A HOLISTIC APPROACH TO SCORING IN COMPLEX MOBILE LEARNING SCENARIOS

Marcel Gebbe, Matthias Teine and Marc Beutner
Chair of Business and Human Resource Education II, University of Paderborn, Germany

ABSTRACT

Interactive dialogues are key elements for designing authentic and motivating learning situations, and in combination with learning analysis they provide educators and users with the opportunity to track information related to professional competences, but mind-sets as well. This paper offers exemplary insights into the project NetEnquiry that is concerned with the development of an m-learning app which provides learners with complex learning scenarios. In these scenarios, we combine different approaches to scoring and assessment. A specialty of the assessment is that it adapts to scenario settings given by the lecturer, and runs mainly in stealth mode. These settings consider formal correctness, social and communicative skills, but also norms that influence whether a given answer is correct, or not.

KEYWORDS

Scoring, dialogues, decision-making, assessment, formative learning analysis

1. INTRODUCTION

Assessment in e-learning in general, and digital game-based learning in particular reached a point where its opportunities are far beyond what would be possible with traditional testing formats. Particularly the user-interactivity and the adaptability of complex, authentic learning environments to the individual learners are the key to both high quality formal and summative assessment (cf. Scalise, K. 2006, p. 4). To realize such benefits, it is necessary to design task- and interaction-formats that allow us to gather meaningful evidence about the learners' performance, based on an effective scoring and score reporting (cf. Scalise, K. 2006, p. 4). Hence, underlying competences (see Roth, H. 1971; Reetz, L. 1984) can be observed and scored by focusing on the learners' visible performance in action (see Sloane, P. / Dilger, B. 2005, p. 7). Therefore, it is inevitable to create an evidence model that defines the connection in-between (see Shute, V. J. et al. 2009, pp. 309). The data and information gathered via such a stealth assessment can ultimately improve learning (cf. Shute, V. J. 2009). Tracked information about the learners' actions, his movements, events, made decisions, etc. can be used to design personalized debriefing sessions where the learners get detailed feedback, and where implemented solutions can be reflected based on sample or expert solutions (see Loreto, I. D. et al. 2013, pp. 169, Raposo, F. et al. 2013, p. 37, Reinmann-Rothmeier, G. 2005, p. 217). This does not only provide the learners with valuable information regarding how to improve, but it also allows the teachers to become the student's learning process facilitator, rather than the learners' opponent (see Shute, V. J. et al. 2009, pp. 298, Sykes, J. 2006, p. 4).

In the research and development project NetEnquiry, funded by the German 'Federal Ministry of Education and Research' (project ID 01PF10008) we are building a mobile learning app framework that allows to offer complex, role based and cooperative learning scenarios, including a holistic scoring system. Thus, the app is observing several learner processes to score and assess learning and offer a base for individual and team oriented reflection on the learning processes. For example, we use interactive, adaptive dialogues to observe the learners' competences and mind-sets. Following, we want to offer several insights into the scoring system and some challenges behind it.

2. THE NETENQUIRY APPROACH

NetEnquiry is a research and development project that aims to conceptualize, develop, implement, and evaluate both an m-learning tool (iOS app) and its web-based version that provide learning scenarios for the banking and finance sector. The Chair II for Business and Human Resource Education of the University Paderborn leads the project, and collaborates with representative organizations from all parts of the German universal banking system, to safeguard that their needs are met. This means that private sector commercial banks, saving banks, and industrial and agricultural credit co-operations are involved particularly into the creation process of the learning scenarios. Overall, NetEnquiry provides three complex learning scenarios:

1) Specific complex credits and handling of complicated clients,
2) Investment banking and consultancy concerning securities, and
3) Documentation of counselling and talks with clients in the field of investment.

The scenarios are usable for initial or further vocational education and training, and we conceptualized the app for the use in blended-learning scenarios. Thus, NetEnquiry can be used integrated into the everyday work-routine, in enterprise settings, or in formal educational settings in vocational schools as well. However, the learning scenarios focus on authentic situations from the world of work, and the learners have to cope with tasks that typically occur in reality (simulated practice). The learning situations and the task base on real processes which we observed by shadowing employees of the co-operating banks, analyzing internal process data (e. g. handbooks, field of activity-descriptions, organigrams, tools) and problem centered discussions.

