Support interoperability and reusability of emerging forms of assessment using IMS LD with IMS QTI

Support Interoperability and Reusability of Emerging Forms of Assessment: Some Issues on Integrating IMS LD with IMS QTI

Yongwu Miao, Jo Boon, Marcel van der Klink, Peter Sloep, and Rob Koper
Centre for Learning Sciences and Technologies,
Open University of the Netherlands
Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands

ABSTRACT

In comparison with traditional assessment, emerging forms of assessment (e.g., self-/peer assessment and 360 degree assessment) involve multiple phases and multiple roles/persons, which are process-oriented assessment. IMS Question and Test Interoperability (QTI) is an open technical specification for task-oriented assessment, which has insufficient expressiveness to specify emerging forms of assessment. Meanwhile, existing software tools supporting emerging forms of assessment lack interoperability and reusability. In this chapter, we claim that a combined use of QTI and IMS Learning Design (LD) is able to support interoperability and reusability of emerging forms of assessment. In order to support this claim, we analyze the characteristics of four emerging forms of assessment from the perspective of process technologies and present the method to specify emerging forms of assessment using QTI and LD. Furthermore, we present the difficulties and problems that we encountered when modeling emerging forms of assessment and propose possible solutions to solve the problems.

INTRODUCTION

There is a marked tendency to place more and more emphasis on competences in education and, therefore, in assessment. Competence is defined as 'effective overall performance within an occupation, which may range from the basic level of proficiency through to the highest level of excellence' (Cheetham and Chivers 2005). A competence is the ability to handle a complex professional task by integrating the relevant cognitive, psychomotor and affective skills. Information gathering for the assessment of competences is increasingly based on qualitative, descriptive and narrative information, in addition to quantitative, numerical data. Such qualitative information cannot be judged against a simple, pre-defined standard (Vleuten and Schuwirth 2005). Some emerging forms of assessment have gained in acceptance and popularity in education. Examples of such forms of assessment are self- and peer assessment, accreditation of prior learning, and 360 degree assessment. These forms of assessments address complex traits of students and foster deep learning and the development of competences (Topping 1998; Boud, Cohen et al. 1999; Gipps 1999).

Assessment consists of making judgments (task aspect) and carrying out administrative activities (process aspect). In comparison with traditional assessment, both of these aspects of assessment are much more problematic in emerging forms of assessment. In particular, emerging forms of assessment usually involve multiple phases and multiple roles/persons. The difficulties and the potential for errors and

omissions increase in a non-linear fashion as the number of candidates and assessors involved grows (Rosbottom 1994). As Bartram pointed out, 360 degree assessment by its very nature is an administrative nightmare to manage. People involved in the process tend to be geographically dispersed but also need close supervision in order to ensure that the ratings are carried out to schedule and that sufficient raters are obtained for each focus of the assessment (Bartram 2005).

In order to make emerging forms of assessments work effectively and efficiently, many software tools have been developed and are increasingly being used. For example, MUCH (Rada, Acquah et al. 1993; Rushton, Ramsey et al. 1993), Peers (Ngu, Shepherd et al. 1995), Peer Grader (Gehringer 2001), SPARK (Freeman and McKenzie 2002), and ESpace (Volder et al. 2007) are multi-user tools that support self- or/and peer-assessment. The eSPRAT system (Lockyer 2003; Davies and Archer 2005) and Appraisal360 (Appraisal360 home page) are example tools that support 360 degree assessment. In self- and peer assessment, with the help of software tools, the tutor, freed from administrative chores, is able to provide a useful, added-value service to students by acting as a problem solver. Student-assessors can concentrate on the clarity, correctness and completeness of each individual exercise without worrying about the relationship with other exercises (Rosbottom 1994). Similarly, for supporting 360 degree assessment, the software tools manage the workflow associated with the 360 degree assessment process, from initial set-up and preparation of the people involved, through the management of the rating process (including delivery and scoring of questionnaires), to the production of reports and their delivery to feedback providers (Bartram 2005).

However, existing software tools supporting emerging forms of assessment are stand-alone and offer limited or no support for interoperability of systems and reusability of assessment resources. They each have their own data representation and their data are not interpretable and operable by other application tools. This prompts the question of whether existing e-learning technical specifications can be used to support emerging forms of assessment. The leading specification for the exchange and interoperability of assessments is IMS Question and Test Interoperability (IMS QTI, 2003). However, the QTI specification addresses the task aspect of assessment. Examples of specified assessment tasks are multiple choices, fill-in-the-blank, and matching items. QTI provides no means to support the design and management of assessment processes. Specifically, it ignores who will be involved and what roles they will play at process level, what kinds of activities should be performed by whom and in which sequence, what assessment resources will be produced and used in which activities, and what dynamic changes may take place in the assessment process and under which conditions. In short, it provides insufficient support for the representation and execution of an assessment plan (Miao et. al. 2008). Therefore, QTI can not independently support emerging forms of assessment.

