Expert Systems With Applications 248 (2024) 123370

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Developing and validating interoperable ontology-driven game-based assessments

Manuel J. Gomez *, José A. Ruipérez-Valiente, Félix J. García Clemente

University of Murcia, Calle Campus Universitario, 30100, Murcia, Spain

ARTICLE INFO

Keywords:
Serious games
Game-based assessment
Interoperability
Ontologies
Data mining

ABSTRACT

Video games have assumed an important place in our daily lives. This has led to an increasing interest on the use of games for non-entertainment purposes, introducing the concept of Serious Games (SGs). In particular, SGs are being explored because of their potential to provide reliable assessments, but also because they can measure competences that would be difficult to measure using traditional forms of assessment. However, one of the key issues is that assessment machinery has to be designed specifically for each game, increasing the time and effort when designing and implementing Game-Based Assessments (GBAs). In this research, we introduce a novel approach to develop interoperable GBAs by: (1) designing and creating an ontology that can standardize the GBA area; (2) conducting a validation study on literature metrics to replicate them and designing novel metrics using data from different SGs; (3) conducting a case study that illustrates how our approach can be used in a real life scenario with real data. Our results confirm that the designed ontology can be used to effectively perform GBAs, along with the metrics replicated and designed in the system. We expect our work to solve the current limitations regarding GBA interoperability, thus allowing the deployment of Game-Based Assessments as a Service (GBAaaS).

1. Introduction

Nowadays, technology plays a very important role in our life, making our work much easier and less time consuming. One of the most prominent examples of technology is the use of digital games for learning (De Freitas, 2006). In recent years, video games have assumed an important place in the lives of children and adolescents, impacting on various aspects of everyday life such as our consumption, communities, and identity formation (Daniel & Garry, 2018; Gros, 2007). In fact, two thirds of adults and three quarters of kids under 18 play video games weekly, and during the pandemic, 71% of parents saw video games as a much-needed break for their children (ESA, 2021). While video games are usually associated with entertainment and leisure, they have recently emerged as powerful tools for learning and skills development. This has generated an increasing interest on the use of games in non-entertainment contexts during the last decade (Susi, Johannesson, & Backlund, 2007). Specifically, the concept of Serious Games (SGs) was first coined by Abt (1987), and probably the most common definition is: "games that do not have entertainment, enjoyment, or fun as their primary purpose" (Laamarti, Eid, & El Saddik, 2014). SGs are mainly used in education; however, they are also used in many other domains (Laamarti et al., 2014), including training, well-being, advertisement, interpersonal communication, or assessment, among others.

SGs are increasingly being explored for use as assessment tools in broad domains, in particular for their potential to provide more valid assessments compared to traditional assessment approaches, also providing more meaningful and authentic contexts for assessments through interactive immersive environments (Kato & de Klerk, 2017). Specifically, Game-Based Assessment (GBA) is a specific application of SGs, referring to a type of assessment that uses players' interactions with the game as a source of evidence to make meaningful inferences to reveal knowledge, skills, and attributes of users and students that are "invisible" or hard to detect when assessed with more traditional assessment methods (de Klerk & Kato, 2017; Gomez, Ruipérez-Valiente, & Clemente, 2022). However, some limitations are still present, hindering the use of GBAs in real world environments. For example, little is known as to what degree of design complexity is required for meaningful learning to occur, and many games are simple designs targeting low level literacy and providing drill and practice methods (Qian & Clark, 2016). Moreover, there is a lack of sound empirical evidence on the effectiveness of GBAs due to different outcome measures for assessing effectiveness, varying methods of data collection and inconclusive or difficult to interpret results (All, Castellar, & Van Looy, 2014). Usually,

* Corresponding author.
E-mail addresses: manueljesus.gomezm@um.es (M.J. Gomez), jruiperez@um.es (J.A. Ruipérez-Valiente), fgarcia@um.es (F.J.G. Clemente).

https://doi.org/10.1016/j.eswa.2024.123370
Received 30 August 2022; Received in revised form 16 November 2023; Accepted 28 January 2024
Available online 1 February 2024

data collected from SGs are totally different between them, and there is no interoperability between different data sets (Serrano-Laguna et al., 2017). Thus, GBA machinery (including metrics, dashboards, and other type of analytics) are usually designed for each game separately, which increases the time and effort needed building each model from scratch. Since this is one of the key issues open in the area, there is an urgent need to work on data interoperability in order to re-use assessment models and machinery from one game to another.

Previous literature has proposed standard data formats, trying to address these data heterogeneity issues. One example is the work in Serrano-Laguna et al. (2017), which proposed Experience API (xAPI), an interaction model to track user activities within learning environments. However, this and other similar approaches are not supported by most SGs. Other technologies that can help us to address data interoperability are Semantic Web technologies. In particular, ontologies capture knowledge about a certain domain and offer an explicit common conceptualization on it (Fathy, Gad, & Badr, 2019). Ontologies are content theories about the classes of individuals, properties of individuals, and relations between individuals that are possible in a specified domain of knowledge (Panov, Džeroski, & Soldatova, 2008). Although the use of games for assessment has enjoyed great popularity and success in recent years, there is a distinct lack of a generally accepted framework that would describe and unify the area of GBAs.

In this research, we introduce a new approach that uses ontologies in order to develop interoperable GBAs, using in-game metrics that are automatically computed processing the provided data. With that purpose in mind, we propose the design of a new GBA ontology, and we validate it using previous GBA metrics present in literature, as well as new metrics proposed to show the potential that this approach has to perform interoperable GBAs effectively. To validate our ontology, we use data from many different SGs, and we present a case study to demonstrate how this approach could be used in real world situations and environments. Specifically, we have the following objectives:

1. To develop an ontology that can standardize the GBA area, creating a common knowledge model that can integrate the log events from a wide variety of games into our ontology model.
2. To conduct a validation study on previous metrics in literature, as well as the design and implementation of novel metrics using data from different SGs, demonstrating and validating that our approach can effectively perform interoperable GBAs.
3. To conduct a case study that illustrates how our ontology, along with previous and newly developed metrics, can be used in a real scenario.

The rest of the paper is structured as follows: Section 2 reviews background literature on SGs and assessment, GBA metrics and models, and ontologies. Section 3 describes the methodology followed to conduct the research, as well as the games and the data collection used. Next, Section 4 present the results (including the ontology developed), a set of interoperable in-game metrics, and finally the case study conducted. Then, we finalize the paper with discussion in Section 5 and conclusions and future work in Section 6.

2. Related work

In this section we present a review of the literature in some areas which are related to our work: in Section 2.1 we present literature related to serious games and assessment; in Section 2.2 we review some GBA and metrics background and previous studies; and finally in Section 2.3 we review literature related to ontologies and their use in interoperable environments.

2.1. Serious games and GBA

In recent years, many new ways of teaching academic and professional skills to children and adults have been tested using multimedia technologies in the form of software products, educational computer games or video games (Girard, Ecalle, & Magnan, 2013). Reyes-Chua and Lidawan (2019) reported a summary of benefits of using games for learning, including increased learner motivation, reduced learning anxiety, or encouraged creativity and cooperation, among others. In addition, games can promote user engagement through fantasy, interactivity, and non-linear narratives in visual and multisensorial environments that take advantage of advancing technologies (Kato & de Klerk, 2017).

SGs are being used in many different contexts: in education, interest in educational games is continuously growing, but their integration in teaching is still somewhat unexplored area of study (Kangas, Koskinen, & Krokfors, 2017): for instance, some studies have reported difficulties when obtaining an optimal game design, since it is an interdisciplinary task, requiring the contribution of experts from many different areas such as graphic design, product design, programming, or animation (Theodosiou & Karasavvidis, 2015). Moreover, the strict educational system and the fact that some teachers refuse the idea of using "toys" in classroom also entails an added challenge when incorporating games in the classroom (Lee, Luchini, Michael, Norris, & Soloway, 2004). Furthermore, SGs and GBAs have been promoted for use in employee selection as a potential method to improve the user experience, and the use of games in the workplace is a growing phenomenon with SGs being increasingly used as evaluative tools (al Qallawi & Raghavan, 2022). Regarding healthcare, SGs, particularly adventure and shooter games, already play an important role in education, prevention and rehabilitation (e.g. to enhance health-related physical activity, improve sensory–motor coordination, prevent asthma, change nutrition behavior and alleviate diabetes and prevent smoking or HIV) (Wiemeyer & Kliem, 2012). Concerning employee training, SGs are being used by corporations of all sizes (Larson, 2020) to train, for example, financial indicators (Donovan & Lead, 2012) or call center assistants (Hinton, 2016; Mollick & Werbach, 2015).

Games are being explored in particular for their potential for assessment, providing promising possibilities for more valid and reliable measurement of users' skills as compared to the traditional methods of assessment like paper-and-pencil tests or performance-based assessments (de Klerk & Kato, 2017). GBAs can provide more detailed and reliable information, and the emerging interest in this field reflects the need for alternative assessment tools to overcome limitations that are present in classic methods (Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013). In contrast with traditional methods, digital GBA methods have the following merits: (a) they are fun and can reduce test anxiety; (b) they allow for recording users' interactions in detail (i.e., via the accumulation of log data generated by keystrokes and mouse clicks); and (c) they can be designed to provide real-time learning supports (Shute & Ventura, 2015). In literature, we can see that previous studies have integrated GBAs in classrooms in order to assess students' skills or knowledge in many different domains: mathematics (Chiu & Hsieh, 2017), art (Basu et al., 2020), language (Song & Sparks, 2019), or soft skills (Nikolaou, Georgiou, & Kotsasarlidou, 2019) are only a few examples of knowledge domains where researchers have conducted studies using games for educational assessment. Apart from education, we can also find previous literature applying GBA in other different contexts, such as healthcare (Vallejo et al., 2017) or employee selection (Georgiou, Gouras, & Nikolaou, 2019). However, the GBA potential to perform valid assessments is mitigated by the time and effort that designing these types of assessments require. Therefore, a common way to conduct GBAs without going through the entire design and implementation process would alleviate these issues and promote more assessments using SGs. In our work, we introduce an intermediate layer that acts as a common knowledge model in order to be used with different data formats so that GBAs can be performed and visualized by simply adapting the data available.

2.2. Game-based assessment models and metrics

Once we have gathered data from users' interaction with a specific game, how to perform a valid and reliable assessment? Although some game and learning analytics can indeed be used in GBAs, they lack specific metrics and methods that outline their effectiveness. SGs analytics need to provide (actionable) insights that are of values to the stakeholders (Loh, Sheng, & Ifenthaler, 2015). In education, new techniques such as Learning Analytics (LA) are trying to provide insight about the educational processes and improve the common educational scenarios benefiting from data-driven approaches (Alonso-Fernández et al., 2019). Its aim is to understand learners and their environments, and improve the learning process through analysis of the data collected from students' interactions with the learning environment to assess students, predict future events and act consequently to refine educational actions (Alonso-Fernandez, Calvo-Morata, Freire, Martinez-Ortiz, & Fernández-Manjón, 2019; Serrano, Marchiori, del Blanco, Torrente, & Fernández-Manjón, 2012).