The learners form teams of three each of them with a different role (e.g. loan consultant, loan manager, and head of department). For example, the loan consultant is the person that has direct contact to the clients; therefore, he is responsible for providing the team members with the needed information to solve the learning scenario successfully. Due to this, providing wrong information or missing out to provide information lead to wrong decisions of the whole team at succeeding stages of the learning process. The role of the loan manager is to provide feedback on loan applications, particularly in difficult cases, and he has specific instruments to review such applications. The information provided by the loan consultant influence this decision directly. However, the head of the department knows about strategic decisions, new strategies, and he has the competence to make decisions on a level of higher importance. Therefore, if the head does not provide all strategic information, or when he decides in a wrong way, this effects directly the work of all team members.

The quintessence is that the learners are required to co-operate, to support each other, and to communicate effectively because all decision-making and problem-solving processes are interdependent. Otherwise, they cannot solve the given problems successfully. Moreover, the result and the provided solutions for the learning scenario depend on the individual competences of each of the team members, and the solutions can only be as good as the weakest part of the team. Thereby, on the one hand, we address collaborative learning, and the development of social and communication skills. On the other hand, this offers multiperspectivity, and fosters the learners' understanding for the complex interdependences of their real-life actions.

Within the app, the learners have three different kinds of tools that help them to deal with the scenarios: *orientations tools*, *workplace tools*, and *learning support tools*. Depending on the role, the scenario, and the progress the learners can access only those tools that really are necessary to solve the scenario successfully:

- The orientation tools aim to help the learners getting involved into the scenario (e.g. a role-specific introduction video, a description of the own profile, etc.).
- The workplace tools help the learners to cope with the business challenges of the scenario (e.g. a telephone mailbox, an email tool, a loan calculator, etc.).
- The learning support tools help the learners to manage their learning processes (e.g. an interactive calendar, a mind-mapping tool, a tool to take notes, a photo documentation tool etc.).

However, particularly the loan consultant has to deal with clients, and therefore we had to design a complex dialogue system. Generally spoken, dialogues can be observing vs. interactive or sequential vs. interlinked and based on a parallel network-structure. To create authenticity and realism, dialogues need to depict interaction with real people and conditions of (operational) reality, such like information overload or the lack of information, as well as emotional reactions of the dialogue partner (e. g. clients). Interactivity then means that the learners are involved actively, they make decisions by clicking on given answer-options, and thus they change the way the dialogue goes. This is particularly the case with dialogues based on a "interlinked parallel network structure [which] is a decision tree in which each branch leads to a different

outcome" (Beutner, M./ Pechuel, R. 2014, p. 175). Moreover, the opportunity of the learners to actively influence how the story of the learning scenario goes, and their variability based on the already made decisions are main drivers of the learners' motivation (see Westphal, A. 2009, p. 134). In NetEnquiry we use interactive and interlinked dialogues in which the client´s responses are provided by short videos. So, the structure of the dialogue is individualized.

Following, we will describe how the interdependence within the roles, the use of the three different kinds of tools, and the dialogue structure found their way into designing an assessment and scoring system.

3. ASSESSMENT AND SCORING IN NETENQUIRY

In NetEnquiry, we paid a lot of attention to the design of the assessment- and scoring-mechanism to support motivational aspects and feedback based reflection. The main principle behind the assessment is to judge the player´s actions based on their estimated real-life effectiveness (cf. Abt, C. C. 1970, p. 51, 58), and we use scores to provide the learners with a numerical measure of their performance (cf. Sykes, J. 2006, p. 4). This shows that we pursue different aims with our assessment and scoring system, and therefore we combined different scoring- and data collection-approaches as we will show now (cf. for similar approaches e.g. Raposo, F. et al. 2013).

As different teams are playing the scenarios at the time, there is an individual scoring, and a team scoring. Since a number of activities of the players affect the scoring opportunities of the other players; we decided that the team scoring is more important than the individual scoring. This decision had a strong influence on the actual design of the scoring system. Individual scores are related to individual learning tasks, tests, and assessments. Team scores are related to their decision-making, their working processes, their information transfer resp. communication, and the quality of their overall solution of the learning scenario. Moreover, as we will show during this article, the scoring provides an estimation related to the following three aspects (cf. Prensky, M. 2007, p. 121):

1) Did the learners stay within, or did they broke the rules of the scenario?
2) How close did the learners come to the ideal solution of the scenario?
3) How did the learners do, compared to their competitors?

