AN AUTOMATED SYSTEM FOR EVALUATING 21ST CENTURY SKILLS USING GAME-BASED ASSESSMENTS

Nafisa Awwal, Claire Scoular, Monjurul Alom
Melbourne Graduate School of Education, The University of Melbourne (AUSTRALIA)

Abstract
Games are gaining momentum in the field of assessment for evaluating student performance and other complex 21st century skills. Gaming environments have the potential to provide learning activities and the opportunity to collaborate, problem solve and apply critical thinking skills. As part of the Assessment and Teaching of 21st Century Skills (ATC21S™) project, an assessment system for assessing cognitive and non-cognitive skills have been developed. The project focuses on assessing and teaching 21st century skills and the assessment system measures these skills in an online capacity. The design of the assessment system is innovative; using data analytic techniques, Rasch analysis and automated scoring to produce instant formative reports for educators. As the first assessment system of its kind to measure 21st century skills a discussion here surrounds the requirements and deployment techniques for the design of such complex assessments from delivery to reporting, making it meaningful and useful for teachers' use in everyday classroom.

Keywords: assessment, measurement, learning analytics, automated scoring, education, technology.

1 INTRODUCTION
Games are gaining momentum in the field of assessment for evaluating student performance and other complex 21st century skills. Educational web based game is being designed to support collaboration and assess different cognitive and social abilities among students. The educational games are developed to capture student responses or actions, both shared and unshared, within the game environment and extrinsic resources. The notion of web based game learning is to combine methods of gaming and learning in order to take benefit of the motivation that is inherent to game-playing. Web based game, a newly evolving medium, involves students in learning that intends to raise their higher order thinking. Games can be viewed as systems that integrate simulation, education and entertainment to create engaging learning. This paper discusses the requirements and deployment techniques for the design of ATC21S game like complex assessments from delivery to reporting, with the purpose of making it meaningful and useful for teachers' use in everyday classroom.

2 ASSESSMENT DESIGN
The assessment games developed during this research mirrored real world problem scenarios allowing the problems to have multiple possible outcomes or routes in reaching outcomes. This encouraged individuals to explore ideas previously considered incorrect or partially correct to facilitate their higher order thinking [1]. In addition, problems presented in the game are ambiguous, require multiple resources (skills, knowledge, artefacts) and engagement of individuals who are dependent on one another for successful resolution of the problem [2, 3]. Tasks were designed with either symmetrical or asymmetrical views, i.e., participants saw either identical or different screens, but in both cases they accessed or controlled different information and resources, reflecting the nature of the real world problems for which the construct is deemed relevant [4]. As with all games constructed for this study, game features are separated among participants so that they are not solvable without partner to ensure the collaborative elements are intrinsic part of the design. The initial levels operates as an exploration space for participants to learn and understand the problem space and the game mechanics. It offers individuals an investigation space to trial options by giving them the opportunity to work together with the aim to explore patterns, share information and discuss rules about the problem. The final level of the game tests the participants’ individual learning outcomes; that is, whether they have understood the rules and are able to apply their learning in a different context. The ability of

---

1 The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout this paper as ATC21S.
individuals to successfully demonstrate the possible correct solution paths for the task indicates evidence of learning.

Games developed in this study has assessment values in its core design. The components of ATC21S assessments, including the log stream data structure and scoring procedures, are game dependent. This dependency made initial game design laborious and intricate and limits its capacity to incorporate new games easily and efficiently. Presentation of the log stream for capturing data is crucial, particularly in the streamlining of scoring procedures. Most of the games developed in the earlier phases have similar data events but the naming convention under which these are recorded varies. Naming them differently has led to complications and additional work during the scoring procedures, although both naming essentially means the same thing. In addition, considerations for scoring procedures are required due to the variability of the games. The games range from a single to eight pages in length, with time varying on average between 5 and 45 minutes for the completion of each game. No one game is expected to capture all of the behaviours and greater variability between games is generally better for capturing a wide range of behaviours. However, in this scenario shorter, fewer page games were too short to include many game characteristics and therefore, little or no opportunity was presented to students in which they could demonstrate particular skills or behaviours.