LA and other techniques, such as data mining (and educational data mining), can be used to fuel the advancement of games research through leveraging the rich data streams enabled by digital GBAs (Owen & Baker, 2019). These areas are applied to explore models and techniques for making efficient and effective use of these data: capturing, tracking, aggregating, analyzing, and visualizing/utilizing information about users' interactions with learning content and their learning progress (Shoukry, 2020). The use of data from games can be collected while users are playing to analyze not only the impact the game is making (in their learning), but also the appropriateness of the game design and its mechanics (Alonso-Fernández et al., 2019).

Gathered data should help get inferences about general traits and abilities of the learner, his general knowledge state, his situation-specific state, his behaviors and his outcomes (Shoukry, Göbel, & Steinmetz, 2014). A frequent approach is to use a set of metrics (or indicators) calculated from users' data. In fact, assessment mechanics help game designers select or design game mechanics that generate useful game metrics (Plass et al., 2013). Many studies have been using metrics to measure students' interaction with educational games, measuring persistence (DiCerbo, 2014) or engagement (Ruiperez-Valiente, Gaydos, Rosenheck, Kim, & Klopfer, 2020), among others. In a survey conducted by Gris and Bengtson (2021) on assessment measures in game-based learning research, 91 studies were analyzed, and results showed that learning aspects are much more assessed than engagement and usability features. Moreover, metrics can be used for other purposes rather than to report users' knowledge. For example, Martınez, Gómez, Ruipérez-Valiente, Pérez, and Kim (2020) developed a series of metrics related to students' activity (e.g., active time, number of different events), but also to the difficulty of different levels in the game, so teachers can adapt their teaching based on these data. Finally, we can highlight the work by Hamdaoui, Khalidi Idrissi, and Bennani (2016), who used in-game metrics to define the students' learning and playing style, but also to adapt gameplay and learning content based on those metrics. Although each environment may have specific metrics some are more common across environments, such as those related to the activity with numbers of events or active time (Ruipérez-Valiente, Gomez, Martínez, & Kim, 2021).

In Liu, Kang, Liu, Zou, and Hodson (2017), authors performed a systematic review on the use of LA for assessment in games, and the results highlighted the promise of using multiple data sources, as well as combining emerging techniques such as visualization and traditional analyses such as statistical and qualitative analyses. When done correctly, visualization can reveal information otherwise unobtainable through traditional statistical analysis. Information visualization is a field of study in its own right and increasingly includes new approaches to visualize spatial and temporal data for reporting and communication purposes (Loh & Sheng, 2015). In recent years, several dashboard applications have been developed to support learning or teaching. Such dashboards provide graphical representations of the current and historical state of a learner or a course to enable flexible decision making using visual elements (Podgorelec & Kuhar, 2011; Verbert, Duval, Klerkx, Govaerts, & Santos, 2013). They allow the data to be processed so that they can be visualized in a way that enables the teacher or learner rather than the software to make sense of them, converting the abstract and complex to the concrete and visible by amplifying human cognition (Card, 1999; Duval, 2011).

These metrics and visualization systems present in literature are designed specifically for each game and study, which creates interoperability issues. An objective in our research is to review previous GBA metrics, as well as designing novel interoperable metrics to demonstrate that is possible to perform GBA with different SGs and data formats. In addition, our case study also includes an example of how to use visualizations to graphically represent all these interoperable metrics.

2.3. Interoperability, standards and ontologies in games

While the use of SGs has extended rapidly to a variety of domains, their design, development and later analyses of results remains a challenging individual process for both developers and teachers/trainers (Stănescu, Stefan, Kravcik, Lim, & Bidarra, 2013). Interoperability is a key requirement for organizations regardless of the field they operate. People, organizations and software systems must communicate between and among themselves. However, due to different needs and background contexts, there can be widely varying viewpoints and assumptions regarding what is essentially the same subject matter (Uschold & Gruninger, 1996). The way to address these problems is to reduce or eliminate conceptual and terminological confusion and come to a shared understanding. Previous studies have proposed approaches trying to standardize analytics into games. Alonso-Fernandez, Calvo, Freire, Martinez-Ortiz, and Fernandez-Manjon (2017) proposed an interaction model (xAPI) that can be used to describe streams composed of actors performing with actions in a specific context. Each xAPI statement represents a learning activity and has three main attributes: an actor, a verb and an object: who did what action, with a target of the action and certain additional attributes. Moreover, Perez-Colado, Rotaru, Freire, Martinez-Ortiz, and Fernandez-Manjon (2018) proposed a method that comprises the specificities of location-based games, as an extension of the xAPI standard to support location-based SGs.

Another area that could help us to establish a common model and remove conceptual confusion is the area of ontologies. Ontologies are often defined as a set of concepts, their definitions and their inter-relationships about certain domain (Uschold, 1996). In computer science, the concept "ontology" is interpreted in many different ways and concrete ontologies can vary in several dimensions, such as degree of formality, authoritativeness or quality (Happel & Seedorf, 2006). Researchers in many areas have all recognized the need for ontologies to clearly define specialized vocabularies for these domains (McDaniel & Storey, 2019), and nowadays we can see ontology-based applications in areas as diverse as customer support and engineering of cars (Staab & Studer, 2010). Several approaches have been proposed for developing ontologies (Corcho, Fernández-López, & Gómez-Pérez, 2003): following a bottom-up strategy, on the basis of an application Knowledge Base (KB); to reuse large ontologies to build domain specific ontologies and KBs; collaborative construction (agreeing new pieces of knowledge with the rest of the knowledge architecture, which has been previously agreed)... However, it is not usually necessary to implement ontologies manually, as most of the available ontology tools are able to generate ontologies in many different ontology languages. Although we identified some studies that have proposed approaches trying to standardize the use of games for assessment (e.g., Said, Cheniti-Belcadhi, & El Khayat, 2019 with an ontology for personalization in serious games for assessment, or Tang & Hanneghan, 2011 with an ontology for serious game design), we did not found any study trying to standardize the GBA area using an ontology-based approach.

As a result of students' interaction with games, large repositories of data are generated. When it comes to the relationship between ontologies and this vast amount of data, data is usually stored in the computer main memory; thus, some problems exist when manipulating a large amount of ontology-based data (Dehainsala, Pierra, & Bellatreche, 2007). Trying to solve these issues, some studies have incorporated the use of big data technologies when managing ontology-based data. Some examples are the use of MongoDB (a NOSQL database) and modular ontologies (Abbes & Gargouri, 2016), or the use of Spark and Flink (two data processing frameworks) for the construction of an engine for scalable processing of large-scale RDF data (Lehmann et al., 2017). In this research, we plan to go beyond literature, trying to standardize the GBA area defining a common knowledge model. With this purpose in mind, we combine the use of ontologies and GBA metrics in order to address the interoperability issue in this particular area.

3. Methodology

In this section, we present the selection of games and metrics that have been used for validation, as well as the construction method that we have followed to build our ontology. In addition, we introduce the framework where our case study and validation has been conducted.

3.1. Serious games selection

Since we wanted to test our ontology in a real scenario, we selected a set of different SGs within different knowledge domains to test the interoperability of our approach. Field Day (Field Day Lab, 2022) is a research lab based at the Wisconsin Center for Education Research at the University of Wisconsin - Madison. Field Day designs learning games that bring contemporary research to the public, making their game data available to the public. Exploring this open game data (Gagnon & Swanson, 2023), we made a selection of SGs that we use in our study:

- *Magnet hunt*: a game where learners have to use magnetic waves to find a set of magnets hidden throughout the yard. This game addressed the topic of magnetism, a class of physical phenomena that are mediated by magnetic fields. Moreover, it also addresses other topics like forces and interactions, magnetic poles, and magnetic fields.
- *Wave combinator*: a game where students learn how waves interact using a mysterious wave combinator found in the yard. This game addresses wave properties, amplitude, offset, wavelength, frequency, and more topics.
- *Crystal cave*: a game where students learn all about crystal molecules and dig up some sweet crystals for their collection in the museum. This game addresses crystals, geometric arrangement, molecular charges, and molecular stability topics.
- *Wind simulator*: using this simulator, students learn how wind travels from high to low pressure systems, but moves in a spiral due to the coriolis effect. The game addresses concepts related to earth's systems, air masses and weather conditions, and weather patterns.
- *Antibiotic resistance*: Playing this game, learners will acquire knowledge about heredity, inheritance, variation of traits, mutation, genes, and antibiotics.
- *Earthquake*: in this game, students learn about real earthquakes, with concepts such as earth's materials and systems, scale proportion and quantity, S waves, P waves and triangulation.
- *Nitrogen cycle*: learners have to figure out how nitrogen atoms move around the world to win the game. In this process, they will learn concepts such as the nitrogen cycle, bacteria digestion, plant death, plant assimilation, or herbivorism.

Table 1
Dataset sizes.

Game	Size (MB)	# of events	# of triples
Magnet	50.8	100,000	1,788,937
Waves	63.0	100,000	1,741,566
Crystal	67.9	100,000	1,724,822
Wind	64.1	100,000	1,646,667
Bacteria	57.1	100,000	1,711,218
Earthquake	44.6	100,000	1,528,762
Nitrogen Cycle	55.2	100,000	1,708,911
Carbon Cycle	52.2	100,000	1,719,099
Shadowspect	44.0	100,000	1,682,333
Lakeland	71,3	100,000	1,821,143

- *Carbon cycle*: learners have to figure out how carbon atoms move around the world, mastering the carbon cycle in order to collect enough carbons to beat the final opponent. Learners will acquire competences related to the carbon cycle, systems and systems models, cycle of matter, and energy transfer in ecosystems.
- *Shadowspect*: a geometry game designed explicitly as a formative assessment tool to measure math content standards (e.g. visualize relationships between 2D and 3D objects). It aims to provide metrics related to geometry content and other behavioral and cognitive constructs.
- *Lakeland*: in this strategic building game, learners decide to build a new town called Lakeland, explore the dynamics of the nutrient system and recognizing the impact humans have on the world. The game addresses the Next Generation Science Standards essential practice of Modeling alongside the cross-cutting concepts of patterns, cause and effect, and systems and system models.

All these games (and their data) can be checked and played in https://fielddaylab.wisc.edu/opengamedata/. As an example, we can see two screenshots of two different SGs from this research lab in Fig. 1. Datasets are usually in TSV format, and, although each dataset has some specific columns, most of them have columns in common that are essential in game log data, such as the "session id", "game id", "timestamp", "game version", or "user id". Data format employed in these datasets has been considered as the base format that is used later as data input in our experiments.

3.2. Data collection

To test the interoperability of our approach, we used ten data sets from the SGs presented in Section 3.1. The size of each dataset is presented in Table 1. As we can see, each one of the datasets contains a total of 100,000 game events derived from real players' interaction with the different games. The number of triples in our experiments varies from 1,528,762 to 1,821,143, depending on each game.

