**ABSTRACT**
Effective formative assessment can be characterized by integrating assessment and instruction, engaging both teachers and students, using various assessment methods, and providing constructive and personalized feedback. In this paper, we propose a standard-based and service-oriented approach to support web-based formative assessment. We present how to develop online formative assessment by using an integrated IMS Learning Design (LD) and IMS Question and Test Interoperability (QTI) authoring tool. Moreover, we present how a formative assessment can be executed in an integrated LD and QTI run-time environment. In comparison with software development approaches, our approach can facilitate the development and execution of web-based formative assessment with characteristics leading to effectiveness.

**KEY WORDS**
Formative assessment, e-learning technical standard, IMS LD, IMS QTI, educational process modelling, service-oriented approach

1. Introduction

Formative assessment refers to activities undertaken by teachers, and by the learners in assessing themselves, which provide information to be used as feedback to adapt the teaching and learning activities in which they are engaged [4]. Many researchers ([6], [7], [8]) have emphasized the importance of formative assessment in student learning achievement. Black and Wiliam [5] concluded that formative assessment that precisely indicates student strengths and weaknesses and provides frequent constructive and individualized feedback leads to significant learning gains when compared to traditional summative assessment.

Although some studies ([9], [10], [12], [25], [27]) report that web-based formative assessment has many benefits to learners, researches on embedding formative assessment in learning environments and its effect are not plentiful [3]. A key reason is that existing web-based formative assessment strategies have some limitations (see section 5): loose integration of instruction and assessment, lack of full involvement of teachers and students, limited types of assessment measures, pre-defined feedback. In addition, most web-based formative assessment support environments can not support interoperability and reusability.

In this paper, we propose a standard-based and service-oriented approach to support web-based formative assessment. We claim that such a technical approach can facilitate the development of online formative assessment with characteristics leading to effectiveness.

2. Characteristics of Effective Formative Assessment

Characteristics of effective formative assessment are analyzed and identified in literature ([3], [22]). In this paper we identify and summarize characteristics from the perspective of process support. Meanwhile, technical requirements to develop effective web-based formative assessment are derived:

- **Assesses what is actually taught at right time:** Formative assessments must be aligned with learning objectives and directly relate to learning activities. While formative assessments may be very short and informal, activities should be purposeful and goal-directed. To be more effective, formative assessments must be ongoing. By continually assessing and providing opportunities for correction, instructors can guide students toward desired learning outcomes. As Black and Wiliam [5] pointed out, formative assessment is carried out frequently and is planned at the same time as teaching.

- **Actively involves both teachers and students:** To be more effective, broad and ongoing teacher discussions and involvement should occur throughout the assessment process. Meanwhile, the active engagement of students in the assessment process is seen as critical. Self-assessment is also considered an essential tool in self-improvement.

- **Uses multiple and varied measures:** Traditionally, multiple-choice and true/false items have been popular
methods. However, these are limited in scope and typically test each student’s capacity for rote memorization. Assessments of the cognitive domain should reflect, at least partially, its higher levels, such as synthesis and evaluation. Hence, teachers should use a variety of assessment methods such as essay, cooperative research projects, and performance test. For showing performance, special application tools (e.g., concept-mapping tool and simulators) may be needed.

- Provides constructive and personalized feedback: Feedback is a key element in formative assessment, and is usually defined in terms of information about how successfully something has been or is being done [22]. Feedback may be either peer or teacher directed as long as it is specific to the learning activity and assessment results. Students can remedy weaknesses in their learning abilities with the help of feedback. The teacher can employ the results of assessment to modify and adjust his or her teaching practices to reflect the needs and progress of his/her students. Additionally, formative assessment should offer opportunities to provide individualized feedback in a flexible and cost-effective manner [9].