The construction of games intended for assessment and learning purposes can be difficult because the goals of the games do not always align with the goals of assessment or learning. For games to be effective evaluation tools, they need to offer ample opportunities to participants to create, acquire, rehearse and apply skills and knowledge through active exploration within the collaborative virtual space [5]. Together with good game characteristics and assessment principles the games in this study are able to engage and motivate individuals actively to critically experience, practice and reflect on their ideas in a problem based context – providing educators a means to evaluate student performance on complex cognitive and non-cognitive 21st century skills. However, some games may have been challenging compared to others to successfully balance the contribution and engagement from both players within a collaborative environment [6]. In terms of development, it is difficult to design challenging but impartial games that require collaboration, while at the same time preventing all single player solutions.

3 DATA DESIGN

Game-based assessments are known to be able to capture large amount of rich information generated while the participants are engaged in the game environment. The games developed during the course of this research also provided rich evidence of student performance. The process data (processes by which students reach their answers) derived through such assessments are considered richer than traditional data as they can describe the type, order and quantity of interactions within such tasks [7, 8]. In the present study, this form of data is essentially collected through the capture of distinct key-strokes and mouse events such as typing, clicking, dragging, cursor movements, hovering time, action sequences and so on, in a process discussed in more detail by Adams, Vista, Scoular, Awwal, & Griffin [9]. Design of these session logs (a systematic records of user–game interactions) can provide salient solution processes in an unobtrusive way [7, 10]. The format in which data have been captured ensures efficient interpretation of participant responses for establishing reliable scoring rules. Session log files are structured to allow human interpretation without obscuring their understanding of the activity that occurred during gameplay. This is also important to validate the scoring rules. For example, each participant’s response is recorded as separate attributes as single instances with corresponding user identification, present state, timestamp, record index and other data as deemed necessary by researchers for the task analysis. In addition, optimum level of detail is captured for both analysis and processing of data for automating the scoring process. Consistent event types have been used across different games to uniformly describe the meaning for understanding the actions they represent. Timestamping and a form of record indexing is used for logging response data from the assessment games. Such measure is useful in sequencing user interactions within the game environment. User responses are recorded in corresponding seconds and milliseconds. Recording of precise times (i.e. milliseconds) of user responses are often required to differentiate sequences of events as intricate and accurate as possible in identifying behaviours and the sequences in which they occur [9].

Multiple actions can be recorded as occurring at the same time when recorded in seconds, but in games activities do occur consecutively and this should be more accurately reflected in the way responses are captured and arranged to ensure that more accurate interpretations of events occurrence. It is important to capture all or as many allowed activities as possible during a game.
environment. If more user interactions can be captured, it can provide meaningful understanding of participant’s salient response processes and allow measuring of any underlying 21st century skillset that are likely to emerge through analysis of the session logs. In short, any clicks or activities on screen by users should be captured regardless of their effectiveness, because invalid, ineffective and tentative actions may prove to be more informative in later analysis [11]. One way to deal with such large session logs or datasets is by visualising behavioural paths and patterns as described by Vista, Awwal & Care [12] to link it to cognitive processes involved while working with complex games. Each of the games have been developed with the capacity to automatically score responses collected for assessment purposes.

4 GAMES & GAME ENGINE DESIGN

Games are now popularly being used as assessments to capture information-rich student performances. The technological requirements of such assessments are those for the capture, storage and processing of time-stamped click events. Various forms of client and server configurations can be used for this purpose. During this study, the use of multiple game technologies have been adapted and used to cater various schools, delivery and research requirements as the system evolved and established into multiple versions over time. The graphical components and other elements of the games were designed using Flash2 with ActionScript 33. The use of Flash limited the range of devices, such as tablets with iOS that could be used for the games. Moreover, to access the tasks at a client browser with relatively up-to-date Flash Player plug-in was required. HTML54 was then used as it is deemed to be suitable for online games development, and also due to its cross compatibility with major web browsers, canvas5 element and rich graphic library like Phaser6 to mention a few.

