

Service-Oriented Flexible and Interoperable Assessment: Towards a Standardized e-Assessment System

Mohammad AL-Smadi¹ and Christian Guetl^{1,2}

¹ Graz University of Technology, Graz, Austria

² Curtin University of Technology, Perth, WA, Australia
msmadi@iicm.edu; cguetl@iicm.edu

Biographical notes:

Mohammad AL-Smadi is a Ph.D. Candidate at the Institute of Information Systems and New Media, Graz University of Technology, Austria. His research is in Technology-Enhanced Learning (TEL) in general and Computer Based/Assisted Assessment in particular especially flexible and interoperable e-assessment.

Christian Guetl is an Assistant Professor and Key Researcher at the Institute of Information Systems and New Media, Graz University of Technology, Austria; and Adjunct Research Professor at School of Information Systems, Curtin University of Technology, Western Australia. Dr. Guetl is author and co-author of more than 100 peer-reviewed publications and involved in numerous program committees and editorial boards. He is also the head of GÜTL IT Research & Consulting and an expert involved in national and international review and evaluation activities.

Abstract

Free-text answers assessment has been a field of interest during the last 50 years. Several free-text answers assessment tools underpinned by different techniques have been developed. In most cases the complexity of the underpinned techniques has caused those tools to be designed and developed as standalone tools. The rationales behind using computers to assist learning assessment are mainly to save time and cost, as well as to reduce staff workload. However, utilizing free-text answers assessment tools separately from the learning environment may increase the staff workload and increase the complexity of the assessment process. Therefore, free-text answers scorers have to have a flexible design to be integrated within the context of the e-assessment system architectures taking advantages of Software-as-a-Service architecture. Moreover, flexible and interoperable e-assessment architecture has to be utilized in order to facilitate this integration. This paper discusses the importance of flexible and interoperable e-assessment. Moreover, it proposes a service-oriented flexible and interoperable architecture for futuristic e-assessment systems. Nevertheless, it shows how such architecture can foster the e-assessment process in general and the free-text answers assessment in particular.

Keywords.

E-assessment; Free-text answers assessment; Learning Tools Interoperability, Middleware; Interoperable Architecture; Software-as-a-Service.

1 Introduction

Using Computers to assist the assessment process has been an important research for the last decades. Useful assessment methods highly depend on the learning objectives. Learning objectives have been classified in Bloom's Taxonomy into six levels: knowledge, comprehension, application, analysis, synthesis and evaluation (Bloom, 1956). Consequently, a variety of question types which can be used to assess different objectives' levels should be applied. Questions can be classified according to the nature of the users' response into, open questions (*fixed response*) and closed questions (*free response*). Fixed response questions -

also called *objective* questions- force students to have a fixed response by selecting an answer from a pre-prepared list of solution alternatives. By using free response questions -which also called *non-objective*- unanticipated answers formulate the user's response, moreover skills like programming, essays writing and meta-skills can be assessed (Dochy & McDowellb, 1997; Culwin, 1998; Hearst 2000).

The IMS Question and Test Interoperability (QTI) is the leading specification for assessment content metadata. The recent version of IMS QTI 2.1 specification has metadata for representing attributes for different question types. Moreover, for questions that have pre-defined responses (closed-response) the specification is designed to handle automatic scoring and feedback where items metadata incorporate passing score, minimum score and the default score, metadata for learner response declaration and response matching, as well as feedback model and timing so that feedback can be provided in an formative way during the test or at the end of the test. However, for the sake of reliable and valid assessment different types of items have to be provided (Brown and Race, 1996). The so called open-ended questions have to be used to evaluate higher levels of skills such as synthesis (Bloom, 1956). However, QTI provides metadata structure to represent open-ended questions but it lacks a response-processing template that can be used to automatically score student response and provided valuable feedback. Therefore, for QTI-conform e-assessment systems it is required that third-party tools for free-text answers scoring to be integrated within the context of the system.

Free-text answers assessment has been a field of interest during the last 50 years. Several applications based on different techniques have been developed during this period. Page (1994), has distinguished between assessing content and style of the answer. Content refers to the answer's body, what does the answer say, where style is related to the syntax and mechanics of writing. As cited in (Valenti, Neri and Cucchiarelli, 2003) in order to grade a free-text answer, both content and style are important (Christie, 2003). However, there are different techniques that only assess one of them during their grading process. Free-text answers scorers can be classified according to the technique they utilize. Some of these systems are based on natural language processing methods (NLP), where some of them are based on statistical methods. A combination between both methods can be found in other systems. As cited in (Pérez-Marín, Pascual-Nieto, and Rodríguez, 2009) another classification can be found in Chung and O'Neill (1997), where such systems are classified into systems that depend on documents classification, systems of this category are multilingual and do not perform any linguistics processes. The other category is the systems that assess the text meaning where a semantic, morphological and/or syntactic analysis is performed.

According to Hearst (2000) using computers to assess free-text answers can support the educational community with effective instructional material for improving reading, writing, and communications abilities. As e-assessment main rationales are to save time and costs, to reduce staff workload, and to provide valuable feedback, the scoring process of free-text answers should be automated and integrated to the learning process (Valenti, Neri and Cucchiarelli, 2003). Therefore, both of the free-text answer scorer and the e-assessment system have to be flexibly designed to support interoperability on both levels of content and services. To this end, this paper discusses the following research questions: Why do we need a flexible and interoperable e-assessment? How e-assessment systems can be designed to be flexible and interoperable? How flexible and interoperable e-assessment architecture can foster free-text answers assessment?

To this end the rest of the paper is organized as follows: section 2 provides insights free-text answers assessment tools and provides a historical overview of the available tools and research. Section 3 discusses the importance of using standards and specification in designing e-assessment systems as well as possible e-assessment content standards and specifications. Based on that, it addresses the importance of having flexible and interoperable e-assessment

system. Moreover, it discusses possible challenges and proposes a flexible and interoperable architecture for a futuristic e-assessment system. Section 4 discusses a learning scenario and shows how the proposed flexible and interoperable e-assessment architecture can be used to foster free-text answers assessment. Section 5 concludes the paper.

2 Historical Review of Free-text Answers e-Assessment

Several research publications have elaborated the history of using computers to assess short free-text answers (Wresch, 1993; Whittingdon and Hunt, 1999; Hearst, 2000; Darus and Stapa, 2001; Williams, 2001; Valenti, Neri and Cucchiarelli, 2003; Gütl, 2008; Pérez-Marín, Pascual-Nieto, and Rodríguez, 2009; Karanikolas, 2010). The use of computers to assess free text answers goes back to the 1960s where a pioneer system Project Essay Grader (PEG) was developed by Page (1994). PEG was based on the deployment of the computers statistical capabilities in the process of textual features detection. Page identified some variables related to the text features such as, “word length, essay length in words, number of commas and number of uncommon words”. He also believed that some of these features could not be directly extracted by computers but they could be approximated and he referred to them by “proxes”, and termed the ones evaluated by human raters as “trins” (Page, 1994). According to Wresch (1993), most of the teachers did not know that there was software for automatic assessment of students’ essays at that time. In the 1970s, Slotnick and Finn had some improvements in the AEG arena. Slotnick used Page’s approach with little changes in identifying “trins” and “proxes”, while Finn evaluated the correlation between the low frequency words and the writing quality (Wresch, 1993).