In order to help the reader to understand the these characteristics, we describe a fictitious example of formative assessment, which is a part of an online course. The whole process consists of seven phases: 1) a tutor presents information about a topic and provides several papers relevant to the topic; 2) students are grouped in pairs and each student in a pair selects a different paper and reads it; 3) each student then write an article to analyze the paper; 4) pairs of students then exchange articles and review each other’s work by (i) deciding whether the article is acceptable without change or whether minor/major revision is required (ii) providing comments; 5) each student sends evaluation sheet to the original author, who has then to consider his/her response to the review, using a response form. Then each student hands in all documents to the tutor; 6) the tutor reviews the original version of the article, the student’s review and response to peer review, and then provide feedback to the student and suggest follow-up activities; 7) the student perform activities to remedy weaknesses.

This example has most characteristics mentioned above. In order to support such online formative assessment, at least technically, it is required to support R1): seamless integration of assessment and instruction, R2): full involvement of both teachers and students, R3): various assessment methods, and R4): providence with constructive and personalized feedback.

3. A Standard-based and Service-oriented Approach

Rather than developing a new web-based learning environment to meet the four identified requirements, we propose a standard-based and service-oriented approach to develop effective web-based formative assessment. Concretely speaking, we use IMS Learning Design (LD) [13], IMS Question and Test Interoperability (QTI) [14], and assessment-specific services to model formative assessment processes. The resulting formative assessment process models with necessary resources can be played in any standard-compatible web-based run-time environment. In this section, after briefly introducing QTI and LD, we present how our approach can meet four identified requirements.

QTI is an international e-learning technical standard which describes a data model for the representation of assessment item and assessment test and their corresponding results. It defines a set of interaction types which can be used to define basic question types (e.g., multiple-choice, match, slide, filling_in_blank, and etc) and complicated question types through combination. However, QTI can not support R1, R2, and can only partially support R4 by providing pre-defined feedback. It can support many assessment method (R3) except for those assessment that needs special application tools.

LD is an international e-learning technical standard, which is based on the Educational Modeling Language developed in Open University of the Netherlands [17]. LD is a pedagogical neutral language which can be used to model a wide range of pedagogical strategies [16]. LD can support R1, R2, and R4. However, it can not fully support various types of assessment (R3) except for a simple open-question type.

Through an analysis, we found that LD and QTI can be used complementarily to support formative assessment with the four emphasized characteristics. This section presents our approach to meet the four derived requirements.

- Using LD and QTI to support seamless integration of assessment and instruction: A teaching/learning process can be formally modeled as a unit of learning using LD. A unit of learning has a learning design specified in the manifest file and associated resources. A learning design consists of a set of components such as roles (e.g., learners and teachers), activities (e.g., learning activities and support activities), environments (containing learning objects and services), and properties. They are organized by using theatrical metaphors such as plays, acts, and role-parts as a hierarchically structured and process-oriented method. It is important to note that an assessment activity (e.g., providing evidence or providing feedback) can be defined and assigned to a role as a role-part. Such an assessment-specific role-part can be specified in any act in a play. QTI v2 provides guidelines for the integration of LD and QTI by including a QTI test/item document as a LD resource and coupling a LD property to an outcome variable of QTI item/test. Thus, it is required that an assessment activity either...
directly refers to a QTI document as an item of the activity description or is indirectly associated with a QTI document through an environment, where an item of a learning object refers to the QTI document. Therefore, assessment activities and learning activities will be treated in the same way and QTI documents and learning materials will be wrapped together within the same unit of learning.

- **Using IMS LD to model multi-role/user involved assessment processes**: IMS LD provides two pre-defined roles: learner and staff. Each role can be refined further in a hierarchical structure if necessary. Roles in a formative assessment process (e.g., a tutor, a student, or other stakeholders) will be assigned to different activities (e.g., teaching, reading, discussing, asking questions, answering, writing an essay, simulating, reviewing, providing feedback, responding, exploring, conducting experiment, summarizing, testing, and so on) performed in parallel or/and in sequence constituting a learn-flow. Note that a user can have more than one role and a role can be played by one or more users. Therefore, all users participating in the same formative assessment process with the same roles or different roles will interact with each other following the defined learn-flow towards the learning goal.

