# LEARNING ANALYTICS FOR AND IN SERIOUS GAMES

Joint workshop of the
GALA Network of Excellence and the LEA's BOX
project at EC-TEL 2014
September 17, 2014, Graz, Austria
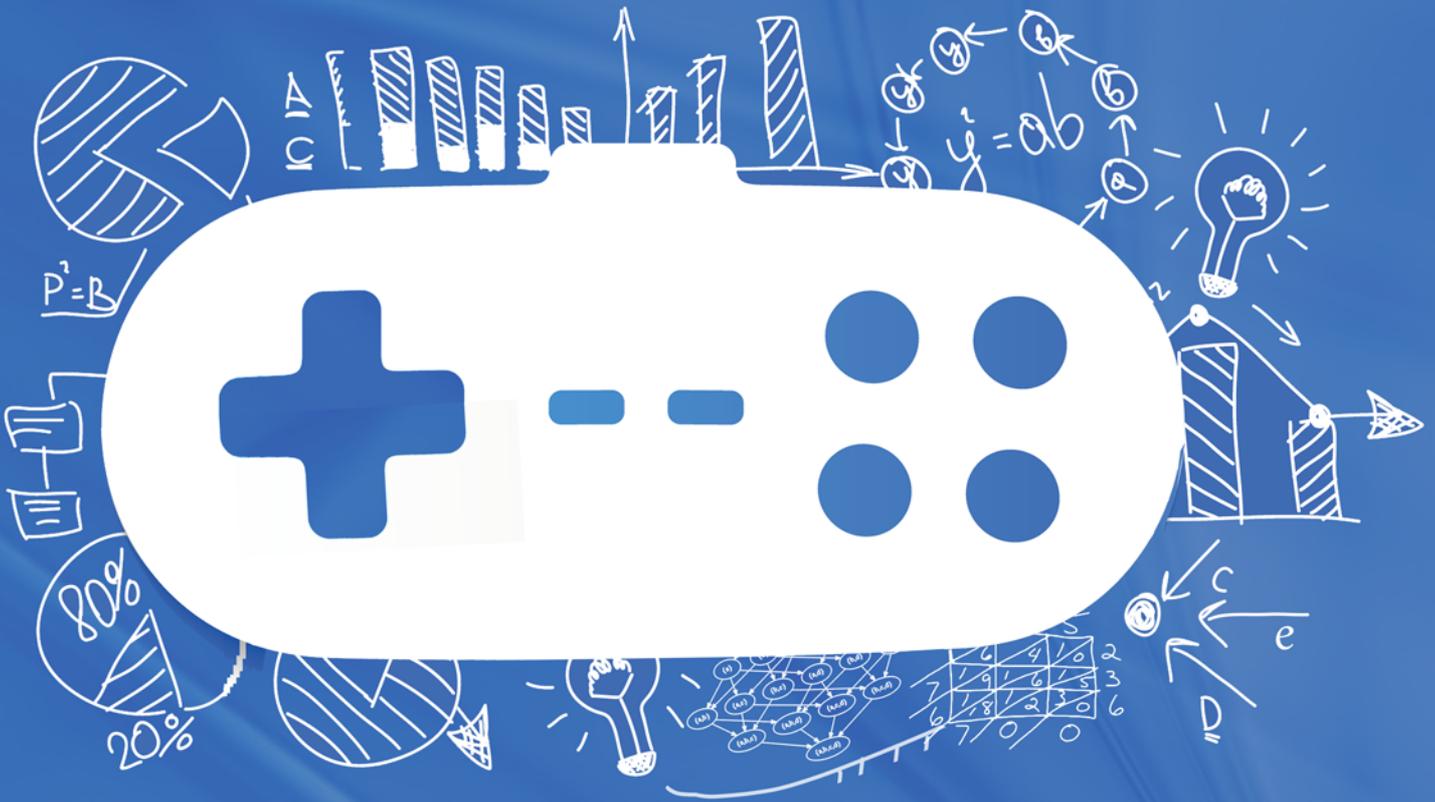
GALA

www.galanoe.eu

LEA'S BOX
Learning Analytics
Toolbox

www.leas-box.eu

Learning Analytics for and in Serious Games
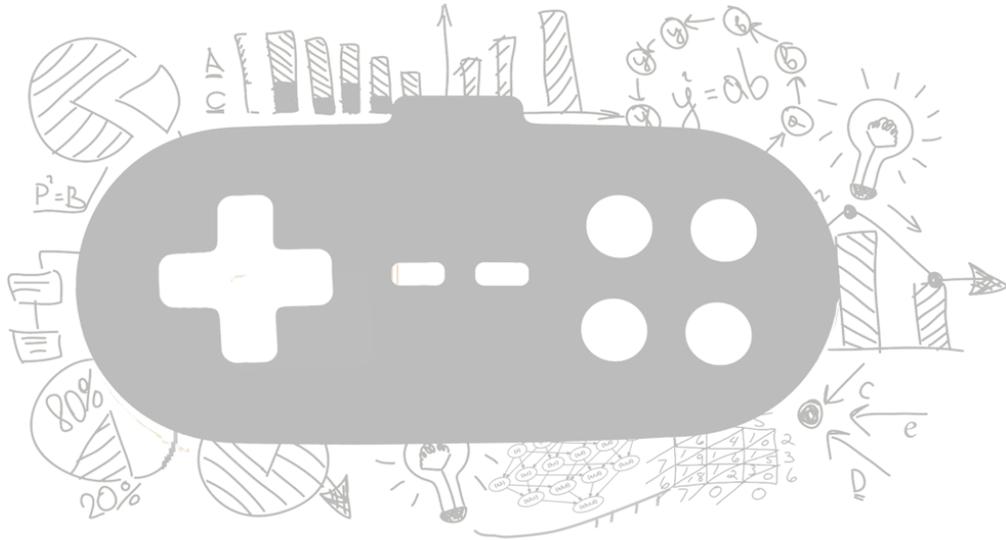
www.galanoe.eu

www.leas-box.eu

Table of Contents

Learning Analytics for and in Serious Games

Different Perspectives

Over the last decade, serious games have become accepted educational tools and the idea of using the great strength of modern computer games for educational purposes experienced a significant boost. From an educational perspective, computer games offer a promising approach to make learning more engaging, satisfying, and probably more effective.

However, playing experience and learning motivation are fragile assets; to be enjoyable, a computer game must be balanced well, meaning the game must match an individual player's playing preferences, playing styles, and playing capabilities in a suitable way in order to too one-sided gameplay. An appropriate adaptation is of crucial importance in order to reach and maintain fun and enjoyment on the one hand and effective, successful learning on the other hand.

The starting point of an educationally suitable adaptation and good game-balancing is to equip the game with and understanding of the learning domain, aspects and characteristics of the player and, in particular, an understanding about what is going on in the game, for example, motivational states or learning performance. Thus, seamless user performance assessment is a major research topic. It is not a trivial to assess and interpret activity data coming from the game in an unobtrusive manner in order not to harm the gaming experience and perhaps 'flow' and requires intelligent technologies.

A recent trend in educational technology is educational data mining (EDM) and learning analytics (LA). The fundamental idea of learning analytics is not new, in essence, the aim is using as much information about learners as possible to understand the meaning of the data in terms of the learners' strengths, abilities, knowledge, weakness, learning progress, attitudes, and social networks with the final goal of providing the best and most appropriate personalized support.

At this point educational adaptation, game balancing, seamless assessment and EDM/LA meet. New educational technologies leverage the potential of serious games and increase their educational depth.

The workshop is organized around and out of 2 European projects; the GALA Network of Excellence (www.galanoe.eu) and the ICT project LEA's BOX (www.leas-box.eu). The goal of the worship is bringing together different research disciplines, technological approaches as well as practitioners in order to discuss this broad conceptual area from a broad range of perspective.

Learning Analytics and Educational Data Mining:
An Overview of Recent Techniques

Christina M. Steiner, Michael D. Kickmeier-Rust, Dietrich Albert
Knowledge Technologies Institute,
Graz University of Technology,
{christina.steiner, michael.kickmeier-rust, dietrich.albert}@tugraz.at

ABSTRACT

This paper summarizes the state-of-the-art on learning analytics and educational data mining, elaborating on key concepts, objectives, data and analytics methods used, visualisations, and key applications. An overview is given on how these techniques are applied to the genre of serious games. Existing challenges in this broad field of research are discussed and a novel, competence-centred approach to learning analytics is outlined that is being developed in the context of the LEA's BOX project.

Keywords

Learning Analytics, Educational Data Mining, Visualisation, Serious Games

1. INTRODUCTION

Assessment plays an important role in education; it is crucial to identify to what extent learning objectives have been met, and in order to be able to provide supportive or remedial interventions to learners and to inform and direct teaching. Assessment or evaluation, as it is also commonly denoted in an educational context, consists in the application of a range of methods for gathering and evaluating information about learning and instruction, with the purpose of making judgments of learners work regarding courses of units of learning [1]. While summative assessment can be characterized as 'assessment of learning', i.e. measuring the achievement of learners in a systematic way and at certain intervals (e.g. at the end of an educational unit), formative assessment can be described as 'assessment for learning', with the positive intent of recognizing progress, promoting learning and planning next steps [2]. The relevance of formative assessment as part of teaching becomes evident when considering the increasing numbers of dropouts in education; in Austria, for example, recently published statistical figures indicate that 154,000 adolescents of age 15 to 24 years are educational dropouts.

Learning analytics (LA) and educational data mining (EDM) are emerging fields of research that may considerably advance formative assessment in educational practice and support establishing a deep understanding of the learning process. This is especially true since the use of a variety of mobile, digital, and educational technologies are used for educational purposes. The important question is how to exploit these different sets of educational data and make sense of them for assessment? Most of the learning management systems used today support basic level analytics, like average usage time, number of educational resources visited per learner etc. [3]. For a holistic understanding of the learning process and progress, more sophisticated analyses are needed, and this is where approaches of LA and EDM come into play.

LA are considered to have immense potential in fostering an evolution of education from a one-size-fits-all delivery approach to a flexible and responsive approach of instruction tailored to learner needs and interests [4]. In essence, it can be described as "big data applied, to education" [5] and has been described as a development that will dramatically shape the future of education [6].

Digital learning games are another development increasingly recognized by educational practitioners as useful educational tools with highly motivating character [4][5]. With the application of games to support instruction and learning, there comes a need of acknowledging learners' learning game experiences also in the context of educational assessment, and initial steps of translating LA to the use in serious games have been taken.

This paper wants to give an overview of LA and EDM, in general, and on developments in the context of educational games, in particular. This provides a basis for further elaboration and application of LA; more concretely for a holistic LA approach that is shaped and implemented in the European project LEA's BOX (http://leas-box.eu/).

This paper is structured as follows: First, a general overview of LA and EDM is given, including its general objectives, process, methods, and applications (section 2), followed by a summary of LA approaches in the context of serious games (section 3). Then, current challenges and problems in LA and EDM are discussed (section 4) and the LEA's BOX approach towards LA is presented (section 5) as an attempt to overcome some of these challenges. The paper ends with a short wrap-up and conclusions (section 6).

2. LEARNING ANALYTICS AND EDUCATIONAL DATA MINING

2.1 Definitions, Related Fields and Key Concepts

Widespread deployment and use of learning or course management systems, web-based learning environments and learning tools produce a whole range of learning-related data and lead to educational institutions dealing with increasingly large amounts of data [7]. In educational contexts, thus, a wide range of data about learners is available. A crucial question is how to make sense of these big data sets for assessment, learning, and teaching? Educational institutions so far have been commonly inefficient in making use of this data. In particular, the available data has traditionally been analysed with substantial delays, thus leading to delayed action and missed opportunities for interventions, like taking measures to reduce or avoid dropouts [6].

Learning analytics (LA) and educational data mining (EDM) constitute related areas of research that aim at making sense of learning-related data; they deal with large data sets relating to learners and their contexts in order to understand and develop learning [3]. Both research areas are defined similarly and with similar concerns. LA is defined by the Society of Learning Analytics Research (http://www.solaresearch.org/) as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments

in which it occurs". Very similarly, the International Educational Data Mining Society (http://www.educationaldatamining.org/) describes EDM as a "discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in". Apart from these, a range of different definitions exists. Although there might be slight differences, all definitions refer to the collection and analysis of data from learning processes, and share an emphasis on translating this data into meaningful actions to support and empower learning [8].

EDM and LA are research areas with slightly different perspectives, but a significant overlap in their objectives and methods. The origins of EDM are usually dated back to the late 1990ies; LA emerged especially in the last decade [7][9]. While EDM has a focus on automated methods, in terms of automated analysis as well as applications in automated adaptation, LA also applies human-led methods to make sense of educational data and seeks applications in terms of using the derived information to empower and support learners and teachers [10]. In line with this, sometimes the two areas are described as having different roots, with the EDM community consisting mostly of researchers from the field of intelligent tutoring systems, and LA researchers having a greater focus on traditional learning systems and LMS. Interestingly, Chatti and colleagues [8] describe EDM as focusing on typical data mining techniques, and LA as including also other methods like visualization tools and social network analysis; which is in contrast to Romero [9], who deliberately included in his review on EDM work that includes typical data mining techniques, but also other approaches (like correlation, visualization etc.), which are not considered to be data mining in a strict sense. In any case, EDM is seen as rather focusing on the technical challenge of extracting information from learning-related data, and LA addressing more the educational challenge of optimizing learning [7]. In general, LA can be seen as a more holistic approach [10], with the deployment of results from analytics and (co-) responding action as important components in addition to pure analysis. John Behrens at LAK2012 outlined that EDM concentrates more on learning as a research topic, and LA has a more practical educational focus [11]. Eventually, both fields of research are closely related and share their interest in enhancing educational practice through researching data-intensive methodologies to education research. They are also both described conjointly in introductory and review articles, and sometimes terms seem to be even used interchangeably [9][11]. In the remainder of this paper the term LA will be used for referring to the wider research area and process of LA and EDM.

Analytics has been spreading over the last years and decades in different domains. Researching methodologies on how to extract meaningful information from big data, making sense of large datasets is trend that has long tradition in natural sciences, and more recently has become an important part of business in terms of business and web analytics but reached the field of learning science comparably late (Baker & Siemens, in press). In fact, data mining efforts are described to have their roots in the commercial sector. Web and business analytics serve identifying of consumer activities and preferences, analysing consumer trends etc., with the goal of tailoring product actions and advertising to consumers [5]. The main reasons for the growing interest and application of analytics approaches in educational domains are: Increasing amounts of learning-related data are available through the use

of mobile, digital, and online technologies and their growing use for educational purposes. Information and experience on how to track this data is available, and standardized data formats are available. This, as well as the increasing computational power that is available has nurtured work on analytics tools that support capturing, organizing, and filtering data as well as tools that support the application of analytics and data mining methods [10]. The idea is, in the end, to use learning data for recommendations (of learning resources, activities, people) and to adapt instruction in a similar manner as it is done with books, music, entertainment etc. in e-commerce [4]. LA and EDM therefore have also strong relations to recommender systems [12], which and adaptive learning environments and intelligent tutoring systems [13] and the goals of these research areas [8].

Further research fields in the educational context that are linked to LA and EDM and share similar objectives are academic analytics and action research. Academic analytics emphasizes the exploitation and analysis of educational data for educational institutions and authorities and at regional, national, or international and governmental level. Academic analytics is less specific than learning analytics, since the focus is more on an analysis at institutional level instead of an analysis of the learning process itself [Long & Siemens, 2011]. Academic analytics and LA initially developed conjointly, but in recent years both areas are gradually developing as separate research areas [7], but overlaps between them naturally remain. Action research is generally described as reflective teaching practice, where instructors analyse, self-reflect, evaluate, and regulate their didactical methods and learning resources provided to students [14]. The main purpose is quality assurance and improvement of instruction and starting point is usually research questions arising from teaching practice [8].

An important aspect of LA is timescale: Most current LA approaches focus on data about the past, reporting what has happened. Other analytics link the present situation with a predicted future; using forecasts and predictive modelling to identify indicators of success, failure or student drop out. A third, and actually preferable approach is a more continuous one and monitoring perspective. This means to consider LA as part of the learning and teaching process, and to link LA and learning design. Concretely, this refers to the use of analytics to support educators and learners to produce a desired future result.