In the 1980s, there has been more interest in providing feedback to the students about their essays. Two main tools had been developed for this purpose, The Writer’s Workbench tool (WWB) which was developed by AT&T was used to evaluate students writing abilities in terms of “spelling, diction and readability” (Kukich, 2000). The other one was the Writer’s Helper (WH) developed by Conduit for writing evaluation with reference to “word frequency, sentence variety, and transition word and paragraph development”.

The 1990s was influenced by the ideas of the 1980s (Wresch, 1993). Two efforts were made to advance the free text answers evaluation research. The first one was the Hal Hellwig’s tool to grade business writing by using the idea of Semantic Differential Scale (SDS). Set of 1,000 commonly used words have been used to construct the scale for evaluating the writing quality. The second effort which is based on the Hellwig’s one was the Alaska Assessment Project. The system was based on textual features detection and variable lists building. An expansion to the variables’ lists used by Page’s system with two additional readability indexes. “Fogg readability” and “Flesch readability” indexes had been used to in the process of reading level determination. According to Wresch (1993), the project had better results than Page’s PEG, with a higher correlation between the system score and the human rater’s one.

Newbold (1990) stressed on the importance of using computers for AEG, but with new techniques rather than the ones used for style grading. The deployment of other techniques such as Natural language processing (NLP) and Information Retrieval (IR) has motivated the researchers to develop new ideas. According to Pérez-Marín, Pascual-Nieto, and Rodríguez, (2009) in 1997, Page’s system has become commercially available. Three new systems were introduced in the same year. The Intelligent Essay Assessor (IEA), which was developed at Colorado University in USA to assess the content of the students’ essays via a Latent Semantic Analysis (LSA) (Foltz, Laham, & Landauer, 1999). E-rater, which is an enhanced version of the Educational Testing Service I (ETS I) combines between NLP and statistical

techniques to measure the organization and the sentence structure rather than essay content (Burstein et al., 1998). The Vantage Learning Technologies, which is an American company developed a new system to assess both the style and the content. This system is based on Artificial Intelligence (AI) approach and called IntelliMetric (Vantage Learning Tech., 2000). A year later ETS developed a new system for content grading and they called it C-rater (Burstein, Leacock, & Swartz, 2001). Since 1999, E-rater has been used in the GMAT exam. Two Years later, ETS invested over a million dollars in the Criterion project to produce the Criterion 1.0 web interface, which is based on E-rater. In 2002, Criterion 1.2 has been integrated with Critique and Criterion 2.0 was presented soon later. Over 200 institutions have purchased the system to have approximately 50.000 users that time.

In 1998, Larkey (1998) presented a new system that depends on text categorization techniques, text complexity features and linear regression methods to automatically grade essays. A year later, the Schema Extract Analyze and Report (SEAR) was presented by Christie (1999). SEAR uses pattern matching techniques to automatically grade the essays content. In 2000, Apex Assessor was developed by Dessus, Lemaire and Vernier (2000). The system is similar to IEA where both of them are based on LSA. In the same year Ming, Mikhailov and Kuan (2000) created IEMS based on the Indextron technique (Mikhailov, 1998). A year later the Automated Text Marker (ATM) was developed at the University of Portsmouth (UK) (Callear et al., 2001). The system looks for concepts in the text and their dependencies with two independent scores, one for the content and the other for the style.

In 2002, several systems were presented. Automark is based on deploying NLP techniques to perform an intelligent based search of answers with reference to a predefined scheme of answers. The scheme is a set of answers that were marked by computers (Mitchell, Russel, Broomhead, & Aldridge, 2002). Lütticke (2006) discussed an approach that uses semantic network to map candidate answers, assess the answers against a model of correct ones, identifies wrong and incomplete answers, and provides feedback in natural language. According to Lütticke (2006) the approach has been successfully used to assess free-text answers since 2002. Rudner and Liang (2002), developed another system called Bayesian Essay Test Scoring sYstem (BETSY), based on statistical analyses. In the same year, the Paperless School free text Marking Engine (PS-ME) was developed by Manson and Grove-Stephenson (2002). PS-ME uses Bloom's taxonomy (Bloom, 1956) and NLP to assess the answers.

In 2003, Auto-marking which is based on NLP and pattern matching methods was presented (Sukkarieh, Pulmand, & Raikes, 2003). In the same year CarmelTC was presented by Roše, Roque, and VanLehn (2003) to grade students' writing based on machine learning classification methods. CarmelTC is uses a rule-learning text classification method, and it combines results from syntactic functional analyses of text with "bag of words" classification approach. Moreover, the research of Mitchell, Aldridge, and Broomhead (2003) where they used the Intelligent Assessment Technologies (IAT) a commercial assessment engine to conduct a "progress test" in the Medical School at the University of Dundee in 2003. IAT employs NLP techniques to assess candidate answers against pre-defined computerized mark scheme template of answers.

In 2004, Williams and Dreher (2004) developed a system at Curtin University of Technology. They called it MarkIT which is underpinned by NLP and pattern matching techniques. E-Examiner (Gütl, 2008) is a recent example for automatically grading short free-text answers, the system where developed at Graz University of Technology in the year 2007. E-Examiner is web-based and uses a hybrid approach built on a natural language pre-processing chain and based on ROUGE (Recall-Oriented Understudy for Gisting Evaluation) (Lin, 2004) characteristics. ROUGE defines a set of statistical measures to automatically determine the quality of a summary by comparing it with reference summaries. In 2010, the

same research group from Graz University of Technology presented a prototype for a tool they call Automatic Question Generator (AQC). AQC utilizes an automated process to create different types of test items out of textual learning content, more precisely to create single choice, multiple-choice, completion exercises and open ended questions. AQC is capable to process learning content stored in various file formats, extracts most important content and related concepts, creates different types of test items and reference answers, as well as exports the those items in QTI format. (Gütl, Lankmayr, and Weinhofer, 2010).

3 Standardized e-Assessment System

Learning content reusability and interoperability, learner's information accessibility and share ability, are main matters of quality for any Learning Management System (LMS). Therefore, LMS should be designed and implemented to be standard-conform. E-assessment as an important part of any e-learning system also faces the same challenge and problem. Different standards and specifications have been developed to design and develop e-learning content and components.