In QTI v2, the integration of QTI and IMS Learning Design (IMS LD, 2003) was specified. LD is an open e-learning technical specification that provides a pedagogy-neutral modeling language. It can be used to specify a teaching/learning process as a formal model, which can then be executed in a specification-complaint run-time environment (Koper and Olivier 2004). The integration between QTI and LD provides a possibility to technically model an aligned teaching, learning, and assessment process. However, only little reported work has been carried out on modeling emerging forms of assessment using LD and QTI. The objective of this chapter is to explore the possibility to support emerging forms of assessment by using existing e-learning technical specifications. Concretely speaking, we investigate the expressiveness of LD and QTI in the representation of emerging forms of assessment by using a case-based analytical method. For each emerging form of assessment, we analyze the key features of the assessment from the perspective of process modeling and identify possible alternative scenarios in practice. We intend to share our experiences with readers in modeling emerging forms of assessment in LD and QTI. In addition, we will identify hurdles which may keep educators and assessment designers from using LD and QTI to specify their assessment. Finally, we propose possible solutions to overcome these difficulties.

BACKGROUND

Most open e-learning technical standards for course development and delivery (e.g., IEEE LOM, IMS CP, IMS SS, ADL SCORM) concern learning content (e.g., the description of content and the organizational structure of the content). Only IMS Simple Sequencing specification (IMS SS, 2001), which is also included in ADL SCORM, provides simple mechanisms to represent the sequence of content. In QTI v2, the integration between QTI and IMS SS has been specified as well. The effort has been made by the ASSIS project (ASSIS homepage) to integrate assessment into adaptive sequences of content. This approach enables a seamless integration between instruction and assessment and supports interoperability and reusability. However, such an approach assumes a learning model in which individual learners consume learning content with certain conditional control. It does not support the integration of learning activities with assessment activities. Instead, it just integrates learning materials with questions/questionnaires. The evaluation results of learners' answers are used to control the sequence of the presentation of the content, not the activity sequence. Therefore, it can not support emerging forms of assessment that involve multiple roles/users and complicated interactions among them.

In the development of e-learning technical standards, the release of LD signals an exciting paradigm shift from a content-centric approach to an activity-centric approach. LD provides a framework to express the pedagogical meaning of instructional content and in doing so reflects in a deeper and more creative way on how to design and structure activities (Koper and Olivier 2004). It can be used to specify a collaborative assessment process in which multiple people with diverse roles (e.g., designer, candidate, assessor, decision-maker, and other stakeholders) perform various activities (e.g., design assignment, create/collect evidence, evaluate evidence, and make decision) in sequence and/or in parallel coordinately at process level. However, LD can not explicitly support various types of assessment tasks. Assessment components within the Educational Modeling Language, the base of LD, were excluded when LD was adopted by IMS, because of the existence of QTI. QTI describes a data model for the representation of assessment item/test and the result report. It defines a set of interaction types which can be used to specify basic question types and complicated question types through combination. As mentioned before, QTI provides no support to model a multi-users/roles-involved and multi-phase assessment at process level. It is obvious that LD and QTI have their respective strengths and weaknesses when it comes to supporting emerging forms of assessment at process level and at task level. What is interesting is that their strengths and weaknesses are complementary. In the next section, we will examine whether an integration of LD and QTI can indeed support emerging forms of assessment.

A STANDARD-BASED APPROACH TO SPPORT EMERGING FORMS OF ASSESSMENT

We present a standard-based approach to support four emerging forms of assessment: self assessment, peer assessment, accreditation of prior learning, and 360-degree assessment, because they are more and more important and popular in education. Table 1 shows the number of hits in de period 2004-2009 on the internet, using databases from Google Scholar and EBSCO.

Table 1. References to different assessment forms

	Google scholar (*)	EBSCO (**)
Self assessment	2490	65
Peer assessment	534	31

Accreditation of prior learning	18	0
360-degree assessment	27	2

(*) term in title, period 2004-2009,
(**) with option Keywords, January 2004 – June 2009P, peer reviewed articles, linked full texts

Google scholar is chosen because it gives a good reflection of the academic mainstream in topics, EBSCO is chosen because if reflects a more specialized collection, mainly consisting of empirical research articles. The ten articles found first in both databases are used as background for the descriptions of the assessment forms hereafter. Most of these articles offered some examples of assessment forms and these examples served as the input for the descriptions of the four assessment forms.

Characterizing Four Assessment Forms from Perspectives of Process Modeling

We first analyze the characteristics of the four forms of assessments from the perspective of process support technologies.

Key features of self assessment

Self assessment (SA) refers to a method where an individual assesses his or her own performance regarding a specific topic. The method is largely used both in work situations as in educational settings to initiate self reflection on issues related to performance. Also in many health related situations self assessment is a relevant method for self diagnosis. In work and educational situations the method is often combined with 360 degree assessment or with peer evaluation; in both cases the self assessment is a first step in the procedure, designed to make comparison with assessment of others and reflection on this comparison richer. The function of SA is evaluation or judgment of the worth of one's performance and the identification of one's strengths and weaknesses with the aim to improve one's learning/working/health outcomes. Table 2 lists the key features of SA from the perspectives of process modeling and alternative scenarios.