LA is usually described as a multi-step, cyclical process consisting of three main stages: data collection and pre-processing, analytics and action, and post-processing [8]. Some authors have also used a more fine-granular description of individual steps of LA, e.g. Serrano et al [15] in the context of serious games. Data collection and pre-processing refers to the gathering of educational data from different learning systems and applications. Since the collected data may be very extensive or cover irrelevant information, the data is prepared and translated into an appropriate format for the next step. The analytics and action phase denotes the actual application of analytic methods, to extract meaningful patterns and information from the data. This step also includes visualization of the derived information and action, like prediction, assessment, adaptation, personalization, and reflection. Post processing refers to the idea of continually improving analytics, by refining analytics methods or using new methods, including new data sources etc.

In the subsequent subsections, we will first elaborate on the main stakeholders and objectives of LA, discuss on the

educational data that is used, and will then summarise the existing analytics methods applied. These topics actually reflect the four dimensions of the reference model for learning analytics as proposed by Chatti et al. [8]. In addition, we will present the aspect of visualisations as a separate topic, which we consider highly important in LA, and will then summarise the key applications of LA.

2.2 The Stakeholders & Objectives

LA can be carried out for different stakeholders, who have different expectations, needs and goals towards the analytics process and its outcomes. The key stakeholders of LA are certainly teachers and learners – Dyckhoff et al. [14] provide an overview in which way educators and learners; but there are actually more groups of people involved, with other objectives and perspectives regarding the data. Those other groups of stakeholders are educational institutions and administrators (note: in this case, strictly speaking, it is academic analytics that is carried out – see section 2.1 above), as well as course developers and training providers, but also educational researchers [9].

The objectives for using LA are in line with the different views of its stakeholder groups. Chatti et al. [8] identified the following main objectives. These objectives certainly have overlaps and usually a specific application of LA will serve several of them.

- Monitoring and analysis: Tracking and checking the learning process, which is then used by teachers or educational institutions as a basis for taking decisions, e.g. on future steps, the design of learning activities, improving the learning environment.

- Prediction and intervention: Estimating learners' future knowledge or performance in terms of finding early indicators for learning success, failure, and potential dropouts, to be able to offer proactive interventions and support for learners in need of assistance.

- Tutoring and mentoring: Helping learners with and in their whole learning process or in the context of specific learning tasks or a course, providing guidance and advise.

- Assessment and feedback: Supporting formative and summative (self) assessment of the learning process, examining efficiency and effectiveness of learning, and providing meaningful feedback on the results to teachers and learners.

- Adaptation: Finding out what a learner should do or learn next and tailoring learning content, activities, or sequences to the individual. This idea of carefully calculated adjustments corresponds to the central aim and component of adaptive learning environments and intelligent tutoring systems [4].

- Personalization and recommendation: Helping learners in deciding over their own learning and learning environment, and what to do next by providing recommendations while leaving the control to the learner.

- Reflection: Prompting and increasing reflection or self-reflection on the teaching and learning process, learning progress and achievements made; providing comparison with past experiences or achievements, between learners, across classes etc.

2.3 The Data

The significance of LA naturally always depends on the educational data available and used for the analytics process.

There are a wide range of learning system and tool available used in an educational context or learning purposes (ITS, LMS, concept mapping, social networks), and all of them provide different data [9]. In general, two big categories of data sources can be distinguished [8]: Centralized educational systems, like LMS, provide extensive log data of learner interaction and activities (accessing learning resources, reading, writing, taking tests). Distributed learning environments provide multiple logs from a range of different sources from formal and informal channels and distributed across space, time, and media.

A challenge when dealing with educational data for LA is the issue of data integration from different sources and formats. Another issue is the storage of data, since the analytics processes by nature use 'big data', i.e. large data sets that would not be practicable to deal with for manual analysis [7]. When considering large data sets, a distinction between extensive and intensive data can be made [16]:

- Extensive data refers to data that is collected from a large number of participants on a limited number of variables, resulting in a wide but shallow set of data, which is typically used for data mining techniques.

- Intensive data, on the contrary, is data from a relatively small number of participants, but with detailed observations on a large number of variables, thus resulting in a deep but narrow set of data. Intensive data usually consists in several traces or logs of data; analysis is done across these different traces.

Extensive and intensive data can meaningfully complement each other, for example for triangulation and validation of results or by using intensive data for deciding on the type of extensive data to be collected [16].

Deciding on the kind of data to be captured and the information to be extracted is key in LA. The choice of data used as predictors and indicators immediately influences the quality and accuracy of the analysis. Three broad types of indicators can be distinguished [17]:

- Dispositional indicators are factors that the learner brings to the learning context and are available before the learning episode begins. Examples are age, gender, previous learning experiences etc. Many of them are factual and readily quantifiable.

- Activity and performance indicators refer to data that learners produce as they are engaging in learning activities and making their way through a course. Examples are the number of logins, time spent, number of discussion posts etc.

- Student artefacts denotes data resulting from learners actual work in terms of the products of the learning process, like essays, blog posts, discussion forum contributions. Analysing such artefacts can provide information on the mastery or competence of learners.

The data and indicators that can be selected for data collection necessarily are based on the data that is available from the learning environments and applications. Data tracking occurs without any extra manual effort by the learner. Thereby it is of course crucial that learners are aware that their data and activities are logged [18]. Dyckhoff et al. [14] have conducted a comprehensive review on state of the art LA and collected about 200 indicators currently used LA (e.g. number of threads per student, number of participants per group, clusters of student who made a specific mistake etc.). They

categorized the according to the different perspective one may have on the data (individual learner, group, course, content, and teacher). Additionally, the origin of data (data sources) was differentiated into six categories: student generated data, context/local data, academic profile, evaluation, performance, and course meta-data. This review showed that a large part of the data used in current LA tools is basic usage data (i.e. activity and performance indicators) of learners engaging with a single learning environment. The authors conclude that in order to be able to answer more complex, highly relevant research questions that educational practitioners have in mind, a greater emphasis needs to be put on high-level indicators and teachers should optimally be actively involved in the definition and design of relevant indicators.

In general, data collection in LA is not confined to the pure capturing of learners' traces via different indicators, but may also consist in the combination of data from different sources, as indicated earlier. This is done in data pre-processing and aggregation, i.e. the datasets are merged and transformed into a suitable format for further analytics processing.

2.4 The Analytics

In LA a range of different methods are used to extract meaningful patterns from educational data. The techniques actually used will depend on the objectives of the analytics tasks, but also on the kind of data collected. Baker and Siemens [10] consider methods from data mining and analytics, in general, as well as psychometrics and educational measurement as the main sources of inspiration for LA methods and tools and provide a systematic overview of the key methods currently used in LA, which fall in five main classes: prediction methods, structure discovery, relationship mining, discovery with models, and distillation of data for human judgment (see also [11]):

• Prediction methods: These are the most popular methods in EDM and essentially aim at developing a model to predict or infer a certain variable (e.g. mark, performance score) from a combination of other indicators of the educational data set. Common prediction methods are classification (for prediction of binary or categorical variables), regression (for prediction of continuous variables), and latent knowledge estimation (assessing learner knowledge or skills).

• Structure discovery: Algorithms of structure discovery aim at detecting structure in educational data without an a priori assumption of what should be found (in contrast to prediction methods, where the predicted/dependent variable is known). Methods of this type are clustering (splitting data set into clusters grouping data points together), factor analysis (finding dimensions of variables grouping together) and domain structure discovery (deriving the structure of knowledge in an educational domain).

Another method from this class, which is quite popular in LA, is social network analysis (SNA) [19]. It allows analysing relationships and interactions between learners in terms of collaboration and communication activities, information exchange etc. SNA uncovers the patterns and structure of interaction and connectivity, which can then be visually illustrated and provide the possibility of quantification (e.g. via centrality measures), to identify learners that are very important, represent 'hubs', or are in isolation [8].

• Relationship mining: The aim of this group of methods is to find out relationships between variables and how strong those relationships are. Concrete methods are association rule mining (finding if-then rules), correlation mining (finding positive or negative linear correlations), sequential pattern mining (finding temporal associations between events), and causal data mining (finding out whether one observation is the cause of another).

• Discovery with models: This class does not denote a specific group of techniques but refers to the general approach of using the results of one analytics method within another analysis. A popular way of doing this is for instance the use of a prediction model within another prediction model, but there are a variety of other ways for conducting discovery with models.

• Distillation of data for human judgment: This is an approach that is common in LA, in a narrower sense, but not considered as a method of EDM, since it consists in providing teachers immediate access to reports and visualisations of the learner data, for their interpretation, judgement and to support decision making and pedagogical action. Examples are learning curves or heat maps [15].

A range of tools has been developed for conducting LA. The tools come from the commercial and academic sector, implement the methods outlined above, and provide support in the validation of models and visualisation of data [10]. Although some tools find already application in educational practice, many existing tools, though, are very complex and do not appropriately fit the needs of educators [9][20].

2.5 The Visualisations

LA is not only about collecting and analysing educational data, but feeding back and making use of the results is essential. The results and inferences of LA are usually used ways: either the information is fed into adaptation and recommendation mechanisms, or it is reported back to the learner, teacher, or other stakeholders to empower and support the teaching and learning process. The fine-grained statistics available from LA are oftentimes too cumbersome to inspect or too time consuming to interpret. Visualization can help people to understand and analyse the data [9].

Suitable visualisations play an essential role in making big sets of learning-related data and results better understandable, to gain an insight in the learning and teaching process and the interrelation between teaching and learning. This is, in fact, a prerequisite for achieving the overarching goal of LA in terms of gradually improving teaching and learning processes [21].

The visualizations make LA results actionable, i.e. they enable teachers, mentors, learners to take appropriate decisions and action [17]. Thereby, visualization will differ in the way results are displayed (chart or diagram type) and the way results are presented for different stakeholders.

Visualisations of learning traces are called learning dashboards and are commonly applied in LA [22]. They enable teachers and learners to get an overview of their activities and how they compare to those of others. Different approaches of dashboards exist. "All-at-one-time" dashboards represent different visualisations with different aspects of information side-by-side. Other dashboard approaches start with one visualization and enable the user to access further information and detail from there [17]. A variety of dashboard applications has been developed recently; the learning dashboard approach is considered to have very good potential to foster awareness, reflection, sense making and, in the end, improve learning. However, the evaluation of actual impact of using them is difficult [22].

Considering LA visualizations for learners, teacher, and other stakeholders, this directly relates to the topic of Open Learner Models (OLM), i.e. the idea of opening up the learner model to the user to support reflection and awareness of learning and dynamically update that information for a deeper understanding of the learning and teaching process [23]. This seems similar to the goals of LA, but while LA visualizations today oftentimes are confined to the illustration of activity and interaction data, OLM focuses on the representation of inferences drawn from that data in terms of learners' skills, knowledge, affective states [24].

With the evolution of LA towards responding to more complex educational questions and making inferences to competence and mastery, visualisations in LA are developing more and more towards including or representing this kind of information, e.g. [25]. Such more sophisticated visualisations may also facilitate finding positive evidence of their impact on learning, e.g. on competence scores and persistence rates.

2.6 Key Applications

In terms of the key applications of LA, the Baker and Yacef [26] take a highly research- and development-oriented view in their review of 2009 and highlight the following application areas: improvement of learner models, improving or uncovering models of a knowledge domain's structure, investigation of pedagogical support to find out which types of support are (most) effective, and finding empirical evidence for refining or elaborating educational theories and phenomena for a better understanding of learning and its influencing factors and as an information source for learning system design. In contrast, Romero [9] takes a more education- and practice-oriented position and identifies a set of educational tasks, partly overlapping with [26] that LA may be applied for:

• Providing meaningful information, feedback, and visualisations to support instructors and educational administrators in decision making on instruction and proactive or remedial action

- Providing recommendations to students

- Adapting learning contents, sequences, and interfaces

- Predicting learner performance

- Learner modelling, including e.g. skills, motivation, learning styles

- Detecting undesirable or erroneous learner behaviour

- Supporting the creation of student groups

- Studying the relationships between learners

- Supporting teachers in concept map creation

- Assisting construction and reuse of learning content

- Enhancing educational planning and scheduling

This list gives an idea that beside using LA for understanding what learners do, predicting what they will do or how successful they'll be, and personalising and improving learning experiences, LA can and should have also an important role in terms of transforming the educational landscape. LA may serve a valuable information source for educational administrators and decision makers to shape education and allocate resources and to optimise learning and educational results at national and international levels [6][7].

To summarise on the state of the art in LA, a recent review [8] showed that centralized web-based learning systems (ITS, LMS) and are the most common data source for LA, the majority of current LA applications are carried out towards intelligent tutoring system design and research. The main objectives followed are thereby monitoring, analysis, and adaptation. The focus and use of LA is expected to transform to more open, personalized, networked, lifelong and also game-based learning and learning contexts.

3. LA AND SERIOUS GAMES

Virtual worlds and digital learning games are increasingly used educational tools with highly motivating character. Virtual worlds are highly engaging, offering opportunities for learning experiences that go beyond traditional e-learning environments [27]. With higher education institutions and universities starting to offer their courses in such online environments and enterprises specializing in the delivery of experiential corporate training, educational administrators and training providers are also starting to use analytics in order to get to know about the number of learners signing up these courses and how they are engaging with the course material and with each other. Learning analytics is considered as a way of better understanding the learning pathways of learners in virtual worlds, in order to identify the effectiveness of this kind of training, to foster reflection on the learning and teaching process, to modify teaching approaches etc. Attempts of applying analytics in virtual worlds have been presented e.g. by [27][28][29].

While teaching scenarios in virtual worlds more resemble more traditional teaching environments and provide some formal learning context, in educational games learning is embedded in the context of a game, ideally realizing a stealth learning environments. There is a broad awareness of the educational potential of videogames and game-based learning and serious games are intensively researched [4][5][30]. Games have a highly motivating character; the interactive and immersive learning experiences that can be created by the use of learning games establish authentic learning tasks and meaningful, situated learning. Games have proven to support skill acquisition related to collaboration, procedural and critical thinking, creative problem-solving, observation, and reasoning, collaboration [5].

Despite the theoretical and empirical evidence for the potential of educational games, there is still some reluctance among teachers towards their broader take up and use in educational practice [31]. This is mainly due to the fact that it is difficult to integrate assessment procedures in terms of tests or question and answers in games, which would be experienced highly disruptive and breaking the flow of the gaming experience [32]. Assessment routines built in commercial games, even if they are developed for educational purposes, are usually black boxes and not tangible for teachers. Serrano-Laguna et al. [33] highlight that there is a need to implement approaches for reliable formative and summative assessment in educational games, which are easy to use and provide teachers useful educational information and evaluation.

Learning games, just as commercial video games, may produce large amounts of data by recoding user (inter)actions on a micro level. This results in another type of learning-related big data that may be used for LA. A crucial question is how to harness and make sense of this data in an effective and efficient manner. LA is currently in the process of initiating the elaboration of analytics that can be used on serious games. By using and combining ideas from gaming analytics, web

analytics and learning analytics, it is possible to establish meaningful analytics on data from games for educational purposes.

A great challenge with learning analytics in educational games is the wide variety of different games available, which complicates the development of analytics tools that are applicable independent of a concrete game. To overcome this, Serrano-Laguna et al. [31][33] propose a two-step generic approach to apply learning analytics in educational games, which is applicable with any kind of different game. First, generic traces are gathered from gameplay, including game traces (start, end, and quit), phase changes (game chapters), input traces like mouse movements or clicks, and other meaningful variables like attempts or scores (depending on the game). These data give rise to reports with general and game-agnostic information, like the number of students who played the game, average playing time, game phases in which users stopped playing etc. This information can be visually reported and may provide initial useful information on how learners interacted with a game. In a second step, additional information may be extracted by letting teachers define game-specific assessment rules based on and combining the generic game trace variables to obtain new information (e.g. setting maximum time thresholds, comparison between actual and expected/required values of variables). These rules clearly need to be closely defined in line with each game to match the educational objectives; however since the building block of this kind of rules are elements from the basic set of traces the creation and provision of template rules to support teachers in defining their own is conceivable.

To make use of learning analytics in educational games, necessarily a game platform needs to be used that allows collection of the relevant data, and that holds a representation of game variables. The data for learning analytics will likely need to be stored and processed separately and remotely [33]. To technically implement such kind of analytics in an educational game, the definition of a learning analytics model and implementation of a learning analytics engine, which is separate from the game engine but communicates with it, has been proposed [15]. The learning analytics engine is conceived as comprising a set of modules enabling the different steps of the learning analytics process, from capturing data, via aggregating end reporting to evaluating in terms of transforming information in educational knowledge.