3.1 Learning Content standards, specifications and guidelines

Several organizations and consortia are working on building standards and specifications for the domains of e-learning and e-assessment. Examples of these organizations are: Dublin Core (DC) (DC, 2009), The Instructional Management System Global Learning Consortium (IMS GLC) (IMS GLC, 2008), The Aviation Industry CBT (Computer Based Training) Committee (AICC) (AICC, 2009), The Alliance of Remote Instructional Authoring and Distribution Networks for Europe (ARIADNE) (ARIADNE, 2008), Advanced Distributed Learning (ADL) (ADL, 2008), and IEEE Learning Technology Standardization Committee (IEEE LTSC) (IEEE LTSC, 2008).

In e-assessment and e-learning domains, standards, specifications, and reference models can be classified according to their applications into the following (Devedžić, Jovanovic & Gašević, 2007): *Metadata Standards*: a set of standards used to describe Learning objects' (LO) attributes, Such as the authors, title and languages. This description can be published with the LOs to facilitate their search and retrieval such as, IEEE Learning Object Metadata (LOM) (IEEE LOM, 2008). *Packaging Standards*: describes the assembly of LOs and other complex learning units (e.g. online courses) from various texts, media files and other resources. Such assembly can be stored in a Learning Object Repository (LOR) and imported in a Learning Management Systems (LMS) such as, IMS Content Packaging and IMS Learning Design) (IMS CP, 2008). *Learner Information Standards*: Formulates the description of the learner information and used to exchange that information between several systems, rather than their use in users modeling and personalization such as, IMS LIP (Learner Information Package) (IMS LIP, 2008) and PAPI Learner (Public and Private Information). *Question and Test Standards*: Special types of standards which are used in the assessment systems to represent questions and tests. IMS QTI (Question and test Interoperability) (IMS QTI, 2008) is an example of such standards. *Communication Standards*: specify the users' access to the LMS content, assessments, collaborative tasks and services communication. *Quality Standards*: specify the pedagogical, technical, design and accessibility perspectives for the LOs' quality. *Semantic Standards*: specify how we can organize content and refer to it in the semantic web.

According to (Shepherd, 2006, p.80) e-learning standards and specifications can be grouped into the following categories: *Authentication*: specifications or standards on how

systems can provide single-sign-on access to individuals and tools. *Content Packaging*: specifications or standards for packaging e-learning or e-assessment content in order to provide sharable content as well as to facilitate content transmission between tools and systems. *Data Definitions*: specification and standards that provide some kind of schema that represent logical data structures of content items such as courses, assessment items, or learner information. *Data Transport*: specifications or standards that explain and describe how data can be transferred among systems. *Launch and Track*: specifications or standards that explain how content (courses, assessments, etc) can be launched and tracked by LMSs. *Metadata*: specification or standards that describes data-about-data which mainly used by LMS for content tagging so to facilitate content search and retrieval. *Philosophical*: specification or standards that represent a framework for describing the overall learning process, materials, services and tools.

3.2 e-Assessment Content Standards and Guidelines

Despite the variance in e-learning content specifications and standards, e-assessment content has a limited number. The IMS QTI represents a data model for describing question (assessmentItem), test (assessmentTest), and their corresponding results reports. Unified Modelling language (UML) has been used to abstractly describe the data model which facilitates the binding with programming tools via the industry standard eXtensible Markup Language (XML) which provides a platform independent interchange and interoperability between different assessment tools and LMSs.

IMS QTI is designed to provide a well-formed assessment content where questions can be created, stored, and exchanged independently from the authoring tool. Moreover to support the deployment of item banks that can be used among several assessment authoring tools and LMSs. Similar to questions the specification is designed to provide a well formed representation of tests so that they can be created by selecting questions form item banks, stored, and exchanged between different assessment delivery tools and LMSs. Moreover, QTI specification supports systems with the ability to report test results.

The IMS QTI information model consists of two main data structures:

- ASI (Assessment, Section, and Item) data structure: for assessment content representation as *Assessment* represents the test unit; *Section* is a group representation of sub-sections and assessment items that may share common learning objectives; *Item* is the fundamental structure that holds information about the question and how to score it. Scoring is handled within the model by transforming the candidate (student) responses into outcomes using pre-defined response processing rules.
- Results Reporting: represents the results from the candidate interactions as *Context*: holds information session variables such as participant username, ID, and institution; *Assessment Results*: used to report the results of candidate's interaction on both levels test (testResult) and item (itemResult).

QTI items are classified according to their points of interaction into: *simple items*, and *composite items*. Simple item only have one point of interaction (e.g. single-choice, multiple-choice, cloze, match, hotspot, graphic-order), composite item is the item that contains more than of point of interaction where multiple instances of the same type of interactions or different types of interactions can be provided. QTI item has also an optional set of rules for converting the candidate's responses based on the interaction type into assessment outcomes. The process of converting the candidate responses into outcomes is called *response processing*. Response processing is used for some item types automatic scoring and may provide immediate or timely feedback based on the candidate response. Response processing is handled by applying a set of *responseRules* to evaluate expressions of item variables using

responseConditions (i.e. *responseIf*, *responseElseIf*, and *responseElse*). Moreover QTI has standard response processors called *response processing templates*.

QTI terms the open-ended questions as *extended text item* and uses a specific interaction type called *extendedTextInteraction* to handle responses. However, the process of response processing for *extended text item* requires external tool for free-text answer scoring or to be done manually by the teacher. Therefore, for QTI-conform e-assessment systems it is required that third-party tools for free-text answers scoring to be integrated within the context of the system. This will save costs and time, and maintains consistency in questions scoring. The following sections propose a solution for this problem via enhanced flexible and interoperable e-assessment system architecture.

Although QTI is the leading e-assessment content metadata, it has some limitations and challenges. For instance, the so-called impedance mismatch between the features offered by the standard and the ones needed in a particular application domain (Helic, 2006). IMS QTI has some difficulties in some application domains (such as, foreign languages teaching). One of these difficulties is that the IMS QTI is designed to formulate general types of questions and does not take into consideration some specific questions (e.g. Crossword puzzle) and test types for a particular domain (Milligan, 2003). According to (Smythe, 2000) the QTI specification is not related to didactical issues and tries to be didactically neutral as possible. Moreover, it has proved to have high complexity during assessment authoring and delivering, it does not cover cognition aspects, as well as it has no text and item analysis, in the other hand it has a model for results reporting (Chang et al., 2004). Chang et al. (2004) have proposed a SCORM 1.3 metadata extension for e-assessment content. The metadata model is called MINE SCORM and has been designed to cover cognition level, discrimination, instructional sensitivity and difficulty, different question types, as well as feedback provision and item/test analysis. The authors used Bloom's taxonomy (Bloom, 1956) of the cognitive domain to classify questions and assessments, and an assessment analysis model that provides useful statistical data about the items to teachers, students, and the system.