3.3. Metric selection

In this section, we describe the process followed to select and develop the set of metrics that are used to perform GBA and test our approach's capabilities. We can divide these metrics into two different subsets: *literature* metrics, and *author proposed* metrics. First, we wanted to demonstrate that our framework can replicate any metric present in literature. With that purpose in mind, we used the selection of papers of our previous systematic review on the GBA area (Gomez et al., 2022), carefully reviewing each paper and selecting the metrics described. Since our objective was to select metrics in literature, we excluded calculations over data that included Machine Learning (ML), Deep Learning (DL) and similar models/algorithms. Second, we wanted to demonstrate the possibilities that our approach provides creating a set of metrics that go beyond the state of the art, introducing new ways to perform GBA using ontology-based data and SparQL queries. To validate that our ontology-driven metrics were correct, we implemented some of the metrics directly using the collected primary data to check that results were the same using both approaches.

(a) Magnet hunt screenshot. (b) Lakeland screenshot.

Fig. 1. Screenshots of two of the SGs selected.

3.4. Ontology development process

In general, methodologies give you set of guidelines of how you should carry out the activities identified in the ontology development process, what kinds of techniques are the most appropriate in each activity and what products each one produces (Fernández-López, Gómez-Pérez, & Juristo, 1997). When constructing an ontology, there are two different methods that we can follow, one is to build a new domain ontology directly, and the other is to expand an existing domain ontology. When choosing ontology construction methods, we should choose the most appropriate method according to the actual situation, or even integrate the advantages of various methods (Sun, Hu, Li, & Wu, 2020). To the best of our knowledge, there is not an existing ontology meeting our requirements and that could be used to expand it, so we decided to built our own ontology from scratch.

For building our ontology, we decided to use Methontology (Fernández-López et al., 1997), a structured method designed to build ontologies from scratch, reusing others as they are, or by a process of re-engineering them. Methontology was stated as the most mature approach for building ontologies, being recommended by the Foundation for Intelligent Physical Agents (FIPA) for the ontology construction task (Corcho et al., 2003). In Fig. 2 we can see the complete development process that we have followed to build and validate our ontology, which is an adaptation of the Methontology original methodology and the one proposed by Olszewska et al. (2020), which is also based on Methontology. Next, we explain each step in detail.

3.4.1. Pre-development activities

The pre-development activities include planification, the environment study, and the knowledge acquisition:

- **Planification and environment study**: this is the first phase of the process. These activities consisted in identifying the problem to be solved with the ontology, the applications where the ontology will be validated and integrated, and verifying that the ontology was possible to build, also considering the limitations of the project (Olszewska et al., 2020).
- **Knowledge acquisition**: was thought as an independent activity in the development process. However, it can be conducted simultaneously with other activities, as most of the acquisition is done with the requirements specification phase. It deals with the acquisition of knowledge from experts or other sources, that can include books, figures, tables, brainstorming techniques, or even other ontologies, among others.

3.4.2. Development states

The development states constitute the main core of the methodology, consisting of:

- **Specification**: the goal of the specification phase is to produce an informal, semi-formal, or formal ontology specification document written in natural language, including (a) the purpose of the ontology, (b) the level of formality of the implemented ontology, and (c) the scope, using a set of intermediate representations or competency questions.
- **Conceptualization**: captures the relevant domain knowledge building a conceptual model describing the problem and its solution. The core concept dictionary must meet the requirements of being unambiguous while covering the entire domain. Moreover, we must take into account the concepts' relationships and attributes. In this vein, we can build a set of intermediate representations such as a glossary of terms, a verb dictionary, or tables of rules and formulas, if needed.
- **Integration**: it explores the use of other ontologies to speed up the construction of your ontology, reusing certain terms or definitions.
- **Implementation**: in this phase, the formal models built previously are converted into a computable model. As the ontology development environment, we decided to use *Web-Protégé*, a web-based lightweight ontology editor which combines the Google Web Toolkit for the user interface, and Protégé for supporting ontology services. It is open source, and also provides collaborative features to facilitate discussions and annotations between different contributors (Tudorache, Vendetti, & Noy, 2008).
- **Validation in a real scenario**: since the main goal of our study was to perform interoperable GBAs in a real context, this phase includes the validation of the ontology using a real dataset, collected as a result of the users' interaction with different games that have been used in real life. In this state, we validate the ontology by trying to represent the information contained in the dataset and checking if the initial objectives defined are accomplished.
- **Formal evaluation**: this stage means to carry out a technical judgment of the ontology and their software environment with respect to a frame of reference. The formal evaluation includes (1) *Verification* (i.e., the technical process that guarantees the correctness of an ontology) and (2) *Validation* (i.e., guarantee that the ontology and the software environment correspond to the system that they are supposes to represent). To perform the *verification*, we used OOPS! (Ontology Pitfall Scanner!) (Poveda-Villalón, Gómez-Pérez, & Suárez-Figueroa, 2014), a tool for detecting pitfalls in ontologies, which operates independently of any ontology development platform and is available online. For example, OOPS! warns you when the domain or range of a relationship is defined as the intersection of two or more classes, or when a cycle between two classes in the hierarchy is included in the ontology, which could lead to reasoning problems.

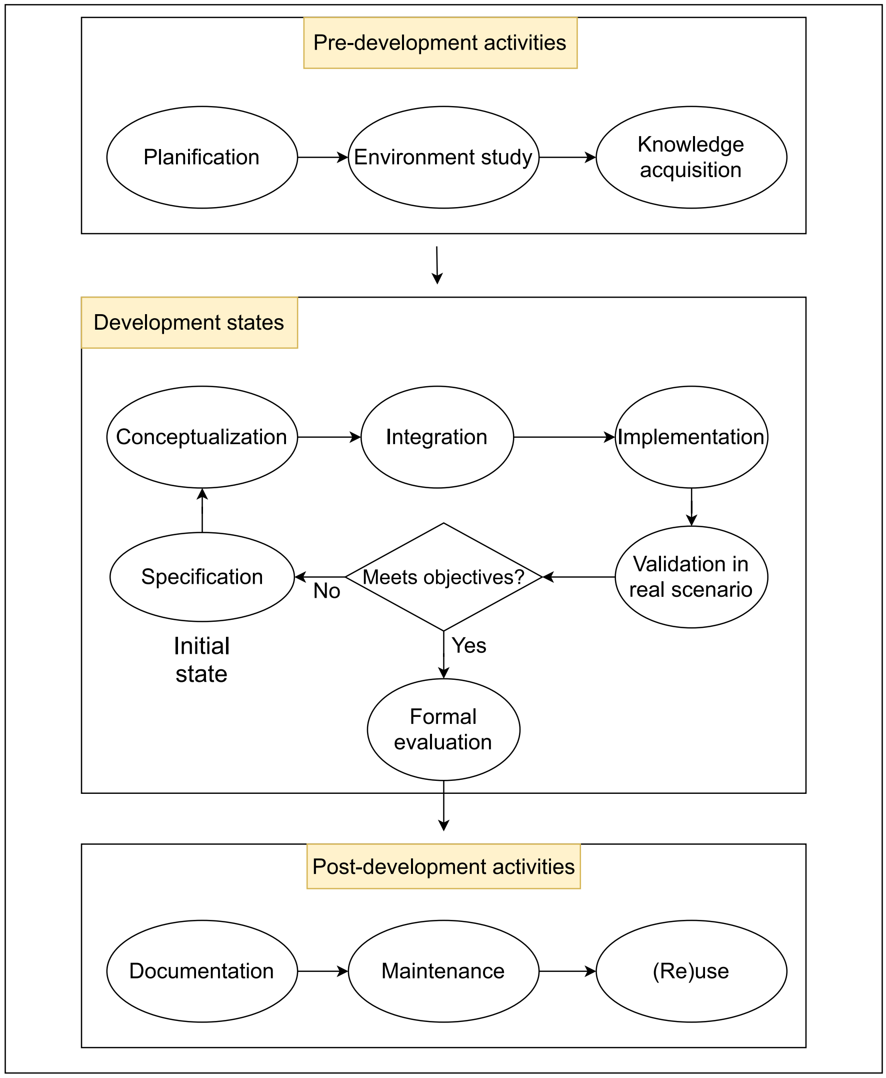


Fig. 2. Ontology development process.

Note that this is an iterative process: after the ontology is initially constructed, it can be evaluated and improved. If in the "Validation in real scenario" stage we discover any type of inconsistency or any objective that is being not satisfied, we then go again to the initial development phases to fix any problem and keep improving the ontology.

3.4.3. Post-development activities

This part of the development is done only when the previous phases have been finished. The post-development activities include:

- **Documentation**: this phase gather all the documents produced during the previous stages in order to create an appropriate ontology documentation.
- **Maintenance**: updates the ontology creating/removing concepts or relationships, allowing the ontology to evolve based on new applications that it could have.
- **(Re)use**: it considers the use of the ontology for the original purpose, but also the reuse of the developed ontology in other ontologies and/or applications.

Although these activities are only executed once, the maintenance and (re)use phases can be repeated if necessary after the ontology development has been accomplished.

3.5. Framework to support interoperable game-based assessments as a service

Once the ontology has been created and validated, we need a powerful tool capable of integrating our ontology and using ontology-based data to develop useful metrics. Consequently, we decided to use a framework developed to support interoperable Game-based Assessments as a Service (GBAaaS) (Gomez, Ruipérez-Valiente, & García Clemente, 2023). The complete framework's architecture in shown in Fig. 3.

As we can see, the framework uses standard formats for data input (e.g., CSV, TSV), transforming these data into ontology-based data in Resource Description Framework (RDF)/XML format. RDF is a general-purpose language for representing data and metadata on the web, and it is supported by its own query language SparQL, enabling the extraction and transformation of RDF data (Gandon, Bottollier, Corby, & Durville, 2007). In the next step, SANSA framework (Lehmann et al., 2017) is used as a base to process these ontology data and infer new information from it, as well as perform queries over the inferred data. All metrics designed and developed have been implemented in form of SparQL queries. Then, thanks to the metric output module, query results can be exported using several formats, such as plain text or CSV. In addition, the framework also provides a REST API module, allowing to use it as an online service. Finally, the authentication and authorization module