Table 2. Key features of Self Assessment

Roles	- the individual the representative of the learning or working context; this might be - the teacher - the manager - the peers
Artifacts	- goals, criteria, procedures - scoring list or questionnaire on the relevant topics - evidence on performance using the scoring list or the questionnaire - the answers to the questionnaire or scores
Activities	- define the goal and the rules, criteria of the assessment. - score performance - report assessment result

Interaction	1. Preparation: Representatives of the context together with the individual define the goal and the rules, criteria of the assessment. 2. Assessment The individual scores him/herself on the relevant issues 3. Finalization The individual communicate with others about the scores
Alternative scenarios	- Self assessment can be conducted by every individual without feedback to the organization. It is possible that a SA scenario has no the final phase. - Self assessment is often used as a first step in a process of 360 degree assessment or peer assessment as a part of an overall assessment process.

Key features of peer assessment

Peer assessment (PA) can be characterized as the process in which students collaborate and evaluate their own performance as well as those of fellow-students (Sluijsmans et al., 2004; Gulikers, Sluijsmans, Baartman & Bartolo, 2009). Most implementations of peer assessment are not restricted to evaluating a peer's performance as such. In many educational contexts the basic idea is that it is essential that both actors, the candidate who undergoes the assessment (the assessed student) and the peers who conduct the peer assessment, should benefit from the peer assessment experience. Peer assessment is primarily used in professional and vocational education. In some professions there is growing interest in peer assessment (e.g. teaching profession) as a tool to enhance continuous professional development. Peer assessment is mostly used for formative assessment purposes, to provide students feedback on their performance that subsequently enables them to consider points of improvements for future learning experiences. Table 3 lists the components of a PA, the main procedures of PA, and alternative scenarios.

Table 3. Key features of peer assessment

Roles	- teacher - candidate - assessor
Artifacts	- instruction - standards and criteria - evidence - assessment form - feedback - improvement
Activities	- inform students - group students - create evidence - assess evidence - evaluate feedback - compose points of improvement

Interaction	1. Preparation Teacher informs students about goals, procedures, timelines etceteras; Teacher groups students in pairs, trios or larger groups. 2. Creating evidence Candidate uses instruction, the standards and criteria to create the evidence. 3. Assessing evidence Peers use the instruction, the standards and criteria to evaluate candidate's performance; Peers fulfill assessment form and write feedback. 4. Reaction Candidate evaluates the feedback and composes points of improvement;
Alternative scenarios	- Peer assessment often has a reciprocal nature, meaning that after the first round, roles shift, and that the candidates subsequently become peers and vice versa. - Not always there is written evidence to be judged afterwards. In some cases, peers observe the behavior of the candidate, which then is the evidence to be judged (for example student teachers who assess each other during internships in schools) - In many cases peers are required to reflect on their role as peer assessor. - Sometimes the candidate informs his peers about the quality of the received feedback

Key features of accreditation of prior learning

Accreditation of prior learning (APL) supports lifelong learning by assessing and recognizing someone's competences obtained informally through (paid and unpaid) work experiences (Joosten-ten Brinke, 2008). APL is most often offered by educational providers who promote APL in order to attract non-traditional student groups. APL is primarily used in education as a means to determine the content and size of student's study program prior to study entrance. It is a sound instrument only for employees who posses sufficient work experience in the domain they want to be educated for. APL is mainly used for summative assessments. Outcomes of the APL procedures are utilized by examination boards to determine what needs to be learned by prospective students in order to receive a particular certificate or diploma. In Table 4, we present the key features of APL from the perspectives of process modeling and alternative scenarios.

Table 4. Key features of procedures for accreditation of prior learning (APL)

Roles	- mentor - assessor - employee's (prior) employer - employee (hereafter candidate) - examination board

Artifacts	- description of set of competencies, including standards and requirements for portfolio - evidence and portfolio - form to check candidate's portfolio - rubrics and scoring forms for assessors - APL certificate - form to notify candidate on study program reduction - form for candidates to appeal against the outcome of their APL procedure
Activities	- discusses - select the competences - collect evidence and store in a portfolio - check portfolio - assessed portfolio using rubrics and scoring forms. - write report (APL certificate) - decide to what extend it is allowed to reduce the candidate's study program.
Interaction	1. Candidate-profiling Candidate discusses with mentor the possibilities for APL; Candidate receives description of set of competencies, including standards and requirements for portfolio. 2. Evidence gathering Candidates collect and classify evidence about their previous experience; Mentor checks the content of candidate's portfolio. 3. Assessment Assessors review the quality of a candidate's evidence using assessment standards and rubrics; Candidate receives a report that describes to what extend the candidate master the competences that are included in the competence profile. 2. Recognition Assessors compose APL certificate and send to candidate; Candidate send APL certificate to examination board; Examination board notify candidate about decision on study program reduction.
Alternative scenarios	- Candidates assess their own prior experience in light of the standard and include the outcomes of this self-assessment in the portfolio; - Besides portfolio assessment one or more additional assessment activities usually will take place, such as a criterion-based interview, demonstration, knowledge test. - …