The multiplayer components of the games are supported with the use of the various Web Socket application as it facilitates bi-directional real-time communication between servers and clients, allowing sharing of graphical or textual information and resources for promoting in-game collaborations without restricting information flow. This presents opportunities for each student to interact with the objects or resources presented to them. When an action is completed by the student, the game sends a message through the Web Socket to the server, and the server then sends that message to the other client, the other student, or partner to inform them of the action. SmartFoxServer 2X7 and Flash Media Interactive Server8, considered a standard choice for streaming media and shared object synchronisation for swf-based clients, were used as the socket server technology to support sharing of communications and synchronisation of shared objects between the multiple users. In addition, other forms of web sockets (e.g. socket io) that are standards within HTML5 have been used. It is important to consider the appropriate choice of technology, as it may limit the reach of such development only to a niche audience rather than a larger population, which in the context of education can have important consequences. The design of games need to be considered carefully. Not all actions or clicks in a game environment triggers changes at a given time. Developers and researchers need to plan which activities is to remain active and what are their possible resultant outcomes ascertain user behaviour it generates. In this study, the games being collaborative, no user is allowed to move forward without their corresponding partner in the game sequence or if at any point one partner’s session stops, the other partner is not allowed to proceed to a new stage. It is essential to preserve reasoning coherence in the game design for maintaining the game flow and its intended assessment purpose [13].

---

2 Flash is an Adobe’s multimedia product used for creating vector graphics, animation, games (online, desktop, mobile) and various other rich applications (online, desktop, mobile).

3 ActionScript is an object-oriented programming language, used primarily for the development of applications developed using Adobe Flash product.

4 HTML5 is a markup language used for structuring and presenting content on the World Wide Web. It is the fifth and current version of the HTML standard.

5 The canvas is an HTML element used to draw graphics using JavaScript. It has several methods for drawing paths, boxes, circles, text, and adding images.

6 Phaser is an open source game framework for Canvas and WebGL powered browser games.

7 SmartFoxServer is a middleware for creating multi-user applications, communities and games with Adobe Flash/Flex/Air, HTML5, Android, Unity3D, Apple iOS, Java etc. 2X refers to its community edition.

8 Flash Media Interactive Server (FMIS) is an Adobe product that works with the Flash Player runtime to create media driven, multiuser RIAs (Rich Internet Applications). It uses Real Time Messaging Protocol (RTMP) to connect to Flash Media Server Hub to synchronize complicated data structures and call remote methods on multiple clients in one go by having clients subscribe to a shared object.
5 PLATFORM & DELIVERY DESIGN

Considering the fact that schools are the potential users of such development, design choices on technologies are crucial as often educational institutions experience limited benefits despite the widespread use of internet and other advanced technologies. In this study, the games are used in educational settings, most commonly in schools, thus were designed to deal with unexpected interruptions due to internet connection failures or other technical issues. During such events, it allowed students to continue access from their previous responses. During a collaborative session, if one student of the pair encounters a technical problem or closes their window (or hits refresh/back button on their browser) by mistake, the games allowed both users to re-enter to the page on which they were last collaborating. In addition, as collaborators from a session may be distantly located, the partner is provided with a system notification of partner loss if their session stops at any time. If a partner closes or refreshes their browser window, the games will detect this activity and inform the other partner of it, as long as the network connection remains active. This is to notify partners of the status of each other’s progress. Both students can move, together, through the pages of the game in a linear fashion. It enacts students’ motivation and engagement with the overall task by avoiding frustration on repeatedly doing same thing, thus reducing the risk of contamination in from student responses. The aim of any assessment development is to reduce student frustration and anxiety and other type of noise when collecting data to evaluate student performances accurately.

Figure 1. Schematic diagram of the platform

Although multiple platforms co-existed for delivering the range of assessment games developed at the beginning of this project, an integrated system was realised and later deployed to consolidate multiple portals, reduce the use of differing technologies, automated scoring of responses and feedback mechanism to teachers and to deliver to a wide range of schools and students. The use of open standards (e.g. W3C standards) were adhered where possible, for a cost effective solution for such deployment. Schools are often limited in IT support and knowledge, which is all the more reason to create and deliver systems in the most accessible form to schools. Other key considerations catered during such development was be able to access for large numbers of simultaneous users, dispersed user location and devices in use, multi-lingual content capability (if required) and results of user attempts to feedback in real-time. The aim of such feedback mechanism is to allow teachers to monitor their student progress over time and link that progress to successful teaching strategies [14]. To support a collaborative process and synchronous communications in these interactive assessments, HTML5 has been heavily used as unlike Flash, it is supported by all devices (including