Assessment in a learning game may have two main purposes. Firstly, just measuring the success of the student – this will serve providing teachers and students the information derived as a basis for action, like selection of new educational resources, decision on additional support or learning tasks etc. Secondly, the derived information may be used for realizing dynamic adaptation during game time through an adaptation model and adapter (part of the learning analytics framework) communicating with the game engine.

LA in terms of on-line assessment and adaptation has been elaborated and implemented in the educational game demonstrators developed in the ELEKTRA and 80Days projects [34]. In a nutshell, learner actions during a complex problem-solving situation are monitored and interpreted in run-time to enable a continuous and non-invasive assessment of learning progress and motivational state. The formal framework of Competence-based Knowledge Space Theory [35][36], originating from the field of adaptive personalized learning, was thereby used as a theoretical basis. Observations of learner actions are interpreted as evidence for available and lacking skills and feed a probabilistic assessment and continuous update of the learner's competence state [37]. The information coming from this assessment is used to provide adaptive interventions tailored to the individual's current state and needs (chosen from a menu of different intervention types), to support and guide the learner in the game and learning task and to retain motivation [38]. The analytics applied in these educational games were dedicated to realizing continuous assessment and automatic live adaptation and support. This application could be broadened in terms of reporting and feeding back the information on skills acquired and learning progress also to learners and teachers, thus leveraging the educational value of the analytics processes of the games and translating the learning data from the game into educational actions outside the game context.

Another example of using learning analytics in a serious game has been presented by [39], who also realized skill assessment in an educational action game by using game events as evidence for users' mathematical skills to analyse study gains in accuracy over time and in speed over time with learning curves. This approach proved useful for formative assessment in educational games and may also be used to inform re-design and improvement of intelligent tutoring systems. Another very recent LA attempt has been made towards elaborating an automated detector of engaged behaviour in a simulation game [40]. The aim thereby was to identify and model which learner actions give evidence of user engagement and, in the end, are predictive for success in the game. An integration of the engagement detector in the game will enable to report the results back to learners and teachers for reflection.

On the whole, LA has started to grow into the field of serious games, but there is still much more work to do to fully exploit the potential that LA may bring to optimize learning experiences with educational games. LA can also be exploited to refine or improve educational games – in terms of using analytics for an evaluation of the game artefact itself [33].

It needs to be taken into account that games may be part of multiple learning activities that learners carry out in parallel and, potentially, on the same educational objective [41]. A learning game will usually provide educational content as a complement to other and more traditional technology-enhanced and classroom learning activities. LA for assessment of skill acquisition should therefore actually not consider one an educational game in isolation, but should optimally utilize and integrate information from multiple sources.

4. DRAWBACKS AND CHALLENGES

The focus of LA research until now has been on the methods of data collection and analysis instead of the actual application in educational practice. Although a recent trend of moving from a technological focus towards a more educational direction can be identified [7], Siemens [2012] highlights that there is still a 'research and practice gap' that exists in learning analytics in terms of a lack of translating analytics research into educational practice: "The work of researchers often sits in isolation from that of vendors and of end users or practitioners" (p. 5). A transition in analytics from technical orientation to one that emphasizes sense-making, decision-making, and action is required to increase interest among educators and administrators. A second transition needed is one that moves LA research and implementation from at-risk identification (which is only a small aspect of what analytics can do to improve education) to an emphasis on learner success and optimization.

Many questions relevant for teachers in educational practice (see [42] for an overview) cannot be answered with analytics tools that are currently available. This became evident from a

meta-analysis identifying the kind of data or indicators used in state of the art learning analytics tools and identifying which kind of teachers' research questions they are able to help to answer [14]. In particular, questions targeting learners' satisfaction and preferences are not sufficiently supported. At the moment there is little literature available elaborating on how learning analytics influence practical educational situations and the behavioural reactions of teachers and learners. Studies often focus on analysing how different e-learning tools or features affect learning, while studying the interactions between learners or between learners and instructors and how it relates to learning strategies and theories is widely disregarded [43]. The consideration of how learning analytics affects teaching and the evaluation of this impact has been highlighted as an important as crucial issues that need to be tackled in learning analytics research in the future [14]. LA tools need to be deployed in educational practice in continuous exchange of teachers and learners with LA researchers, to enable a continual improvement and further elaboration of LA strategies and methods.

With respect to analytics techniques and tools another problem is that findings from marketed analytics software are most often proprietary and therefore not available to other researchers and hinder quick and iterative improvement cycles of learning analytics methodologies. Corporate analytics products are closed to researchers and do not allow to access and scrutinize, change and improve algorithms and there are only few tools providing the openness, accessibility, and transparency desired by researchers (one example is the software package R) [20]. For application in educational practice, tools are needed that are suitable for teachers and non-expert users – tools with intuitive user interfaces that are easy to use and provide features that are meaningful to educators and provide positive end-user experience [9][20].

Another challenge in LA is given by the fact that the amount and diversity of educational data is ever growing, which makes the analysis of educational data increasingly complex and increases demands on data storage and computational power. This data explosion in the educational sector itself impacts analytics – with increased quantity of data the methods and approaches used for analysing and making sense of this data necessarily need to change [6].

Even more important, though, is the need for an integration of learning data form different sources. When learners are dealing with online learning environments and tools, they will most likely not be engaged with on single learning activity, but will instead carry out various different activities or learning tasks at a time [41]. LA, like e.g. approaches of assessing learners' knowledge or skills, today usually is still confined to only one activity, also if the same skills are involved in different parallel learning activities.

An increasing awareness of this need of integrating learning data from different sources to build more comprehensive and conclusive learner models and to derive more targeted conclusions for supporting or optimizing teaching and learning. Existing approaches in this line of thought, however, concentrated on predicting learner performance on one activity on the basis of data from another activity or, respectively, investigating whether learning transfers to new contexts. An example is research carried out by Miller et al. [41], who integrated information from conceptual instructions, problem-solving, and mini games for predicting student performance using Bayesian Knowledge Tracing. While this work used learning data on a small set of different activities, but stemming from one and the same learning system, there have also been initial attempts of tracking meta-

skills of science inquiry across different knowledge domains but within the same activity [44].

"One big problem around learning analytics is the lack of clarity about what exactly should be measured to get a deeper understanding of how learning is taking place" ([18], p. 16). Currently used indicators mostly are limited to interaction data [14], but how much information can these actually provide of the learning process? LA results sometimes still consist in very basic measures and indicators, while lacking a deeper consideration of how to translate the educational data in meaningful information and to measure the complex processes of knowledge and skill state or acquisition of learners [20]. LA needs to go beyond tracking and reporting basic usage data to make inferences on the knowledge and competence of learners, their affective states etc., and it needs to include more than data from only one centralized learning system [45], but include also data from more informal learning, to establish an accurate and deep understanding of learning and teaching. To ensure that LA develops towards answering the complex research questions that are relevant for practitioners, emphasis needs to be put on such high-level and combined indicators and educators should optimally be actively involved in the definition and design of relevant indicators.

5. A LEARNING ANALYTICS TOOLBOX

In doing learning analytics a thorough understanding needs to be established on what needs to be known and what data is most suitable to provide this information. Especially, since the amount and diversity of educational data is ever growing, new approaches are needed to exploit the informational potential residing in this data [6]. The European project LEA's BOX aims at establishing a novel approach of competence-centred and theory-grounded LA that will help advancing LA research and application towards overcoming the challenges and problems outlined above. This approach will extend existing LA and EDM techniques by methods on the basis of Competence-based Knowledge Space Theory (CbKST) and Formal Concept Analysis (FCA) [46][47]. CbKST originated in the field of intelligent tutoring systems and was elaborated to non-invasive formative skill assessment in line with LA ideas, as a basis for automated adaptation and support of learning experiences. FCA originated from applied mathematics as an attempt to formalize concepts and concept hierarchies. In LEA's BOX a clearly learning-focused perspective is taken and a holistic framework for modelling, structuring, assessment, and feedback is developed.

A well-known dilemma in learning analytics is that of using top-down versus bottom-up analytic approaches. Commonly a top-down approach is used in learning analytics, when it comes to data collection - gathering learning data over a period of time and then analysing it in order to extract valuable information patterns [26]. Instead of purely data-driven approaches of pattern recognition in learning-related data, there are increasingly claims for a more bottom-up-like strategy, i.e. that reasoning about data requires robust and well-elaborated psycho-pedagogical foundations. Chatti et al. [21], for example, argue that LA needs to start from a research question, in a first step, and the selection of suitable analysis methods should be done in a second step. Starting analytics from questions and psycho-pedagogical theory and models of teaching and learning, from conceptions of knowledge, of how learning and learning success take place, is considered one of the main challenges in the emerging field of learning analytics.

The competence-based LA framework of LEA's BOX consists in a hybrid approach building upon and harmonizing bottom-up and top-down procedures and aiming at realizing feasible, efficient, effective and pedagogically meaningful analysis and sense-making of learning related data. The general idea is to start with psycho-pedagogical considerations, information, and consultation to establish. These lead to a coarse, 'theory-driven' competence models in tradition of CbKST, which provide a meaningful representation of competences with pedagogically and semantically rich underlying descriptions and their structure. These models serve as the theoretical foundation and pedagogical hypothesis for carrying out LA. Through data-analytic methods of domain structure discovery, for example grounding on FCA as described in [48], a more fine-grained structure of the knowledge domain can be established. The established structures can be exploited for continuous formative skill assessment and monitoring purposes.

In terms of data, for the learning analytical process in LEA's BOX a blend of different educationally relevant dispositional, activity and performance indicators, and student artefacts from various learning systems and educational tools (with an openness to data from educational games) are gathered and triangulated for strong and nuanced interpretation. The information derived from LA will be reported back to teachers and learners. To this end, existing dashboard approaches will be enriched by visual and graph representations, such as Hasse diagrams or concept lattices known from CbKST and FCA, to come up with a set of visualizations that optimally support understanding of learner models and progress.

These visualizations shall also serve the collection of human-contributed feedback and corrections to be used for further refinement of the competence and LA models, in addition to automated validation and refinement. This corresponds to the claim of giving users the possibility to influence LA [20] and the idea of continuous improvement of LA as a final step of the LA process [8].

In LEA's BOX also the issue of missing analysis of the impact of learning analytics on teachers and learners, their behaviour, metacognition, teaching and learning experiences as highlighted by Dyckhoff et al. [14] is addressed. This is tackled by a continuous engagement with educational practitioners and schools in several European countries throughout the project, from requirements analysis via pilot studies to summative evaluations. That in an important prerequisite to ensure that the analytics tools developed in the project and deployed via a generic LEA's BOX platform, are not only usable, but also useful for the targeted end users.

6. CONCLUSIONS

This paper provided an overview of the broad field of LA research, in general, and in serious games. LA has great potential to empower educational processes, but to fully exploit this potential LA needs to advance the handling of high-level indicators and complex educational questions and needs to be incorporated in the daily workflow of educational practice. In this way, LA may have a valuable impact on the optimization and support teaching and learning experiences and the evolution and refinement of educational structures. The learning analytics framework and toolbox under development in the LEA's BOX project provides a holistic approach to effectively assess, monitor, and promote skills integrating educational data stemming from multiple activities and data sources. This approach is characterized by a combination of theory- and data-driven methods as a basis for competence-centred learning analytics and will help to build a

more comprehensive and accurate understanding of learning and progress that can be reported to learners and teachers and will serve supporting learning, optimizing teaching, and refining LA.

8. REFERENCES

[1] Taras, M., "Assessment – Summative and formative – Some theoretical reflections". British Journal of Educational Studies, 53, 466-478, 2005.

[2] Harlen, W. & James, M., "Assessment and learning: Differences and relationships between formative and summative assessment". Assessment in Education, 4, 365-379, 1997.

[3] Ferguson, R., "Learning analytics for open and distance education". In S. Mishra (Ed.), CEMCA EdTech Notes. New Delhi, India: Commonwealth Educational Media Centre for Asia (CEMCA), 2013.

[4] Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A., "NMC Horizon Report: 2014 Higher Education Edition." Austin, Texas: The New Media Consortium, 2014

[5] Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Ludgate, H., "NMC Horizon Report: 2013 Higher Education Edition". Austin, Texas: The New Media Consortium, 2013

[6] Long, P., & Siemens, G., "Penetrating the fog. Analytics in learning and education." EDUCAUSE Review, 46, 30-40, 2011.

[7] Ferguson, R., "Learning analytics: drivers, developments and challenges". International Journal of Technology Enhanced Learning, 4, 304-317, 2012.

[8] Chatti, M.A., Dyckhoff, A.L., Schroeder, U., & Thüs, H. "A reference model for learning analytics." International Journal of Technology Enhanced Learning, 5, 318-331, 2012.

[9] Romero, C., " Eductional data mining: A review of the state of the art." IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews, 40, 601-618, 2010.

[10] Baker, R., & Siemens, G., "Educational data mining and learning analytics". To appear in Sawyer, K. (Ed.) Cambridge Handbook of the Learning Sciences: 2nd Edition, in press.

[11] Baker, R., Inventado, P.S. "Educational data mining and learning analytics". To appear in J.A. Larusson, B. White (Eds.) Learning Analytics: From Research to Practice. Berlin, Germany: Springer.

[12] Adomavicius, G. & Tuzhilin, A., "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17, 734-749, 2005.

[13] Brusilovsky, P., & Peylo, C., "Adaptive and intelligent Web-based educational systems". International Journal of Artificial Intelligence in Education, 13, 159-172, 2003.

[14] Dyckhoff, A.L., Lukarov, V., Muslim, A., Chatti, M.A, & Schroeder, U. "Supporting action research with learning analytics". In: Proceedings of the International

Conference on Learning Analytics and Knowledge (220–229). New York: ACM Press, 2013.

[15] Serrano, A., Marchiori, E.J., del Blanco, A., Torrente, J., and Fernandez-Manjon, B.A. "A framework to improve evaluation in educational games". In Proceedings of the 2012 IEEE Global Engineering Education Conference (1-8). IEEE, 2012

[16] Homer, B.D., "Introductory Talk to the Learning Analytics and Educational Data Mining Workshop", CREATE Lab, New York University, April 2013.

[17] Brown, M., "Learning analytics: Moving from concept to practice". EDUCAUSE Learning Initiative. July 2013. Retrieved July 7, 2014 from http://net.educause.edu/ir/library/pdf/ELIB1101.pdf

[18] Duval, E., "Attention please! learning analytics for visualization and recommendation". In Proceedings of the International Conference on Learning Analytics and Knowledge (9-17). New York: ACM, 2011.

[19] Bakharia, A. & Dawson, S., "SNAPP: A bird's-eye view of temporal participant interaction". In Proceedings of the International Conference on Learning Analytics and Knowledge 2011 (168-173). New York: ACM, 2011.

[20] Siemens, G., " Learning analytics: Envisioning a research discipline and a domain of practice. In Proceedings of the International Conference on Learning Analytics and Knowledge (4-8). New York: ACM, 2012.

[21] Chatti, M.A., Dyckhoff, A.L., Schroeder, U., & Thüs, H., "Forschungsfeld Learning Analytics. Learning Analytics Research Challenges." i-com – Zeitschrift für interaktive und kooperative Medien, 1/2012, 22-25, 2012.

[22] Verbert, K, Duval, E., Klerkx, J., Govaerts, S., & Santos, J.L., "Learning analytics dashboard applications". American Behavioral Scientist, 57, 1500-1509, 2013.

[23] Bull, S. & Kay, J., "Open Learner Models". In R. Nkambou, J. Bourdeau and R. Mizoguchi (eds.), Advances in Intelligent Tutoring Systems (318-338). Berlin: Springer, 2010.

[24] Bull, S., Kickmeier-Rust, M., Vatrapu, R., Johnson, M.D., Hammermueller, K., Byrne, W. et al., "Learning, learning analytics, activity visualization and open learner model: Confusing?" In Hernández-Leo et al. (eds.), EC-TEL 2013. LNCS 8095 (532-535). Berlin: Springer, 2013.

[25] Grann, J. & Bushway, D., " Competency map: Visualizing student learning to promote student success". In Proceeding of the International Conference on Learning Analytics and Knowledge (168-172). New York: ACM, 2014.

[26] Baker, R., & Yacef, K., "The State of Educational Data Mining in 2009: A Review and Future Visions". Journal of Educational Data Mining, 1(1), 3-17, 2009.

