments,” in *Handbook of Research on Educational Communications and Technology*, New York, NY: Springer New York, 2014, pp. 401–412.

[4] V. Shute *et al.*, “Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games,” *J. Comput. Assist. Learn.*, vol. 37, no. 1, pp. 127–141, Feb. 2021.

[5] M. Freire, Á. Serrano-Laguna, B. M. Iglesias, I. Martínez-Ortiz, P. Moreno-Ger, and B. Fernández-Manjón, “Game Learning Analytics: Learning Analytics for Serious Games,” in *Learning, Design, and Technology*, Cham: Springer International Publishing, 2016, pp. 1–29.

[6] M. S. Calabor, A. Mora, and S. Moya, “The future of ‘serious games’ in accounting education: A Delphi study,” *J. Account. Educ.*, vol. 46, pp. 43–52, Mar. 2019.

[7] I. Perez-Colado, C. Alonso-Fernandez, M. Freire, I. Martinez-Ortiz, and B. Fernandez-Manjon, “Game learning analytics is not informagic!,” in *IEEE Global Engineering Education Conference, EDUCON*, 2018, vol. 2018-April.

[8] R. Yáñez-Gómez, D. Cascado-Caballero, and J.-L. Sevillano, “Academic methods for usability evaluation of serious games: a systematic review,” *Multimed. Tools Appl.*, vol. 76, no. 4, pp. 5755–5784, Feb. 2017.

[9] I. J. Perez Colado, V. M. Perez Colado, I. Martinez Ortiz, M. Freire Moran, and B. Fernandez Manjon, “Simplifiying serious games authoring and validation with uadventure and SIMVA,” *Proc. - IEEE 20th Int. Conf. Adv. Learn. Technol. ICALT 2020*, pp. 106–108, 2020.

[10] D. Faouaz Santillana, A. García Cárdenas, and Á. Poyatos Morate, “Serious Games to Promote Computational Thinking and Coding,” 2021.

[11] A. Calvo-Morata, C. Alonso-Fernández, I. J. Pérez-Colado, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, “Improving teacher game learning analytics dashboards through ad-hoc development,” *J. Univers. Comput. Sci.*, vol. 25, no. 12, 2019.

[12] C. Alonso-Fernández, A. Calvo-Morata, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, “Game Learning Analytics:,” *J. Learn. Anal.*, vol. 9, no. 3, pp. 1–18, Dec. 2022.

[13] I. J. Perez-Colado, V. M. Perez-Colado, I. Martinez-Ortiz, M. Freire, and B. Fernandez-Manjon, “Using e-Learning Standards to Improve Serious Game Deployment and Evaluation,” in *2022 IEEE Global Engineering Education Conference (EDUCON)*, 2022, pp. 2077–2083.