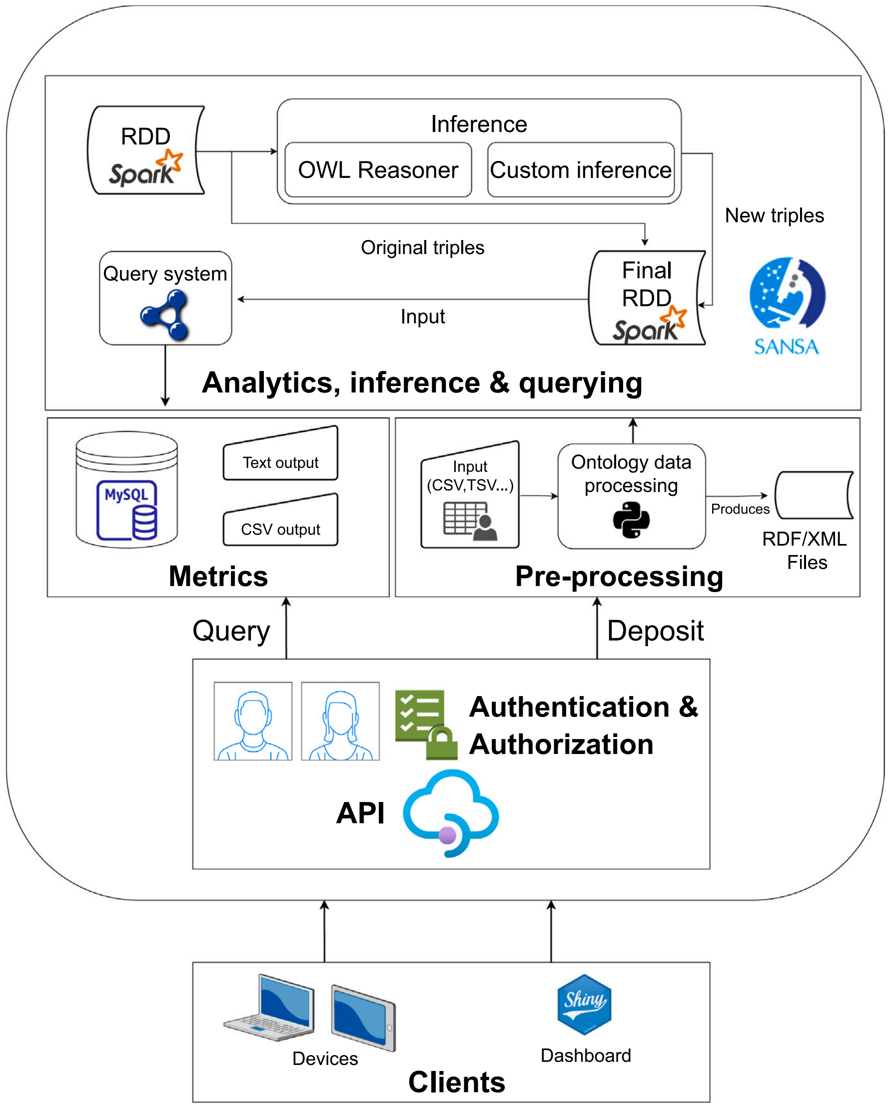


Fig. 3. GBA framework's architecture.

manage the different roles and authorizations in the system, making sure that clients making petitions through the API module are properly logged in and have the right permissions to operate. A comprehensive description of this framework and its components can be consulted in our previous work (Gomez et al., 2023).

4. Results

4.1. Proposed ontology

The ontology has been developed following the methodology described in Section 3.4. We only include the results derived from the phases "Specification", "Conceptualization" and "Implementation", since the rest of activities and phases do not have a specific output.

4.1.1. Specification

This activity produces a specification document as an output. As stated in our methodology, the specification document must address the most vital questions related to the domain we are interested in, the ontology purpose, and its scope. In addition, we have proposed a set of Competency Questions (CQs), which are the questions that are later used at the evaluation phase to assure that the ontology is appropriate for the purpose originally thought. In Table 2 we can see the specification document proposed for our ontology.

4.1.2. Conceptualization

This activity aims to capture the relevant knowledge with a set of intermediate representations. For our ontology, we built a core concept dictionary (which includes the main terms and concepts involved in the GBA domain, along with possible synonyms), a table of attributes (including the attributes of each concept), and a binary relation table (which includes the relationships between the different concepts and its cardinality). In Table 3 we can see the core concept dictionary of our ontology, including relevant terms such as "Game", "Game session", "Player" or "Learning outcome". Moreover, in Table 4 we can see two examples of the binary relation table, indicating the source and target concepts and their cardinality.

4.1.3. Implementation

In this step, we prepared and converted the GBA Ontology into machine readable format, using an ontology development editor. As we stated in the methodology, we use *Web-Protégé* and *Protégé* as ontology modeling tools. The classes constructed and their relationships are shown in Fig. 4.

As we can see, the final ontology model includes the core concepts previously identified in form of classes, as well as a set of relationships that aim to represent the links between the different classes. For example, we see that a user can have a relationship with certain user

Table 2
The ontology requirements specification document.

Specification Document	
Domain	Game-based Assessment.
Date	Nov, 9th 2021.
Conceptualized by	Research author.
Purpose	Ontology about game-based assessments to be used in different contexts and with different types of data. The ontology could be used to infer knowledge about the existing data related to users' assessment, creating new information such as the level of the player, or different play styles.
Level of formality	Semi-formal.
Scope	Users' assessment using data from serious educational games.
Competency questions	• How to assess or measure that the required learning objective has been achieved? • Which users have a specific play style (e.g., persistence)? • Which levels have been completed by a user in a game? • How much time has a user spent playing different games? • How users have interacted with different games?

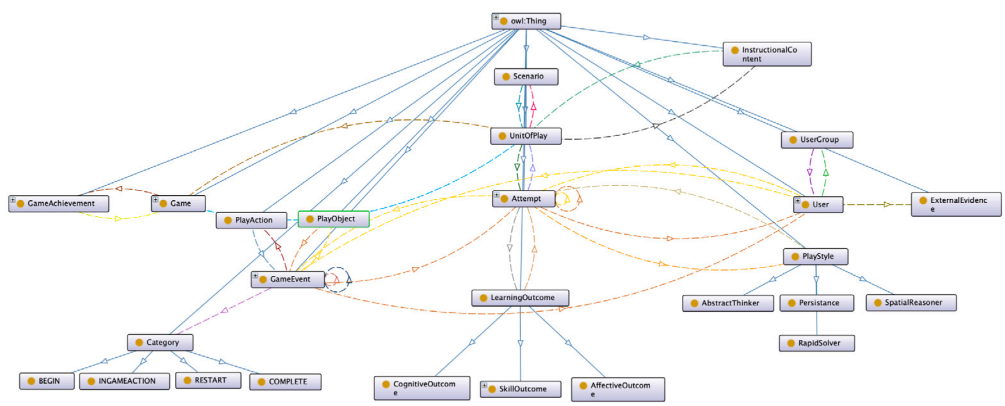


Fig. 4. GBA Ontology classes and relationships visualized via Protégé Ontograf.

Table 3
Core concept dictionary.

Core concepts and terms	
Concept	Synonym
Game	–
Game event	Assessment statement
Game achievement	–
Instructional content	–
Unit of play	Level
Scenario	Environment
Instructional content	–
Attempt	–
Learning outcome	Capability
Game session	–
User	Player
User group	Player group
Play style	Behavior
External evidence	–
Sources of knowledge	Nouira, Cheniti-Belcadhi, and Braham (2018), Plass, Homer, and Kinzer (2015), Rocha and Zucker (2015), Said et al. (2019), Tang and Hanneghan (2011), Yusoff, Crowder, Gilbert, and Wills (2009)

Table 4
Binary relation table sample.

Binary relation table	
Relation name	Has
Source concept	Game
Source cardinality	(1,n)
Target concept	Game achievement
Target cardinality	(1,n)
Relation name	Has
Source concept	Game session
Source cardinality	(1)
Target concept	Attempt
Target cardinality	(1,n)

group, but also with game sessions and play styles. These relationships contribute to improve the knowledge and they can be used to perform queries from the ontology with reasoning tasks. For example, they can be used to ask for the game events that certain user has on certain game(s). In order to view the complete ontology, the reader can create a new account on the *Web-Protégé* web page, and then use the following link: https://webprotege.stanford.edu/#projects/d0adbf6c-e703-4f81-bbfa-9fcd9a974a07/edit/Classes.

4.2. Metric ontology validation

In this Section we aimed to validate the ontology proposed by using two different groups of metrics: metrics that emerged from previous literature and newly designed metrics. Finally, we also explain how these metrics have been implemented in the framework.

4.2.1. GBA literature metrics

As we stated previously, we did a careful review of the GBA literature to search and replicate metrics that have been implemented in previous studies. After reviewing and grouping each metric, we created six groups:

- **Activity indicators**: This metric is computed for each game, group and user, and includes the total amount of time spent in the game, the total number of events, and the frequency of events (number of events/total time).
- **Persistence indicators**: This metric is computed for each game, group and user, and includes the total amount spent in units (levels), the number of units completed, and the maximum time spent in a single unit.
- **Event types**: This metric is computed for each user and game, and includes the number of events of each user grouped by event type (e.g., "Complete", "Retry", "Interaction"). In addition, this group also includes the interaction level, which is defined as interaction events divided by the sum of the rest of events.
- **User performance**: This metric is computed for each game, group and user, and includes the percentage of success (which is defined as the number of units completed divided by the number of units started), and the maximum unit reached by the player.
- **Levels of activity**: This metric is computed for each game, group and user, and include straightforward metrics to compute based on a feature engineering process, such as the active time, inactive time, number of events, and the number of different types of events.
- **Funnel by user**: This metric is computed for each game, group and user, including the percentage of units that the user has started, the percentage of units that the user has interacted with, and finally the percentage of units that the user has completed. This funnel seeks to provide a quick overview of the current status and progress for each user and game.

These metrics have been implemented in our framework in form of SparQL queries, and they will be used to illustrate how we can use this approach in a real scenario. In addition, the granularity of these metrics can be changed very quickly, being able to aggregate results by whole groups or even the whole game.

4.2.2. Additional metric proposal

Moreover, going beyond this type of operations over data, we designed and implemented a set of metrics that require of more complex calculations, such as standardization, normalization, or more complex methods (e.g., machine learning techniques). Next, we explain each metric in detail:

- **Levels of difficulty**: This metric is computed for each game and unit of play (level), and provides a set of parameters that are related to the difficulty of the different units (Ruipérez-Valiente et al., 2021), namely: *completed_time*, which is computed by dividing the amount of time invested in the game by the number of completed units; *actions_completed*, which is computed by dividing the number of actions by the number of completed units; and *p_abandoned*, computed by dividing the number of started units by the number of completed units. Then, a standardized and normalized measure of the three previous parameters together in a single value is computed, representing the difficulty score of each unit.
- **Persistence**: This metric is computed for each game, user, and unit of play (level). Although there are few indicators on how to calculate persistence, it can be observed that time, both for completed and uncompleted activities, and the number of attempts are essential characteristics for persistence. Following some related works, it can be seen that the rest of the parameters are more linked to the specifications of each scenario where it has been implemented (Valiente, 2022). In our specific metric, to see if a user has been persistent or not, we consider several metrics: if the unit has been completed or not, active time spent, number of events triggered, and number of attempts. Accordingly, we consider percentiles of each of the parameters considered (time, attempts, events), supposing that the user was persistent if the respective value exceeds the value of 75%. Lastly, for each user, we identify the units in which he has been persistent, and we calculate if the user globally has been persistent or not according to the number of units in which he has been persistent.
- **Play styles**: This metric is computed for each game and user. To identify different play styles that users can have when playing, we perform clustering by using k-means algorithm, which commonly uses a set of continuous variables as input. We use as input the following indicators: total active time, different days played, number of different events triggered, number of interaction events, and completed units. Based on these indicators, we can obtain higher level profiles so that we can analyze each cluster separately and determine different play styles.