Key features of 360 degree assessment

360 degree assessment is also known as multi-source performance assessment or 360 degree feedback. The method refers to the process by which performance appraisals are collected from different sources, such as supervisors, peers, subordinates and sometimes also customers - rather than from a single source. This should provide the feedback recipient with a unique combination of information which is not otherwise available. It is assumed that the feedback givers chosen are in the best position to observe and evaluate certain types of behaviors. The method can be used for assessing performance and designing professionalization or development paths, sometimes the method is used to analyze interpersonal behavior

(Whitehouse et al., 2007) or for training evaluation (Jellema, Visscher and Scheerens, 2006). It is used sometimes as a decision making tool (for example on career advancements or salary increases). 360 degree assessment is usually used at workplaces, both private and public. It can also be used in a class situation for educational purposes, but this is less likely. Table 5 shows the key features of 360 degree assessment.

Table 5. Key features of 360 degree Assessment

Roles	- feedback receiver (or target employee)
	- responsible for process (RFP), can be a HRM representative
	- feedback giver:
	  - supervisor
	  - peers/co-worker
	  - subordinate

Artifacts	- form with closed and open questions on issues and criteria to be used as a questionnaire or a guide for an interview - mission statement of organization with competency map - appraisal and feedback - summary and priorities
Activities	- define assessment goals - instruct - formulate appraisal - structure feedback - communicate feedback
Interaction	1. Preparation - HRM representative define assessment goals; - HRM representative instructs all participants on procedure, roles, goals and criteria. 2. Assessment - Downward appraisal from supervisor - Lateral appraisal from peers/co-workers - Upward appraisal from subordinates - Inward appraisal from target employee 3. Finalization - HRM representative summarizes feedback - HRM representative formulates next steps trajectory

Alternative scenarios	- The target employee formulates improvement goals at the beginning of the process and the different other roles react on these - The input from each appraisal is discussed consequently with the target employee - The target employee gives feedback on improvement goals to the superior, peer or subordinate - The self assessment is not always part of the procedure. Some authors argue that a previous self assessment optimizes the process (Garbett et al., 2007) - More than one employee from each role-group is appointed (more than 1 supervisor, peer, subordinate) - Feedback can be given during a group session; this could reinforce the effects of reflection (van der Heijden and Nijhof, 2004). - Feedback can be given anonymous or anonymously - A group of employees instead of a target employee can be the feedback receiver - A training is given to participants if necessary - Some companies collect feedback from the customer

Modeling Emerging forms of Assessment Using LD and QTI

QTI v2 specified integration of LD with QTI by coupling an LD property to a QTI outcome variable. The original motivation for integrating LD and QTI stems from use cases involving formative assessment and summative assessment using assessment items with traditional question types. Here we try to extend the application areas of the integration of LD and QTI and to improve the benefit of their combined use. As a consequence, the emerging forms of assessment can be modeled as a unit of assessment, a process-oriented assessment model represented in the form of a specific unit of learning. Thus, such a unit of assessment can be executed in an LD and QTI compliant run-time environment. Furthermore, a unit of assessment can be instantiated as a complete model many times and can be customized or partially reused by different groups/organizations.

When analyzing the emerging forms of assessment, we have created a table for each form of assessment in the last sub-section. There are five rows in each table: roles, artifacts, activities, interaction, and alternative scenarios. The first three rows are components of a process. The interaction describes how participants with diverse roles perform activities in sequence and/or in parallel and how artifacts are used, produced, and transferred in/between activities. Alternative scenarios describe some variations in assessment practices. In this sub-section, we present how to model them through a combined use of LD and QTI.

Modeling multiple roles

As we have seen in each table, multiple roles are involved in each form of assessment. When modelling an emerging form of assessment, it is required to explicitly define multiple roles. The QTI specification is concerned with individual learners. Although QTI does not prohibit use in contexts involving other actors (e.g., instructors, supervisors, and peers), it does not explicitly support defining other roles or sequencing behaviors that result from participation of other actors. However, LD can support a multi-role/user teaching-learning process. In LD, two primary roles (learner and staff) are pre-defined. Each role can have sub-roles defined by designers to fit the context of the learning design. A role is bound with certain activities as role-parts. At run-time a person with a certain role will have privileges and responsibilities which allow him or her to perform the activities and to access certain learning resources according to the definition of the learning design. With LD, multiple roles as listed in the four tables can be modeled. The

hierarchical structure of roles (e.g., in 360 degree assessment the role of feedback giver has three sub-roles: supervisor, peers/co-worker, and subordinate) can be modeled as well. Note that in LD each role can be played by multiple users at run-time. Thus, it can be modeled that more than one employee from each sub-role of feedback giver can be appointed in 360 degree assessment.

