Developing an E-assessment environment for the 21st century

35th IAEA Annual Conference
Brisbane, Australia, September 13~18, 2009

Peter Hermans
Cito, Arnhem, The Netherlands

Abstract

Cito, the Dutch National Institute for Educational Measurement, has experimented with online and on demand testing, digital portfolio, adaptive tests and digital student monitoring systems. Cito also developed a variety of computer based and computer assisted National Examinations for Dutch secondary education, in cooperation with the National Examination Board and the schools.

ICE (Integrated Cito Environment), the codename for Cito's third generation e-assessment software, is built on a modular architecture that makes it possible to generate a working environment that is tailor made for any type of computer based assessment, high stakes or low stakes, from student tracking systems to national examinations, stand alone and online. In the new system a wide variety of item types, from straightforward text based multiple choice items to complex question types using interactive multimedia can be processed.

This paper addresses the changing design paradigms for e-assessment software and the growing demand for IT systems that can handle the construction, production and delivery of cutting edge computer assessments in an efficient, reliable and cost effective way.

Past, present and future of e-assessment

As a leading provider of assessment solutions, Cito has invested much time and energy in research and development of innovative e-assessments in the past 10 years. From the beginning of it's involvement in e-assessment, the institute has taken the position that computer use in assessment should contribute to the improvement of the quality of assessment in terms of validity, effectiveness and efficiency and furthermore, that the needs of the test taker should be leading, not the technology. Cito has never limited its efforts to one specific type of computer use. Instead a number of innovative new formats that provide versatile assessment options for modern education, learning and training were developed: computer supported assessments, computer based tests, computer adaptive tests and web based tests. Incorporating state of the art IT in tests and assessments requires significant research and investment in fields such as multimedia design, psychometrics, and information and computer technology. Research is focused on secure administration and reporting, interactivity and rich media, validity and efficiency of innovative question formats, automated scoring algorithms, computer adaptive test design, but also on screen design and typography, interaction design and the design of item templates.

Since its founding, in 1968, Cito has based a growing part of its core business on information technology. In the first decades, computers were used mainly for programming, data analysis and data processing. Then, soon after the introduction of the personal computer in the office environment, the first prototypes of item banking systems were developed, such as the TSS (Test Service System), a partnership between the Foundation for Educational Research (SVO), the University of Twente and Cito. The idea was to resolve a series of psychometric, administrative and organizational issues related to an expected demand for individualized, tailor made paper based tests, through integrated and efficient computer use. It was also expected that these item banks would serve as a source of inspiration for test developers. The development of this system was delayed, however, in anticipation of a large-scale introduction of IT in school curricula and training programs. In the meantime, the existing item banking systems were extended by modules for the registration of candidates and planning of test administration, a digital environment for test delivery and modules for scoring, data return and reporting and evolved into full grown assessment management systems.

After some initial experiments in 1998-2000, Cito had rapidly become a forerunner in the field of innovative e-assessments. After successful the development of two experimental nationwide computer exams, the introduction of computer-assisted and computer-based testing in all test programs had guided test development as well as our research program. The first experimental products were developed using third party test authoring software, but it

soon became apparent that the capabilities of these applications were too limited to handle the innovative assessments that needed to be developed.

Policy makers in education at that time were convinced that the use of computers in assessments and examination would speed up the stagnant introduction of computers in the classroom, provided these tests were cutting edge and innovative and thus revealing the full potential of computer use and multimedia in education At the same time there was a false assumption that the large variety of paper based tests and testing programs, which is typical for Dutch education, could be simply converted into interactive digital assessments. But the first generation of test authoring software could only handle a limited number of item formats, had little or no possibilities for use of (moving) images and sound, let alone the use of interactive multimedia.

Extensive comparative research led to the conclusion that there was no system on the market that would support all demands of innovative test design and development. A second conclusion was that every testing program gave preference to the system that would closest meet the specific demands of its sponsors, the different levels of education and training and different education systems. This made it impossible to come to a common choice for one third party system. Adapting or customizing existing systems turned out to be more expensive and time consuming then building a new system from scratch. Another major conclusion was that the need to anticipate future changes and innovations in information technology as well as in education and training, would make dependence on third parties in development an unacceptable risk.
The obvious choice at that time (2000) was to develop a genuinely new assessment management system (Cito Tester / Cito Bank) in cooperation with a number of specialized partners. However, soon after the introduction of the first prototype of this second generation assessment software, it became apparent that the increasing variation in test and exam formats inside and outside education and training, resulted in an increasing pressure on the flexibility and versatility of new systems. Meeting all of these requirements in one single application would eventually result in large, sometimes risky investments. On the other side, ignoring the limitations of the new system would eventually frustrate the future development of new and innovative item types and assessments. The new software would end up as a legacy application before it was completely implemented, just like its predecessors

ICE: building assessment systems for tomorrow

The construction of the next generation assessment management software (codename ICE) is based on a number of important notions for sustainable, long term application development.