A special focus was also on the scoring and assessment of decision-making processes within the learning scenarios, which is a topic difficult to grasp. In discussions with our partners from the banking sector it became clear that effective economic processes and methods of 'business handbooks' do not always match with the real life circumstances, especially in those cases where situational human competence is needed. For example, strategic habits and moral-ethical norms can lead to a decision that diverges from idealistic economic processes. Moreover, two or more conflicting economic goals can come into case within one situation. Hence, decision making in such situations needs to be trained, scored and assessed transparently for both learners and coaches. This will lead to a basis for discussion and reflection on learning and teaching processes (see Reinmann-Rothmeier, G. 2005 o. 141). Of course, trivial scoring systems and assessment methods alone won´t fit this challenging task. Following we will show how we addressed the different outlined challenges.

3.1 Activity and Quality-Based Quantitative Scoring

Contest and competition is what drives us when playing, as it is about winning or losing against others (cf. Rieber, L. P. 1996, p. 44.). Therefore, learners need information about how they are currently performing compared to the others, and scores provide us with an easy way to quantify the performance of learners. Thus, scores are a way to provide feedback and to let the learners know if they are progressing towards their goal (cf. Becker, K. 2007, p. 29), and thereby they enhance the learners' engagement (cf. Sykes, J. 2006, p. 4). The value of the score assigned to the learners depends on the quality of the performance (see Juul, J. 2005, p. 36), and can be visualized e.g. by using high-score tables (see Prensky, M. 2007, p. 121).

Within the NetEnquiry project, however, we use high-score tables as well. As we provide high-score tables for both an individual score and a team score, the learners can always see how they are performing compared to other learners, and other teams. In the projects' approach, we provide the learners with an absolute, and a relative score. Thereby, the current best absolute score is the benchmark for the relative score,

which is an easy to interpret indicator for the learners then to see how they are doing. Within the app, the learners can always see their score, and they can see the data either in a tabular form, or as a bar chart. Anyway, following we want to describe how the learners can get score points, and how scoring takes place on an individual, and team level.

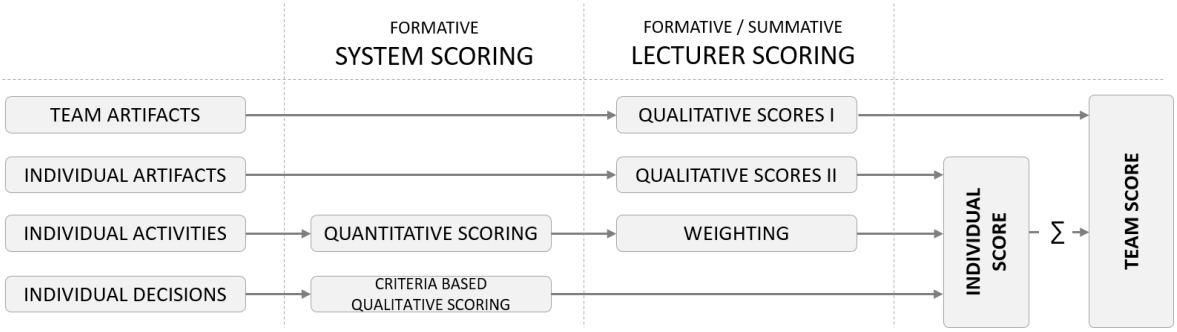

Figure 1. Overview about the NetEnquiry scoring system

The above standing Figure 1 provides you with a comprehensive overview about the NetEnquiry scoring system, which we will describe more in detail now.

3.1.1 Scoring of the Tools

In each of the different scenarios, all three players get individual scores for making use of the different orientation, work, and learning support tools. Moreover, the learners get team scores particularly for the use of work-related tools such like for making use of the mailing systems, the client information system, or the loan management system.