Modeling artifacts

In each emerging form of assessment various types of artifacts are created and/or used in activities. Some are represented in the form of questions (e.g., some assignment forms for creating evidence and some assessment forms with rubrics) and some are normal documents for different purposes (e.g., assessment goal and feedback). Usually, an artifact in the form of question/questionnaire can be modeled using QTI, which can represent many types of questions such as multiple-choice/response, Likert-scale, open-question, fill-in-blank, hot-spot, matching, ordering, association, slider, drag&drop, and upload-file. QTI also provides sufficient flexibility to grow into the advanced constructed-response items and interactive tasks we envisage as the future of assessment (Almond, Steinberg et al. 2001). Furthermore, it provides mechanisms to design structured assessment and control branches and calculate weighted scores. That is, all standard questions and structured tests/exams that form the core of current practice can be supported by using QTI. In addition, LD can be used to represent non-question artifacts. Although LD has no concept of "artifact" in the specification, it enables to define a property with a data type, such as string, text, Boolean, integer, real, url, time, duration, and file. A kind of artifact can be modeled as a property using an appropriate data type. For example, an assessment goal or a feedback item can be defined as a property with the string or text type. A structured document can be modeled as a file-type property. Note that reusable documents can be put on the web and can be accessed by many assessment processes through using URLs of the web pages.

Modeling activities

In each emerging form of assessment, various activities are performed by diverse roles. LD provides constructor (i.e. activity and environment) to define an activity with some attributes (i.e., title, description, and completion). Most activities listed in the tables can be easily modeled in LD through specifying the values of attributes. For modeling some assessment activities, the question/questionnaire should be modeled as a QTI document as described above, which has to be referred to by an information item within the activity or in the associated environment. It is important to note that a corresponding LD property should be defined in such a way that its identifier is a combination of the identifier of the QTI document and the identifier of the corresponding outcome variable, such as a score. When a candidate/assessor accesses the activity or the environment at run-time, the question/questionnaire will be presented to the candidate/assessor by the QTI engine. After the candidate/assessor submits the answer(s), the QTI engine will evaluate the response and transfer the result to the LD engine. Then LD engine can then adapt the teaching/learning process to the assessment result. For supporting some online activities, such as interview, monitoring, and group meetings, additional services are needed. Fortunately, LD provides some built-in services such as conference and monitor, which can be used to support online communications and monitoring works of participants with a given role.

Modeling interaction

As illustrated in the tables, emerging forms of assessment are phase-based processes, in which multiple participants with diverse roles perform various activities in sequence and/or in parallel and artifacts are transferred from one activity/role to another.

QTI allows candidates to answer questions in a pre-defined sequence or in any order to finish an assessment test. However, such control of the sequence of the tasks is restricted within an individual assessment test. LD can support the modeling of a learning flow with complicated process controls. Activities can be arranged as a sequence or a selection structure. A set of role-parts can be performed in parallel within an act, and acts within a play will be carried out in sequence. Multiple plays can be

executed as concurrent threads. The termination of one activity may trigger the start of another activity. In addition, conditions and notifications provide more powerful mechanisms to control the process. The support provided at LD levels B and C makes it possible to trigger the start and termination of activities in a data-driven manner as well.

Some artifacts such as evidence and feedback are intermediate products, which are transferred from one activity/role to another. Some are pre-defined and assessable in the assessment process. QTI provides mechanisms for declaring outcomes. The outcome of an item, a section or a test can be processed as the output of an assessment. QTIv2 specifies how an outcome variable of QTI can be coupled to a LD property. With the help of this mechanism, an item response and an assessment score can be transferred to relevant participants. That is, the data produced by a participant (e.g., a candidate) can be presented to another one (e.g., an assessor). Additionally, scores given by all assessors can be processed according pre-defined calculation rules as a final result. This result can be transferred to a candidate or even can be used to control the branching. Furthermore, LD provides rich mechanisms to produce and transfer artifacts that are modeled as properties. For example, set-property, change-property, and view-property are basic mechanisms to create, modify, and retrieve artifacts. The local property and global property allow one to transfer artifacts within a learning design and across learning designs. The monitor service can support to view the artifacts produced by other roles.

In summary, both LD and QTI have certain strengths and weaknesses in their support of emerging forms of assessment. They cannot model all features of emerging forms of assessment independently. However, they complement each other on task and process aspects. Thus a combined use of LD and QTI can model most of the features of emerging forms of assessment listed in the tables. In the next sub-section, we will use this standard-based method to model an example of an emerging form of assessment.

An Example

In this sub-section we describe a 360 degree assessment scenario. Then we model it with LD and QTI and present how to execute it.

Description of a 360 degree assessment scenario

Professor Hiks works at department C of a university and is responsible for the coordination of one of the sections of this department, focusing on the theme of consumer education. He develops research proposals and acquires research funds, supervises young researchers and has contacts with paying clients outside the university who want to have his advice on consumer education. He has three senior researchers who support him in his job.