---

9 W3C stands for World Wide Web Consortium, which is the main international standards organization for the web.
10 HTML5 is a markup language used for structuring and presenting content on the web, including detailed processing models to encourage more interoperable implementations.
iOS devices). It is not constrained by licensing limits for the use of additional server based software which is required for the scalability. In addition, it provides a means for building collaborative services to support synchronous communication among collaborating users by the use of only a standard web browser. Employment of these technologies is likely to provide a consistent experience for users across all browsers and platforms. However, a disadvantage of these design choices is that older browser versions do not adequately support the HTML5 standards. The multiplayer architecture is hosted on a remote server and includes several components including Linux, Apache HTTP Server, MySQL and PHP [11]. The database is presented as a relational structure and the application packages are configured to support the various target languages and output the resultant game view at the client’s end. The application Node.js has been used to create a cross-platform runtime environment to allow the client and the server to communicate freely in real-time at efficient speed. The event-driven architecture employed by Node.js allows for good scalability and contains a built-in library which allows applications to act as a Web server without requiring additional software.

The underlying architecture for of the games and processing engine remained similar with changing technology to match the research and other requirements. The assessments were created to be executed in a similar fashion to that of a game within an online multi-user gaming architecture (Figure 1) [11]. The collaborative environment followed some essential steps of authentication and other teacher defined stages (for example, the target language to view the assessment games) to give students access to the assessment games. On successful login, the student arrives at a virtual room, referred to as the Game Lobby and is presented with a selection of games as selected by their teachers from an assessment pool. Once students select a game they are provided with options to select a role, A or B, which, unknown to them, provides alternate views of the game. After selection of their role the assigned pair are moved through to the virtual Game Room on the basis of predefined rules. While paired under a session, students can collaborate or operate among themselves and within the problem space. At the completion of a task, students are redirected back to the dashboard containing the assessment list applicable to an individual pair. The whole process steps users through a shared session while collecting their responses for collaborative behaviour.

6 INTERPRETATION AND USE

The key purpose of the use of assessment is to advance student’s learning and inform teacher’s teaching interventions. The form of complex assessment developed during this study are intended to be used by teachers as one of the many sources of evidence that reflect student learning. Reports that are generated from these assessments are presented to maximise their use within a developmental model of learning. These reports focus on the skill level where the student is ready to learn. They provide teachers with a means to track student progress for the purpose of informing and evaluating teaching. The test results are used to identify the point of learning readiness or zone of proximal development (ZPD) and they are reported as a level on the corresponding developmental progression for collaborative problem solving.

It can also provide teachers with vital analytical information that they can use in the classroom. This form of timely and accurate information about student ZPD can then be used as part of the evidence base against which teaching can be planned. A great advantage of such assessments is that it delivers instant responses and analysis of results to teachers and students, so that immediate information is available for intervention purposes. Moreover, such an approach to test administration can control many factors that introduce error into measurements. Game like assessments are generally believed to be engaging for students if designed correctly and can also be designed to match students’ ability levels. Teachers are given access to test results in time to inform their teaching and monitor the progress of their students over time.

7 CONCLUSION

A crucial benchmark of the ATC21S project has been to develop interactive games like assessments that use synchronous communications to provide support for a collaborative process for evaluating student performance. As the first assessment system of its kind to measure 21st century skills, the design of this assessment system is innovative; using data analytic techniques, Rasch analysis and automated scoring to produce instant formative reports for educators to inform their teaching interventions. The availability of many new and novel technologies such as HTML5 and Web Socket

11 Node.js is an open-source, cross-platform JavaScript run-time environment for executing JavaScript code server-side.
have made such complex assessment development possible while at the same time ensuring consistent game flow with a positive user experience while maintaining the integrity of student evaluation techniques. Multiplayer games designed in this study, allowed two students to collaborate by sharing resources and communicating through an embedded chat box. Online multiplayer game has been used to support collaboration and to assess cognitive and social skills among students. Designing multiplayer games for the purpose of evaluating students’ performance that demand collaboration is time consuming and resource intensive task. However, careful considerations on all the design elements that are employed in planning such collaborative games may be inevitable to measure complex skills and to respond to the needs of working life in the 21st century.

REFERENCES