3.3 Interoperability

Interoperability has always been a challenge for e-learning software designers and developers. How to integrate third-party tools with LMSs in a way to extend the LMS provided services. Several definitions have been provided to the term interoperability. The Oxford English Dictionary (<http://oxforddictionaries.com>) defines the word "interoperable" as: "(of computer systems or software) able to exchange and make use of information". The IEEE defines interoperability as: "the ability of two or more systems or components to exchange information and to use the information that has been exchanged". Taking into consideration the integration point of view, Merriman (2008) defines interoperability as: "The measure of ease of integration between two systems or software components to achieve a functional goal. A highly interoperable integration is one that can be easily achieved by the individual who requires the result". (O.K.I, 2010). Merriman discussed the aforementioned two definitions and argued that both of them do not take integration into consideration. Moreover, the author stressed on the integration goal level of achievement as a main measure for interoperability. Based on that, interoperability is not only the ability of sharing information, rather than it goes beyond that to cover the ability of sharing functions and services in flexible way of integration. Bull and McKenna (2004, p. 112) defines interoperability as: "interoperability describes the capacity for different systems to share information and services such that two or more networks can communicate with each other to exchange data in a common file format". Similar to Bull and McKenna definition, Crisp (2007, p. 158) defines interoperability as: "interoperability is the ability of a system, content or activity to be exchanged or used in a variety of situations with the confidence that it will

function in a predictable manner. Interoperability allows efficient use of resources and avoids the necessity to design a system, content or activity *de novo* for every context”.

Consequently, interoperability has two main levels: *information* (content, user data) level and on *tools level* (tools interoperability). Information interoperability has been a major research area for years. Several specifications and standards have been published. For content examples are IEEE LOM, IMS Meta-data, SCORM, and IMS QTI. For user data examples are IEEE PAPI, and IMS LIP. Some other supportive standards are IMS CP for content packaging and IMS LD for the learning process design and workflow. Tools interoperability is an emerging research where limited examples of specifications are available. Among these specifications we can mention the Open Knowledge Initiatives (O.K.I.) and its Open Service Interface Definition (OSID) (O.K.I., 2010), and CopperCore Service Integration (CCSI) (Vogten *et. al.*, 2006). A more recent research is the IMS TI (Tools Interoperability) specifications by which tools and LMSs are provided guidelines of how they can be designed to flexibly be integrated with each other. This decoupling of content and tools as well as building systems using Service-Oriented Architecture (SOA) supports the comprehensive idea of interoperability.

3.4 Service-Oriented Flexible and Interoperable e-Assessment

In the world of Technology Enhanced Learning (TEL) extending the LMS services via third-party tools holds a great promise and challenge in the same time. To what extent third-party tools and LMSs are flexibly designed to be integrated with zero-line codes? What are the levels of integration (integration goals)? What are the main requirements in order to reach such flexibility? These and some other questions have been considered during the research of tools and content interoperability. According to (Thorne, 2004), the following elements have to be considered when it comes to have an interoperable tools and services:

- *Data & information (content)*: e-assessment content has to be represented using common specifications and standards (e.g. IMS QTI, IMS LIP) so that different tools can share and reuse their content in a flexible manner.
- *Communication (transport & protocols)*: tools have to use common platform independent communication protocols (SOAP, HTTP) so that they can easily communicate to share functions, activities, or content.
- *Software Interfaces*: that forms as a contract between service provider and consumer (e.g. OKI OSID). Moreover, interfaces represent an abstraction level to tools and services which make them easily integrated into LMSs. Interfaces decouples between services implementation and access where service providers are free to evolve and improve their services without affecting consumers as well as consumers can switch between different service providers in case those providers share common semantic definitions for their services.
- *Domain Models*: provides a common conceptual understanding of the problem domain in general and e-learning domain in particular. Domain models help developers to have common understanding with input/output data, data representation, possible services, and their workflow to achieve specific goals. Examples of this are the e-learning Framework (ELF) and Framework REference Model for Assessment (FREMA).

Similar to Thornes' interoperability aspects, AL-Smadi and Gütl (2010) suggested the following requirements for a flexible e-assessment system: clear guidance represented by a well-formed *framework, Standards and specifications* that represents the whole process of assessment as well as the communication between the services and components, *Cross-domain requirements* analysis in order to define the specific requirements for each application

domain (such as, educational editor in the mathematic domain), and *Web services* that provide the cross-domain requirements and interact through well-defined interfaces.

Dagger et al. (2007) discussed the flexibility and interoperability challenges for what they called “next-generation” LMSs. Dagger et al., stressed on the importance of that LMS should exchange both information’s syntax and semantics which goes in line with IEEE definition as systems have to be able to share information and to use them as well. Moreover, they argued that semantic exchange is not enough, LMSs have to have control on the shared tools and services so that they can keep their workflows, internal representations, and tracking mechanisms. They also recommend a shared dynamic semantic view about services (such as Semantic Web) instead of APIs as in OKI OSIDs so that services can be easily selected, orchestrated, and consumed based on a common understanding of the learning process.

Based on that, e-assessment systems should be designed to work as standalone applications or to be flexibly integrated with other LMSs. In order to have a “pluggable” and flexible e-assessment system we have to distinguish between two levels of standardization by which we can tackle the challenge of information and tools interoperability (AL-Smadi, Guetl, & Helic, 2009). The authors discussed internal and external levels of standards-conformation. The internal level is usually used where assessment content and user information are designed to adhere to specifications and standards such as IMS QTI and IMS LIP for instance. The internal level is suitable when assessment systems are used as standalone systems and they only share their content and user information. In order to tackle the challenge of tools and services interoperability the authors discussed what they called external level where they recommended a standard-conform interface. This interface represents an abstraction level of adaptation and could conform to anyone of the tools interoperability specifications discussed before.

Taking into consideration those two levels of standards-conformation what is the suitable architecture that e-assessment should use to be pluggable and flexible? What are the challenges in designing such architecture? Moreover, how this architecture will foster e-assessment with integrating free-text answers scorers as third-party tools?

3.4.1 e-Assessment Flexible an Interoperable Architecture

The ultimate goal in having a flexible and pluggable e-assessment system is to design e-assessment Software-as-a-Service (SaaS). According to Dagger et al. (2007) future LMSs will satisfy wide range of needs by providing what they called “interoperable architectures”. What could be a possible architecture for the next generation e-assessment systems? Figure 1 depicts a flexible and interoperable architecture for a futuristic e-assessment system. The architecture reflects the idea of the two levels of standards-conformation based on SOFA (AL-Smadi, Guetl, & Helic, 2009). Those two levels go inline with Dagggers’ levels of interoperability (*intradomain* and *interdomain*) where this architecture fosters e-assessment systems to be flexibly used as standalone systems or to be integrated with other tools and systems.

