Innovative international eAssessments at the completion of the IB's Middle Years Programme (11-16 years).

Rebecca Hamer, Sarah Manlove & Carolyn Adams[1]
International Baccalaureate

Abstract

Young adults and students of high school age in developed countries currently have unrestricted access to internet and the vast majority type more often than they write. However exam models have been slow to change: many models depend heavily on writing skills under timed conditions, with negative effects on student performance, especially when the students are asked to write long hand. The IB is developing an onscreen examination interface for its Middle Years Programme (MYP) that will change the way students are assessed with little backwash on teaching. This paper presents an initial user perception trial of the first generation of the online eAssessment. Thirty students aged 11 to 16 participated. Initial reactions of students and teachers are positive and constructive with specific feedback being utilized in the 2nd generation of the online eAssessment package. While the eAssessment will provide the sole pathway to the MYP certificate, the IB will be keeping multiple assessment models in place in the Diploma Programme to accommodate the still existing digital divide that prevents a complete transfer to eAssessment within the next decade.

Key words: Computer Assisted Assessment, Perceived Ease of Use, Interface Design

"They are the Nisei of cyberspace—the first generation born into a world that has never not *known digital life and so never had to adjust to it as the rest of us settlers have. Like all Nisei, they understand the new world in ways their parents never will and speak its language with far more fluency. If you want to understand the past two decades, they are perhaps the perfect subjects. The drumbeat of disruption and technological advance that has defined the past 20 years is their natural rhythm."(Jerry Adler in Wired Magazine)*[2]

Background

Walk through the halls of a high school or middle school, or for that matter look around at students on trains and buses and it's obvious that technology is a powerful tool for learning. One student chats on her cell phone while simultaneously sending an assignment from her laptop. Another student uses his laptop to post to a learning forum while another watches a lecture through a MOOC. The proliferation of media and information available to today's students is beyond human history's reckoning. The generation called the "Nisei of

[1] contacting author: carolyn.adams@ibo.org
[2] http://www.wired.com/magazine/2013/04/genwired/

cyberspace", the first generation born into a world that has never *not* known a digital life, are in their twenties and entering university and working life. Their little brothers and sisters are in junior high school or high school and following in their footsteps. The ones that have access, the lucky, supported ones, are leveraging every digital possibility available to them for their future success and some expectation on their part that education institutions are catching up and creating viable good designs for learning and assessment that match the modern age aren't, on the surface, unreasonable. The International Baccalaureate (IB) seeks to meet the demands of these modern digital students through development of an onscreen examination interface (OEI) for its Middle Years Programme (MYP). This paper seeks to describe the rationale of this project through first discussing the unique context of modern digital learners, and of the IB and the MYP as lending itself well to an OEI and finally presents the qualitative results of a small pilot trial of the system with students to get at their perceptions of the items and the interface usability

The IB Diploma Programme (DP) is an academically challenging and balanced programme of education with final examinations that prepares students, aged 16 to 19, for success at university and life beyond. It has been designed to address the intellectual, social, emotional and physical well-being of students. The continuum programmes of the Primary Years (PYP for ages 3 to 12) and Middle Years programmes (MYP ages 11 to 16), while not necessary for entrance to the IB DP programme, do represent the IB's dedication to develop inquiring, knowledgeable and caring young people who help create a more peaceful world through intercultural understanding. From 2010-2012 the MYP undertook a review of its curriculum in order to move towards innovation that enables students to be successful in further IB studies while also facilitating schools in combining the MYP with requirements of national/state systems.

In upcoming years, as part of the MYP: Next Chapter, IB will offer an optional e-Assessment that has disciplinary and interdisciplinary components. As with all pedagogically sound assessment designs the starting point was the curriculum. One of the hall marks of the MYP is its emphasis on the processes of inquiry, problem solving, critical thinking, analysis and the ability to use knowledge in unfamiliar situations. Long held as a challenge for assessment, these higher order learning goals have seen a proliferation of ICT for teaching and learning to develop them, including the explosion of online learning environments, simulations and virtual worlds, not to mention the simple word processor. And yet, assessment, as Ridgway, McCusker & Pead (2004) state in their literature review of e-assessment, has often lagged behind in the use of ICT. Although progress has been made to some extent what they wrote in 2004 still holds true; "We are approaching a bizarre situation where students use powerful and appropriate tools to support learning and solve problems in class, but are then denied access to these tools when their knowledge is assessed. (p.9)"