[27] Camilleri V, de Freitas S, Montebello M, & McDonagh-Smith P., "A case study inside virtual worlds: Use of learning analytics for immersive spaces". In: D. Suthers & K. Verbert (eds.), Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (230–234). New York: ACM Press, 2013.

[28] Fernández-Gallego, B., Lama, M., Vidal, J.C., & Mucientes, M., "Learning analytics framework for educational virtual worlds". Procedia Computer Science, 25, 443-447, 2013.

[29] Kickmeier-Rust, M. & Albert, D. "Learning analytics to support the use of virtual worlds in the classroom". In: A. Holzinger & G. Pasi (eds.), Human-computer interaction and knowledge discovery in complex, unstructured, big data. LNCS vol. 7949 (358-365). Berlin: Springer; 2013.

[30] de Freitas, S., "Learning in immersive worlds. A review of game-based learning". JISC E-learning programme, 2006. Retrieved March 1, 2013 from http://www.jisc.ac.uk/media/documents/programmes/elearninginnovation/gamingreport_v3.pdf

[31] Serrano-Laguna, A., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B., "Application of learning analytics in educational videogames". Entertainment Computing, 2014.

[32] Van Eck, R., "Digital game-based learning. It's not just the digital natives who are restless". Educause Review, 16-30, 2006.

[33] Serrano-Laguna, A., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B., "Tracing a little for big improvements: Application of learning analytics and videogames for student assessment". Procedia Computer Science, 15, 203-209, 2012.

[34] Kickmeier-Rust, M.D., & Albert, D., "Micro adaptivity: Protecting immersion in didactically adaptive digital educational games". Journal of Computer Assisted Learning, 26, 95-105, 2010.

[35] Albert D. & Lukas J., "Knowledge spaces: Theories, empirical research, applications". Mahwah: Lawrence Erlbaum Associates, 1999.

[36] Heller, J., Steiner, C., Hockemeyer, C., & Albert, D., "Competence-based knowledge structures for personalised learning". International Journal on E-Learning, 5, 75-88, 2006.

[37] Augustin, T., Hockemeyer, C., Kickmeier-Rust, M., & Albert, D., "Individualized skill assessment in digital learning games: Basic definitions and mathematical formalism". IEEE Transactions on Learning Technologies, 4, 138-148, 2011.

[38] Kickmeier-Rust, M.D., Steiner, C.,M. & Albert, D., "Apt to adapt: Micro- and macro-level adaptation in educational games". In T. Daradoumis, S. Caballé, A. Juan & F. Xhafa (eds.), Technology-Enhanced Systems and Tools for Collaborative Learning Scaffolding. Studies in Computational Intelligence vol. 350 (221-238). Berlin: Springer, 2011.

[39] Baker, R.S.J.d., Habgood, M.P.J., Ainsworth, S.E., & Corbett, A.T., "Modeling the acquisition of fluent skill in educational action games". In: Proceedings of User Modeling 2007 (17-26), 2007.

[40] Stephenson, S., Baker, R., Corrigan, S., "Towards building an automated detector of engaged and disengaged behavior in game-based assessments". Poster presented at the Annual Conference on Games+Learning+Society, 2014.

[41] Miller, W.L., Baker, R.S., & Rossi, L.M., "Unifying computer-based assessment across conceptual instruction, problem-solving, and digital games". Technology, Knowledge, and Learning, 19, 165-181, 2014.

[42] Dyckhoff, A.L., "Implications for learning analytics tools: A meta-analysis of applied research questions. International Journal of Computer Information Systems and Industrial Management Applications, 3, 594-601, 2011.

[43] Mohamad, S.K. & Tasir, Z., "Eduational data mining: A review". Procedia – Social and Behavioral Sciences, 97, 320-324, 2013.

[44] Sao Pedro, M.A., Baker, R.S.J.d., Gobert, J., Montalvo, O., & Nakama, A., "Leveraging machine-learned detectors of systematic inquiry behavior to estimate and predict transfer of inquiry skill". User Modeling and User-Adapted Interaction, 23, 1-39, 2013.

[45] Johnson, L., Adams, S., & Cummins, M., "The NMC Horizon Report: 2012 Higher Education Edition." Austin, Texas: The New Media Consortium, 2012

[46] Wille, R., "Restructuring lattice theory: An approach based on hierarchies of concepts". In I. Rival (ed.), Ordered sets (445–470). Dordrecht: Reidel, 1982.

[47] Wille, R., "Formal concept analysis as mathematical theory of concepts and concept hierarchies". In B. Ganter , G. Stumme and R. Wille (eds.), Formal Concept Analysis (1-34), Berlin: Springer, 2005.

[48] Ganter, B. & Glodeanu, C.V., "Factors and skills". To appear in C.V. Glodeanu, M. Kaytoue, & C. Sacarea (Eds.), Formal concept analysis. LNAI vol. 8478, in press.

Towards Mobile Multimodal Learning Analytics

Laila Shoukry, Stefan Göbel, Ralf Steinmetz

Multimedia Communication Lab - KOM

TU Darmstadt; Germany

{laila.shoukry; stefan goebel; ralf.steinmetz}@ kom.tudarmstadt.de

EXTENDED ABSTRACT

Smartphones nowadays are equipped with an increasing number of sensors which can offer rich information for analytics unobtrusively and their connectivity enabling natural data collection. As a naturalistic assessment of learning experiences should no longer ignore new interaction paradigms and data sources, we argue that a multimodal approach combining different logging information and sensor readings and adapted to mobile learning contexts is required. This can enable deeper insight into interactions in novel learning settings involving smartphones as pure logging of traditional interaction patterns is becoming insufficient. We call this next-generation Learning Analytics (LA) [1] Mobile Multimodal Learning Analytics. Our purpose is to investigate how different smartphone sensors can be used for collecting information which can be useful for LA and what challenges are associated with this approach. Different studies showed the use of smartphones for eye tracking, facial feature extraction, voice analysis and other techniques useful in recognizing cognition states which are considered valuable for LA. Migrating LA, from traditional settings, where they have proven successful [7, 10, 4], to mobile environments to make assessments in natural, non-stationary settings requires considering many new factors influencing the learning process like dynamic context, device capabilities and social interactions [11]. However, not only are capabilities of mobile devices on the rise, but there are also other opportunities offered by smartphones which can be exploited to cope with or even eliminate these challenges. Another considerable challenge associated with gathering data about smartphone users is getting ethical clearance as collecting and disseminating sensor data raises serious privacy and security issues [8]. The front-facing cameras of smartphones can be used for a variety of techniques to measure cognition which can also be used in LA, like eye tracking and facial feature extraction, despite their generally lower resolution in comparison to the back-facing cameras [5]. All mobile phones have built-in microphones which can be used for voice analysis. Studies have shown modest accuracy at measuring emotion and high accuracy at estimating stress [3, 6]. Instead of mouse and keyboard, users of smartphones and tablets predominantly use touch interactions with touch strength and movement additionally introducing new sources of sensory data. In addition, affect can also be measured on smartphones using phone interactions and app usage [9]. As affect detection in an intelligent tutoring environment has already been proven to improve learning effectiveness [2], we argue that collecting multimodal data from smartphones offers unprecedented opportunities for the design of adaptive learning games and applications on mobile devices.

REFERENCES

[1] R. S. J. D. Baker and K. Yacef. The State of Educational Data Mining in 2009 : A Review and Future Visions. 1(1):3-16, 2009.

[2] I. Arroyo, D. G. Cooper, W. Burleson, B. P. Woolf, K. Muldner, and R. Christopherson. Emotion sensors go to school. In AIED, volume 200, pages 17-24, 2009.

[3] K.-h. Chang, D. Fisher, J. Canny, and B. Hartmann. How's my mood and stress?: an efficient speech analysis library for unobtrusive monitoring on mobile phones. In Proceedings of the 6th International Conference on Body Area Networks, pages 71-77. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011.

[4] U. J. E. H. Beth Dietz. Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective. Journal of Interactive Online Learning, 2013.

[5] C. S. Ikehara, J. He, and M. E. Crosby. Issues in implementing augmented cognition and gamification on a mobile platform. In Foundations of Augmented Cognition, pages 685{694. Springer, 2013.

[6] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pages 351-360. ACM, 2012.

[7] L. P. Macfadyen and S. Dawson. Mining LMS data to develop an early warning system for educators: A proof of concept. Computers & Education, 54(2):588-599, 2010.

[8] D. McMillan, A. Morrison, and M. Chalmers. Categorised ethical guidelines for large scale mobile hci. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, pages 1853-1862, New York, NY, USA, 2013. ACM.

[9] R. Rana, J. Reilly, R. Jurdak, W. Hu, X. Li, and J. Soar. Affect sensing on smartphone-possibilities of understanding cognitive decline in aging population. arXiv preprint arXiv:1407.5910, 2014.

[10] A. Skinner, C. Russo, L. Baraniecki, and M. Maloof. Ubiquitous augmented cognition. In D. Schmorrow and C. Fidopiastis, editors, Foundations of Augmented Cognition. Advancing Human Performance and Decision-Making through Adaptive Systems, volume 8534 of Lecture Notes in Computer Science, pages 67-77. Springer International Publishing, 2014.

[11] J. Traxler. Defining, discussing and evaluating mobile learning: The moving finger writes and having writ...The International Review of Research in Open and Distance Learning, 8(2), 2007.

Validation of Game Scenarios for the Assessment of Professional Competence: Development of a Serious Game for System Managers in Training

Hans G.K. Hummel[a], Rob J. Nadolski[a], Desirée Joostenten Brinke[a],
Liesbeth K.J. Baartman[b]

[a]Welten Institute
Research Centre for Learning, Teaching and Technology
Faculty of Psychology and Educational Sciences
Open University of the Netherlands
[hans.hummel, rob.nadolski, desiree.joosten-tenBrinke]@ou.nl
[b]University of Applied Sciences of Utrecht
liesbeth.baartman@hu.nl

ABSTRACT

Serious games hold potential for fostering the acquisition of more complex problem solving skills in professional practice. However, until now the empirical evidence on these workplace learning effects of serious games has remained rather scarce. Therefore such games have hardly been adopted for assessment purposes. This article argues why a validation method is needed that points out and controls what and where learners are learning from games. The core of the method entails mapping the learning activities on the performance indicators and outputs, as derived from the formal attainment levels in vocational education. In this study we have elaborated and applied a validation method for the development of a scenario-based assessment game for system managers in (secondary vocational) education. The method provides a general procedure, practical guidelines, and assessment forms, that can be used beyond this educational context and domain by those interested in more dynamic and motivating ways to assess the acquisition of complex skills in workplace learning.

Keywords
Serious games, seamless assessment, validation method, professional competence, game scenarios

1. INTRODUCTION

Several authors have argued the strength of games as assessment engines [1, 2]. Gee and Schaffer argue that games are good learning engines because they are first good assessment engines [1]. Interest in and use of serious games for learning has grown over the last decade, but until now the empirical evidence on the professional learning effects of serious games has remained rather scarce. As a consequence such games have hardly been adopted for assessment purposes yet.

For the true adoption of serious games for both learning and assessment, we first of all need to employ some type of validation method that makes us understand better what a learner is learning from playing the game, to what degree, and in which contexts while at the same time no sacrifices are made to reliability, and validity of assessment and to the core essence of the highly dynamic interactive nature of games.

The study presented in this article will describe a method to validate game scenarios for the assessment of professional competence, and describe the application of this method on an assessment game that was developed for system managers within secondary vocational education. Core to the approach and developed game is that all performance indicators (as were derived from the formal attainment level) have been clearly mapped on the learning activities and outputs (within the game scenario). We will describe this validation method and argue why our approach can be useful beyond this educational context and domain for those interested in more dynamic and motivating ways to formatively assess professional competence in action.

The remainder of this introduction will now further explain the need for seamless assessment using scenario-based gaming (section 1.1), explain the need for more transparency using a validation method (section 1.2), and introduce the educational context and assessment game (and its scenario) we have used for this study (section 1.3). The validation method itself (section 2) and the game obtained by applying the method (section 3) will then be elaborated and presented in subsequent sections. We will conclude (section 4) with an evaluation of this validation method and suggest future research.

1.1 Seamless assessment in games
The main challenge involved with creating games that assess competencies key to workplace learning is to consider their highly dynamic interactive nature, being unobtrusive to the player, while not sacrificing reliability and validity in the assessment process. The integration of formative assessment within game play should be 'seamless'. Gee and Shaffer expect games to reform current educational assessment (mainly facts and knowledge), and lead to radical transformation towards learning for 21st century skills [1]. As they and other educationists state it: "Assessment is the tail that wags the dog of learning". Assessment of learning is the process of using data to demonstrate that stated learning objectives are actually being met by a learner [3, 4]. Creating scenarios with learning activities closely aligned with the learning objectives is key in ensuring learning goals will be met. In other words, assessments need to be aligned with learning objectives and with the learning activities (i.e. constructive alignment [5, 6]). As a consequence, the domain of assessment is in transition from a perspective with an emphasis on summative assessment to a more balanced assessment program in which summative assessment is balanced with formative assessments. Redecker et al. describe the stepwise development from 1st generation in the 1990s (automated administration and scoring) and 2nd generation in the 2000s (more adaptive) to 3rd generation from 2010 (continuous, unobtrusive, more formative assessment), which is supposed to further include behavioral tracking in immersive and game-based environments [7]. For several formative assessment methods, like giving feedback, feed up

and feed forward, working with rubrics or self and peer assessment, evidence is available that formative assessment is effective for learning [8]. However, as stated before, a major impediment for exploiting games for the formative assessment of more complex skills purposes is the current lack of proof on the efficacy and impact of serious games on learner achievement [9, 10]. This type of learning and assessment requires more complex, seamless but also transparent validation methods and assessment procedures, which we will present in this article. According to Corti: "Serious games will only grow as an industry if the learning experience is definable, quantifiable and measurable" [11].

1.2 Validation methods and assessment

Validation is the process of building arguments to support the claims and decisions that are made from assessment scores [12]. Validation methods evaluate whether assessment achieves its purposes, i.e. the fitness for purpose[13]. Fitness for purpose encompasses the way results of an assessment are interpreted and used by the educators and students. A validation model provides information whether the assessment is in line with the learning objectives and the learning scenario. This implicates that assessments are representative for and balanced over the learning objectives. Validation has to be argument-based using two kinds of arguments [12]. Interpretive arguments specify the proposed interpretations and uses of scores and are used as a starting point for validation. This includes the analyses of performance indicators and the learning activities. Validity arguments then evaluate the plausibility of these interpretations and uses by evaluating to which extend performance indicators are covered by learning activities and the availability of assessment procedures, instructions and forms. A validation method to assess complex skills therefore has to involve different kinds of evidence, like the implementation of assessment procedures, the translation of the learning objectives into the learning scenario, the expert judgments and the documentation.

With the implementation of competence-based education comes the need for other, more dynamic forms of assessment. More classical forms of testing and assessment have gradually been replaced by so called competence assessment programs (CAP), where the mere application of classical criteria for reliability and validity no longer suffices. Such programs and the examination projects within also need to comply to the new demands of competence-based assessment, like acceptability, authenticity, meaningfulness, cognitive complexity, fairness, fitness for purpose, reproducibility of decision, educational consequences, self-regulated learning, transparency, comparability and costs and efficiency [14]. For this educational context, the general quality of education and assessment is considered to be problematic by the various stakeholders involved [15]. Evidently, serious games offer great potential for CAP as they provide highly engaging and dynamic environments with authentic tasks at the core for the development of professional competence.

The qualitative problem with assessment is largely caused by the lack of clear design criteria and standards for examination which make that the various examination projects differ largely and are hard to compare. Another important aspect that has hampered the uptake of more dynamic forms of education and assessment (like serious games) is the lack of sufficient evidence-based research into these innovations, even though research did reveal that the way assessment is conducted is a major determinant of reaching graduation.