Orientation Tools: We decided to score the learners' use of the orientation tools to reward them for dealing with the app actively, even if these tasks are rather insignificant for the work process itself and not match winning. Anyway, they are important for the learner to understand the own role within the process and to be able to work in the right direction. Due to the nature of the orientation tools, the learners can receive a score here only once. For example, the learners receive 2 points when they fill in their own players' profile, or when they read the description of the others player profiles. These role-related scores are also summed up to a team score. As all teams consist of three learners, it would be possible to receive 6 points as a team by filling in all player profiles.

Workplace and Learning Support Tools: In contrast to the orientation tools, the workplace and learning support tools are designed to be used multiple times during the problem-solving process. The scores here assigned to each of the learners, not the team as a whole – they are role-related. However, partially the score bases on the quantity of use of the tools, but the learners have to reach different threshold to get higher scores. E.g., a learner will get a score of 1 for writing 5 work related emails to his team members, for writing 2 emails to his lecturer, or for writing 1 email to a client. Therefore, 15 mails written to team members result in a score of 3. Based on the importance of the different tools for the learning process, and the purpose of their use, weight varies.

Unfortunately, this quantity based approach leads to a serious problem, because the quantity of use does not say anything about the quality. Therefore, we decided to give the lecturer the opportunity to rate the quality of some quantitative oriented scores, where the quality becomes highly important and the system is not able to evaluate the quality on its own. This weighting of selected quantitative scores can be done formative or summative. Regarding the communication via email and the so-called quick notes tool (in-team communication), where arguments should be shared, we included such a quality evaluation opportunity. A lecturer can do so by assigning a weight from 0-times - which means bad quality of communication, to 3-times - which means excellent quality of communication (cf. Figure 1). The trainer determines the weight by using a slider, and the default value is 1-times. Thus, the final score for the communication is the product of the weight x original score. Moreover, some of the process-oriented tools are explicitly documenting the learners' decision-making processes and their results. Here, NetEnquiry uses not only quantitative but two kinds of qualitative data as well:

a) choosing an alternative through selection, or
b) by entering data (numbers, words, arguments).

Depending on the kind of data, the system is able to automatically score and evaluate the data put into the tools, by comparing it with the complex scenario settings, or to ask the lecturer to bring in his expertise (criteria based qualitative scoring, cf. Figure 1). Unfortunately, we do not have a semantic analytic system for complex texts at the moment. However, the scenarios can include conflicting solutions, where e.g. the strategic and moral-ethical norms come into case (see 3.2 for details).

Right now we would like to showcase an example regarding the so-called 'KDF-Tool' (Kapitaldienstfähigkeit, financial capacity to pay a monthly rate), which is offered next to a loan management system, a customer management system and a collateral tool. The KDF tool is available in a loan scenario and an investment scenario. It is useful to calculate the debt service coverage, and the income available for regular investments on a monthly base. However, for each information put into the KDF-tool via the different drop-down menus, the learners will receive a score of 2, regardless whether the solution was right, or not. This shall engage them, to move on, as they see progress. As the quality of decisions matters most, the system will check the quality of the solution, and the learners will receive a score higher than 2 (up to 6, depending on the importance) when they put into the correct solution.

As explained above, the right solutions depend on processes, rules, norms etc. of the actual institution (e.g. bank). For an adequate score, each institution or lecturer is able to pre-define what is the 'right' decision regarding the institutional norms. Thereby, they define, e.g., if the credit manager has only little or ample scope to agree on a credit when the financial capacity of the clients is low, but the purpose of the credit could be to pay for hospital-costs (ethical-moral dilemma). Moreover, we have variables that can be set within the scenario (e.g. income, spendings, residual debt, kind and amount of collaterals, instalment, external shocks (devaluation), expected chance of survival (cause of disease), etc.). Thereby, the rules for the scenario are set, and the app evaluates if the learners broke the rules, e.g. by not using their scope, or if they stayed within the rules. This information is particularly useful for reflection- and debriefing-sessions.

3.1.2 Scoring of Tasks and Tests

Within the different scenarios, the learners have to solve learning tasks and testing tasks. Learning tasks need to be solved by all learners together, and they are not automatically evaluated but by the lecturer. The learners receive a score only for completing the task. The test tasks are evaluated by the system automatically, and the team will receive 1 score point, independent from the quality of the solution. For a correctly dissolved task, the team will get 3 score points. Testing tasks are classical instruments like multiple choice and an optional addition for lecturers, but obviously not the core aims of the NetEnquiry-system.