- **Using QTI to model traditional assessment methods and using LD services to model special assessment methods**: QTI can support various classic item types such as open-question, multiple choice, multiple response, fill-in-blank, hotspot, drag&drop, slide, match, and so on. It also provides sufficient flexibility to grow into advanced constructed-response items and interactive tasks. Furthermore, it provides mechanisms to design structured assessment and control branches and calculate weighted scores. That is, all standard assessment tasks and structured assessment that form the core subset of current practice can be modeled by using QTI. However, QTI can not support demonstrative assessment items in some special assessment methods, in which special application tools may be needed to demonstrate certain competences. LD offers an opportunity to integrate application tools as services. Any software tool used for the purpose of assessment (e.g., a concept-mapping tool or a simulator) can be integrated into a learning design as a specific assessment tool.

- **Using QTI and LD to model feedback**: In QTI feedback can be specified. However, this mechanism is not sufficient to support formative assessment, because it is based on predictable responses and pre-defined feedback. For example, in a multiple-choice question, for each correct or incorrect choice, a corresponding feedback may be pre-defined. By using LD properties and QTI outcome variables together, it can be supported that tutors or/and peer students provide personalized and constructive feedback based on unpredictable answers and performance results. Furthermore, by using LD conditions and notifications, the feedback can be used not only to recommend content for remedy but also to adapt learning activities or/and their sequences.

In summary, complementary use of LD and QTI makes it possible to model formative assessment processes with the four emphasized characteristics.

### 4. Using an Authoring Tool to Develop Effective Online Formative Assessment

Using the approach described above, a crucial step is to formally specify the process of formative assessment and necessary documents exactly following LD and QTI in the form of XML documents. Once such a formal model called a unit of learning is created, it then can be interpreted and executed in any standard-compatible runtime environment. However, it is quite difficult to code units of learning with assessment tests/items directly in XML.

In recent three years, several IMS LD authoring tools such as RELOAD [21], ASK-LDT [23], and CoSMoS [19] have been developed. These tools make it easy and efficient to create learning designs without needs to directly edit XML code. However, by using these tools to create an online formative assessment, users have to manually manage all complexity of the integration of LD and QTI documents. Such a task is time-consuming and error-prone.

In order to facilitate the development of online formative assessment efficiently by adopting this approach, we developed an integrated LD and QTI.v2 authoring tool by extending CoSMoS. Fig. 1 shows a screenshot of the user interface of the tool, which is used to specify the formative assessment example described in section 2. Although some functions (e.g., assessment test and image-based interactions) are still under development, the tool can facilitate the creation of formative assessment with the four characteristics.

- **The tool enables to model a unit of learning that seamlessly integrates instruction and assessment**: Users can define a formative assessment process consisting of learning activities and assessment activities through interacting with a user-friendly interface. As shown in Fig. 1, learning designs and associated assessment items can be defined separately so that a learning design can refer to multiple assessment items and multiple learning designs can share the same assessment item. When a formative assessment process is defined, the tool can automatically generate XML code according to the definition of the learning design, assessment items, and their associations. Then the tool can wrap all files as a unit of assessment, a specific unit of learning containing assessment-specific elements such as assessment items and assessment-specific services.
The tool enables to model learn-flow with the involvement of multiple roles: Users can create and define multiple roles in a hierarchical structure easily. A role-part can be defined by using drag&drop operations to associate a role with a learning/assessment activity. Then, learning activities and assessment activities can be arranged in parallel or/and in sequence to form a formative assessment process according to LD.