1.3 Example game: Events Agency Galema

Secondary vocational education is (in the Netherlands) largely offered by so called Regional Education Centres, large training institutes that on the average serve about 30,000 students each. The attainment levels for each profession and educational level (of which there are four) are documented in so called Qualification Dossiers which have been accredited on a national level. The Stichting Praktijk Leren (SPL) is the Dutch Foundation on work-based learning that operates closely together with branch organizations for various professions, and has the responsibility to stimulate, coordinate and coach the development of more innovative ways of professional training and assessment in secondary vocational education. Recently SPL decided to aim for an integral, transparent and proven system of examination projects that covers all Core Tasks within the Qualification Dossiers. To validate such assessments currently two instruments are available and used for their design: quality criteria for CAP [14], and frameworks of the educational inspection [15]. The development of the learning and assessment games is done by applying the EMERGO game platform [16]. Eventually SPL strives to have each core task assessed by a game. The curriculum for training System Managers on attainment level 4 has been taken as first pilot, one learning game and one assessment game have been developed so far. This study deals with the developed assessment game which is called 'Events Agency Galema' (name of the case and virtual contractor).

The examination project 'Events Agency Galema' is based on a practical case that has to be done within a virtual company 'ITadvice4U'. This means that students are largely assessed while carrying out tasks on their computer. The game is based on a scenario with consecutive learning activities that have to be carried out within the virtual company by guidance of a virtual coach, and partly by having face-to-face talks with the teacher in real life. The main task that is given to the student: develop a new system for project management for a agency that organizes events. For this, the student performs a needs-analysis, distills a functional and technical design of the new system, draws up a plan for developing the new system, tests a first version, and writes a test report. This all yields a total study load of about two days to pass the assessment game.

2. METHOD

This section will briefly introduce the validation method we used and its four steps (section 2.1), then explain the first two steps (Performance Indicators and Game Scenario) in section 2.2, and on the last two steps (Mapping and Assessment Procedures) in section 2.3. The next section will present the results of applying this validation method on the Galema game.

2.1 Validation method

The validation method essentially is comprised of executing following four steps procedure: (1) Analyze the Qualification Dossier, with having Performance Indicators as its outcome; (2) Develop learning activities, with having a detailed Game Scenario as its outcome; (3) Evaluate to which extend performance indicators are covered by learning activities, with having a Mapping of intended performance on activity; and (4) Distill Assessment procedures, instructions and forms. The method is not merely consecutive, but iterative as well. For instance, evaluation takes place in various rounds, leaving opportunity to adjust the game scenario. The core of the method can be depicted as in Figure 1.
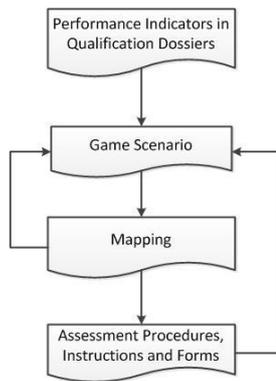
Figure 1. Stepwise validation method

2.2 Performance indicators and game scenario

For Step 1 we analyze the Qualification Dossier. As stated before, the attainment levels and performance indicators of vocational education for various professionals and levels are nationally documented and accredited in so called Qualification Dossiers. The structure of each Qualification Dossier is comprised of Core Tasks, that each contain Work Processes. Each Work Process is described with Performance Indicators and Wanted outcomes. The assessment game under study aims at the core task 1 'Develop (parts of) information -or media systems' which is comprised of five work processes. For brevity reasons, we only look at the first work process ('Analyze the needs of the contractor'). This process has two outcomes (i.e., a full and correct overview of (O1): the information needs of the contractor organization; and (O2): the conditions and possibilities within the organization) and six performance indicators P1 up till P6 (see Table 1). Step 1 ends by filling a validation table with four columns: performance indicator; place of occurrence within the scenario, information the game - if applicable - contains for the assessment, and information the document output or face-to-face talk - if applicable - contain for the assessment (see Table 1 which is already filled for the game example that is further described in Section 3). The third and fourth column of this table will reveal if and which performance indicators have to be assessed beyond the digital part of the game (i.e., computer program) and how. The second and third column will reveal which activities of the scenario will be used for assessment purposes. The third column describes the information the computer program contains for assessment purposes, like logging data on progress, sent mails and document outcomes.

For Step 2 we need to have a fully elaborated and adjusted game scenario. At this point it is good to further define scenario-based serious games as simulated task environments, which have been modeled after real-life situations that often include a sequence of learning activities that involve complex decision making, problem solving strategies, intelligent reasoning and other complex cognitive skills. Such games are often based on professional or academic role adoption and modeled after expert behavior. Students are left in charge to deal with complex problems according to professional or scientific standards. Real-life situations display ambiguity and conflicting information and offer a large degree of freedom. The EMERGO approach and toolkit is dedicated towards such scenario-based games, and has been used for the development of the scenario and game under study [16]. Before game development actually starts, for each activity is identified how

students are expected and allowed to perform: what does the student do, with whom, with what tools and resources, and with which support (teacher, fellow student, or embedded in the game)? Does task performance result in a product, and if so, how will this be evaluated? Is a sufficient result needed before students can carry on? Which interactions with other participants and the digital part of the game are foreseen during and after carrying out activities? All (possible) interactions for each activity are exhaustively described, also in terms of required tools and resources.

2.3 Mapping and assessment procedures

For Step 3 a number of iterative evaluation rounds to establish the content validity are carried out in which the performance indicators will be mapped on the game scenario. The performance indicators for core task 1 (Develop (parts of) information- or media systems) were used as they could be derived and formulated by SPL based on the Qualification Dossier. Two assessment experts mapped indicators on activities and outputs as contained in the game scenario, using Table 1 independent from each other. In case not all indicators could be mapped, this was reported back to the project team which then decided either to incorporate the assessment of more indicators in the scenario or leave them out.

For Step 4, clear instructions are needed for the teachers / assessors that will be using the assessment game. In this case some performance indicators are left out of the digital part of the game and will be assessed during face-to-face talks. As results of Step 4, Assessment forms are developed for each core task (and the individual scoring on performance indicators for each work processes), as well as for the overall assessment that refers to a weighted sum of the performance scores on all five work processes and constitutes the final output of the validation method.

3. RESULTS

This section provides the results of applying the validation method on the Galema game. Again, we first describe the first two steps in section 3.1, and then the last two steps in section 3.2.

3.1 Game activities for assessment

Two assessment experts found that most performance indicators could be mapped on activities in the (adjusted version) of the game scenario. Some Work Processes could only be partly mapped on the scenario. And for some Performance Indicators it was decided they could better be assessed completely beyond the computer program (but still as integral part of the game scenario) by means of a face-to-face talk with the teacher (i.e., the game role that is indicated with the label 'Mr. Jonkman'). The Validation table for work process 1.1 is provided in Table 1.

Eventually, a detailed scenario of about 50 pages containing 55 learning activities could be agreed upon (Step 2), which could be used for the evaluation / mapping rounds in Step 3.

3.2 Assessment procedure and instructions

During Step 3, for each work process, a scoring model could be derived after it was decided what performance indicators were assessed (where and how). Such scoring models also clarify to what extend the assessor can use information obtained from outcomes (like written needs analysis) or contained in the computer program (like reports sent or logging of actions).

Attainment of each performance indicator is assessed by either I (insufficient), S (Sufficient) or G (Good). It was further decided

and documented (in the assessment manual) that several criteria should be considered by the teacher when assessing work processes (for example: task is clearly described; the current way of working in projects is clearly described; problems of the current system are clearly mentioned; demands on the new system are clearly mentioned; wishes (may haves) and requirements (must haves) are clearly distinguished). Furthermore, the assessment manual contains example questions for the face-to-face talks and provides information for the game-role the teacher has to fulfill.

Table 1. Validation for work process 1.1 (Analyze the needs of the contractor)

Performance indicators	Content validation (place in scenario / activity student)	Assessment Information (system)	Assessment Information (in documents or by Jonkman)
(P1) Collect sufficient information by both interviewing and document analysis.	Virtual talks with employees Galema; F2F talk with Mr. Jonkman: Must prepare questions		F2F talk with Mr. Jonkman: Does student pose relevant and sufficient question?
(P2) Ask for the ideas and needs of employees to get a good overview of the information need within the organization	Virtual talks with employees Galema; F2F talk with Mr. Jonkman: Must prepare questions.		F2F talk with Mr. Jonkman: Does student pose questions about opinions, ideas and needs?
(P3) Consider the wishes of the client in relation with the possibilities when determining the information needs	Make a needs-analysis		Needs-analysis: Does student weigh the wishes and possibilities?
(P4) Show plan to relevant others and adjust them when appropriate	Send report talk with Mr. Boekhorst to him ; Send reports of all talks to coach; F2F-talk with Mr. Jonkman: discuss ideas and adjust analysis; Send needs-analysis to Jonkman, coach and Galema	Report talk with Boekhorst been send to him?; All reports sent to coach?; Has needs-analysis been send to Jonkman, coach and Galema?	F2F-talk with Mr. Jonkman: Does student respond adequately to comments?
(P5) Acquire a full and correct overview of business processes and information streams	Make needs-analysis		Needs-analysis: Does it show practice correctly and completely?
(P6) Verify correctness of acquired information, structure information, and consider conclusions by using available facts and weighing pros and cons.	Make needs-analysis; Report talk with Boekhorst: Verify with him if it is a correct reflection of actual practice	Report sent to Boekhorst requesting him to check for correctness?	Needs-analysis: is document correct and complete with clear structure?

4. CONCLUSION AND DISCUSSION

This study shows it is indeed possible to develop and apply a validation method to validate game scenarios for assessment purposes. Preliminary experiences reveal that an assessment game that results from this validation is indeed more transparent, better documented, and can be more effectively compared and organized. Both students and teachers find this more dynamic way of assessment more motivating and effective. Two teachers that used this assessment game over the last months (with 20 students) report that both the preparation and execution of the examination project is now less labor-intensive.

However, some of the performance indicators were not suitable for e-assessment (i.e., the digital part of the game). Therefore, the face-to-face component is still required. A blended approach (both virtual and face-to-face) with students and teacher "stepping in and out" of the digital part of the game did not appear to be problematic for students and teachers. Current gaming platforms do not yet cater for valid and reliable in-game assessment of all types of activities. For example, the assessment of the more 'soft' communication competence is beyond scope although there are some promising developments with respect to speech recognition and emotion recognition that alleviates the work of the teacher and can prevent students from struggling too long on

ineffective learning paths [17].Validating the content of game scenarios seems to be an important line of future research, and can ensure that serious games are better warranted against the current criticism of not being transparent enough for assessment purposes. The assessment in this case study seems to result in comparable and more efficient assessments. Such advances in adaptive serious games with "embedded assessment" make better visible how learners develop skills and monitor their success, and thus provide teachers with new insights that help them improve their teaching and tutoring. It has remained beyond the scope of this study (which is mainly descriptive) to investigate the impact of different design mechanism upon students' and teachers' opinions with respect to assessment and students' skill development and success. We are currently preparing a study with a larger group of participants in which we will examine the impact of different game guidance mechanics towards students' success. Another limitation of this study is that we do not have enough proof that such assessment games are sufficiently warranted towards fraud on the long run, when larger numbers of students study the same cases. Although we cannot fully exclude such risks, it needs more attention in the design and exploitation. Furthermore, the positive effects of studying just one assessment game (of two days) will be snowed under when the remainder of the curriculum is still classically tested. For this reason SPL is now developing assessment games for all core

tasks within the piloted curriculum. Finally, we also have to see if results found within the domain of system management are generalizable towards other domains.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Gee, J.P., & Shaffer, D.W. 2010. Looking Where the Light is Bad: Video Games and the Future of Assessment. *Phi Delta Kappa International EDge*, 6(1).

[2] Ifenthaler, D., Eseryel, D., & Ge, X. 2012. *Assessment in Game-Based Learning: Foundations, innovations and perspectives*. New York: Springer.

[3] Black, P., & Wiliam, D. 1998. Assessment and Classroom learning. *Assessment in Education, 5*, 7-74.

[4] Chin, J., Dukes, R., & Gamson, W. 2009. Assessment in simulation and gaming: A review of the last 40 years. *Simulation and Gaming, 40*(4), 553-568.

[5] Biggs, J. 1996. Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-364.

[6] Biggs, J. 2003. *Teaching for quality learning at university*. Maidenhead: SRHE.

[7] Redecker, C., Punie, Y. & Ferrari, A. 2012. *eAssessment for 21st century learning and skills*. In: Ravenscroft, A., Lindsteadt, S. Kloos, C.D. & Hernandez-Leo, D. (Eds.) 21st Century Learning for 21st Century Skills. Proceedings EC-TEL, Saarbrücken, 2012, 292-305. Heidelberg: Springer.

[8] Sluijsmans, D. M. A., Joosten-ten Brinke, D., & Van der Vleuten, C. P. M. 2013. *Toetsen met leerwaarde. Een reviewstudie naar effectieve kenmerken van formatief toetsen*. [Formative assessment. A review study on effectiveness of formative assessment]. Den Haag: NWO.

[9] Young, M.F., Slota, S., Cutter, A.B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. 2012. Our Princess Is in Another Castle: A Review of Trends in Serious Gaming. *Review of Educational Research, 82*(1), 61-69.

[10] Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. 2009. Melding the Power of Serious Games and Embedded Assessment to Monitor and Foster Learning: Flow and Grow. In U. Ritterfeld, M. Cody, & M. Vorderer (Eds.), *Serious Games: Mechanisms and Effects* (pp. 295-321). New York: Routledge.

[11] Corti 2011. Proof of Learning: Assessment in Serious Games. Last accessed: 4 May 2013, http://www.gamasutra.com/view/feature/2433/proof_of_l earning_assessment_in_.php

[12] Kane, M. 2006. Validation. In R. Brennan (Ed.), *Educational measurement*, 4th ed. (pp. 17–64), Westport, CT: American Council on Education and Praeger.

[13] Van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartmand, L. K. J., & Van Tartwijk, J. 2012. A model for programmatic assessment fit for purpose. *Medical Teacher, 34*, 205-214.

[14] Baartman, L. K. J. 2008. *Assessing the assessment: Development and use of quality criteria for Competence Assessment programs*. Thesis. University Utrecht: Utrecht, The Netherlands.

[15] Inspectie van het Onderwijs. 2009. Boekhouder of wakend oog. Verslag van een onderzoek bij examencommissies in het hoger onderwijs over de garantie van het niveau. Inspectierapport 2009-16 (april). Available at: www.onderwijsinspectie.nl/actueel/ publicaties/Boekhouder+of+wakend+oog.html

[16] Nadolski, R. J., Hummel, H. G. K., Van den Brink, H. J., Hoefakker, R., Slootmaker, A., Kurvers, H., and Storm, J. 2008. EMERGO: Methodology and Toolkit for Efficient Development of Serious Games in Higher Education. *Simulations & Gaming*. 39, 3 (Sep. 2008), 338-352. DOI = http://sag.sagepub.com/content/39/3/338.full.pdf+html

[17] Bahreini, K., Nadolski, R., and Westera, W. 2014. Towards Multimodal Emotion Recognition in E-learning Environments. *Interactive Learning Environments*. May 2014, 1-16.

Multimodal Emotion Recognition for Assessment of Learning in a Game-Based Communication Skills Training

Kiavash Bahreini, Rob Nadolski, Wim Westera
Welten Institute
Research Centre for Learning, Teaching and Technology
Faculty of Psychology and Educational Sciences
Open University of the Netherlands
{kiavash.bahreini, rob.nadolski, wim.westera}@ou.nl

ABSTRACT

This paper describes how our FILTWAM software artifacts for face and voice emotion recognition will be used for assessing learners' progress and providing adequate feedback in an online game-based communication skills training. This constitutes an example of in-game assessment for mainly formative purposes. During this training, learners are requested to mimic specific emotions via a webcam and a microphone in which the software artifacts determine the adequacy of the mimicked emotion from either face and/or voice. Our previous studies have shown that these software artifacts are able to detect face and voice emotions in real-time and with sufficient reliability. In our current work, we present a software system architecture that unobtrusively monitors learners' behaviors in an online game-based approach and offers timely and relevant feedback based upon learner's face and voice expressions. Whereas emotion detection is often used for adapting learning content or learning tasks, our approach focuses on using emotions for guiding learners towards improved communication skills. Herein, learners need to have an opportunity of frequent guided practice in order to learn how to express the right emotion at the right time. We assume that this approach can address several issues with the current trainings in this area. We sketch the research design of our planned study that investigates the efficiency, effectiveness and enjoyableness of our approach. We conclude the paper by considering the challenges of this study.

Keywords

Formative assessment; communication skills; multimodal emotion recognition; serious gaming; software development; feedback provision.