- A high quality assessment system is a prerequisite for the diversified and individualized learning concepts of tomorrow. Continuous assessments will become

more and more integrated with learning and training, along continuous learning tracks, and smooth horizontal and vertical transitions will require a transfer of high quality data.

- The next generation assessment system must facilitate the inevitable transition from paper based testing (PBT) towards computer based testing (CBT).
- Future learning arrangements with integrated assessments, tests and monitoring of progress will require further standardization of existing open standards for the exchange of content (IMS QTI), personal details, results (including portfolio and other qualitative documentation tools), metadata and feedback formats.

This type of ambition is at odds with the general trend in the world of test development where the urge to invest in major innovations is low. Because Cito already employed highly committed, qualified system analysts and software developers, with years of experience in developing high level computer based assessments, as well as the technical knowledge of the Microsoft.NET Framework, the decision was made to develop assessment system software for 21st century entirely in-company.

A new perspective

The history of computer-based testing has shown that systems need to be extensible because of constant developments in education, in IT and in test development itself. Any new system should allow for innovation in the broad area of test development (item construction, psychometric models, item types, distribution models) and support computer use in tests, examinations and assessments. At the same time, the costs of development of all kinds of extensions to the system, need to be reduced. To address these problems, third generation computer based assessment systems are based on a modular and highly extensible architecture, with support for multiple distribution models, such as the “Smart Client” application deployment method.

Smart client applications combine the robustness of the "rich clients" that were developed in large numbers two decades ago (including the first generation of computer based testing software) with the flexibility of the 'thin client' applications developed as a result of the internet boom.

As the complexity of a rich client application increases, it becomes increasingly difficult to implement them in a reliable and secure way on the client computer. Rich clients used to be prone to ‘application fragility’: the installation of a new application could have disastrous consequences for an application already installed on the client computer, especially when shared libraries were used. The maintenance and modification of a rich client application is not really effective, any change to the program requires the production of an entirely new version and a complete re-install on the client computers.

Thin clients are browser-based applications hosted on a server with a direct connection to the Internet. Applications are implemented, maintained and updated on the server; installation on the client side is not necessarily required. The only disadvantage thin clients have, is that execution of the program at the client side depends on the quality of a network or internet connection. Thin clients continuously exchange information and data with the server and are therefore vulnerable to crashes, network time-outs and delays. Furthermore, due to security considerations, browsers have limited control over the client operating system, which makes it virtually impossible to automatically secure the environment during high stakes tests.

Smart clients combine the benefits of rich and thin clients, because they use both local sources and network/ internet resources. They serve each user whether the application is used for a long or for a short time, continuously or interrupted. Installation and maintenance can be handled in an intelligent way and smart clients can easily be made suitable to do their work under successive generations of an operating system. Smart clients are considered 'lightweight' which makes it possible to run computer based tests from a USB flash drive or on portable devices (smart phones, game computers). Due to their nature, smart clients can have full access to the hosting operating system, which allows them to secure the environment, force the intended presentation and cache content.

Efficiency as a motive

Computer based testing evolved quickly from a screen representation of 40 multiple choice questions with black and white illustrations to multimedia driven interactive test environments. In the same time span, the demand for a greater variety of item types increased. The variation in test use has also increased, partly because of a growing demand for professionally-developed, effective assessment tools and services outside basic public education. This greater variety of products, services and media calls for operational excellence in test construction, test production
Assessment management systems for the 21st century allow for customizing item and test production and pre-testing, data return and analysis and reporting for any giving assessment model or test. It must meet the highest requirements in terms of security and technical reliability. The new assessment management system uses an integral data model, enabling seamless integration of different systems. It also allows for true, real-time item preview at any time of test production (without intermediate routines as is the case in most current systems) and, in the near future, direct linking of psychometric data to items and tests. Test production for specific programs or purposes is guided is by program-specific templates.

Application of these principles will lead to a greater satisfaction and commitment with all stakeholders. It will also require less training and experience to work with the system. The number of possible errors and mistakes can be greatly reduced and security is increased

because the number of copies of items and tests in circulation, is reduced to an absolute minimum.
In the new assessment management system, psychometric research can be integrated with data return within the system. This allows for customized processing and reporting of results. Computer adaptive testing and other forms of branched testing can be easily facilitated, but also criterion referenced, random and optimal item selection is supported.