In the department where he works a competence map is developed that describes all the competences relevant for different staff members in different jobs. In the beginning of the year, the management team decides that a new round of 360 degree assessments will be organized. The staff member who is responsible for the coordination of sections sends mister Hicks a mail explaining the procedure, and setting a time frame for about when he will have a talk with his manager, in his case the director of the department.

First, professor Hiks uses the competence map to perform a self assessment. Using the map he rates his score on the relevant competences and decides on which topics he would like to have more formal and informal training in the coming year. Second, he invites a coordinator of another section, one of the young researchers he is supervising, as well as one of the clients he worked for during the last months. He asks all three feedback givers to fill in a short questionnaire with questions on his commitment, the quality of his output, the degree to which he keeps his appointments and the quality of his functioning as a team

member. The questionnaire leaves room for other remarks on his performance. Three feedback givers send their reactions to the director of the department and send a copy to professor Hiks himself.

At the agreed date, the director of the department receives the self assessment and the information of the three feedback givers and the report of the 360 degree assessment of last year. He uses all this information to have a discussion with professor Hiks about his performance. In the self assessment Hiks indicates some competences on the management level where he wants to have some training, especially on the field of supervision and time management. It turns out that his colleague coordinator is very positive on all points and only mentions that sticking to appointments is sometimes a problem; professor Hiks often comes late in meetings and has to leave early. The young researcher is also very positive but mentions that she has to wait sometimes for weeks before receiving feedback on research proposals. The client is very satisfied on all the points and mentions that for the next contract he wants professor Hiks to advise him on a specific new topic. During the discussion with the director of the department appointments are made about training in time management, delegation of tasks and setting of priorities. The appointments are formalized in a short report and stored in the personnel portfolio of professor Hiks.

Modeling the 360 degree assessment scenario

We can develop a descriptive model that formally specifies the scenario with LD and QTI. A descriptive model abstractly describes how a process is performed in a particular environment in an inductive manner. In the model, five roles are defined: feedback receiver, manager, and three feedback givers including colleague, subordinate, and client. The competence map is modeled as a QTI test document including a list of Likert-scale questions. Three short questionnaires for feedback givers are modeled as QTI test documents as well. The reports of the 360 degree assessment of last year and this year are modeled as file-type properties. Five activities are defined: one self-assessment, three assessment activities of feedback givers, one discussion. The whole process consists of three phases: self assessment, assessment of feedback givers, and discussion and decision. Self assessment result and all feedbacks created in the first two phases will be used in the discussion. A short report will be produced in the discussion.

Execution of the model and reuse of the model

The model can be published in a LD and QTI compliant run-time environment. If the assessment would be conducted in the computer-supported environment, the process will be carried out as below.

The staff member who is responsible for the coordination of sections, instantiates the model by creating a new run of the model. S/he has to prepare settings for this run through assigning the role of feedback receiver to Hiks and assigning the role of manager to the director of the department. The staff member will arrange a conference service if the discussion is an online activity. Otherwise, a meeting room should be arranged with a scheduled duration for the discussion. After that, the staff will inform all about the start of the assessment. Professor Hiks can access the first activity in which the instruction about how to carry out the assessment and the competence map are available. The expected output of this activity is the self assessment result. Then, he invites three participants by assigning the role of colleague to the coordinator, the role of subordinate to the young researcher, and the role of client to the person for whom professor Hiks has worked during the last months. The invited feedback givers will be informed and can find an assessment activity in their to-do list. After accessing the activity, s/he can read the instruction and the short questionnaire. After having answered the questionnaire, s/he can simply submit it. All assessment results and the report of 360 degree assessment of the last year can be accessed in the discussion activity. In the time scheduled, Hiks and the director of the department can access the activity work space and discuss results either using the online service or face-to-face. The director can write a short report in the activity work space and the report will be send to professor Hiks. This then terminates the execution of the assessment.

It is important to note that this model can be reused for assessing other colleagues of the department. For this purpose, the staff member only needs to create other runs and to assign the role of feedback receiver to other colleagues. The model can also be reused for assessing the performance of professor Hiks in the next year. Finally, it can be customized by other departments through modifying the competence map and questionnaires.

FUTURE RESEARCH DIRECTIONS

When modeling emerging forms of assessment, we encountered some difficulties and problems. Firstly, it is difficult to perform statistical analyses (By statistical analyses we do not refer to the usual analysis of assessment results, but rather data analyses that lead to an adaptation of the assessment process itself), if the number of role members is not fixed in an assessment process. Even if the number of candidates is predictable, the degree of complexity of the model will increase as the number increases. For example, if the number of peers is unpredictable, the score given by each peer can only be modeled as a personal property. However, LD provides no means to express the calculation of the mean of the scores given by all peers. Secondly, the adaptation of an assessment is currently restricted within the definition of the assessment and the assessment can be adapted only to candidates' responses to the questions. It is difficult to adapt assessment to the learners' characteristics and environmental information. For example, the competence map cannot be adapted to the position/function of the feedback receiver. Thirdly, assignments and/or assessment forms, sometimes, have to be developed by the participants at run-time, not by the designer at design-time. It is difficult to include new assessment after a UoL has been published. For example, in APL it is unpredictable what additional questions are required to answer. The assessor may need to create a questionnaire for collecting additional evidence at run-time. Fourthly, it is difficult to integrate assessment-specific services in LD. For example, in APL additional assessment activities may be needed in which assessment-specific services such as certain simulators and concept-mapping tools are needed.