As depicted in the figure the architecture addresses four main areas -represented with dashed borders- of *user agents*, *middleware*, *third-party tools*, and *e-assessment system*. The user agents represent the potential users for the flexible and interoperable e-assessment system. Futuristic e-assessment system should adopt SOA to represent its services (Millard et al, 2005; Al-Smadi & Gütl, 2010). Based on SOFA, two levels of services are available in e-assessment systems: e-assessment services, and common services. E-assessment services represent the common assessment services provided by any e-assessment system such as authoring, scheduling, delivering, grading, and reporting. Whereas the common services address the required services that can be available in different systems such as, security services (authentication & authorization) and infrastructure services such as database

management and network management. Moreover the architecture shows e-assessment content that should be standard-conform. As discussed before little number of specifications is used for e-assessment content, the more widely used is the IMS QTI. Learner's information should also adhere to specifications and standards such as IMS LIP and PAPI Learner so that to be sharable among different tools and systems.

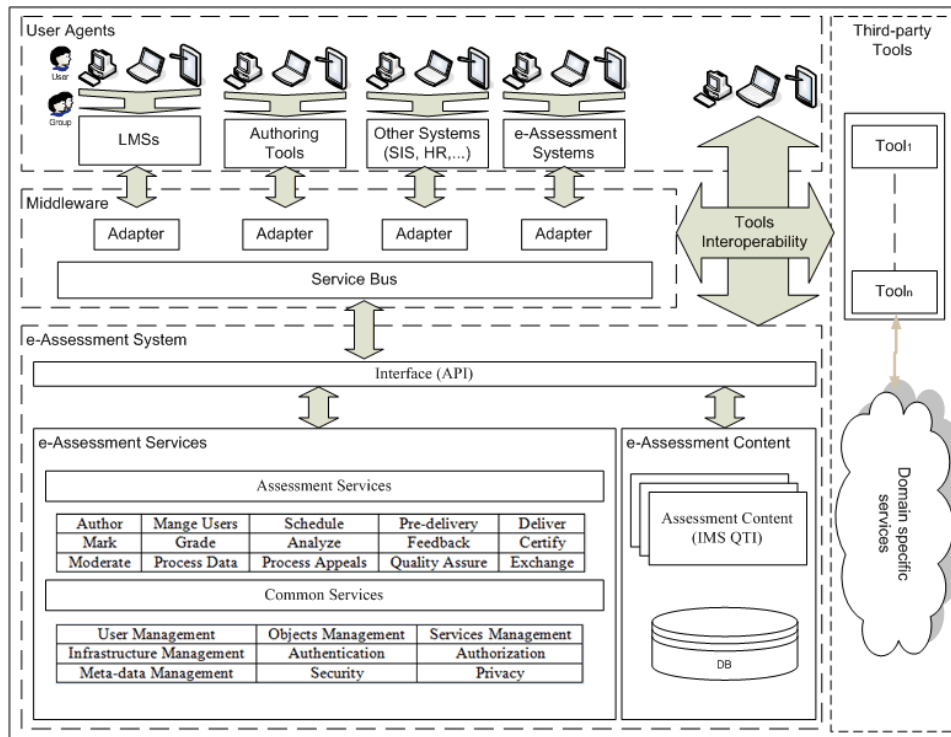


Figure 1. Architecture for a Service-Oriented Flexible and Interoperable e-Assessment system.

The middleware layer mainly contains a service bus and adapters to integrate the service bus services with other tools and systems. The adapters should adhere to one or more of the previously discussed tools interoperability specifications (i.e. O.K.I OSIDs, CCSI, or IMS TI). It is worth mentioning that the middleware should contain a *service registry* so that services from the e-assessment system as well as from other domain-specific services can be registered. This will foster the middleware to search, compose or orchestrate suitable services based on the demands of the user agents. The service registry can be part of the middleware architecture such as using UDDI (Universal Description, Discovery and Integration), or can be provided by an external service such as the JISC's IESR (Information Environment Service Registry) project (JISC IESR, 2010), which focuses on improving resource discovery mechanisms for electronic resources, that is, to make it easier to find materials to support teaching, learning and research.

3.4.2 Middleware for Flexible and Interoperable e-Assessment

Futuristic e-assessment systems should be flexibly designed: to be used as standalone systems where users of different devices can have a secure-direct access, moreover to be integrated with other tools and systems such as LMSs and Student Information System (SIS). The whole e-assessment system should behave like a service so that to be used by other related tools and services. This highlights challenges and aspects such as *single-sign-on (SSO)*

and *software instances management*, moreover, integration aspects such as *process integration, control integration, data integration, and presentation integration*. In a step towards tackling such challenges a middleware layer has been added to the architecture. Figure 2 depicts the middleware architecture in general and the service bus services in particular. The service bus services mainly related to the challenges and aspects discussed above. Generally, the middleware is suggested to foster third-party tools interoperability with LMSs moreover, to support the integration of e-assessment systems with other related tools and services.

As shown in the figure, the service bus services cover the following aspects: (1) *Single-Sign-On (SSO) services*: in order to support SSO, a set of services have been suggested. Those services mainly cover security, authentication, and authorization aspects. Mechanisms of delegated authentication and authorization have to be considered. The same problem has been discussed in (González et al., 2009) where the authors discussed different *Delegation Permits* initiatives such as OAuth (OAuth specification, 2009), Delegation Permits (Hasan et al., 2008), Shibboleth (Shibboleth, 2009), and OpenSSO (OpenSSO, 2009). Moreover, the authors have suggested an adapted technology of OAuth which they called “Reverse OAuth” as a useful model to support SSO among e-learning systems. (2) *Integration services*: refereeing to O.K.I. adopted definition of interoperability as measure of ease of integration between to systems to achieve a functional goal, several integration aspects have to be considered. Among these aspects the following have been highlighted:

- **Process Integration**: how much tools are flexibly integrated to support the performance of a process. Moreover, how much they agree on required events for this service as well as how much they enforce constraints. In order to achieve that tools have to share a common understanding of the application domain, services, learning activities, tasks, standards and specifications, and outcomes. This stresses the requirement for a framework and reference models for e-assessment in particular and for e-learning in general. Moreover, a need to the so-called “semantic exchange” among services in order to explain what services can do, what are their inputs/outputs, how they can be managed and used (Dagger et. al., 2007).
- **Control and Presentation Integration**: to what extent tools exchange control on their services, processes, interactions, appearance, and behaviour. How much flexible the tools can adapt their appearance and behaviour according to LMS style. How much flexible to change the tools internal workflow and to utilize it within the LMS workflow.
- **Data Integration**: do tools share the same specifications of their data (content and information)? To what extent they cooperate to maintain data consistency? Do they share the same semantic constraints on the data they manipulate?
- **Management and Supportive services**: this group of services handles aspects related to session management, tools instances management, services registration, services management in general such as services search and orchestration, data and process backup.

For each tool an adapter (interface) has to be provided in the middleware. Those adapters have to be designed to adhere to one or more of the tools interoperability specifications discussed before. This fosters third-party tools in general and e-assessment systems in particular to expose their services in a flexible way, moreover in order to adapt their class’s interfaces according to clients needs.