3.1.3 Lecturer Scoring

In addition to the scores that are assigned automatically by the NetEnquiry app, the trainer has the opportunity to assign scores for

(1) the communication with the trainer,
(2) the quality of the mind-maps created by the learners,
(3) the quality of the solutions of the learning tasks,
(4) the quality of the whiteboard-use in the team, and
(5) for the quality of work in face-to-face situations, as the tool has a strong focus on blended-learning.

In these cases, the trainer has to provide scores between 0 and ten for each aspect. This offers the opportunity to combine the quantitative scores of the system for the online phases with scores for activities in face-to-face situations. Moreover, the teacher can score weights for the automatic scoring of communication by the system which is multiplied with the system scores, so that intensive and high quality communication gets more scores than only small talk of senseless communication.

3.2 Dialogue Related Scoring

A key element of the NetEnquiry learning scenarios are video-based, interactive dialogues with NPC-clients. Within the dialogues the customer consultant implements decisions that affect the whole problem-solving process, by choosing an answer-alternative. That also means, that the dialogue system always offers multiple

alternatives to answer (see Beutner / Pechuel 2014). As the content of the alternatives is designed within the dialogue writing process, there is an opportunity to offer answers that vary in direction and quality.

Direction means, that learners can structure the topic and the path of the ongoing dialogue. For example, they can decide to lead the dialogue into a small talk, ask for information or give information. Of cause, leading the dialogue in the right direction is an indicator for quality as well. In some cases, it is not clever to provide the learners with information beforehand (e.g. conditions), but to let them request the information through the dialogues. Hence, we clustered and structured dialogues, which makes it possible to use them for scoring. By offering clusters at special moments within the process, we can check if the learners are able to structure the learning process in the right way, and to score this decision.

Quality in contrast, refers to the way of articulation, the formal correctness of answers, and the strategic quality. The latter point describes decisions between different formally correct alternatives. However, as we have several variables in scenario settings, the correct strategic decision can vary, too. For example, it could be economically right to give the customer a loan, but depending on the scenario settings, it could be better to offer a bullet loan rather than an annuity loan. Moreover, norms can be the guidance value for the 'correctness' of decisions in the dialogue. This can be massively important, as the whole scenario result can be right or wrong, only caused by a single norm. In the loan scenario, the lecturer not only has to set economic variables of the case, but also strategic norms that influence the 'correctness' of moral-ethical decision making. For example, the lecturer has to answer the following question within the scenario settings: "Could it be right to give a loan, while the single economic decision would be wrong (e. g. high risk taking)?" For the scoring process a 'yes' means, that economically wrong decisions are not scored negative by default, because the learner in a role of the loan consultant decided correct from an ethical-moral point of view – he used his scope. Thus, decisions can drift from clearly black or white into a 'grey area'. And as described at the beginning of this paper, these are learning goals that matter and should find their way into the reflection process.

But how do we score the quality of decision making and what is about quantitative scores within the dialogue system? We do not offer scores for simple action, but for quality. As reference system for quality, we use three sources: a) expertise of sales professionals, b) data based on economic scenario-settings, and c) data based on lecturer settings (variables) regarding moral-ethical and strategic aspects.

- *Expertise based scores:* To know what is right or wrong, we observed sales training courses and documented how to structure customer dialogues, how to articulate questions, and how to offer information in a customer and sales oriented way. Through the writing process, we were able to bring that information into the dialogue as answer-alternatives for learners. As only certain parts of the texts were variable (e.g. the income), we could easily give individual scores for choosing a certain statement. In contrast to the other scoring system described for the other tools, we decided to give negative, neutral and positive scores. The scores vary from minus 4 (massively wrong) to plus 4 (difficult, highest quality answers and statements).
- *Setting based scores:* The flexibility of the scenario through variables (e.g. income of the client which is pre-defined by the lecturer) is relevant for the dialogue scores. As described above, some statements were filled with variables. Thereby, in some cases, answers switch from formal correct into wrong caused by the interdependency of variables. In consequence, scores have to switch from positive to negative, depending on scenario settings. For example, with the income-spending-ratio the correctness of the statement "Your monthly debt service is positive." can change. Therefore, we identified those answers that vary in correctness, depending on scenario settings. We combined those statements with our logical-mathematical algorithms and defined dependent scores. Moreover, these scores are independent from institutional settings. The amount of scores also varies between plus and minus four.
- *Normative Scores:* As written above, moral-ethical and strategic aspects can massively influence the correctness of the learners' solution. So, next to the economic driven answers, we also identified statements that can drift into the 'grey area' of decision making. Here, logical-mathematical algorithms are not sufficient. Hence, we had to define the circumstances under which a 'switch' into the 'grey area' of decision-making is given. Moreover, we had to check, if the institutional norms allow bringing in aspects next to economic ones into case. If this is confirmed, you can earn scores for economically wrong decisions, because they meet the institutional 'corridor of decision making'. Here also, scores vary between minus and plus four.