In the near future, research should target solving the problems just identified, if we want genuinely to support emerging forms of assessment in an interoperable and reusable manner. Firstly, LD would have to be able better to deal with personal properties (e.g., the sum of scores given by multiple peers when the score is modeled as a personal property); this can be done by extending the specification of the expression element. Secondly, the concept of 'income variable' should be introduced in QTI, so that the information can be transferred from teaching-learning activities to assessment. The adaptation can be defined in such a way that it adapts assessment to the value of income variable. Thirdly, QTI editor had better be specified as a built-in service in LD, so that LD can handle the QTI documents created at the run-time. Fourthly, a more generic solution (like BPEL4WS in business process management) should be developed to integrate third-part services in LD, so that the external services can be specified in the design-time, can be configured at instantiation-time, and then can be invoked at the run-time easily.

Finally, the standard-based approach for modeling emerging forms of assessment described in this chapter suits technical developers only, who have a sound knowledge of process modeling and technical specifications. As pointed in (Miao and Koper 2007), it is very difficult if not impossible for practitioners to model a complicated teaching, learning, and assessment process with LD and QTI. In order to support 'ordinary' teachers and assessment designers to specify and customize an assessment plan, a high-level assessment modeling language is needed; this we are currently working on (Miao et al. 2008 and Miao et. al. in press). For the sake of interoperability and reusability, an assessment plan represented in such a high-level modeling language will be transformed into an executable model represented in LD and QTI. Thus the assessment process can be supported by using existing LD and QTI complaint run-time environment.

CONCLUSION

Emerging forms of assessment become increasingly important. The importance of four emerging forms of assessment is underpinned by presenting some findings on how broadly these forms are represented in contemporary academic writings.

Through an analysis of key features of four emerging forms of assessment from the perspective of process technologies, we found that all these forms of assessment 1) involve multiple roles/participants; 2) deal with various artifacts; 3) consist of various activities; and 4) include a complicated control-flow and data-flow. Although many software tools have been developed to support emerging forms of assessments, these software tools are stand-alone and lack interoperability and reusability. QTI, the leading specification for the exchange and interoperability of assessments, supports task-oriented assessment, but cannot support process-oriented assessment. LD, a process-oriented modeling language, can be used to model multi-role/user and multi-phase processes, but lacks facilities to model various assessment tasks. That is, neither of them can fully support emerging forms of assessment.

In this chapter we developed and presented an approach to support interoperability and reusability of emerging forms of assessment. The approach is based on the existing open e-learning standards LD and QTI. Through a combined use of LD and QTI, emerging forms of assessment can be modeled as units of assessment, which then can be executed in any LD and QTI compliant run-time environment. That is, an emerging form of assessment represented as an executable model can be reused and customized by other groups/organizations. Meanwhile, the components of a model can be reused as well. Because the model is represented in LD and QTI, all standard-compliant tools (irrespective of the authoring tool, repository, simulator, or engine) can interoperate on the assessment model.

We also indicated some difficulties we met when modeling emerging forms of assessment with LD and QTI. We proposed solutions to overcome them. As part of that, we are working on a high-level, assessment process modelling language. It is designed for practitioners to allow them to specify or customize emerging forms of assessment. Using this language, the emerging form of assessment can be specified as a high-level assessment process model, which can be automatically transformed into an executable model represented in LD and QTI. Once this goal is achieved, practitioners will be able to reap the benefits from using technical standards without the need to handle technical complexity.

REFERENCES

Appraisal360, home page, Retrieved July 12, 2009, from http://www.appraisal360.co.uk/home

ASSIS, home page, Retrieved July 12, 2009, from http://www.hull.ac.uk/esig/assis.html

Almond, R. G., L. Steinberg, et al. (2001). A sample assessment using the four process framework. CSE Report 543. Center for study of evaluation. . Los Angeles, University of California.

Bartram, D. (2005). Testing on the Internet: Issues, Challenges and Opportunities in the Field of Occupational Assessment. *Computer-Based Testing and the Internet: Issues and Advances*. D. Bartram and R. Hambleton (Eds), Wiley.

Boud, D., Cohen, R. et al. (1999). Peer Learning and Assessment. *Assessment and Evaluation in Higher Education* 24(4): 413-426.

Cheetham, G. and G. Chivers (2005). *Professions, Competence and Informal Learning*. p. 54, Cheltenham, Edward Elgar Publishing.

Davies, H. and J. Archer (2005). Multi-source feedback: development and practical aspects. *Clinical Teacher* 2(2): 77-81.