The tool enables to model basic and combined assessment items and to couple an outcome variable to a property: Users can define an assessment item being guided by the user interface. A response variable (e.g., RESP-comment in Fig. 1) will be automatically created and defined by the tool when each interaction (e.g., the extended text interaction at the bottom of the window in Fig. 1) is defined. Based on IMS QTI.v2, the tool enables to define complicated assessment items consisting of multiple and varied interactions. The identifier of a property coupling to an outcome variable can be automatically created through combining the identifier of the assessment item and the identifier of the outcome variable.

The tool enables to model personalized feedback and adaptive learning: Users can not only specify the pre-defined feedback associated to predictable responses (supported by QTI), but also model the feedback provided by tutors and peer students at run-time based on observed or/and unpredictable responses (e.g., an article) and performance information. In Fig. 1, the tool is used to define an extended text interaction, which enable a peer student to comment on his/her article by providing constructive and personalized feedback. Moreover, users can also model adaptive learning in a way that not only content but also activities will be adjusted by machine and/or human users (e.g., teachers and students) by using LD conditions and notifications.

5. An Integrated Execution Environment

This section presents an integrated LD and QTI compatible run-time environment.
In order to support for the execution of units of learning, we developed a generic integrative service framework, called CopperCore Service Integration (CCSI) [26]. This work was done in a project as a joint effort of both the Open University (UK) and the Open University of the Netherlands. The project extended earlier work which involved building a LD run-time service called CopperCore [11] and a corresponding web based client application called Service-based Learning Design Player (SLeD) [24]. The CCSI framework provides an extensible solution for the tight integration of loosely coupled services. The cross service concerns in particular are targeted by CCSI, alleviating the calling process from the burden of dealing with these concerns. In order to support formative assessment, a QTI service called Assessment Provision through Interoperable Segments (APIS) [1] was integrated by using CCSI. As shown in figure 2, an integrated execution environment to deliver online formative assessment has been established. This diagram depicts three clients (SLeD) and various services such as a LD enactment service (CopperCore), a generic assessment service (APIS) and assessment-specific services implemented in a service-oriented architecture. It is important to note that it is not a MUST to run all services in the same machine.

The integration of LD and QTI is realized through the synchronization of QTI outcomes and LD properties. How these variables and properties are matched is defined by the IMS Tools Interoperability Guidelines [16]. This synchronization is realized via an APIS assessment adapter. Clients depicted in Fig. 2 will use this adapter for making calls to APIS. Whenever a call by the client results in a change of a score outcome variable the adapter triggers an event notifying CCSI about it. Other service adapters may register as listeners to these events. In the example shown in Fig. 2, the LD enactment service adapter has registered itself as listener and will pick up on the events send by the APIS adapter by setting the associated LD property. This way the QTI outcome variables and the LD properties are kept in synchronization allowing the adaptation of the learning flow on the basis of the assessments results. This effectively allows the support for formative assessment.

So far, we have successfully executed many formative assessment examples in this run-time environment. In order to test whether the four derived requirements can be met, three formative assessment examples (including peer assessment and 360 degree feedback) are meticulously designed, which cover the four characteristics leading to effectiveness. The test results reveal that all formative assessment processes can be executed appropriately following the formative assessment models. The execution environment guided multiple users to interact with each other as expected. Fig. 3 shows a screenshot of SLeD, which is used to execute the formative assessment example described in section 2. A user was assessing the article of his/her peer student by answering two questions. These questions were modelled using the tool as shown in Fig. 1. The test results also reveal that sometimes the
SLeD user interface provides insufficient information. For example, when reviewing an article of the peer, the UI does not enable the user to view the article while answering questions. In addition, after a user answered a question and submitted, the system responded by showing the same question form again without the input text. The user will be normally confused whether the answers has been successfully submitted or not. Of course, these problems can be easily solved in the future version of the client. They are not the problems of our approach.