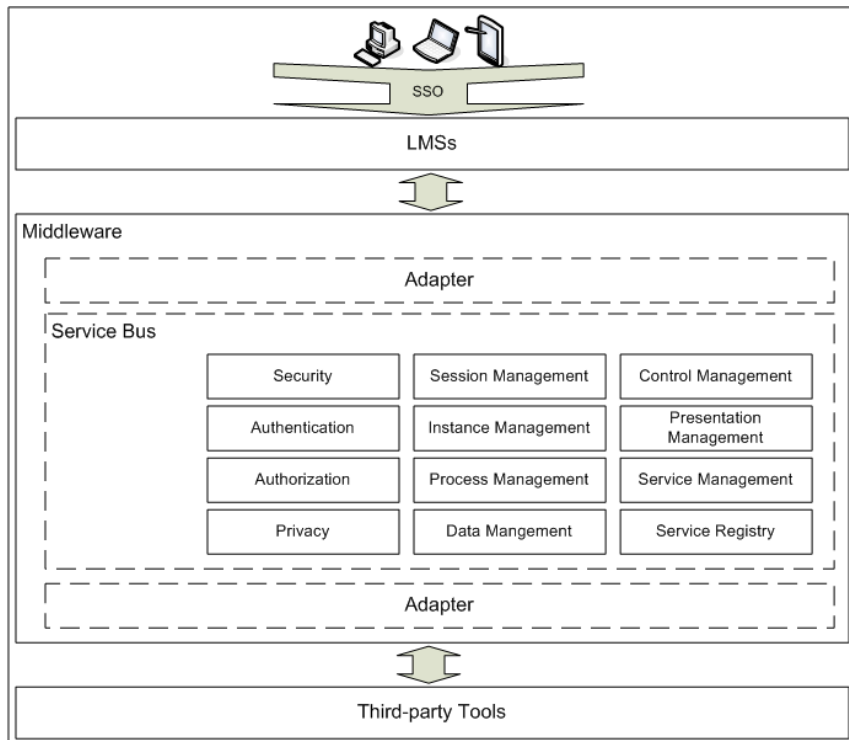


Figure 2. Architecture for a Middleware to foster Learning Tools Interoperability.

4 Discussion

In order to show how the proposed flexible and interoperable e-assessment architecture can foster free-text answer assessment the following scenario will be discussed. Let us assume that we are assessing the English writing skills for students. As part of this learning activity the instructor (1) selects an article for instance “History of Austria”. Moreover she is interested in providing her students QTI-conform automatically created open-ended questions based on the selected article. Therefore, (2) she uses the Automatic Question Creator (AQC) tool mentioned in section 3. Then (3) she uses the test authoring tool as part of the university e-assessment system to author a test based on the automatically created items. (4) The questions are delivered to the student using the e-assessment system as part of the university learning environment. (5) The students answer the test and provide free-text answers for the questions. (6) The instructor uses an automatic free-text answer marking tool to mark the students’ answers for instance e-Examiner form section 3. (7) The instructor uses the results from e-Examiner to grade the learning activity and provides feedback to her students.

Based on this scenario, the instructor is using four stand-alone tools or systems to perform the learning activity, a Content Management System (CMS) to select the learning material, AQC to automatically create open-ended questions, e-assessment system to author, schedule, and report the test, and the e-Examiner to automatically mark the students’ free-text answers. Moreover, the instructor has to be familiar with those tools and she has credentials to access and use all of them. Nevertheless, she has to take the output from one tool to the other manually based on the scenario sequence. The created test is not adaptive and the student didn’t receive timely feedback and the learning activity didn’t consider learner model and preferences. In order to tackle all of these problems let us assume that previous scenario is performed using our proposed flexible and interoperable architecture. The AQC, e-Examiner, as third-party tools will be integrated with the e-assessment context using the Middleware services. Moreover, the e-assessment system will be integrated with the university e-learning environment using the middleware services. The instructor uses the LMS to create the whole

learning activity, designs the learning workflow, selects candidate students, schedule the tasks. The LMS through the proposed middleware utilizes the AQC to create the questions, uses the e-assessment system to author and deliver the test within the context of the LMS, uses the e-Examiner to automatically mark the students' free-text answers, provides feedback to students and instructor, and updates the learner model and the learner knowledge state.

5 Conclusions

Computer Based/ Assisted Assessment has been an important research field over last five decades. Free-text answers assessment has been a field of interest over the last 50 years. Several free-text answers assessment tools underpinned by different techniques have been developed. The complexity of these underpinned techniques has caused those tools to be designed and developed as standalone tools. The rationales behind using computers to assist learning assessment are mainly to save time and cost, as well as to reduce staff workload. However, utilizing free-text answers assessment tools separately from the learning environment may increase the staff workload.

Futuristic e-assessment systems have to be standard-conform moreover, they have to have a flexible and interoperable architecture in order to be used as standalone systems or to be flexibly integrated with LMSs. IMS QTI terms the open-ended questions as *extended text item* and uses a specific interaction type called *extendedTextInteraction* to handle responses. However, the process of response processing for *extended text item* requires external tool for free-text answer scoring or to be done manually by the teacher. Therefore, for QTI-conform e-assessment systems it is required that third-party tools for free-text answers scoring to be integrated within the context of the system. This will save costs and time, and maintains consistency in questions scoring.

Using the Software-as-a-Service (SaaS) model may foster e-assessment systems to be pluggable and flexibly integrated with other related tools and systems (for instance free-text answer scorers). However, this highlights the challenge of Learning Tools Interoperability (LTI). LTI does not only consider information (content, and user information) interoperability moreover it goes beyond that to discuss the capability of third-party tools integration with LMSs. When it comes to LTI, several aspects such as single-sign-on (SSO), session management, tool instances management, process integration, control integration, data integration, and presentation integration have to be considered. Therefore, futuristic e-assessment systems have to have service-oriented flexible and interoperable architectures. Nevertheless, such architecture can be enriched with a middleware by which some of the aforementioned aspects can be tackled.

Designing futuristic e-assessment systems based on the proposed architecture will foster the assessment process in general and free-text answers assessment in particular. Moreover, it will flexibly integrate the available free-text answers assessment tools and other related third-party tools within the context of learning environment. Moreover, it will realize e-assessment goals and rationals of saving time and money, reducing assessment load and staff workload, and maintaining assessment validity and scoring transparency.

References

- ADL. 2008. The Advanced Distributed Learning, <http://www.adlnet.gov/about/index.aspx>, last retrieved October 7th 2008.
- AICC, 2008. The Aviation Industry CBT (Computer Based Learning) Committee, <http://www.aicc.org/index.html>, last retrieved October 7th 2008.