The new system will form the basis for research into new context-rich, media rich, interactive item types for testing of skills in addition to knowledge. Until now, research and development of innovative item types was divided and fragmented. The new system allows for correlation of test designs, experimental data and test results. Because the new system will become the working environment for every employee involved in test development, test construction and test analysis, the system will also function as a shared knowledge base and facilitate knowledge management across the different departments and programs

Characteristics of ICE: a third generation assessment system

Focus on CBT: ICE is not developed as a medium independent authoring system and this is contrary to the strategic perspective that has dominated the development of test publishing software for years. Computer Based Testing is 'first citizen' in the new system. Paper versions will only be available as draft quality proof prints or listings. High quality paper based output will require specific PBT templates or (XML) exports to other Desktop Publishing (DTP) environments.

An integrated working environment for everyone involved: a shared workspace, which provides real-time preview (reliable, truthful presentation) and WYSIWYG handling of computer based testing content. ICE enables test developers to preview questions as it will be eventually presented to the candidate, include multimedia, interaction and scoring options, in real-time the course of item writing and test construction.

Modular: ICE is modular in nature. For example, the TestPlayer component for real-time previewing, can be seamlessly integrated in many environments on the Internet, on CD-ROM, but also within a SharePoint sever environment. Links between the various front-and back-end systems can easily be established. Special modules, such as the scoring module for responses to open ended questions for the national examinations in secondary education can be incorporated at any time. This way it is possible to assemble customized assessment authoring solutions for every purpose or testing program .

Generic where applicable, specific where needed: The item bank and TestPlayer, the interface for test delivery, are the most generic building blocks of the system. This core is expandable with a variety of plugins and add-ons, enabling customized front-end < > back-end relations (e.g. different psychometrics, logistics, distribution, planning or reporting services).

Extensible: Item types or item selection mechanisms (computer adaptive testing) are no longer regarded as properties of a system or application, but of the test itself. These differences are also regarded as forms of customization which, if required, can be added to the authoring and delivery environment of the system. They are considered to be plugins, rather than updates of applications. This way, the core system can remain lightweight and "dumb", whereas specific, complex logic can reside in the *content*. Analogous to a web browser: you do not need to download a new browser to visit a new website, plugins like Abobe Flash and Silverlight can be easily used to extend existing functionality.

The modular nature of ICE and the extensibility of the system, allows for custom test design and delivery without having to code, test and install new applications every time.

High stakes, large volumes: One of the biggest challenges in computer based testing arises when large numbers of students take the same online test within the same, restricted time frame. Smart client technology also facilitates offline test delivery and administration, thus guaranteeing continuity. Using client resources, results in a lower server load which makes it easy to handle large volumes.

Scalable: a redesign of the server makes it possible to mount servers at strategic locations (e.g. schools and test centres) which improves the scalability of the system. Smaller sites (e.g. home use, some computer labs) can communicate directly with the central servers using the internet.

Multi lingual: the system is prepared for multi lingual use. The GUI for test delivery uses as little written language as possible, most of the linguistic customization is embedded in the different test definitions. Although the system itself and its modules and plugins are developed in English and have English as their primary language, it is relatively easy to convert to any other language.

Online & Offline: ICE can deliver tests online (live via Internet) and offline (on workstations with a blocked internet connection during test delivery). The Internet can be used for continuous test delivery but also for updating test and locally hosted software. Secured local storage of multimedia content in advance of the actual

deliveries of the test, the separation of test content and multimedia content in place and time, guarantees a minimal network load.

Template oriented: test developers create items using predefined item-templates. These templates are in turn assembled from a library of smaller building blocks: from simple controls (e.g. buttons) to a text-to-speech module. By using templates, items with complex logic (e.g. a animated character showing up on the screen after ten seconds of inactivity) can be assembled with relative ease.

ICE is **not** browser based: this third generation assessment software can run in a web based environment and use the infrastructure of the internet. The components of the system can be launched *from* a web browser, but they are not a native browser based applications, because it does not run within a browser (e.g. Internet Explorer or Mozilla Firefox) .