In many schools during exam sessions, room after room is filled with students writing answers on paper under timed conditions, raising the question if we are not increasingly disadvantaging these digitally literate students. For example when handwriting becomes an unfamiliar skill, handwriting style and speed deteriorate influencing legibility (Graham et al., 1998). Stress has been shown to reduce undergraduate students' writing fluency to that of an

eleven year old, with serious negative impact on the quality of responses and the marks awarded (J. Connelly et al., 2005), implying that the motor skill of fluent and legible handwriting interferes with the expression of the higher order thinking skills necessary for high quality essay writing (Peverly, 2006). Focusing on essay writing under timed conditions, typed responses seem to be slightly longer and score marginally better on readability measures (Hartley, 2013). Although these studies focus on undergraduates these finding are very relevant to assessment of younger students as well, as almost all of these younger students use the computer every day, or at least weekly for school (Cranmer, 2006). In the pilot presenting more fully below about 80 percent of the MYP students indicated that they do at least 60 percent of their class work and over 70 percent of their homework on a mobile device such as a laptop. Over 70 percent of the students preferred typing to handwriting and felt that the quality of their work was higher if they typed, commenting that "using a keyboard is much more in line with what [I] do every day."

The suitability of an OEI for measuring MYP's problem solving, inquiry, critical thinking and analyses skills necessary for our new digital learners is heralded then as a promise of computer assisted assessment (CAA) (Quellmalz et al., 2013, Quellmalz, Timms, Silberglitt, & Buckley, 2012). Seen as an opportunity, at its best CAA can go beyond paper-based assessment in its assessment of higher order skills in its allowance of "...more complex item types compared with paper-based assessment, including the use of audio-visual materials and more complex interactions between learner and computer (Conole & Warburton, 2005 p. 21)." At its most challenging however CAA is confounded by interface and design issues (Ricketts & Wilks, 2001), institutional infrastructure problems (Warburton, 2013) as well as parental, student, teacher, and when warranted (as in the case of summative high stakes testing-) external examiner perceptions (Deutsch, Herrmann, Frese, & Sandholzer, 2012; Terzis, Moridis, & Economides, 2013), as well as copyright issues if proprietary media is used in item development. Faced with these challenges, the adoption of CAA has progressed slowly and through an iterative formative development and feedback cycle across many themes and issues such as user perception, strategic development, results handling, quality assurance, item design and site implementation (Ashton, Schofield, & Woodger, 2003; Ras, Maquil, Foulonneau, & Latour, 2013; Warburton, 2013; Zakrzewski & Steven, 2000). The road to adoption of an OEI within IB's MYP has run and will continue to follow the same model of iteratively cycling development, testing and piloting. This cycle is necessary in order to take careful consideration of IB's curricula, its stakeholders and its unique context

An important aspect of that context is the international nature of The IB. The MYP programme currently operates in 1,011 schools in 91 countries, 70 percent in developed nations and 30 percent in less developed regions (International Baccalaureate, 2013). Take for example access to the internet as a measure of technological infrastructure. In North America, Europe and parts of the Asia-Pacific region, household access to internet increased from less than ten percent to coverage levels of 70 to 90 percent in little more than a decade, see Figure 1 (Internet World Stats, 2013). Internet access of households with dependent children in the European Union is currently approaching full coverage, with only two of the

new member states providing less than 80 percent coverage to children of high school age (Eurostat, 2013).