- AL-Smadi M., and Gütl C. 2010. SOA-based Architecture for a Generic and Flexible E-assessment System, Proceedings of IEEE EDUCON Education Engineering 2010 – The Future of Global Learning Engineering Education, April 14 - 16, 2010, Madrid, SPAIN.
- AL-Smadi M., Guetl, C., Helic, D. 2009. Towards a Standardized e-Assessment System: Motivations, Challenges and First Findings, International Journal of Emerging Technologies in Learning (IJET), Volume 4, Special Issue 2: "IMCL2009".
- ARIADNE, 2008, ARIADNE Foundation, <http://www.ariadne-eu.org/>, last retrieved October 7th 2008.
- Bloom, B. 1956. Taxonomy of educational objectives: the classification of educational goals. Handbook I, Cognitive Domain. Longman, Whiteplains (New York); Toronto: Longmans, Green.
- Bloom, B.S. 1956. Taxonomy of educational objectives, Handbook I: the cognitive domain. David McKay Co Inc., New York.
- Brown, S., Race, P. 1996. *500 Tips on assessment*, Cogan Page, London, UK.
- Bull, J. & McKenna, C. 2001. *Blueprint for Computer Assisted Assessment*. London: RoutledgeFalmer.
- Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Bradenharder, L. & Harris, M. (1998). Automated scoring using a hybrid feature identification technique. Proceedings of the Annual Meeting of the Association of Computational Linguistics, Montreal, Canada.
- Burstein, J., Leacock, C. & Swartz, R. 2001. Automated evaluation of essays and short answers. In Proceedings of the 5th International Computer Assisted Assessment Conference, Loughborough, UK.
- Callear, D., Jerrams-Smith, J. & Soh, V. 2001. CAA of short non-MCQ answers. In Proceedings of the 5th International Computer Assisted Assessment conference, Loughborough, UK.
- Chang, Wen-Chih., Hsu, Hui-Huang., Smith, Timothy K. & Wang, Chun-Chia. 2004. Enhancing SCORM metadata for assessment authoring in e-Learning. *Journal of Computer Assisted Learning* 20:pp 305-316.
- Christie, J. 1999. Automated essay marking--for both style and content. In Proceedings of the 3rd Computer Assisted Assessment International Conference, Loughborough, UK.
- Chung, G. & O'Neill, H. 1997. Methodological Approaches to Online Scoring of Essays. Technical Report 461, UCLA, National Center for Research on Evaluation, Student Standards, and Testing, USA.
- Crisp, G. 2007. *The e-Assessment Handbook*. Continuum International Publishing Group, New York.
- Culwin F.1998. Web hosted assessment--possibilities and policy., Proceedings of the 6th annual Conference on the Teaching of Computing/3rd Annual ITiCSE Conference on Changing the. Delivery of Computer Science Education, 1998, 55–58.
- Dagger D., O'Connor A., Lawless S., Walsh E., and Wade V.P. 2007. Service-Oriented E-Learning Platforms From Monolithic Systems to Flexible Services, Published by the IEEE Computer Society 2007.
- Darus, S. & Stapa, S. 2001. Lecturers' expectations of a computer-based essay marking systems. *Journal of the Malaysian English Language Teachers' Association (MELTA)* 30, 47-56.
- Davies W. M. , and , Davis, H. C.2005. Designing Assessment Tools in a Service Oriented Architecture, In Proceedings of *1st International ELeGI Conference on Advanced Technology for Enhanced Learning*, Napoli, Italy, 2005.
- DC, 2008. The Dublin Core Metadata Initiative, <http://dublincore.org/>, last retrieved October 7th 2008.
- Dessus, P., Lemaire, B. & Vernier, A. 2000. Free text assessment in a virtual campus. In Proceedings of the 3rd International Conference on Human System Learning, Paris, France, 61-75.
- Devedžic V., Jovanovic J., Gašević D.2007, The Pragmatics of Current e-Learning Standards. *IEEE Internet Computing, Special Issue on Distance Learning*, Vol.11, No.2, May/June 2007, pp. 16-24.
- Dochy, F. J.; & McDowell, L. 1997. Introduction. Assessment as a tool for learning. *Studies in Educational Evaluation*, 23 (4), 279-298.
- Foltz, P., Laham, D. & Landauer, T. 1999. The intelligent essay assessor: Applications to educational technology. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning* 1(2). Retrieved from: <http://imej.wfu.edu/articles/1999/2/04/index.asp>
- Gütl C., Lankmayr K., and Weinhofer J..2010. "Enhanced approach of automatic creation of test items to foster modern learning setting." In Proc. of the 9th European Conference on e-Learning, Porto, Portugal, 4-5 November 2010, 225-234.
- Gütl, C. 2008. Moving towards a Fully-Automatic Knowledge Assessment Tool. *ijET International Journal of Emerging Technologies in Learning*, Vol 3, No 1 (2008).
- HasanR, ConlanR, Slesinsky B,Ramani N,Winslett M. 2008. Please permit me: stateless delegated authorization in mashups. In: Annual computer security applications conference (ACSAC).
- Hearst, M. 2000. The debate on automated essay grading. *IEEE Intelligent Systems*, 15(5), 22-37, IEEE CS Press.
- Helic D. 2006. "Template-based Approach to Development of Interactive Tests with IMS Question and Test Interoperability". In Proceedings of ED-MEDIA 2006, pages 2075-2081, AACE, Charlottesville, USA, 2006.
- IEEE LTSA, 2008, IEEE Standard for Learning Technology-Learning Technology Systems Architecture, <http://www.ieeeltsc.org/standards/1484-1-2003/>, last retrieved October 7th 2008.
- IEEE LOM, IEEE Standard for Learning Object Metadata, <http://www.ieeeltsc.org/standards/1484-12-1-2002/>, last retrieved October 7th 2008.