For these metrics, the initial calculations required have also been implemented in form of SparQL queries. However, other advanced techniques (such as ML) have been included in the system by developing separated scripts that use the SparQL queries results to perform complex calculations. Moreover, the granularity in these metrics can also be changed easily so that metric results are displayed not only by individual users but also by group or game.

4.2.3. Metric implementation

To implement the metrics presented, we used the framework described in Section 3.5. Thanks to its capabilities, we have integrated the ontology developed along with the different metrics. Once the log data from learners' interaction with SGs has been transformed into ontology-based data, the system uses SparQL (the standard language for querying RDF data) queries to gather information from these data. Therefore, we implemented all the basic metrics in form of queries, including metrics emerged from literature as well as the first stages of the additional metrics proposed. Regarding the additional metrics, more complex calculations were executed using specific scripts integrated into the framework, which also used the results from the SparQL queries. Next, as an example, we show the query implemented for the "Activity indicators" metric:

```
SELECT ?game ?group ?user
  ?totalTime ?totalGameEvents
  ((?totalGameEvents/?totalTime) as ?freqEvents)
 WHERE{
  ?user rdf:type m:User.
  {SELECT ?group ?user ?game
    (SUM(?individualTime) as ?totalTime)
    (COUNT(DISTINCT(?ev)) as ?totalGameEvents)
   WHERE{
    ?user rdf:type m:User.
    ?user m:has ?attempt.
    ?user m:hasGroup ?group.
    ?attempt rdf:type m:Attempt.
    ?attempt m:playedInUnit ?unit.
    ?game m:hasUnitOfPlay ?unit.
    ?attempt m:has ?ev.
    ?ev rdf:type m:GameEvent.
    ?ev m:timeBetweenEvents ?individualTime.
   }
   GROUP BY ?group ?user ?game
  }
}
```

As stated in the metric definition, it calculates the total amount of time spent in the game, the total number of events, and the frequency of events of each learner separately. As we can see, the total time is calculated by adding the number of seconds that each event lasts, and this number is aggregated by group, game and user. Moreover, we also see that the frequency is calculated dividing the total number of events by the total time (in seconds) once the results have been aggregated by learner.

4.3. Case study

In this section, using our experiment results, we present a case study exemplifying how our ontology-based approach and the computed metrics could be used in a real life environment. To test the interoperability and usability of our work, we followed previous related work (Díaz et al., 2019; Jayapandian, Zhao, Ewing, Zhang, & Sahoo, 2012; Santos, Dantas, Furtado, Pinheiro, & McGuinness, 2017) and conducted a similar case study.

4.3.1. Dashboard overview

In this use case, we present a visualization dashboard system that uses the data analyzed and transformed into metrics, being consumed via visualizations. This enables instructors to monitor what learners are doing while playing, use these data to adapt their interventions when necessary, or even use these metrics as a part of a formative evaluation. Moreover, this dashboard also allows learners to track their own activity within the games. We have developed the dashboard using Shiny's R framework, and we have deployed it on ShinyApps web server. In our implementation, we have two types of users: on the one side, we have instructors (or teachers) that can use the dashboard to visualize what their learners are doing. Therefore, instructors are able to insert new GBA data, but also to query metrics from games and groups where they are participating. On the other side, we have learners, which are only allowed to query their own metric results. This way, we restrict the access to different groups and games data to ensure the privacy of each user.

Fig. 5 shows the dashboard running on the ShinyApps server. The login page is shown in Fig. 5(a), where the user can log into the system by using a username and a password. Depending on the credentials used, each person has access to different features and different data sets within the dashboard, as each user created in the platform has its own roles and permissions. After login credentials have been initially verified, the user gets access to different functionalities: in Fig. 5(b) we can see the file upload page, where authorized users can upload new GBA assessment data to be processed by the framework and incorporated in form of new metrics data. Moreover, the user can choose between different tabs available in the sidebar, either to upload new data if the current user's role permits it, or to query metric results and see them graphically via visualizations.

Furthermore, Fig. 6 shows two illustrations of how metrics are represented in the system. In our dashboard, we have implemented both group-oriented and individual-oriented visualizations, depending on the granularity of each metric. The dashboard takes advantage of the complete interoperability between games and metrics, as the user can manipulate the selection boxes to filter by different games and groups (as shown in Fig. 6(a)), and also by user if the metric allows it (as shown in Fig. 6(b)). That way, when a game is selected among the available options, the system shows the existing groups for that specific game in the corresponding selection box; once all the selection boxes for that metric tab are filled with a choice, the system queries the necessary information and represents it using interactive visualizations. Next, we present a use case using these visualizations with real data.

4.3.2. Group and student analysis

This use case exemplifies how an instructor can use the dashboard to analyze the global group status, but also to monitor individual learners. This can be very useful to track issues related to the whole group and be aware of the learners' current progress. For this use case, we analyze some of the metrics that have been defined with specific examples, using real data from the game and group that we have selected for the analysis.

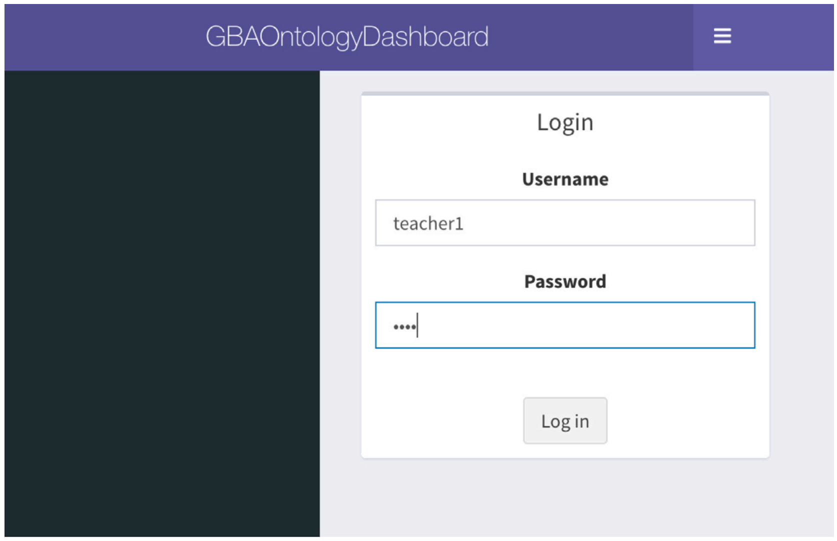
In Fig. 7(a) we see the Funnel by user metric visualization, which is based on the metric we defined previously. In this concrete use case, we selected the group "MainGroup" from the game "CRYSTAL". We see that there are 26 learners, with a funnel corresponding to each learner, showing the percentage of units that have been started, interacted, and completed. For example, we can focus on the user with identifier "210604085851886" (third funnel in the first row), which is a learner with a good performance in the game. This learner has started and interacted 100% percent of the units, and completed correctly 89% of them. Then, the instructor could use the "Persistence" visualization (Fig. 7(b)) to have a more detailed perspective of how each learner has interacted with the different units. In this visualization, a pie chart for each user is shown, indicating the percentage of units in which the user has shown different types of behaviors, such as "productive persistence" or "unproductive persistence". Focusing on the same user as before, we see that has shown "no behavior" in 62.5% of the units, "productive persistence" in 25% of the units, and that has been "non persistent" in 12.5% of the units.

Then, the instructor might want to know more about the activity that the learner has had in each unit. With this purpose in mind, an instructor could use the "Levels of activity" visualization shown in Fig. 8(b), adding a higher level of detail regarding learner's interaction. In this visualization, for a given user, we can see the active time, number of events, and number of different events in each unit. For the selected user, we can see that most interacted units have been "CRYSTAL-7" and "CRYSTAL-8", with an active time higher than 200 s, and a number of events of 115 and 66, respectively. Then, to see which have been the more complicated units for the group, the instructor can take a look at the "Levels of difficulty" visualization (Fig. 8(a), in which we can see the parameters defined, and the final difficulty measure calculated for each unit. As we can observe, the most difficult units for these group have been "CRYSTAL-7", and "CRYSTAL-8", which perfectly matches with the interaction patterns previously seen within the selected learner's data.

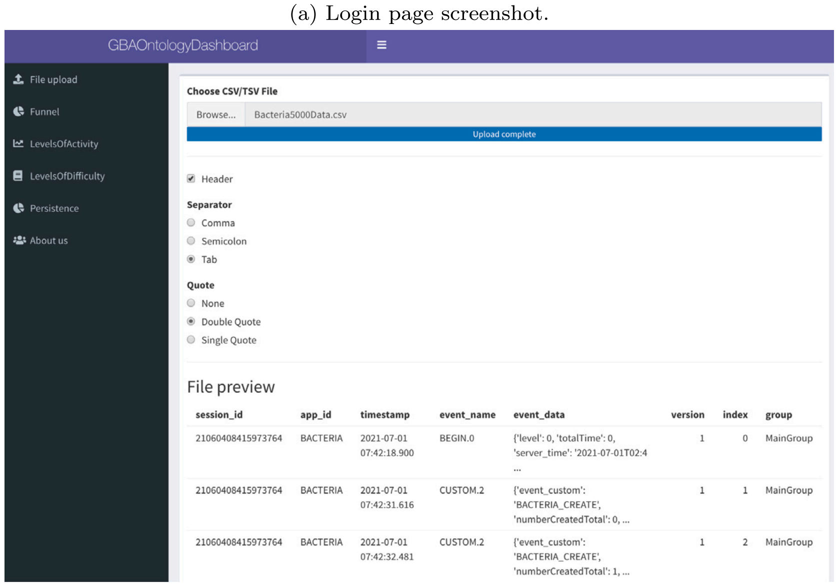
5. Discussion

Although SGs are being considered as useful tools to perform complex and reliable assessments in broad domains (Kato & de Klerk, 2017; Sliney & Murphy, 2011), the implementation of GBAs features is seen as a very time consuming step (Ifenthaler, Eseryel, & Ge, 2012). This is due to heterogeneity issues, since assessment machinery is usually designed specifically in each different game and context. Previous work has addressed this problem by proposing standard models, such as xAPI, comprising the specificities of analytics in games (Alonso-Fernandez et al., 2017; Perez-Colado et al., 2018). In this research, we try to address the interoperability issue by providing a higher level interoperable approach to perform GBAs. We designed and implemented a new ontology that serves as a common knowledge model, being able to integrate log events from any game into a unified data model. This implies the standardization of a wide area, with games designed with different purposes, based on different knowledge areas, and targeting participants with different characteristics.

Comparing our approach with previous standardized data format approaches, which benefits can bring the use of an ontology to this area? First, ontologies provide an organization and reuse of knowledge, allowing to disambiguate or uniquely identify the meaning of concepts in a given domain (Bürger & Simperl, 2008). Second, ontologies allow



(a) Login page screenshot.