As the dialogue system is partly open in result and offers several cross-dialogue pathways (to switch to another topic), we had another challenge to be handled. In some cases, learners can make decisions that lead to opportunities to restart dialogue-paths that they already took. This also means that learners can receive and ask for information twice, or more often. This is not only problematic from a didactical point of view, but it would also have consequences in real life, if a loan officer asks for the same information repeatedly. Therefore, we had to safeguard that the learners always have enough formally correct alternatives to give the dialogue another direction without making a fault. If the learner anyway decides to ask the client a second time, this will lead to a negative score of one. This should be the normal and intended way for the learner.

4. CONCLUSION

As Hattie found out, "the biggest effects on student learning occur when teachers become learners of their own teaching, and when students become their own teachers. When students become their own teachers they exhibit the self-regulatory attributes that seem most desirable for learners (self-monitoring, self-evaluation, self-assessment, self-teaching). Thus, it is visible teaching and learning by teachers and students that makes the difference." (Hattie, J. 2008, p. 22) A visible scoring in this case offers the opportunity for both, teachers and learners, to reflect their own teaching and learning. Moreover, we can recently observe astonishing advances in the field of scoring and assessment in e-learning in general, and serious games and digital game-based learning in particular. This makes it possible to not focus on the evaluation of profession-related competences or closed question formats only (e.g. Hansen, O. B. et al. 2013, Vandewynchel, J. et al. 2013, Loh, C. S. et al. 2014); but rather to successfully evaluate a holistic competence development, with the ability of decision making in dynamic, complex environments as one of them. Moreover, we can foster the learners' ability to reflect on several subjects as well as on moral-ethical norms of institutions and societies, as institutionalized compliance strategies, organizational cultures, and economic changes in the world show. This complexity of the scoring system, and the interdependences in-between the different scoring layers can be visualized as follows (see Figure 2).