1. INTRODUCTION

Communication skills are becoming more important in modern society. This causes a greater demand for communication skills training as it used to be in the past. Furthermore, there are still a lot of people that were educated in an area that communication was not that important in society as it is today [1]. Such people might have insufficient communication skills. Indeed, higher standards for communication skills - increasing the need for extended and varied practice - are applicable for all ages and foster new approaches towards communications skills trainings that are more suited for modern man. A flexible and online training program can offer a solution in which learners are able to practice a lot on a regular basis to enhance their communication skills. Nowadays, learners must attend specific courses that use face-to-face approach, turning them into quite inflexible training programs as far as freedom of place and time is concerned. The courses are costly and often demotivating because of the needed repeated practice. This frequent practice is inevitable for the learners in order to master the skill on a sufficient level. An additional problem is the shortage of trainers that can provide communication skills in face-to-face situations [2]. Serious games are games developed for educational purposes rather than entertainment. Such games seem adequate for addressing issues with online face-to-face trainings and to a certain extent also deal with the shortage of trainers as it can make a more effective use of their limited availability [3]. These games compared to the regular e-learning solutions are 1) engaging, 2) motivating, 3) user centric, 4) goal oriented, 5) more interactive, and 6) more personalized [4]. However, there is also an issue with many serious games with respect to their assessment of learning. They often do not reliably assess the learning [5]. Our approach intends to offer an effective training of communication skills while at the same time dealing with assessment of learning within the game. It is not a replacement of the face-to-face training, yet it offers much more flexibility and scalability.

We describe how our FILTWAM software artifacts for face and voice emotion recognition will be used for unobtrusive in-game assessment of learners' progress of learning. FILTWAM is integrated with a game-engine and is used for the development of a serious game for communication skills training. We call our serious game "Communication Advisor". In this game, frequent feedback is provided for guiding learners towards improved communication skills (i.e., formative assessment [6], [7]). We assume that deploying the FILTWAM artifacts for multimodal emotion recognition can lead to better learning. Automated emotion recognition may compensate for the limited number of trainers that are available for the training of communication skills whereas a serious game is suggested because of its motivational strength for fostering learning in which frequent practice is needed for automatic skill mastery situations [2]. It is commonly acknowledged that emotions are important factors in any learning process, since it influences information processing, memory and performance [8]. Our previous research on face emotion recognition and voice emotion recognition has shown that it is possible to measure emotions from these two software artifacts with sufficient reliability in real-time [9, 10]. The easiest way and the most accessible equipment for gathering data for emotion recognition are webcams and microphones. There is valuable information inside face and voice expressions that can mirror affective aspects of learning, but that in the case of communication skills can also inform learners' progress towards their mastery. In our work, we focus on the latter usage of face and voice emotions, whereas most research deals with using emotions for adapting learning content or learning tasks. This insight has led to the research and development of affective tutoring systems [11]. Adequate

communication is not only about the 'what' (i.e., the content of the communication), but also about the 'how' (i.e., the way this content is delivered). Emotions need to be aligned with the message to have its intended effect. It is important for the learners to learn how to express the correct emotion at the right time. Feedback can guide the alignment between emotions and message and therefore expected to be significant in communication skills training. Also, feedback based on emotional states may enhance the learners' awareness of their own behavior.

Communication Advisor allows learners to practice in so called conversation snippets. In each snippet the learner receives feedback. The feedback is based upon the detected learners' mimicked emotions in their expressed conversation part of the snippet. The content of the message has been chosen from text alternatives. Text alternatives are used so that the serious game can easily detect the content of the message in the snippet. Our approach proposes an example of unobtrusive in-game assessment for providing timely feedback to the learner. This assessment is meant for learning although the game could also be used as part of the setup for the assessment of learning in which case only limited or even no feedback would be given to the learner. It is assumed that deploying the FILTWAM artifacts for multimodal emotion recognition in Communication Advisor lead to better and more enjoyable learning of communication skills. In sum, to characterize the novelty of our work, we propose multimodal emotion recognition for assessment of learning in game-based communication skills training. In this paper, section 2 introduces the software system architecture. In section 3, we sketch the research design of our planned study that investigates the efficiency, effectiveness and enjoyableness of Communication Advisor in which its current prototype is used to illustrate our approach in more detail. Section 4 discusses the challenges and provides few suggestions to conduct that study.

2. SOFTWARE SYSTEM ARCHITECTURE

We propose a loosely coupling system design of FILTWAM and EMERGO in our architecture. EMERGO is used as the game-based engine, and content manager of Communication Advisor. EMERGO is a methodology and open source toolkit for the development and delivery of serious games [12]. To connect Communication Advisor and EMERGO to FILTWAM, we followed the web service approach and developed a web service on both client and server sides of our architecture. The communication protocol between the game, EMERGO, and the software artifacts are established through a developed web service. Figure 1 demonstrates the components of the architecture and the relationships between them. We placed nine components in the data flow diagram: 1) Learner, 2) Browser, 3) EMERGO web service client, 4) Face emotion recognition software (FERS), 5) Voice emotion recognition software (VERS), 6) Real-time data file for FERS, 7) Real-time data file for VERS, 8) EMERGO serious game engine, and 9) EMERGO web service. The first three components are situated within the client side. The other six components have been placed in the server side. The learner opens the browser in the client side (number 1 in the figure 1) and launches the EMERGO serious game engine (number 2). The EMERGO serious game engine component deals with triggering specific feedback messages (i.e., content of feedback) of the game, manages game rules, and influences training content. The engine calls the EMERGO web service component (number 3) and triggers feedback based upon the rules. In the client side,

the EMERGO web service client component that is already executed by the learner (number 4) is responsible to call FERS and VERS components (number 5 and 7). These two components generate a real-time data file for the FERS and for the VERS components (number 6 and 8). FERS and VERS do face emotion recognition and voice emotion recognition from the webcam input data that they receive from the learner. The EMERGO web service client uses both data files (number 9 and 10). It sends the real-time emotion data to the EMERGO web service (number 11) and this data will be sent through this component to the EMERGO serious game engine (number 12), then to the browser and the learner (number 14 and 15). As it is shown in figure 1, the learner will receive all feedback through a single browser on the client side.
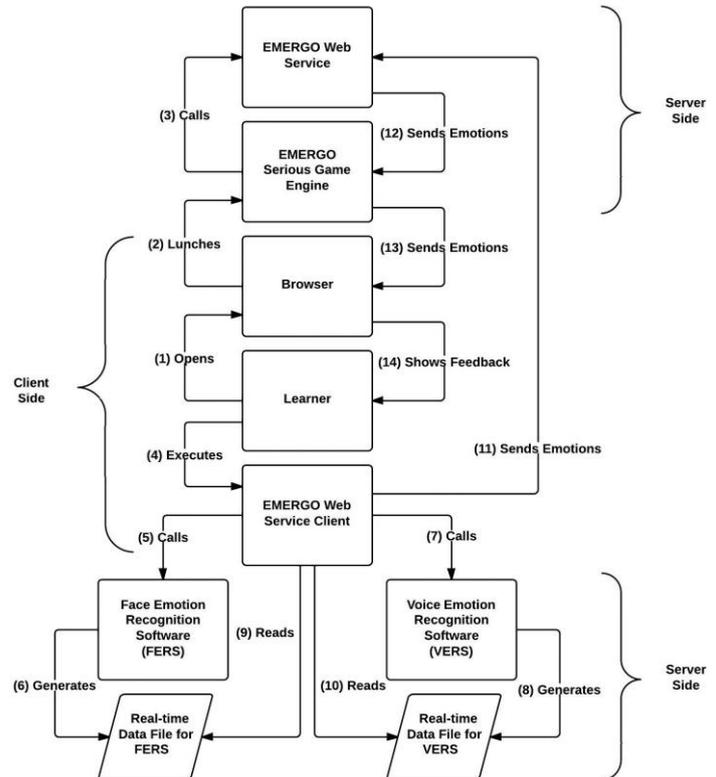

Figure 1. Data flow of the software system architecture.

3. COMMUNICATION ADVISOR

We deploy Communication Advisor for skill-based learning using the EMERGO game engine. During development, the EMERGO method and toolkit is used. During design the EMERGO method is used. For testing and running the game, the EMERGO game engine is used. We offer the learner a varied set of audio-visual micro stories (tasks) in an online game-based setting. This approach intends to improve the alignment between the nonverbal behavior and the verbal message of the learner interaction snippets being part of such micro stories. An interaction snippet consists of (1) choosing an alternative (i.e., spoken text-message and emotion (which is either happy, sad, surprise, fear, disgust, angry, or neutral) from a small list of text-alternatives, (2) speak the message with the chosen emotion in front of the webcam, (3) reaction from the conversation partner, and (4) presenting feedback as text format. The feedback is provided based on the message of the chosen text-alternative and the shown emotion by the learner. Figure 2 illustrates the prototype of Communication Advisor when a task is presented to the learner.
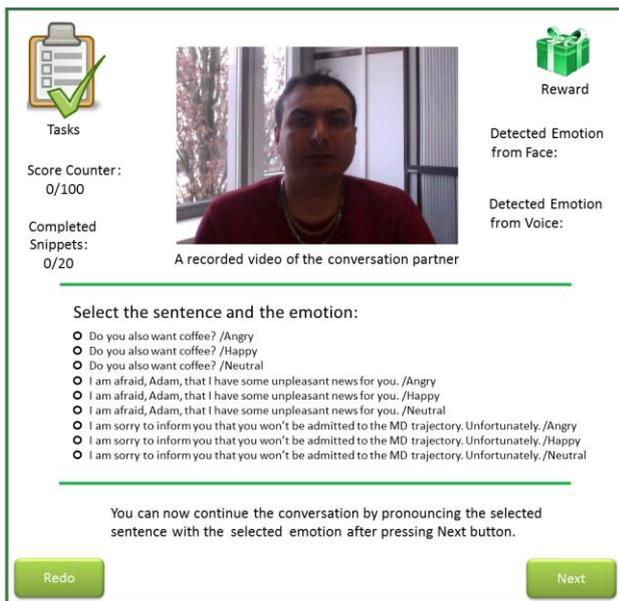
Figure 2. A screen capture of the Communication Advisor**'s prototype when a snippet is given to the learner.**

3.1 Method

We follow the EMERGO approach for game design and development and use several game design guidelines from the literature [13, 14]. Ten micro stories (tasks) including one or more interaction snippets (twenty in total) will be offered to the learner within the game. Each snippet will be sequentially presented through webpages. The learner will see the following components on the screen (see figure 2): 1) a score counter, 2) a tasks' list, 3) a number of completed snippets, 4) a recorded video of the conversation partner, 5) a reward component, 6) the text transcripts and the emotions, 7) an instruction text on how to proceed the game, 8) a button to proceed the game, 9) a button to redo the snippet 10) a message for the detected emotion of the face, and 11) a message for the detected emotion of the voice. The label of the detected emotion will be shown in green when the learner has presented the emotion from the chosen alternative correctly and will be shown in red when the learner did this incorrectly (see Figure 3). The reward mechanism will offer a prize when the learner expresses five out of twenty consecutive snippets correctly. The text transcripts and instructions for the micro stories will be selected from an existing OUNL training course [15] and a communication book [16]. Figure 3 presents an example of the feedback to the learner. The score counter component is increased by 5 and the completed snippets component is increased by 1 when the learner completes a snippet. The selected sentence and the emotion are also marked.

The entire game happens over the course of fifty minutes in an online virtual learning environment that is equipped with a microphone and a webcam. During our research, all learning sessions will be captured through an integrated webcam and a 1080HD external camera to capture and record the emotions of the participants. The interactions of the learner with the mouse and the keyboard on the computer screen will also be captured through Silverback usability testing software version 2.0. We examine five different versions of the game on its learning effectiveness, efficiency and attractiveness. The first version will only provide the feedback based on the chosen text alternative. The second and third versions will provide feedback on the combination of the chosen text alternative and one sensor (either webcam or microphone). The fourth

version will provide feedback on the combination of the chosen text alternative and both sensors. Finally, the fifth version (i.e. control version) of the game does not provide any feedback. Some interaction snippets will be used as initial assessment for pre-skill measurement. Some will be used as summative assessment for post-skill measurement. In both pre-skill and post-skill measurement, no feedback and no reaction from the conversation partner will be shown to the learner. In addition, a pre-questionnaire (before the game) and a post-questionnaire (after the game) will collect other data that might be relevant for explaining individual differences in learning (e.g., attitude towards learning through games, motivation, and effort). We will especially examine the progress of each learner during the game play using assessment data from the sensors (recognized emotions) and chosen text alternatives within the micro stories. We will measure effectiveness of the learning progress after completing the game through comparing pre-skill measurement and post-skill measurement. The efficiency will be measured by measuring the study time. We measure enjoyableness of the learning by a questionnaire.
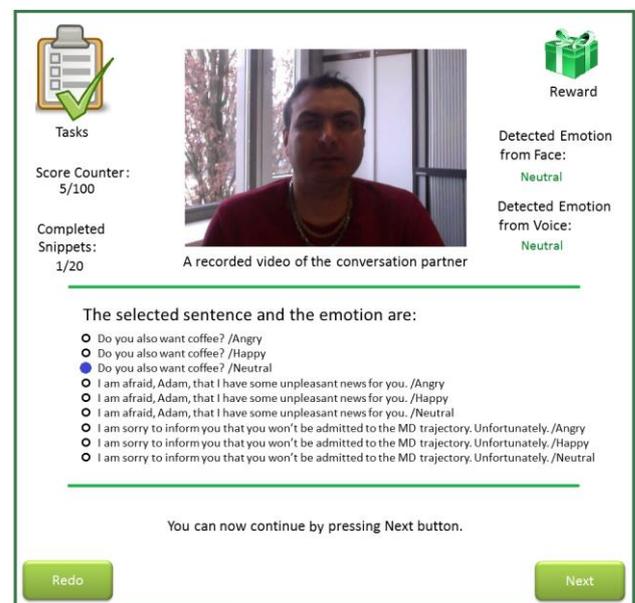
Figure 3. A screen capture of the Communication Advisor**'s prototype when the feedback is provided.**

4. CHALLENGES AND DISCUSSION

In this article we described the integration of our FILTWAM software artifacts with an online serious game using the EMERGO platform and engine. The face and voice emotion recognition software and the game enable unobtrusive in-game assessment of learners' progress are assumed to provide more adequate feedback than a game without those software artifacts. Both software artifacts are able to detect learner's emotions for the purpose of providing timely feedback (i.e. instantly after the learner has pronounced the conversation snippet) within Communication Advisor. We focus on analyzing emotions for improving learner's communication skills whereas previous research on emotions in e-learning has mainly dealt with adapting learning content and tasks based on emotion detection. It is an interesting avenue for future research to combine both approaches. For Communication Advisor this would imply offering different learning tasks based on learners' detected emotions during the game play (for example, easier or more difficult tasks). Speech recognition was left out in our approach in order to keep our research scope as simple as possible. There are still some issues with

speech recognition (validity, reliability, performance) that might obscure our findings. However, our architecture can be easily extended with speech recognition. It is acknowledged that speech recognition would be a very valuable extension to our approach as the training situation is more aligned with the real situation. Such alignment is assumed to be important for transfer (i.e., application of the learning in other contexts). A technical challenge of our approach might be the performance of our game once more content and rules would be added. At this moment of writing we did follow a pretty straightforward approach with web services, but a stronger integration between the two software artifacts of FILTWAM and the EMERGO platform is needed if it occurs to be an issue to deliver the feedback timely.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Brantley, C. P., and Miller, M. G. 2007. Effective Communication for Colleges. *Thomson Higher Education*. Cengage Learning. 11 edition (Sep. 2013).

[2] Hager, P. J., Hager, P., and Halliday, J. 2006. Recovering Informal Learning: Wisdom Judgment and Community. *Lifelong Learning Book Series*. Springer, Dordrecht.

[3] Westera, W., Nadolski, R., and Hummel, H. 2013. Learning Analytics in Serious Gaming: Uncovering the Hidden Treasury of Game Log Files. *Games and Learning Alliance (GaLA) Conference*. (Paris, France, Oct. 23-25. 2013).

[4] Connolly, T. M., Boyle, E. A., MacArthur, E. Hainey, T., and Boyle, J. M. 2012. A Systematic Literature Review of Empirical Evidence on Computer Games and Serious Games. *Computers & Education*. 59, 2 (Sep. 2012), 661-686. DOI= http://dx.doi.org/10.1016/j.compedu.2012.03.004.