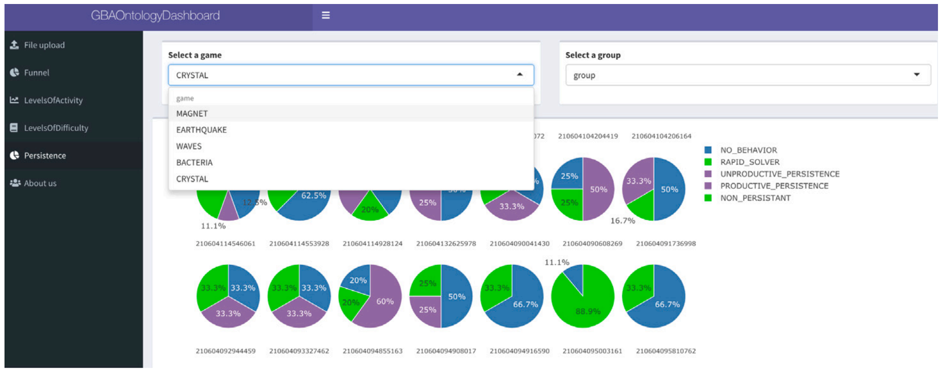
(b) File upload tab screenshot.

Fig. 5. Screenshots of the dashboard developed.

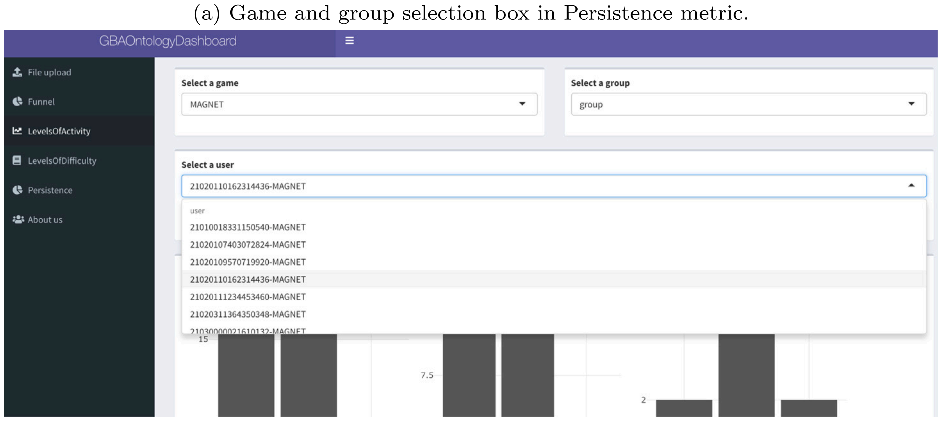
to take advantage of the richness and complexity of relationships between concepts and entities within a domain. By representing these relationships explicitly, ontologies can provide a better understanding of the GBA area and enable a more sophisticated reasoning. This knowledge representation also provides computational inference, which can help to spot logical inconsistencies to indicate modeling errors. Third, it is reasonable to expect a performance gain in precision and recall when using ontology-based approaches compared with the data mining approaches (Dou, Wang, & Liu, 2015). Furthermore, ontology-based approaches can be combined with Big Data technologies in order to obtain great performance results, as shown in Lehmann et al. (2017). In fact, using the framework mentioned in Section 3.5, we were able to compute data from approximately 39 full classrooms during an entire month in 107.2 min (Gomez et al., 2023).

Alonso-Fernandez et al. (2019) conducted a review regarding applications of data science to game learning analytics data, and noted that most papers did not report the format in which they collected the data, so it is unknown if they were using a standard or relying on their own data formats, which leads to reproducibility and reusability problems. Our ontology works regardless of the data format, which can be easily adapted to be incorporated in the model and used for further processing. This enables an easy way to use any type of GBA data and make straightforward assessments by simply adapting the data to our model. In addition, authors also noted that sample sizes used in the studies are, in general, quite low, presenting low statistical power and having a reduced chance of detecting actual effects (Petri & von Wangenheim, 2017). Thus, it is important that future research used larger data samples, in order to improve the results' generalization, and also to enable the use of more complex techniques that usually require a big amount of data. Our work also enables the processing of large data samples, since it has been integrated in a framework that uses Big Data technologies (specifically Spark) to process the ontology-based data.

The availability of real-time information about the learners' interaction and behaviors provides a great opportunity to analyze these data during gameplay. The analysis of those actions and the investigation of more complex series of actions and behaviors can provide key insights into ongoing learning processes in these environments (Kim & Ifenthaler, 2019). GBAs aim to convert learner-generated information into actionable insights, including learners' individual characteristics (e.g., interests, prior knowledge, skills) and learner-generated game data (e.g., time spent, goals or tasks completed). However, these analyses are usually quite simple: we conducted a review on literature to



(a) Game and group selection box in Persistence metric.



(b) Game, group and user selection box in Levels of Activity metric.

Fig. 6. Selection options in the dashboard.

collect which metrics have been used in previous studies, concluding that most of them only use basic ones (e.g., completion times, count of events, general scores), which implies simple calculations, such as additions or averages. While all these metrics previously developed have been implemented in our research, we have also developed novel metrics using more complex calculations, such as normalization, standardization, and ML algorithms, demonstrating that our ontology can replicate previous literature as well as using new approaches to perform interoperable GBAs.

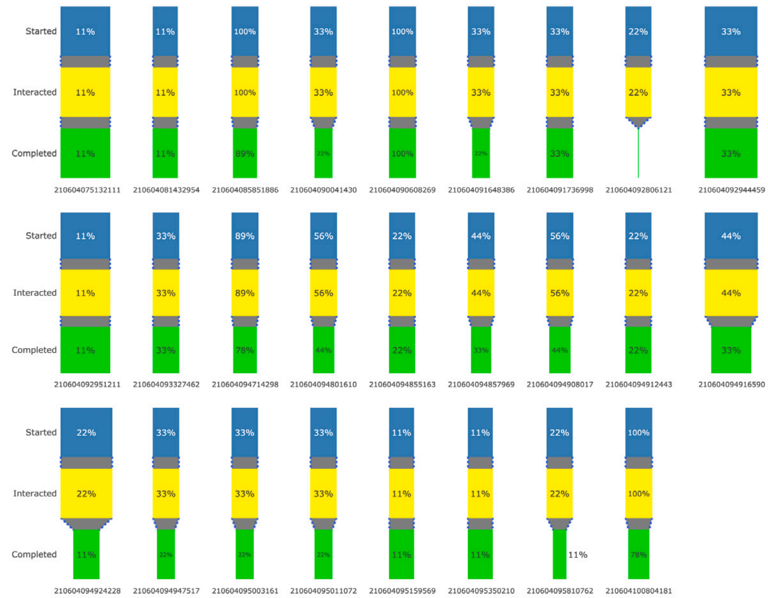
To see how our environment works in a real context, we collected data from 10 different SGs and used the ontology and metrics to perform GBA and conduct a case study. In many fields, dashboards are used as a tool to inform and transmit knowledge, and their importance and usefulness make them the subject of many studies (Ruipérez-Valiente et al., 2021). Although some studies in the field have used visualizations and deployed dashboards as effective tools to represent GBA data (Gomez, Ruipérez-Valiente, Martinez, & Kim, 2020; Kim, Lin, & Ruipérez-Valiente, 2021), our dashboard designed for the case study is the first one in the area supporting interoperable GBAs, since it covers data from different SGs at the same time by simply using the ontology developed.

This work also has also some limitations: first, data has to be incorporated into the ontology providing the data file manually, and although we have defined a data format with columns that almost any log data from the area should have, probably some adaptations to the original data would be necessary in order to meet the input's requirements. Moreover, this approach supports ML techniques, but it does not support more complex methods, such as Knowledge Inference (KI) or Deep Learning (DL). The use of these methods could help to infer more useful information from learners' data, as well as improving the results' validity and reliability.

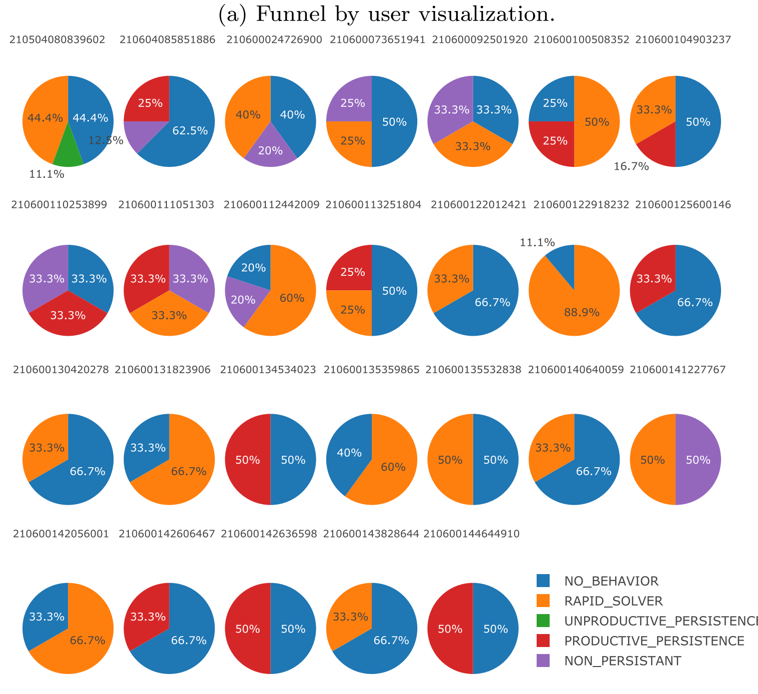
We can see the great potential that GBA have applied in many contexts. Regarding professional environments, companies have begun to use GBA for employee recruitment and selection (Bina, Mullins, & Petter, 2021). In healthcare, games are also being used for assessment, training and rehabilitation (Ferreira-Brito et al., 2019). Moreover, GBA can also be used to measure psychological well-being, by analyzing learners' anxiety, for example (Smits & Charlier, 2011). Despite all these opportunities, the current challenge is (still) to make use of data from learners, teachers, and game learning environments for assessments. Therefore, we firmly believe that the future of games for assessment is promising, and our work can help to alleviate some of the challenges in the area by providing a common knowledge model and provide straightforward interoperable assessments.

6. Conclusions

This research aimed to develop a novel approach to achieve interoperable GBAs using ontologies and in-game metrics that are automatically computed using ontology-based data. With that purpose in mind, we established three objectives: (1) to design and develop an ontology to standardize the GBA area, (2) to conduct a validation study on previous metrics in literature, as well as to design and implement novel metrics, and 3) to conduct a case study illustrating how our approach can be used in a real environment. First, we designed an ontology from scratch using Methontology (Fernández-López et al., 1997) as a base in our methodology, and we conducted a formal evaluation using OOPS! (Poveda-Villalón et al., 2014) to detect possible issues and iterate over the methodology to solve them if necessary. Then, we conducted an study on previous GBA literature, carefully reviewing each paper and noting the metrics developed so that we could replicate them later in our environment. In addition to those metrics, we also designed novel interoperable metrics (including more complex calculations and ML algorithms) to demonstrate and validate the capabilities of our work. Finally, we conducted a case study performing GBA with data from 10 different SGs. Benefiting from the capacities of the framework



(a) Funnel by user visualization.