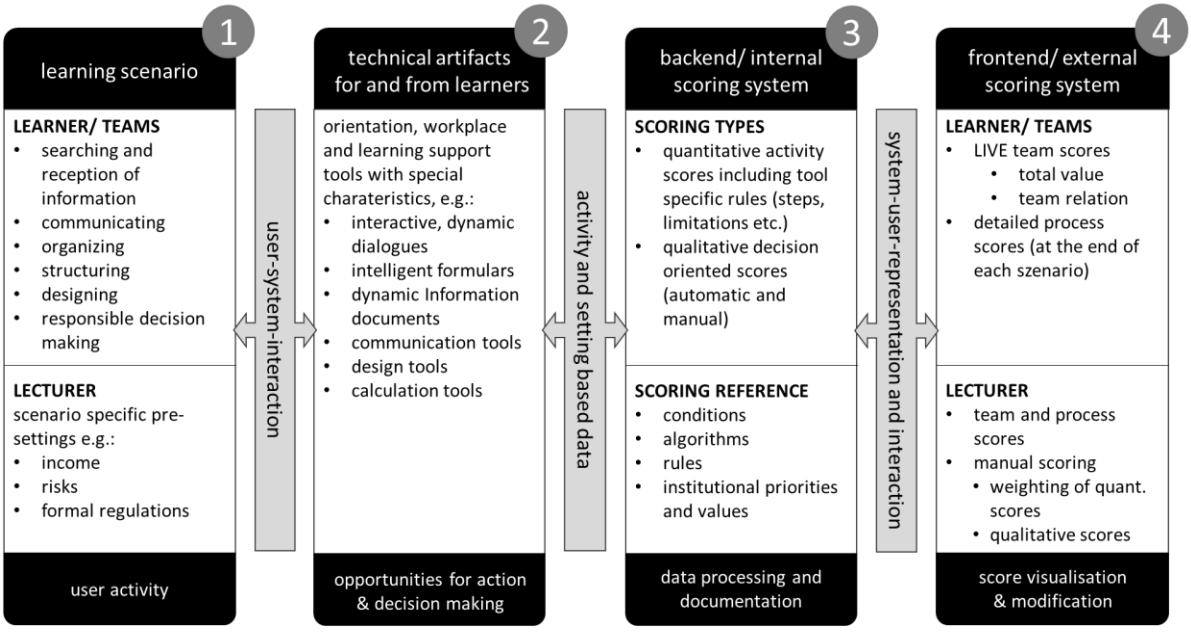

Figure 2. The systems' different layers and their interaction

In this paper, we presented a complex and variable assessment- and scoring system as a part of variable, complex, and partly open learning scenarios. It combines several scoring mechanisms like activity tracking, weighting of quantitative scores, lecturer scores for quality-driven elements, automatic quality scores for economic and strategic decisions as well as a dialogue oriented scoring system with automatic, qualitative

oriented scores. Thereby, we addressed motivational goals, offered live feedback, and supported the learners' reflection on the quality of decision-making. Furthermore, the data helps the lecturer to identify the learners' strengths, and weaknesses, and to provide highly individualized feedback in reflection sessions. A specialty of the system is, that it adapts itself, based on the lecturer´s input.

Anyway, the NetEnquiry system provides data that can massively enhance the learning process by providing a valuable data basis for reflection, which is, referring to Hattie, the golden way to learn. Nevertheless, there are also shortcomings. Referring to Chung (2014) cluster-analyses could enrich our assessment system. The basic idea is to let experts play the different scenarios, which would provide us with the ideal path through the dialogues (best practice). Then, the learners' solutions can be compared with the best practices. Of course, this needs to be done with several experts. Moreover, different institutions have different organizational cultures that influence values, mind-sets, and norms, which will affect the ideal solution. This, however, needs to be considered when the same learning scenario shall be used in different organizations. We would propose to not only create one cluster-analysis per learning scenario, but to make an analysis per dialogue. The dialogues can furthermore be divided into different logical parts like, e.g. welcoming clients, asking for and recording private data, using cross-selling opportunities, etc. The paths of the learners, and thereby their decisions, can then be compared with the ideal paths, which are the results of the cluster analyses. Correlations will then give detailed information, and at the same time, the lecturer can point out and address the weaknesses of the learners more precisely for reflection sessions. However, one of the reasons why we did not make use of this approach are the costs, and the availability of enough experts that would be necessary to create a valid and reliable set of ideal pathways

Another shortcoming of the system is that we are scoring communication and social skills, but within the overall score we do not distinguish any longer. There is no opportunity for the learner to see where they got their scores from, and therefore their reflection processes cannot be as detailed as desired. However, here we have been in the trade-off of keeping the competition system clear, or providing the learners with as much data about his learning as possible. Of course, when the focus is on distance learning only, and when there are no opportunities to have reflection-sessions with the lecturers, the decision should be to provide the learner with an appropriate amount of data about his progress, and performance.

Moreover, we are thinking about clustering the scores by source to monitor, which kind of learning activities in a specific scenario leads to a small or big amount of the scores achieved. Ongoing, we are able to shift the scores given per action source. This would enables us to weight activities in relation to their relevance and learning goals within a specific scenario. However, even if the advantages in the field of assessment are astonishing, research and practice needs to come to a point where this data can be used to automatically adapt learning scenarios to the learners' needs and skills, maybe based on psychophysiological data and detected player-emotions (cf. Breuer, J. 2010, Sykes, J. 2006). Once, such approaches and measures will be taken for granted, digitalization really outscores traditional learning environments.

REFERENCES

Abt, C. C., 1970. Ernste Spiele. Lernen durch gespielte Wirklichkeit.

Becker, K. 2007. Pedagogy in Commercial Video Games. In Games and simulations in online learning. Research and development frameworks. Information Science Pub, Hershey, PA. Pp. 21-47.