[5] Shute, V. J., Ventura, M., Bauer, M., and Zapata-Rivera, D. 2009. Melding the Power of Serious Games and Embedded Assessment to Monitor and Foster Learning: Flow and Grow. *The Social Sciences of Serious Games: Theories and Applications. In U. Ritterfeld, M. Cody, & M. Vorderer (Eds.), Serious Games: Mechanisms and Effects*. 295-321. New York, Routledge.

[6] Frunza, V. 2014. Advantages and Barriers of Formative Assessment in the Teaching-learning Activity. *Procedia - Social and Behavioral Sciences. 4th World Conference on Psychology, Counseling and Guidance (WCPCG-2013)*. 114 (May 2014), 452-455. DOI= http://dx.doi.org/10.1016/j.sbspro.2013.12.728.

[7] Gulikers, J. T.M., Biemans, H. J.A., Wesselink, R., and van der Wel, M. 2013. Aligning formative and summative assessments: A collaborative action research challenging teacher conceptions. *Studies in Educational Evaluation*. 39, 2 (Jun. 2013), 116-124. DOI = http://dx.doi.org/10.1016/j.stueduc.2013.03.001.

[8] Pekrun, R. 1992. The Impact of Emotions on Learning and Achievement: Towards a Theory of Cognitive/Motivational Mediators. *Journal of Applied Psychology*. 41 (Oct. 1992), 359-376. DOI= http://dx.doi.org/10.1111/j.1464-0597.1992.tb00712.x.

[9] Bahreini, K., Nadolski, R., and Westera, W. 2014. Towards Multimodal Emotion Recognition in E-learning Environments. *Interactive Learning Environments*. May 2014, 1-16. DOI= http://dx.doi.org/10.1080/10494820.2014.908927.

[10] Bahreini, K., Nadolski, R., and Westera, W. 2013. FLITWAM and Voice Emotion Recognition. *Games and Learning Alliance (GaLA) Conference*. (Paris, France, Oct. 23-25. 2013).

[11] Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., and Bigdeli, A. 2008. How Do You Know That I Don't Understand? A Look at the Future of Intelligent Tutoring Systems. *Computers in Human Behavior*. 24, 4 (Jul. 2008), 1342-1363. DOI= http://dx.doi.org/10.1016/j.chb.2007.07.008.

[12] Nadolski, R. J., Hummel, H. G. K., Van den Brink, H. J., Hoefakker, R., Slootmaker, A., Kurvers, H., and Storm, J. 2008. EMERGO: Methodology and Toolkit for Efficient Development of Serious Games in Higher Education. *Simulations & Gaming*. 39, 3 (Sep. 2008), 338-352. DOI = http://sag.sagepub.com/content/39/3/338.full.pdf+html.

[13] Schell, J. 2008. The art of game design: a book of lenses. *Burlington, MA: Morgan Kaufmann Publishers Inc*. San Francisco, CA, USA.

[14] Kiili, K., de Freitas, S., Arnab, S., and Lainema, T. 2012. The Design Principles for Flow Experience in Educational Games. *4th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES'12). Procedia Computer Science*. 15 (Dec 2012), 78-91. DOI = http://dx.doi.org/10.1016/j.procs.2012.10.060.

[15] Lang, G., and van der Molen, H.T. 2008. *Psychologische gespreksvoering [Psychological interviewing]*. Open University of the Netherlands. Heerlen, The Netherlands.

[16] Van der Molen H.T., and Gramsbergen-Hoogland, Y.H. 2005. *Communication in Organizations: Basic Skills and Conversation Models. Psychology Press*. ISBN 978-1-84169-556-3. New York, USA.

Identifying Engagement with Learning in Serious Games

Claudia Ribeiro[1] , Élise Lavoué[2] , Karim Sehaba[3] , João Pereira[1] ,
Jannicke Baalsrud Hauge[4]

[1] INESC-ID, Lisbon, Portugal
Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisbon, Portugal
[2] Université Lyon 3, MAGELLAN, LIRIS, UMR5205, F-69355, France
[3] Université Lyon 2, LIRIS, UMR5205, F-69676, France
[4] Bremer Institut für Produktion und Logistik, Hochschulring 20, 28359 Bremen, Germany

ABSTRACT

Within the health sector, the effectiveness is a key factor for the deployment of Serious Games for training purposes. Engagement is central factor relating games to learning. Therefore, identifying player's level of engagement is an important aspect to consider when assessing Serious Games effectiveness. This paper describes an approach to identify player's engaged-behaviours based on users interaction traces. A trace is a history of users' actions collected in real-time during their inter-action with a computer system. The process of identifying the level of engagement consists of transforming low-level behaviours (e.g. clicks) in contextualized high-level behaviours. The proposed approach is exemplified by applying it to the identification of engaged-behaviours in a serious game for training clinical procedures on sepsis treatment protocol.

Keywords

Serious Games, Engagement, Assessment

1. METHODOLOGICAL APPROACH FOR IDENTIFYING ENGAGEMENT IN DIGITAL GAMES

Bouvier et al. [2, 1, 3] have proposed a qualitative approach to identify users' engagement and qualify their engaged-behaviours from their traces of interaction. A trace is the history of users' actions collected in real-time from their interactions with a computer system. The basis of the authors' approach is to transform low-level traces of interaction (e.g. clicks) into meaningful information represented in higher-level traces (i.e. activities). These high level traces correspond to engaged behaviours. A behaviour corresponds to a chain of actions (i.e. an aggregation of actions) actually performed by the user in the interactive system. From an operational point of view, a player is engaged if s/he manifests at least one engaged-behaviour. Considering some chains of actions rather than single actions provides comprehensive contextual information on behaviours and thus, facilitates their understanding.

To decide whether a behaviour reflects, or not, an engagement, the authors considered the question of the learners' motives and needs that determine engagement. For that, based on the Self-Determination Theory [7] four types of engaged-behaviours were identified [2]: 1.) environmental, 2.) social, 3.) self, and 4.) action. To demonstrate how this approach can be applied to Serious Games, we describe next an example of use to identify engaged-behaviours within the sepsis fast track Serious Game. This example is based on real data collected during on-the-job training sessions with doctors in training at an academic hospital.

2. IDENTIFYING ENGAGED-BEHAVIOURS IN SEPSIS FAST TRACK SERIOUS GAME

In the Sepsis Fast Track Serious Game the player assumes the role of a physician and her/his goal is to confirm if the patient is or is not a case of sepsis, fill out the sepsis fast track form in the hospital IT system and carry out the appropriate medical interventions. The sepsis fast track form is composed of three main parts. The first is concerned with the systemic inflammatory response syndrome criteria, which are the body temperature, heart rate, and respiratory rate and is completed by the triage nurse. The second part registers the information confirming or not confirming the sepsis case suspicion is registered. It includes the registration of the arterial blood pressure, checked using the game mechanics Examine ECG Monitor, the exclusion criteria, checked using the game mechanics Examine Patient Chart, the Glasgow coma scale, checked using the game mechanics Examine Patient, and the lactate value, checked using the game mechanics Examine Arterial Blood Gas. In this part the sepsis fast track activation is also validated. Validating a sepsis case means, asserting that the patient identified by the triage nurse is in fact a sepsis case. Finally, the third part of form is concerned with the information about the therapy administered to the patient. This should only be used if a sepsis case is confirmed and validated. Also, the time when the patient had the therapy (hemocultures, antibiotherapy and fluid therapy) should be registered. A detailed description of the sepsis fast track Serious Game and respective game mechanics can be found in [6, 5].

The sepsis fast track is a point-and-click Serious Game meaning that all interactions are traduced by a combination of clicks in game objects (e.g. patient, nurse, ECG monitor). All clickable objects are represented in Figure 1. An example of a medical intervention is to confirm the suspicion that the patient has a sepsis. This medical intervention requires the physician to ask the patient about his/her current symptoms, to verify the Glasgow Comma Scale, to perform a blood test to verify the lactate value and examining the patient. In terms of interaction with the game (primary traces) these actions are represent b a sequence of clicks, moves and choices (game actions) spread throughout a game session (these are the operations in Figure 1). In order to aggregate them into meaningful actions a set of rules have to be defined to transform these primary traces into an intermediate transformed-trace (actions). An example of such rule is (see Figure 1)

(CLICK patient.timestamp - CLICK help.timestamp <= 2000) AND (CLICK ECG.timestamp - CLICK help.timestamp <= 2000) AND (CLICK nurse.timestamp - CLICK help.timestamp <= 2000) AND (CLICK computer.timestamp - CLICK (CLICK clipboard.timestamp - CLICK help.timestamp <= 2000)

This rule identifies actions that the player undertook to understand better how the game works. To identify the actions relevant for the medical intervention previously described a similar rule would have to be implemented. Finally, it is necessary to aggregate these actions in order to identify the high-level traces, meaning the activities. In the context of Serious Games the activities are the unit of gameplay that are closer connected to the learning objectives of the game. In order to identify the activities a second level of transformation rules need to be defined and applied to the intermediate transformed-traces. An example of such rule is:

(consult action list.timestamp - consult help screen.timestamp <= 5000)

With this rule we can identify when and how many times the player tried to in- crease h/her knowledge about the game. By doing this transformation process, it is possible to identify the different dimensions of engaged-behaviours (Environ- mental, Self, Social, Action). In our running example we have shown how to identify environmental-engaged behaviours. A schematic representation of the process of identifying engaged-behaviours in the sepsis fast track Serious Game is depicted in Figure 1.

Combining these two sources of information and comparing it between different game sessions (in the context of sepsis fast track Serious Game, different clinical cases), we will be able to observe how the deviation of optimal performance evolves and in this manner understand how committed and how much effort the player placed in the learning activity (solving a clinical case).

3. CONCLUSION AND FUTURE WORK

In this paper we presented a summary review of definitions proposed to help clarify and understand engagement with learning. Based on these definitions and an approach previously used to identify engaged-behaviours in a social game we have shown how the same approach can be used to identify engaged-behaviours in a Serious Game for clinical education, the sepsis fast track Serious Game. By extracting these high-level behaviours from gameplay activity and comparing it with subsequent game sessions we can potentially have an indication of what was the player level of commitment and effort, and therefore engagement, in the learning activity during gameplay.
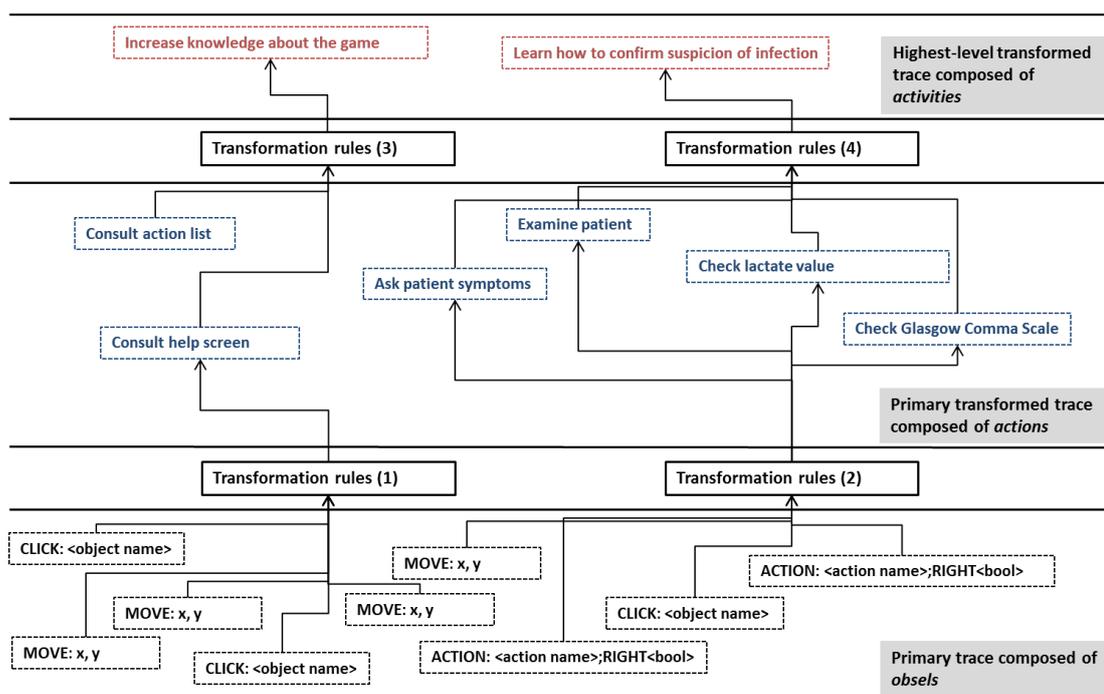

Fig. 1: Identifying engagement with learning: sepsis fast serious game example

By relying on rules to capture the meaning of low-level traces, this also al- low us to identify the situations where there was a deviation of the optimal performance, meaning achieving a certain activity by performing a certain number of wrong actions (intermediate transformed-trace). This can be accomplished by implementing rules that capture sequences of operations that also include operations that are not part of the resulting action.

This work represents the first part of an on-going research work that involves using this approach to identify engaged-behaviours of 17 doctors playing 5 different sessions of the sepsis fast track Serious Game as well as understand is there is a causal effect between the level of engagement and the level of knowledge retention.

4. REFERENCES

[1] Bouvier, P., Lavoué, E., Sehaba, K.: Defining engagement and characterizing engaged-behaviors in digital gaming (in print). Simulation & Gaming

[2] Bouvier, P., Sehaba, K., Lavoué, E., George, S.: Using traces to qualify learner's engagement in game-based learning. In: Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on. pp. 432–436 (July 2013)

[3] Bouvier, P., Sehaba, K., Lavoué, E.: A trace-based approach to identifying users' engagement and qualifying their engaged-behaviours in interactive systems: application to a social game. User Modeling and User-Adapted Interaction pp. 1–39 (2014), http://dx.doi.org/10.1007/s11257-014-9150-2

[4] Garris, R., Ahlers, R., Driskell, J.E.: Games, motivation, and learning: A research and practice model. Simulation & Gaming 33(4), 441–467 (Dec 2002)

[5] Ribeiro, C.: Serious Games for Learning: A Model of Driving Attributes for the Medical Domain. Ph.D. thesis, Lisboa, Portugal: Instituto Superior T´ecnico (2014)

[6] Ribeiro, C., Monteiro, M., Pereira, J.A., Antunes, T., Baalsrud Hauge, J.: Sepsis fast track: A serious game for medical decision making. In: Serious Games Development & Applications (SGDA), 2014 5th International Conference on (2014)

[7] Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 55(1), 68–78 (2000)Wille, R. 2005. *Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies*, In B. Ganter , G. Stumme and R. Wille (Eds.), Formal Concept Analysis (pp. 1-34), Berlin: Springer.

Towards a Hybrid Approach to Learning Analytics, Educational Data Mining, and Personalization for Serious Games

Michael D. Kickmeier-Rust, Christina M. Steiner, Dietrich Albert
Cognitive Science Section, Knowledge Technologies Institute, TU Graz
Inffeldgasse 13 / V
A-8010 Graz, Austria
+43 316 87330636
{michael.kickmeier-rust; christina.steiner; dietrich.albert}@tugraz.at

ABSTRACT

This paper gives a brief overview of approaches to educationally intelligent technologies for serious games and draws the line to the current and future challenges and needs for research and development to increase the quality and impact of learning games. We present expels from our future work and reference an approach to provide educational middleware based on web services. Finally, we outline the foundations of a novel and broader approach to realize in-game data mining, learning analytics, and competence-oriented personalization based on Competence-based Knowledge Space Theory and Formal Concept Analysis.

Keywords

Serious Games, Personalization, Competence-based Knowledge Space Theory, Formal Concept Analysis, Learning Analytics

1. INTRODUCTION

There is little doubt that computer games are an acknowledged educational medium of the 21st century. Serious games capitalize on their core strengths - distilled to the essence: fun, fantasy, curiosity, challenge, and control [15]. These strength lead to an enormous intrinsic motivational potential so that digital games can reach a broad audience and, perhaps most importantly, the can potentially reach learners that are "outsiders" from traditional educational systems. Of course, playing games, in general, is not only one of the most natural forms of human activity but also one of the most natural forms of learning. Children learn to talk by playing with sounds and learn collaboration and strategic thinking when doing role play games.