6. Comparison with Other Approaches

Existing web-based learning environments enabling formative assessment are normally developed in two technical approaches. One approach is to develop web-based applications [9, 27]. Using such systems, individual users learn by browsing web pages organized as an online course and by doing assessment at several points of course. At each point, a user has to repeatedly do test until s/he can correctly answer the presented multiple-choice questions selected from an item bank according to the learning and test records of the user. Rather than showing a correct answer, the feedback is pre-defined links to relevant web pages when an incorrect choice is selected by the user. Such an online formative assessment strategy is limited in four aspects: loosely-integrated learning and assessment, sole individual learning, limited assessment methods, and pre-defined feedback based on predictable responses. Furthermore, they can not support interoperability and reusability.

The other approach [12, 25] is to use standard-compatible assessment tool such as Questionmark Perception [20] and commercial web-based LMS such as WebCT [28]. The formative assessment strategies developed by adopting this approach may support interoperability and reusability to some extent. In addition, they enable to use more item types in formative assessment. However, they can not fully support other requirements to enhance online formative assessment.

The most related technical approach is to integrate IMS QTI and IMS Simple Sequence [15], as the work done in ASSIS project [2]. The development of these new learning experiences includes the extensive use of formative assessment tools and objects and the integration of assessment into adaptive sequences of content. This approach enables a seamless integration both instruction and assessment, variety in assessment types, and supports interoperability and reusability. However, such an approach assumes a learning model in which individual learners consume learning content with certain conditional control. Rather than learning activity, it supports for integrating learning materials with assessment and adapts content chunks to learner’s responses. It can not support a formative assessment process involved by multiple roles/users (R2). It provides no opportunities to use assessment-specific services (partial R3) and to offer personalized and constructive feedback based on unpredictable responses and performance results (R4).

In summary, other approaches can only meet partial requirements identified in section 2. Our approach can facilitate the development of formative assessment with all identified characteristics leading to effectiveness.

7. Conclusions

In this paper, we analyzed characteristics of effective formative assessment from the perspectives of process support and identified technical requirements for supporting online formative assessment. We proposed a standard-based and service-oriented approach to facilitate the development and execution of effective online formative assessment. Through a complementary use of LD, QTI, and assessment-specific services, a formative assessment can be modeled as a unit of assessment, which can be executed in any LD and QTI compatible execution environment. In comparison with other approaches, our approach can fully support seamless integration of assessment and learning activities, full involvement of multiple users/roles, multiple and various assessment methods, and personalized and constructive feedback. Furthermore, because our approach is based on open standards, it can support interoperability and reusability.

In order to facilitate the authoring of formative assessment, we developed an integrated LD and QTI authoring tool. Moreover, an integrated LD and QTI runtime environment are built to support the execution of formative assessment. We have tested the authoring tool and the integrated execution environment through modeling and executing many formative assessment examples. In particular, we have tested the system by using three meticulously designed formative assessment examples, which have the four emphasized characteristics leading to effectiveness. The test results demonstrated the feasibility of the approach and the usability of the authoring tool and the execution environment. After improving the client of the execution environment, we will conduct more serious evaluation.

Rather than programming knowledge and skills for developing software, process modeling knowledge and skills and knowledge about specifications are required for developing formative assessment models by adopting this approach. However, it is still difficult for average teachers and instruction designers to do modeling work. Our future work in this direction is to develop a high-level assessment process modeling language by adopting Domain-specific Modeling (DSM) approach. The assessment process modelling language will be defined through selecting concepts and rules used to describe formative assessment in practice. It is expected that
average teachers and instruction designers can understand and design formative assessment models represented in the assessment process modeling language. We will also develop mapping algorithms to transform the formative assessment models represented in the assessment process modeling language into the executable models represented in IMS LD and IMS QTI in a way described in this paper. Then, average teachers and instruction designers might be able to develop web-based formative assessment themselves instead of asking software developers to do it.

Acknowledgements

The work on this paper has been sponsored by the TENCompetence Integrated Project that is funded by the European Commission's 6th Framework Programme, priority IST/Technology Enhanced Learning. Contract 027087.

References