Beutner, M. 2010. Authentizität als Forderung und Prinzip. Überlegungen zur dramaturgischen PLOT-Gestaltung in Lernsituationen. In: Kölner Zeitschrift für Wirtschaft und Pädagogik. Vol. 25, No. 49, Köln 2010, Pp. 123-146.

Beutner, M./ Pechuel, R. 2014. Didactical Use of Dialogues in Modern Authentic E-Learning Scenarios – Creating Dialogues in a Dialogue-Driven Learning Approach. In Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education. Pp. 171-177.

Breuer, J. 2010. Spielend lernen? Eine Bestandaufnahme zum (Digital) Game-Based Learning.

Chung, G. K. W. K. 2014. Guidelines for the Design and Implementation of Game Telemetry for Serious Games Analytics. In Serious Games Analytics, Advances in Game Based Learning. Pp. 59-79.

Hansen, O. B. et al. 2013. Emotion Recognition for Mobile Devices with a Potential Use in Serious Games for Autism Spectrum Disorder. In Serious Games Development and Applications. Springer, Heidelberg, Germany. Pp. 1-14.

Hattie 2008, Visible Learning. A Synthesis of Over 800 Meta-Analysis Relating to Achievement.

Juul, J. 2005. Half-real. Video games between real rules and fictional worlds.

Loh, C. S. et al. 2014. Serious Games Analytics: Theoretical Framework. In Serious Games Analytics, Advances in Game Based Learning. Pp. 3-29.

Loreto, I. D. et al. 2013. Supporting Crisis Training with a Mobile Game System. In Serious Games Development and Applications. Springer, Heidelberg, Germany. Pp. 165-177.

Prensky, M. 2007. Digital Game-Based Learning.

Raposo, F. et al. 2013. SimClinic – An Auxiliary Tool for Evaluation on Clinical Case Settings. In Serious Games Development and Applications. Springer, Heidelberg, Germany. Pp. 37-50.

Reetz, L. 1984. Wirtschaftsdidaktik: Eine Einführung in Theorie und Praxis wirtschaftsberuflicher Curriculumentwicklung und Unterrichtsgestaltung. Bad Heilbrunn, Germany.

Reinmann-Rothmeier, G. 2005. Blended Learning in der Lehrerbildung.Grundlagen für die Konzeption innovativer Lernumgebungen.

Rieber, L. P. (1996). Seriously Considering Play. Designing Interactive Learning Environments Based on the Blending of Microworlds, Simulations, and Games. In Educational Technology Research and Development. Vol. 44, No. 2, pp. 43-58.

Roth, H. 1971. Pädagogische Anthropologie. Bd. II. Entwicklung und Erziehung. Grundlagen einer Entwicklungspädagogik. Hannover, Germany.

Scalise, K. & Gifford, B., 2006. Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms. In The Journal of Technology, Learning, And Assessment, Vol. 6, No. 5.

Shute, V. J. et al., 2009. Medling the Power of Serious Games and Embeedded Assessment to Monitor and Foster Learning. Flow and Grow. In The Social Science of Serious Games: Theories and Applications. Pp. 295-321.

Sloane, P. F. E./ Dilger, B. 2005.: The Competence Clash. Dilemmata bei der Übertragung des 'Konzepts der nationalen Bildungsstandards' auf die berufliche Bildung. In: Tramm, T. / Brand, W. (Eds.): Prüfungen und Standards in der beruflichen Bildung: bwp@ ed. 8 (32 p.). Online: http://www.bwpat.de/ausgabe8/sloane_dilger_bwpat8.pdf.

Sykes, J. 2006. Affective Gaming: Advancing the Argument for Game-Based Learning. In Affective and emotional aspects of human computer interaction. Game-based and innovative learning approaches. IOS Press, Washington, USA. Pp. 3-7.

Vandewynchel, J. et al. 2013. Training Adapted to Alzheimer Patients for Reducing Daily Acitivity Errors. In Serious Games Development and Applications. Springer, Heidelberg, Germany. Pp. 28-36.

Westphal, A., 2009. Die Einbettung von Lerninhalten in Serious Games. In Herzog, M. A./ Sieck, J. (Eds.): Kultur und Informatik: Serious Games. Pp. 115-142.