Freeman, M. and J. McKenzie (2002). Implementing and evaluating SPARK, a confidential web-based template for self and peer assessment of student teamwork: benefits of evaluating across different subjects. *British Journal of Educational Technology* 33(5): 553-572.

Garbett, R., Hardy, S., Manley, K., Titchen, A., McCormack, B. (2007). Developing a qualitative approach to 360-degree feedback to aid understanding and development of clinical expertise. *Journal of nursing management, 15,* 342-347.

Gehringer, E. F. (2001). Electronic peer review and peer grading in computer-science courses. *Proceedings of the 32nd SIGCSE Technical Symposium on Computer Science Education, Charlotte, North Carolina.*

Gipps, C. (1999). Socio-cultural perspective on assessment. *Review of Research in Education* 24: 355-392.

Gulikers, J., Sluijsmans, D., Baartman, L. and Bartolo, P. (2009). The power of assessment in teacher education. In A. Swennen and M. van der Klink (eds.) *Becoming a teacher educator. Theory and practice for teacher educators*, 173-188. Dordrecht: Springer

IMS LD. (2003). IMS Learning Design Specification. Retrieved July 12, 2009, from http://www.imsglobal.org/learningdesign/index.cfm.

IMS QTI. (2003). IMS Question and Test Interoperability Specification. Retrieved July 12, 2009, from http://www.imsglobal.org/question/index.html.

IMS SS. (2003). IMS Simple Sequencing Specification. Retrieved July 12, 2009, from http://www.imsglobal.org/simplesequencing/index.html

Jellema, F., Visscher, A., Scheerens, J. (2006). Measuring change in work behavior by means of multisource feedback. *International Journal of Training and Development, 10*(2), 121-139.

Joosten-ten Brinke, D. (2008). Assessment of prior learning. Heerlen, The Netherlands: Open Universiteit Nederland

Koper, R. and B. Olivier (2004). Representing the Learning Design of Units of Learning. *Journal of Educational Technology & Society* 7(3): 97-111.

Lockyer, J. (2003). Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof* 23(1): 4-12.

Miao, Y. & Koper, R. (2007). An Efficient and Flexible Technical Approach to Develop and Deliver Online Peer Assessment. Clark A. Chinn, G. Erkens, S. Puntambekar (Ed.) Mice, Minds, and Society, Proceedings of the 7th Computer Supported Collaborative Learning (CSCL 2007) conference, p. 502-511, July 16-21, 2007, New Jersey, USA.

Miao, Y., Sloep, P., & Koper, R. (2008) Modeling Units of Assessment for Sharing As-sessment Process Information: towards an Assessment Process Specification, Proceedings of the 7th International

Conference on Web-based Learning (ICWL 2008), pp. 132-144, Jinhua, China. Lecture Notes in Computer Science 5145 Springer 2008.

Miao, Y., van der Klink, M., Boon, J., Sloep, P., & Koper, R. (in press). Enabling Teachers to Develop Pedagogically Sound and Technically Executable Learning Designs. Distance Education, 30(2), 2009.

Ngu, A. H. H., J. Shepherd, et al. (1995). Engineering the 'Peers' system: the development of a computer-assisted approach to peer assessment. *Research and Development in Higher Education* 18: 582-587.

Rada, R., S. Acquah, et al. (1993). Collaborative learning and the MUCH System. *Computers and Education* 20: 225-233.

Rosbottom, J. (1994 September). *Software Tools for Peer Assessment*. Paper presented at the 2nd All Ireland Conference on the Teaching of Computing, Dublin, Ireland.

Rushton, C., P. Ramsey, et al. (1993). Peer assessment in a collaborative hypermedia environment: A case-study. *Journal of Computer-Based Instruction* 20: 75-80.

Sluijsmans, D., Brand-Gruwel, S., van Merriënboer, J. and Martens, R. (2004). Training teachers in peer assessment skills: effects on performance and perceptions. *Innovations in Education and Teaching International, 41*, 59-78.

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research* 68: 249-276.

Van der Heijden, B. and Nijhof, A. (2004). The value of subjectivity: problems and prospects for 360-degree appraisal systems. *International Journal of Human Resource Management, 15*(3), 493-511.

Vleuten, C. P. M. v. d. and L. W. T. Schuwirth (2005). Assessing professional competence: from methods to programmes. *Medical Education* 39: 309-317.

Volder, M.D., Rutjens, M., Slootmaker, A., Kurvers, H., Bitter, M., Kappe, R., Roossink, H., Goeijen, J., and Reitzema, H. (2007). *Espace: A New Web-Tool for Peer Assessment with In-Built Feedback Quality System.* Paper presented at ABR & TLC Conference, 2007, Hawaii, USA.

Whitehouse, A., Hassel, A., Bullock, A., Wood, L. and Wall, D. (2007). 360 degree assessment (multisource feedback) of UK trainee doctors: Field testing of team assessment of behaviors (TAB). *Medical Teacher, 29*, 171-176.