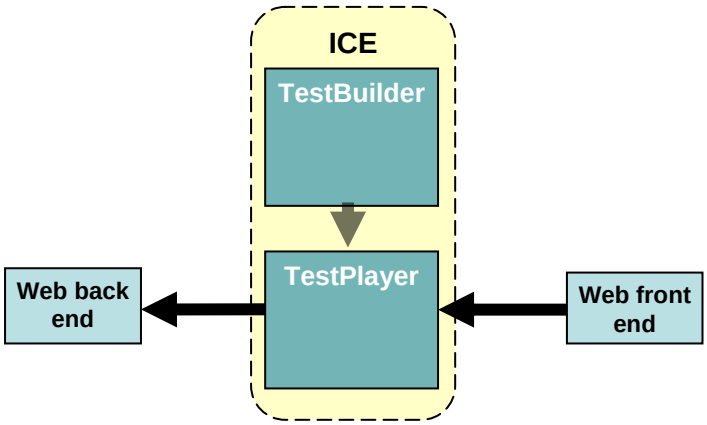
ICE is **not** platform independent. The new application will run on any current version of Microsoft Windows (98 SE and upward) and only requires the Microsoft.NET Framework 2.0 or up. Depending on assessment content, additional plugins may be required.

ICE is **not** a monolithic one-size-fits-all comprehensive system, it is not a 100 piece Swiss Army Knife. It is generic where possible, custom made where necessary.

Scenario's

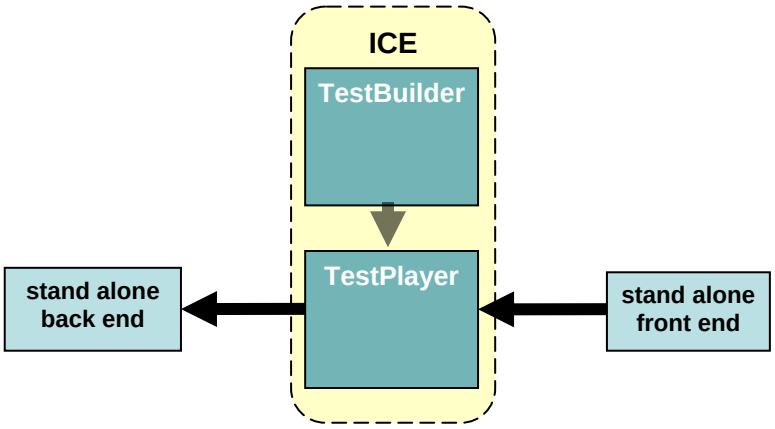
The models below, represent only three out of a multitude of possible scenarios.

Online testing scenario:



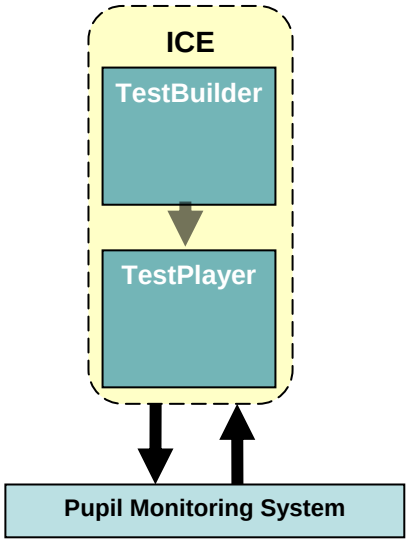
Tests are planned and launched the test is done online at the test location; after data processing on the central server at Cito, feedback and reporting are done online

Offline testing scenario



Tests are planned at the test location, using a separate planning module; tests can be pre-installed offline as well as online but tests are launched in an offline environment, on the local computer. Data processing, feedback and reporting are all handled by the local computer

The monitoring system scenario:



Monitoring Systems can handle all front and back end activities with ICE. Monitoring system software is located at a local machine or server, which connects to the main server if necessary. The online implementation of this scenario is the perfect example of the power of smart client technology, because the system uses online as well as offline resources in a very efficient way.

Rethinking and redesigning the way we use IT in test development, also led to major advances in theory and practice of computer based testing. Computers can handle scoring algorithms that are just too complicated or time consuming for a human rater . This has given rise to new research into existing and new algorithms for certain question types. Growing expertise in the field of multimedia and digital animation has made it possible to build interactive questions and tests where on-screen actions of test takers are recorded, translated into meaningful data and scored.

IT will remain one of the major driving forces for innovation in assessment and testing. It will affect the way we design and produce tests as well as the way we collect, analyse and report results. The fact that nine out of ten new assessment products are in part or entirely dependent on computers, implies a continuous and growing investment in products, knowledge and networks.

IT lies at the heart of the key competencies in test development and is therefore of strategic importance to any enterprise that takes the future of testing and assessment seriously.