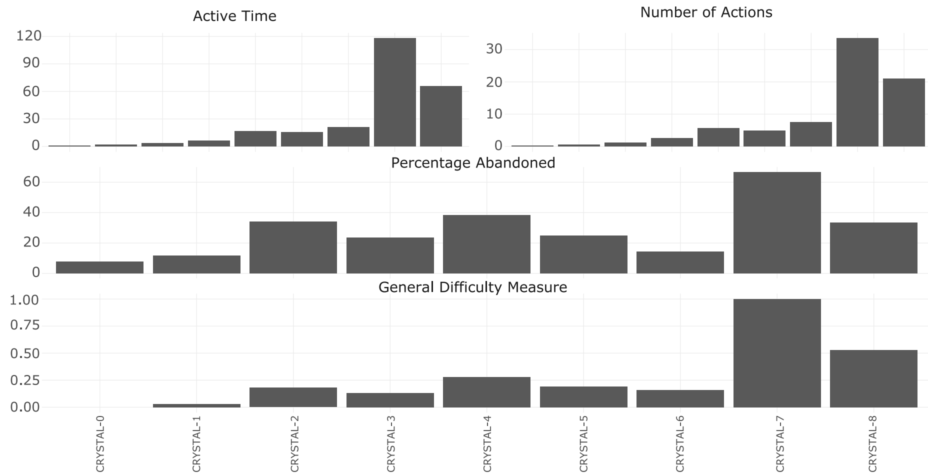


(b) Persistence visualization.

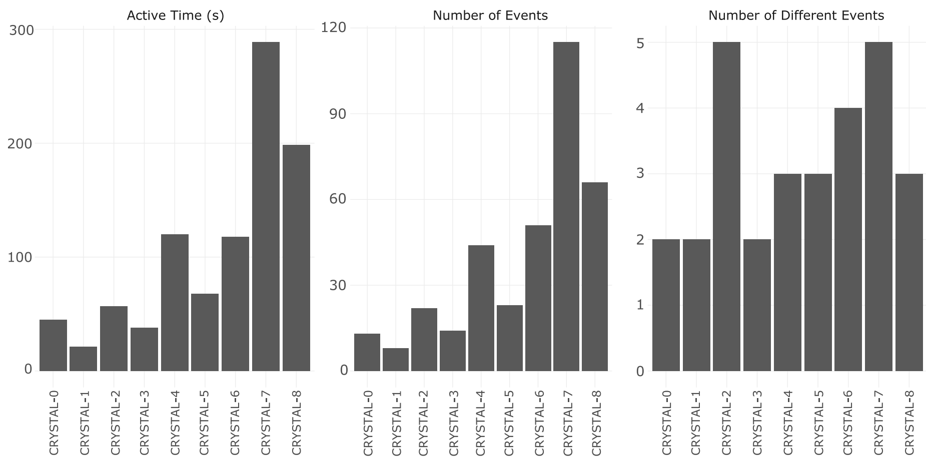
Fig. 7. Funnel by user and Persistence visualizations for the selected game and group.

used to integrate our ontology (Gomez et al., 2023), a dashboard was developed using Shiny's R framework and deployed via ShinyApps server to test our approach with real data.

As part of our future work, we would like to validate our approach by conducting case studies (using the framework showed in this research) and collecting data in real time from learners and instructors. Moreover, we will be developing new metrics to continue expanding the system and its possibilities. Future studies should also consider integrating more complex algorithms (KI,DL) that are quite important within the GBA machinery literature in form of metrics that could unveil the full potential that interaction data from learners have in assessment. Finally, novel uses of our ontology and metrics, such as reports generation, will be explored in order to perform new case studies, giving our research even more practical application. This work provides significant contributions to the literature, including a new ontology designed as a common knowledge model to unify the GBA area, aiming to provide interoperable GBAs using in-game metrics to monitor learners' interaction with games and provide useful insights. In addition, since we developed the metrics in a modular way, integrating them into a powerful ontology-based framework, our work can be easily expanded with new metrics using novel approaches, such as ML techniques. We expect the use of our work (including the ontology and GBA metrics) along with its integration in the framework to solve the current limitations regarding GBA interoperability, reducing the cost and effort of developing specific GBAs, and therefore allowing the deployment of GBAaaS.



(a) Levels of difficulty visualization for the selected game and group.



(b) Levels of activity visualization for the selected game, group, and user.

Fig. 8. Levels of difficulty and Levels of activity visualizations.

CRediT authorship contribution statement

Manuel J. Gomez: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Visualization. **José A. Ruipérez-Valiente:** Conceptualization, Writing – review & editing, Supervision, Project administration. **Félix J. García Clemente:** Conceptualization, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was partially supported by (a) grant 21795/FPI/22 - Séneca Foundation. Cofinanced by Innovatiio Global Educación. Region of Murcia (Spain), (b) REASSESS project (grant 21948/JLI/22), funded by the Call for Projects to Generate New Scientific Leadership, included in the Regional Program for the Promotion of Scientific and Technical Excellence Research (2022 Action Plan) of the Seneca Foundation, Science and Technology Agency of the Region of Murcia, and (c) the strategic project CDL-TALENTUM from the Spanish National Institute of Cybersecurity (INCIBE) and by the Recovery, Transformation and Resilience Plan, Next Generation EU. In addition, we would like to express our gratitude to Field Day Lab for sharing their datasets publicly, which have been used in our experiments.

References

Abbes, H., & Gargouri, F. (2016). Big data integration: A MongoDB database and modular ontologies based approach. *Procedia Computer Science*, *96*, 446–455.

Abt, C. C. (1987). *Serious games*. University press of America.

All, A., Castellar, E. P. N., & Van Looy, J. (2014). Measuring effectiveness in digital game-based learning: A methodological review. *International Journal of Serious Games*, *1*(2).

Alonso-Fernandez, C., Calvo, A., Freire, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. (2017). Systematizing game learning analytics for serious games. In *2017 IEEE global engineering education conference* (pp. 1111–1118). IEEE.

Alonso-Fernandez, C., Calvo-Morata, A., Freire, M., Martinez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, *141*, Article 103612.

Alonso-Fernández, C., Cano, A. R., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Lessons learned applying learning analytics to assess serious games. *Computers in Human Behavior*, *99*, 301–309.

Basu, S., Disalvo, B., Rutstein, D., Xu, Y., Roschelle, J., & Holbert, N. (2020). The role of evidence centered design and participatory design in a playful assessment for computational thinking about data. In *Annual conference on innovation and technology in computer science education* (pp. 985–991). http://dx.doi.org/10.1145/3328778.3366881.

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: an overview. *Advances in Human-Computer Interaction*, *2013*.

Bina, S., Mullins, J., & Petter, S. (2021). Examining game-based approaches in human resources recruitment and selection: A literature review and research agenda. In *Proceedings of the 54th hawaii international conference on system sciences* (p. 1325).

Bürger, T., & Simperl, E. (2008). Measuring the benefits of ontologies. In *On the move to meaningful internet systems: OTM 2008 workshops: oTM confederated international workshops and posters, ADI, aWeSoMe, COMBEK, EI2N, IWSSA, MONET, onToContent+ QSI, ORM, perSys, RDDS, SEMELS, and SWWS 2008, Monterrey, Mexico, November 9-14, 2008. proceedings* (pp. 584–594). Springer.

Card, M. (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann.

Chiu, F.-Y., & Hsieh, M.-L. (2017). Role-playing game based assessment to fractional concept in second grade mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(4), 1075–1083. http://dx.doi.org/10.12973/eurasia.2017.00659a.

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point? *Data & Knowledge Engineering*, *46*(1), 41–64.

Daniel, M., & Garry, C. (2018). *Video games as culture: considering the role and importance of video games in contemporary society*. Routledge.

De Freitas, S. (2006). Learning in immersive worlds: A review of game-based learning. *JISC ELearning Innov*.

de Klerk, S., & Kato, P. M. (2017). The future value of serious games for assessment: Where do we go now? *Journal of Applied Testing Technology*, *18*(S1), 32–37.

Dehainsala, H., Pierra, G., & Bellatreche, L. (2007). Ontodb: An ontology-based database for data intensive applications. In *International conference on database systems for advanced applications* (pp. 497–508). Springer.

Díaz, A. R., Benito-Santos, A., Dorn, A., Abgaz, Y., Wandl-Vogt, E., & Therón, R. (2019). Intuitive ontology-based SPARQL queries for RDF data exploration. *IEEE Access*, *7*, 156272–156286.

DiCerbo, K. E. (2014). Game-based assessment of persistence. *Journal of Educational Technology & Society*, *17*(1), 17–28.

Donovan, L., & Lead, P. (2012). The use of serious games in the corporate sector. In *A state of the art report*. Learnovate Centre (December 2012.

Dou, D., Wang, H., & Liu, H. (2015). Semantic data mining: A survey of ontology-based approaches. In *Proceedings of the 2015 IEEE 9th international conference on semantic computing* (pp. 244–251). IEEE.

Duval, E. (2011). Attention please! learning analytics for visualization and recommendation. In *Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 9–17).

ESA (2021). *2021 Essential facts about the computer and video game industry*: *Technical report*, Entertainment Software Association.

Fathy, N., Gad, W., & Badr, N. (2019). A unified access to heterogeneous big data through ontology-based semantic integration. In *2019 ninth international conference on intelligent computing and information systems* (pp. 387–392). IEEE.

Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Methontology: from ontological art towards ontological engineering. In *Proceedings of the AAAI97 spring symposium* (pp. 33–40).

Ferreira-Brito, F., Fialho, M., Virgolino, A., Neves, I., Miranda, A. C., Sousa-Santos, N., et al. (2019). Game-based interventions for neuropsychological assessment, training and rehabilitation: Which game-elements to use? A systematic review. *Journal of Biomedical Informatics*, *98*, Article 103287.

Field Day Lab (2022). We're field day. URL https://fielddaylab.wisc.edu/about/. (Last Accessed on 29 April 2022).

Gagnon, D. J., & Swanson, L. (2023). Open game data: A technical infrastructure for open science with educational games. In *Joint international conference on serious games* (pp. 3–19). Springer.

Gandon, F., Bottollier, V., Corby, O., & Durville, P. (2007). RDF/XML source declaration.

Georgiou, K., Gouras, A., & Nikolaou, I. (2019). Gamification in employee selection: The development of a gamified assessment. *International Journal of Selection and Assessment*, *27*(2), 91–103.

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, *29*(3), 207–219.

Gomez, M. J., Ruipérez-Valiente, J. A., & Clemente, F. J. G. (2022). A systematic literature review of game-based assessment studies: Trends and challenges. *IEEE Transactions on Learning Technologies*.

Gomez, M. J., Ruipérez-Valiente, J. A., & García Clemente, F. J. (2023). A framework to support interoperable game-based assessments as a service (GBAaaS): Design, development, and use cases. *Software - Practice and Experience*, *53*(11), 2222–2240.

Gomez, M. J., Ruipérez-Valiente, J. A., Martinez, P. A., & Kim, Y. J. (2020). Exploring the affordances of sequence mining in educational games. In *Eighth international conference on technological ecosystems for enhancing multiculturality* (pp. 648–654).

Gris, G., & Bengtson, C. (2021). Assessment measures in game-based learning research: a systematic review. *International Journal of Serious Games*, *8*(1), 3–26.