However, while success stories are most often told from the side of informal learning, there are clear weaknesses and downsides when it comes to formal, curriculum-oriented, school-type learning. Learning in and with virtual environments and games must become more effective and more importantly must become more supportive to teachers. It is important that games are not isolated realities within which a student might or might not practice and learn and from which little information escapes in order to support formatively inspired educational measures by teachers and trainers. The solution to that might be making games educationally smarter. But to equip serious games with "serious educational power" is not trivial – not from the design perspective and not from the technical perspective. So, despite the undisputed great potential and despite a significant hype - some claim that, *"by 2020, almost every professional practitioner will be doing some of their learning in a serious game"* (Parvati Dev[1]) - serious games have not become serious business so far. And we are far from Dev's vision.

The key to that problem is "educational AI"; intelligent techniques that gather data, perform educationally inspired data mining and learning analytics and utilize the outcomes. One the one hand, it is important to provide teachers and trainers with deeper insight into activities, behaviours, and learning progress, on the other hand, it is highly important to enable a reasonable and educational meaningful behaviour of the autonomous game system.

On the one hand, over the past years, influenced by the general economic downturn, many smaller game studios had to go out of business and there have been some high-profile layoffs. On the other hand, we saw some remarkable diversification and creativity in the whole industry – partly thanks to increasing use of AI techniques like procedural generation and to a rebirth of traditional AI genres like simulation games. Recent trends in conventional game AI focus on emergent AI (distributed machine learning), natural language processing, and credibility of characters.

In principle, the trends and rules are the same for the serous game business; on the other hand, these trends indicate the research fields where SG research and development needs to invest double the amount of efforts to produce serious, credible, fun, and above all effective games. Accordingly, the three most important trends in a technical sense we see are (i) intelligent agents supporting and accompanying the players – in gameplay and story but also in learning and development, (ii) natural language processing, and (iii) educationally controlling emergence (free and emergent sandbox games need to be empowered monitor and control educational intentions – seamlessly integrated in the games flow).

2. EDUCATIONALLY INTELLIGENT

Along with the industrial revolution and the invention of assembly-lines, a new view of learning and teaching emerged. Standardized tests emerged during World War I, teaching machines were invented in the 1920s, instructional films came up in the 1940s, or programmed instruction came up in the 1950s, in the 1960s and 1970s the educational television was invented and educational computer technology became popular in the 1980s and 1990s. [19]. Now it was time to make the mere technology smarter. Thus, the modern approaches to adaptive educational technology can be traced back to Lee Cronbach and Richard Snow [5] and their *Aptitude-Treatment Interaction* theory. They argued that different aptitudes (abilities, skills, knowledge) require different treatments (i.e., ways of teaching,

---

1 http://www.healthcommcapacity.org/429/

instructional methods, curricula designs, or conditions and modalities). Further work in the area was also leveraged by Benjamin Bloom [3] who posed the *2-sigma theory, which*, essentially, states that tailored tutoring results in performance superior by 2 standard deviations (sigma) in comparison to regular teaching. This is trivial on the one hand, but ultimately, the idea is that learners are not overburdened by the educational demands and therefore quickly frustrated but, at the same time, not under-challenged and therefore bored. Thus, the concept is closely related to the ideas of game adaptation. Commenced by Bloom, psychologists, educationists, and technicians attempted to develop technology that is able to take the role of a private teacher and to intelligently provide individual learners with suitable tutoring [6].

A well-elaborated approach to achieve educational intelligent adaptation of autonomous systems is so called *Competence-based Knowledge Space Theory* (CbKST). Most concisely, this theory is an extension of the originally behavioural *Knowledge Space Theory* by Doignon and Falmagne [7] which portrays the idea that a knowledge domain is characterized by a set of problems or test items / tasks. The *knowledge state* of a learner is identified on the subset of problems s/he is capable of solving. Due to mutual dependencies between the items captured by *prerequisite relations*, not all potential knowledge states are supposed to occur. The set of all possible states is called a *knowledge structure*. To account for the fact that a problem might have several prerequisites (i.e., and/or-type relations) the notion of a *prerequisite function* was introduced. Recent updates of this rigorous mathematical approach are described by [8] The principle idea of CbKST is to separate the observable behavioural aspects, i.e., whether a learner masters a problem or test item, from the not directly observable construct of aptitude/ability/knowledge behind the performance. The entities of aptitude matching the concept of problems or items are called skills or competencies. Equal to knowledge structures, prerequisites between competencies establish competence structures that include only meaningful sets of competencies a person can have. To give an example, having the competency to multiply integers but, at the same time, not having the competency to add integers are not meaningful or plausible. The relationships between the competencies and problems/items are established by a *skill function*. Such function assigns a collection of subsets of competencies (i.e., *competence states*) to each problem that are relevant for solving it. By associating competencies with the problems/items of a domain, a knowledge structure on the set of problems is induced. The latent competencies can be uncovered on the basis of a person's observable performance. Due to the mathematical nature of the approach, it perfectly serves as a functional basis for computer systems.

3. FROM ASSESSMENT TO LEARNING ANALYTICS

Obviously, assessment and appraisal of student's achievements and activities is a key task of teachers and the very basis for controlling teaching activities. Recently, [17] of the *Berkeley Evaluation and Research Center* established four principles of assessment: (i) the developmental perspective (viewing learning as a longer process that cannot be assessed by one shot), (ii) instructional fidelity (match between objectives for assessment and important learning objectives), (iii) teacher management and responsibility (valid for classroom assessment), and (iv) quality of evidence (e.g., standards of fairness, etc.). This is the human side. – From the perspective of autonomous computer systems, this is a non-trivial task. Consequently, classifying and assessing the learners was subject of a large body of research

(see [16] for an overview) and has a very long tradition. In the context of autonomous intelligent, adaptive tutorial systems (ITS / ATS), classifying and modelling the learner plays a crucial role and bears one of the most substantial challenges to research. The immersiveness of computer games adds another dimension and challenge for equipping the technical systems with the capability to perform assessments, appraisals, and classifications on an unobtrusive level in order not to harm the gaming experience.

In previous work, we developed and validated an approach named micro adaptation, which is based on the notions of CbKST (e.g., [1] [11] [12]). The principle idea is to monitor each activity a learner or a group of learners exhibits and to interpret the behaviour in terms of available or lacking competencies or cognitive states such as motivation. Originally, this concept was developed in the European ELEKTRA project (http://kti.tugraz.at/css/projects/elektra/) and advanced in the following 80Days project (www.eightydays.eu). In recent projects we developed generic web services around the micro adaptation framework. The service oriented architecture (as described by[4]) is based on a set of recommendations, policies and practices for software architectural design which implements business processes and it is using loosely coupled components which are arranged to deliver a certain level of service or set of functionalities [10]. The services are (partly) available and accessible through the service catalogue platform of the Serious Games Society (http://services.seriousgamessociety.org/).

The service approach has been implemented and evaluated in primary level maths games. One example is the Sonic Divider [13], a tool to practice the formal sequence of solving divisions at the level of 3rd grade. In this case, the micro adaptation framework builds upon a domain model that includes about 100 atomic competencies (including number dimensions, knowledge about sequences, rounding of numbers, etc.). The system identifies correct and incorrect actions of the learners and updates an underlying probability model of available competency states (in the sense of CbKST). This kind of believe model then is used to triggered highly targeted interventions (such as guidance or feedback), perfectly matching the competency levels of the learners. Equally, we used the approach in a multiplication game; in the 1x1 Ninja tool, we established a domain model for 2nd grade multiplication skills including the number dimensions for multiplicand and multipliers. The tool can give tailored feedback and automatically adapts the difficulty level of the multiplication tasks according to the performance of the learners. In school studies we could show that suitable and individualized interventions are superior to no, non-individualized, or simple right/wrong statements through the game system.

A highly interesting application of the micro adaptation framework was realized in the context of the European Next-Tell (www.next-tell.eu) project. In this example we developed a full teacher control suit that allows realizing educational sessions in virtual worlds (such as Second Life or OpenSim). The tool analyses the log files from the virtual world in real-time and – in greater detail – post hoc and provides teachers with activity stats, chat summaries, probabilities over competencies and competence states (based on heuristic-based analyses of activities), and real time messages (e.g., in case of unwanted activities such as using inappropriate language). In example studies with Norwegian and Austrian children, meeting up and learning English together in an OpenSim environment, we could demonstrate clear benefits for the teacher who had the

opportunity to monitor and document activities and, more importantly, to look into language competencies [14].

The presented approaches, and of course many others, can increase the educational value of serious games significantly. However, there are still challenges for research. One of the major problems of serious games in general and in particular what concerns the adoption of "educationally intelligent middleware" is the high costs. One of the major downsides of CbKST related approaches is a comparably high authoring load to establish domain models and to formalize curricula on a small level of granularity. The service approach and open accessible web service catalogue such as the one of the Serious Games Society are a big leap forward. A next step would be making the systems smarter by implementing mechanisms of real time data mining as source of evidence for a deeper and broader individualization of the game and more detailed information for educators.

4. TOWARDS EDUCATIONAL DATA MINING IN GAMES

The title (Helvetica 18-point bold), authors' names (Helvetica 12- In a recent project named LEA's BOX (www.leas-box.eu) we are currently extending the CbKST framework by more bottom-up driven approaches for data mining coming rather from the psychological than the technical side, for example *Formal Concept Analysis* (FCA). FCA, established by [18], aims to describe concepts and concept hierarchies in mathematical terms. The starting point of the FCA is the definition of the "formal context" (also called learning domain) which is defined as a triple composed of objects, attributes, and relationships between them. The meaningful combination of objects and attributes establishes a so-called formal context which results in a formal lattice structure. The advantage is that such approach can be used to recognize patterns in data (e.g., log file data from games) and to identify relationships and superordinate/subordinate concepts [9].

However, mere data mining approaches applied in a single application such as a learning game (instead of being used to analyse huge data sets) is problematic since most often these bottom-driven approaches cannot deliver sufficiently robust and valid information for real-time adaptation. This relates to the well-known dilemma in learning analytics of using top-down versus bottom-up analytic approaches [2]. Instead of purely data-driven approaches of pattern recognition in learning-related data, there are increasingly claims for a more top-down-like strategy, i.e. that reasoning about data requires robust and well-elaborated psycho-pedagogical foundations. Starting analytics from questions and psycho-pedagogical theory and models of teaching and learning, from conceptions of knowledge, of how learning and learning success take place, is considered one of the main challenges in the emerging field of learning analytics. Presently we are working on a hybrid approach that might be suitable and fruitful for establishing the competence-based learning analytics framework for LEA's BOX, thus aiming in realising feasible, efficient, effective and pedagogically meaningful analysis and sense-making of learning related data. The general idea is to identifying rough/preliminary competences based on psycho-pedagogical considerations, information, and consultation. This makes a-priori pedagogical work simpler, more effective, and in the end cheaper. FCA is then used in a learning environment (perhaps a game) to analyse the activity and performance information and the refine the competence models and the learner models on the fly. This give the originally rough "understanding" of the system a more detailed and individualized dimension.
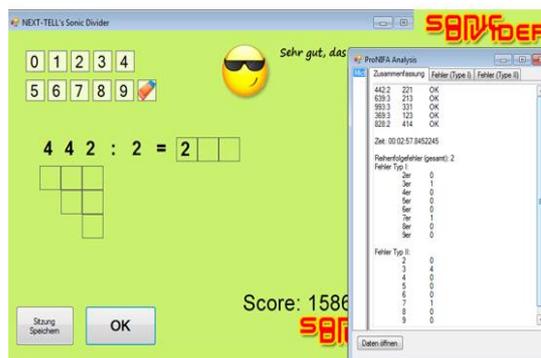

Figure 1. Screenshots of micro adaptive games (from top to bottom): ELEKTRA, Feon's Quest from 80Days, Sonic Divider and 1x1 Ninja from Next-Tell.

5. CONCLUSIONS

In this paper we presented a rough overview of the state of educationally intelligent technologies, about today's and future challenges. We also referenced some role models of how leading adaptive technologies can be used to individualize gaming and learning experiences and, not least, to allow teachers looking into the activities and the performance within those closed little immersive worlds of computer games. Still, major challenges need to be tackled. We argue that real impact of games in education needs smarter, cheaper, and easier to access educationally intelligent middleware. The direction we are pursuing presently, that is developing hybrid approaches of top-down and bottom-up mechanisms shall advance the current technology. We can support the sequence of in-game assessment and data mining > reasoning > learning analytics > individualization and adaption > opening information channels to teachers and learners in a meaningful and effective way.

Presently we are collection activity data in Austrian schools using virtual environments (OpenSim) and teacher-led activity tracking software. This is a joint effort of the Next-Tell project and LEA's BOX. The first results are promising and let us believe that a hybrid and service oriented approach is a significant step forward in making serious games smarter and more effective.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Augustin, T., Hockemeyer, C., Kickmeier-Rust, M., & Albert, D., 2011. *Individualized Skill Assessment in Digital Learning Games: Basic Definitions and Mathematical Formalism*. IEEE Transactions on Learning Technologies, 12 Aug. 2010. IEEE computer Society Digital Library. IEEE Computer Society.

[2] Baker, R. S. J. D., & Yacef, K. 2009. *The State of Educational Data Mining in 2009: A Review and Future Visions,* Journal of Educational Data Mining, 1(1), 3-17.

[3] Bloom, B., 1984. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring, Educational Researcher, 13(6), 4-16.

[4] Carvalho, M. B., Bellotti, F., Berta, R., De Gloria, A., Gazzarata G., Hu, J., & Kickmeier-Rust, M. (in press). *The Journey: a service-based adaptive Serious Game on probability*.

[5] Cronbach, L., & Snow, L., 1977. Aptitudes and instructional methods: A handbook of research on interactions. New York: Inrington Publishers.

[6] De Bra, P. M. E., 2008. *Adaptive hypermedia*, In H. H. Adelsberger, Kinshuk, J. M. Pawlowski, & D. Sampson, (Eds.), Handbook on Information Technologies for Education and Training (pp. 29-46), Berlin: Springer.

[7] Doignon, J.-P., & Falmagne, J.-C. 1999. *Knowledge spaces*. Berlin: Springer.

[8] Falmagne, J.-C., & Doignon, J.-P. 2010. *Learning Spaces: Interdisciplinary Applied Mathematics*. Berlin: Springer.

[9] Ganter, B. & Glodeanu, C.V. in press. Factors and skills. To appear in C.V. Glodeanu, M. Kaytoue, & C. Sacarea (Eds.), *Formal concept analysis*. LNAI, vol. 8478, Berlin: Springer.

[10] Hurwitz, J., Bloor, R., Baroudi, C., Kaufman, M. 2007. *Service Oriented Architecture For Dummies.* {For dummies. Wiley.

[11] Kickmeier-Rust, M. D., Mattheiss, E., Steiner, C. M., & Albert, D. 2011. A psycho-pedagogical framework for multi-adaptive educational games. International Journal of Game-Based Learning, 1 (1. 45-58.

[12] Kickmeier-Rust, M. D., & Albert, D. (Eds.), 2012. *An Alien's guide to multi-adaptive educational games.* Santa Rosa, CA: Informing Science Press.

[13] Kickmeier-Rust, M. D., & Albert, D. 2013. *Gamification and intelligent feedback mechanisms for a division learning tool*. In P. Escudeiro and C. Vaz de Carvalho (Eds.), Proceedings of the 7th European Conference on Games based Learning (pp. 290-296), October 2-3, 2013, Porto, Portugal.

[14] Kickmeier-Rust, M. D., & Albert, D. 2013. *Learning Analytics to Support the Use of Virtual Worlds in the Classroom.* In A. Holzinger & G. Pasi (Eds.), Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data. Lecture Notes in Computer Science, 7947, 358 - 365.

[15] Malone, T. W. 1981. Toward a theory of intrinsically motivating instruction. Cognitive Science, 4, 333–369.

[16] Sampson, D., Karagiannidis, C., & Kinshuk, 2002. *Personalized learning: Educational, technological, and standardisation perspective.* Interactive Educational Multimedia, 4, 24-39.

[17] Sloane, K., Wilson, M., & Samson, S. 1996. *Designing an embedded assessment system: From principles to practice.* University of California, Berkeley: BEAR Report Series, SA-96-1.

[18] Wille, R. 2005. *Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies*, In B. Ganter , G. Stumme and R. Wille (Eds.), Formal Concept Analysis (pp. 1-34), Berlin: Springer.

[19] Wulfeck, II, W. H. 2009. *Adapting instruction*. In D. D. Schmorrow et al. Eds.), Augmented cognition, Lecture Notes in Computer Science, 5638, 687-695. Berlin: Springer.