- IEEE PAPI, 2003, IEEE P1484.2 Learner Model Working Group: PAPI Learner, Draft 7 Specification, IEEE Standards Committee on Learning Technology.
- IMS GLC, 2008, The IMS Global Learning Consortium, <http://www.imsglobal.org/background.html>, last retrieved October 7th 2008.
- IMS LRM, 2008 Learning Resource Meta-data Specification Version 1.3 - Final Specification – HTML, <http://www.imsglobal.org/metadata/index.html>, last retrieved October 7th 2008.
- IMS CP, 2008. Content Packaging, Content Packaging Specification v1.1.4, <http://www.imsglobal.org/content/packaging/index.html>, last retrieved October 7th 2008.
- IMS LIP, IMS Learner Information Package Specification, <http://www.imsglobal.org/profiles/index.html>, last retrieved October 7th 2008.
- IMS TI. 2010, The IMS Tools Interoperability guidelines, <http://www.imsglobal.org/ti/index.html>, last retrieved October 20th 2010.
- IMS QTI, 2008. IMS Question & Test Interoperability Specification, Version 2.1 - Final Specification, <http://www.imsglobal.org/question/index.html>, last retrieved October 7th 2008.
- Fontenla J., Caeiro M., Llamas M., Anido L.. 2009. Reverse OAuth – A solution to achieve delegated authorizations in single sign-on environments, Computers and Security. Last accessed in October, 2010 at: <http://dx.doi.org/10.1016/j.cose.2009.06.002>.
- JISC's IESR 2010. <http://iesr.ac.uk/>, last visited October 30th 2010.
- Karanikolas, N.N. 2010. Computer Assisted Assessment (CAA) of Free-Text: Literature Review and the Specification of an Alternative CAA System. In Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE), 2010 19th IEEE International Workshop on. 116 – 118.
- Koper, E. J. R. 2001. Modelling units of study from a pedagogical perspective: the pedagogical metamodel behind EML. Open University of the Netherlands.
- Kukich, K. 2000. Beyond Automated Essay Scoring. IEEE Intelligent Systems, 15(5), 22–27.
- Larkey, L. S. 1998. Automatic essay grading using text categorization techniques, in 'Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval', pp. 90–95.
- Lin, C.-Y. 2004. Rouge: A package for automatic evaluation of summaries. In Proceedings of ACL 2004 workshop on Text Summarization Branches Out, Barcelona, Spain, pp. 74-81.
- Lütticke R. 2006. "Using Semantic Networks for Assessment of Learners' Answers", Proceedings of ICALT 2006, 1070-1072.
- Mason, O. & Grove-Stephenson, I. 2002. Automated free text marking with paperless school. In Proceedings of the 6th International Computer Assisted Assessment Conference, Loughborough, UK.
- Merriman, J. 2000. Redefining interoperability (Or why the IEEE and Oxford English Dictionary have it Wrong). The Open Knowledge Initiative (OKI), retrieved from <http://www.okiproject.org/view/html/site/oki/node/2916>, last visited October, 19th, 2010.
- Mikhailov, A. 1998. Indextron. Intelligent Engineering Systems Through Artificial Neural Networks 8, 57-67.
- Millard, D., Howard, Y., Bailey, C., Davis, H., Gilbert, L., Jeyes, S., Price, J., Sclater, N., Sherratt, R., Tulloch, I., Wills, G. and Young, R. 2005. Mapping the e-Learning Assessment Domain: Concept Maps for Orientation and Navigation. In *Proceedings of e-Learn 2005*, Vancouver, Canada.
- Milligan C. 2003. Question and Test Interoperability (QTI): Extending the specification for Mathematics and Numerical Disciplines. *Maths CAA Series*, The Maths, Stats & OR Network, University of Birmingham, UK.
- Ming, Y., Mikhailov, A. & Kuan, T. 2000. Intelligent essay marking system. In Learners Together, Cheers, C. (ed.). Ngee ANN Polytechnic.
- Mitchell T., Aldridge N., and Broomhead P., 2003. "Computerised Marking of Short-Answer Free-Text Responses", Manchester IAEA conference.
- Mitchell, T., Russell, T., Broomhead, P. & Aldridge, N. 2002. Towards robust computerised marking of free-text responses. In Proceedings of the 6th Computer Assisted Assessment Conference, Loughborough, UK.
- Newbold, W. W. 1990., Computers and writing assessment: A preliminary view. *Computers and Composition* 8 (1990), pp. 7–22.
- O.K.I., 2010, <http://www.okiproject.org>, last visited October 30th 2010.
- OAuth 2010. <http://oauth.net>, last visited October 30th 2010.
- OpenSSO 2010. [https:// opensso.dev.java.net](https://opensso.dev.java.net), last visited October 30th 2010.
- Page, E. 1994. Computer grading of student prose, using modern concepts and software. *Journal of Experimental Education* 2(62), 127-142.
- Pérez-Marín D., Pascual-Nieto I., Rodríguez P., 2009. Computer-assisted assessment of free-text answers. *The Knowledge Engineering Review* (2009), 24: 353-374 Cambridge University Press.
- Roše, C., Roque, A., D., D. B. and VanLehn, K. (2003), 'A hybrid text classification approach for analysis of student essays', *Build Educational Applications Using Natural Language Processing* pp. 68–75.

- Rudner, L. & Gagne, P. 2001. An overview of three approaches to scoring written essays by computer. Educational Resources Information Center (ERIC) digest, ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.
- Rudner, L. & Liang, T. 2002. Automated essay scoring using bayes' theorem. In Proceedings of the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- Rudner, L. & Liang, T. 2002. Automated essay scoring using Bayes' Theorem. *The Journal of Technology, Learning and Assessment*, 1(2), 3-21.
- Shepherd E. 2006. Definitions, Uses, and Benefits of Standards, Questionmark Corporation, USA, In *Online Assessment and Measurement Foundations and Challenges*, Information Science Publishing, USA, P 64 -84, 2006.
- Shibboleth 2010. <http://shibboleth.internet2.edu>, last visited October 30th 2010.
- Sukkarieh, J., Pulman, S. & Raikes, N. 2003. Auto-marking: using computational linguistics to score short, free text responses. In Proceedings of the 29th IAEA Conference, Theme: Societies' Goals and Assessment, Philadelphia, USA.
- Thorne S. 2004. Conceptual Architectural Framework, *O.K.I. Project Site*. MIT, retrieved from <http://www.okiproject.org/view/html/site/oki/node/1402>, last visited October, 19th, 2010.
- Valenti, S., Neri, F. & Cucchiarelli, A. 2003. An overview of current research on automated essay grading. *Journal of Information Technology Education* 2, 319-330.
- Vantage Learning Technology 2000. A Study of Expert Scoring and Intellimetric Scoring Accuracy for Dimensional Scoring of Grade 11 Student Writing Responses. Technical Report RB-397, Vantage, USA.
- Vogten, H., Martens, H., Nadolski, R., Tattersall, C., Van Rosmalen, P., Koper, R., 2006. Integrating IMS Learning Design and IMS Question and Test Interoperability using CopperCore Service Integration. Proceedings of the International Workshop in Learning Networks for Lifelong Competence Development at the TENCompetence Conference in March 30th-31st, 2006., Sofia, Bulgaria: TENCompetence.
- Whittington, D. & Hunt, H. 1999. Approaches to the computerised assessment of free-text responses. In Proceedings of the 3rd International Computer Assisted Assessment Conference, Loughborough, UK.
- Williams, R. & Dreher, H. 2004. Automatically Grading Essays with Markit. In Proceedings of Informing Science Conference, Rockhampton, Queensland, Australia.
- Williams, R. 2001. Automated essay grading: an evaluation of four conceptual models. In Proceedings of the 10th Annual Teaching and Learning Forum: Expanding Horizons in Teaching and Learning, Curtin University of Technology, Perth, Australia.
- Wresch, W. 1993. The imminence of grading essays by computer—25 years later. *Computers and composition*, 2(10), 45–58.
- WS-I. 2010, “Web services interoperability basic profile v1.1”, <http://www.ws-i.org/>, last retrieved October 20th 2010.