Gros, B. (2007). Digital games in education: The design of games-based learning environments. *Journal of Research on Technology in Education*, *40*(1), 23–38.

Hamdaoui, N., Khalidi Idrissi, M., & Bennani, S. (2016). Adaptive educational games using game metrics. In *International afro-European conference for industrial advancement* (pp. 198–208). Springer.

Happel, H.-J., & Seedorf, S. (2006). Applications of ontologies in software engineering. In *Proc. of workshop on sematic web enabled software engineering"(SWESE) on the ISWC* (pp. 5–9). Citeseer.

Hinton, S. (2016). Applying gamification in New Zealand contact centers. *Special Interest Group on Human-Computer Interaction*.

Ifenthaler, D., Eseryel, D., & Ge, X. (2012). Assessment for game-based learning. In *Assessment in game-based learning* (pp. 1–8). Springer.

Jayapandian, C. P., Zhao, M., Ewing, R. M., Zhang, G.-Q., & Sahoo, S. S. (2012). A semantic proteomics dashboard (SemPoD) for data management in translational research. *BMC Systems Biology*, *6*(3), 1–13.

Kangas, M., Koskinen, A., & Krokfors, L. (2017). A qualitative literature review of educational games in the classroom: the teacher's pedagogical activities. *Teachers and Teaching*, *23*(4), 451–470.

Kato, P. M., & de Klerk, S. (2017). Serious games for assessment: Welcome to the jungle. *Journal of Applied Testing Technology*, *18*(S1), 1–6.

Kim, Y. J., & Ifenthaler, D. (2019). Game-based assessment: The past ten years and moving forward. In *Game-based assessment revisited* (pp. 3–11). Springer.

Kim, Y. J., Lin, G., & Ruipérez-Valiente, J. A. (2021). Expanding teacher assessment literacy with the use of data visualizations in game-based assessment. In *Visualizations and dashboards for learning analytics* (pp. 399–419). Springer.

Laamarti, F., Eid, M., & El Saddik, A. (2014). An overview of serious games. *International Journal of Computer Games Technology*, *2014*.

Larson, K. (2020). Serious games and gamification in the corporate training environment: A literature review. *TechTrends*, *64*(2), 319–328.

Lee, J., Luchini, K., Michael, B., Norris, C., & Soloway, E. (2004). More than just fun and games: Assessing the value of educational video games in the classroom. In *CHI'04 extended abstracts on human factors in computing systems* (pp. 1375–1378).

Lehmann, J., Sejdiu, G., Bühmann, L., Westphal, P., Stadler, C., Ermilov, I., et al. (2017). Distributed semantic analytics using the SANSA stack. In *International semantic web conference* (pp. 147–155). Springer.

Liu, M., Kang, J., Liu, S., Zou, W., & Hodson, J. (2017). Learning analytics as an assessment tool in serious games: A review of literature. *Serious Games and Edutainment Applications*, 537–563.

Loh, C. S., & Sheng, Y. (2015). Measuring expert performance for serious games analytics: From data to insights. In *Serious games analytics* (pp. 101–134). Springer.

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious games analytics*: *vol. 10*, (pp. 3–29). Cham: Springer International Publishing.

Martınez, P. A., Gómez, M. J., Ruipérez-Valiente, J. A., Pérez, G. M., & Kim, Y. J. (2020). Visualizing educational game data: A case study of visualizations to support teachers.

McDaniel, M., & Storey, V. C. (2019). Evaluating domain ontologies: clarification, classification, and challenges. *ACM Computing Surveys*, *52*(4), 1–44.

Mollick, E., & Werbach, K. (2015). Gamification and the enterprise. *The Gameful World: Approaches, Issues, Applications*, *439*.

Nikolaou, I., Georgiou, K., & Kotsasarlidou, V. (2019). Exploring the relationship of a gamified assessment with performance. *Spanish Journal of Psychology*, http://dx.doi.org/10.1017/sjp.2019.5.

Nouira, A., Cheniti-Belcadhi, L., & Braham, R. (2018). An enhanced xAPI data model supporting assessment analytics. *Procedia Computer Science*, *126*, 566–575.

Olszewska, J. I., Houghtaling, M., Goncalves, P. J., Fabiano, N., Haidegger, T., Carbonera, J. L., et al. (2020). Robotic standard development life cycle in action. *Journal of Intelligent and Robotic Systems*, *98*(1), 119–131.

Owen, V. E., & Baker, R. S. (2019). Learning analytics for games. *Handbook of game-based learning*, 513–535.

Panov, P., Džeroski, S., & Soldatova, L. (2008). OntoDM: An ontology of data mining. In *2008 IEEE international conference on data mining workshops* (pp. 752–760). IEEE.

Perez-Colado, V. M., Rotaru, D. C., Freire, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. (2018). Learning analytics for location-based serious games. In *2018 IEEE global engineering education conference* (pp. 1192–1200). IEEE.

Petri, G., & von Wangenheim, C. G. (2017). How games for computing education are evaluated? A systematic literature review. *Computers & Education*, *107*, 68–90.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*(4), 258–283.

Plass, J. L., Homer, B. D., Kinzer, C. K., Chang, Y. K., Frye, J., Kaczetow, W., et al. (2013). Metrics in simulations and games for learning. In *Game analytics* (pp. 697–729). Springer.

Podgorelec, V., & Kuhar, S. (2011). Taking advantage of education data: Advanced data analysis and reporting in virtual learning environments. *Elektronika ir Elektrotechnika*, *114*(8), 111–116.

Poveda-Villalón, M., Gómez-Pérez, A., & Suárez-Figueroa, M. C. (2014). Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *10*(2), 7–34.

al Qallawi, S., & Raghavan, M. (2022). A review of online reactions to game-based assessment mobile applications. *International Journal of Selection and Assessment*, *30*(1), 14–26.

Qian, M., & Clark, K. R. (2016). Game-based learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, *63*, 50–58.

Reyes-Chua, E., & Lidawan, M. W. (2019). Games as effective language classroom strategies: a perspective from english major students. *European Journal of Foreign Language Teaching*.

Rocha, O. R., & Zucker, C. F. (2015). Ludo: an ontology to create linked data driven serious games. In *ISWC 2015-workshop on lINKed eDucation, LINKED 2015, at Bethlehem, Pennsylvania, United states*.

Ruiperez-Valiente, J. A., Gaydos, M., Rosenheck, L., Kim, Y. J., & Klopfer, E. (2020). Patterns of engagement in an educational massively multiplayer online game: A multidimensional view. *IEEE Transactions on Learning Technologies*, *13*(4), 648–661.

Ruipérez-Valiente, J. A., Gomez, M. J., Martínez, P. A., & Kim, Y. J. (2021). Ideating and developing a visualization dashboard to support teachers using educational games in the classroom. *IEEE Access*, *9*, 83467–83481.

Said, B., Cheniti-Belcadhi, L., & El Khayat, G. (2019). An ontology for personalization in serious games for assessment. In *2019 IEEE second international conference on artificial intelligence and knowledge engineering* (pp. 148–154). IEEE.

Santos, H., Dantas, V., Furtado, V., Pinheiro, P., & McGuinness, D. L. (2017). From data to city indicators: A knowledge graph for supporting automatic generation of dashboards. In *European semantic web conference* (pp. 94–108). Springer.

Serrano, Á., Marchiori, E. J., del Blanco, Á., Torrente, J., & Fernández-Manjón, B. (2012). A framework to improve evaluation in educational games. In *Proceedings of the 2012 IEEE global engineering education conference* (pp. 1–8). IEEE.

Serrano-Laguna, Á., Martínez-Ortiz, I., Haag, J., Regan, D., Johnson, A., & Fernández-Manjón, B. (2017). Applying standards to systematize learning analytics in serious games. *Computer Standards & Interfaces*, *50*, 116–123.

Shoukry, L. (2020). *Mobile multimodal serious games analytics* (Ph.D. thesis), Technische Universität.

Shoukry, L., Göbel, S., & Steinmetz, R. (2014). Learning analytics and serious games: Trends and considerations. In *Proceedings of the 2014 ACM international workshop on serious games* (pp. 21–26).

Shute, V. J., & Ventura, M. (2015). Stealth assessment. *The SAGE Encyclopedia of Educational Technology*, 675–676.

Sliney, A., & Murphy, D. (2011). Using serious games for assessment. In *Serious games and edutainment applications* (pp. 225–243). Springer.

Smits, J., & Charlier, N. (2011). Game-based assessment and the effect on test anxiety: A case study. In *European conference on games based learning* (p. 562). Academic Conferences International Limited.

Song, Y., & Sparks, J. (2019). Measuring argumentation skills through a game-enhanced scenario-based assessment. *Journal of Educational Computing Research*, *56*(8), 1324–1344. http://dx.doi.org/10.1177/0735633117740605.

Staab, S., & Studer, R. (2010). *Handbook on ontologies*. Springer Science & Business Media.

Stănescu, I. A., Stefan, A., Kravcik, M., Lim, T., & Bidarra, R. (2013). Interoperability strategies for serious games development. *Internet Learning*, *2*(1), 6.

Sun, Z., Hu, C., Li, C., & Wu, L. (2020). Domain ontology construction and evaluation for the entire process of software testing. *IEEE Access*, *8*, 205374–205385.

Susi, T., Johannesson, M., & Backlund, P. (2007). *Serious games: An overview*: *Technical report*, Institutionen för kommunikation och information.

Tang, S., & Hanneghan, M. (2011). Game content model: an ontology for documenting serious game design. In *2011 developments in e-systems engineering* (pp. 431–436). IEEE.

Theodosiou, S., & Karasavvidis, I. (2015). Serious games design: A mapping of the problems novice game designers experience in designing games. *Journal of e-Learning and Knowledge Society*, *11*(3).

Tudorache, T., Vendetti, J., & Noy, N. F. (2008). Web-protege: A lightweight OWL ontology editor for the web. In *OWLED*: *vol. 432*, (p. 2009).

Uschold, M. (1996). Building ontologies: Towards a uni ed methodology. In *Proceedings of 16th annual conference of the british computer society specialists group on expert systems*. Citeseer.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, *11*(2), 93–136.

Valiente, J. A. R. (2022). Unveiling the potential of learning analytics in game-based learning: Case studies with a geometry game. In *Handbook of research on promoting economic and social development through serious games* (pp. 524–544). IGI Global.

Vallejo, V., Wyss, P., Rampa, L., Mitache, A. V., Müri, R. M., Mosimann, U. P., et al. (2017). Evaluation of a novel Serious Game based assessment tool for patients with Alzheimer's disease. *PLoS One*, *12*(5), Article e0175999.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, *57*(10), 1500–1509.

Wiemeyer, J., & Kliem, A. (2012). Serious games in prevention and rehabilitation—a new panacea for elderly people? *European Review of Aging and Physical Activity*, *9*(1), 41–50.

Yusoff, A., Crowder, R., Gilbert, L., & Wills, G. (2009). A conceptual framework for serious games. In *2009 ninth IEEE international conference on advanced learning technologies* (pp. 21–23). IEEE.