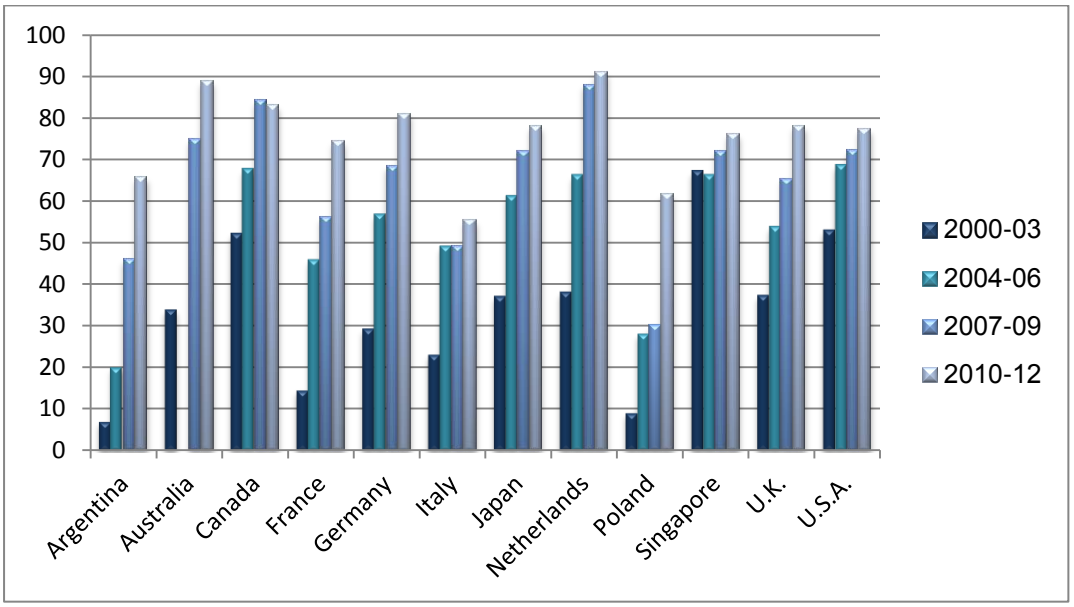

Figure 1: Household access to internet in developed countries (2000-2012)

Currently countries in other regions, e.g. South America, Asia and Africa are attempting to catch-up and provide the same chances to their young people, see Figure 2 (Internet World Stats, 2013). However, a comparison of these two graphs shows clearly that a digital divide still exists and any adoption of CAA within the IB context will take ease of access to computers, internet, and the availability to upload results for external examination seriously. For this reason the MYP eAssessment will be optional in its roll out.

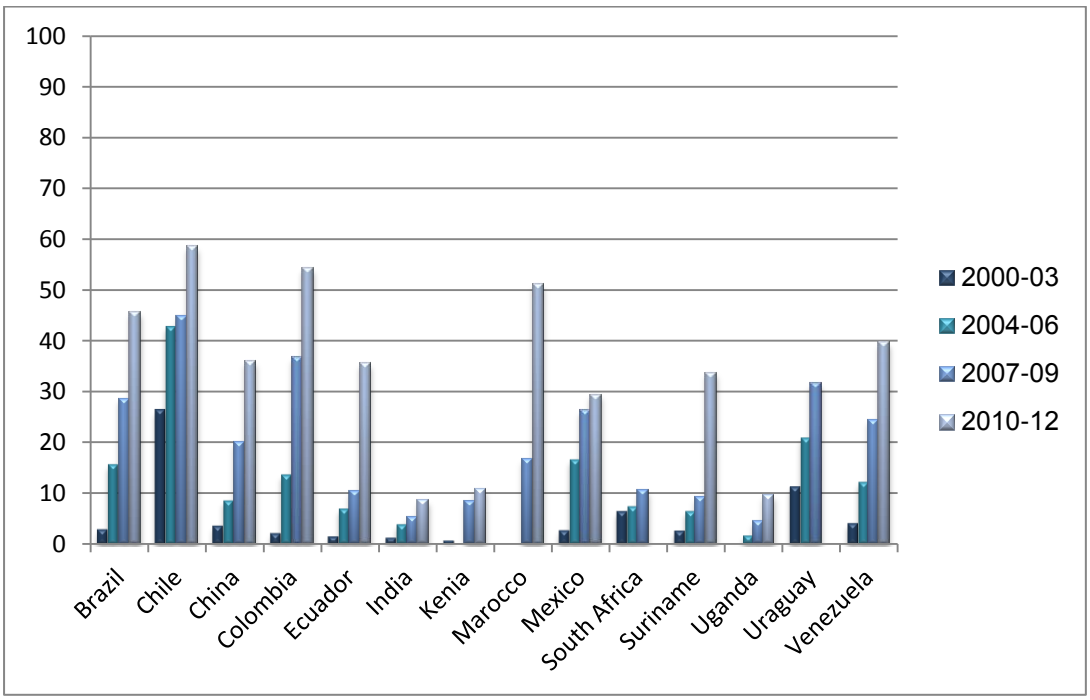

Figure 2: Access to internet in various countries (2003-2011)

Beyond technology accessibility issues the IB faces, there are further infrastructure issues being addressed. Existing IB infrastructure for e-marking in the DP programme relies on two technology platforms for marking and results reporting. The OEI will naturally integrate with those systems. Secondly the IB relies on external examiners all over the world to mark student work. The MYP currently uses a cadre of external examiners to moderate student work if schools choose this option. MYP's OEI will expand this base of external examiners to accommodate the expected growth of schools opting for the MYP: Next Chapter certificate, and eventually participate in future item design during paper setting for subsequent exam sessions. The IB's commitment to reliable and valid assessment within the OEI platform means that standards of economy, usability and familiarity are being established so it can be implemented as seamlessly as possible for its stakeholders. Towards this end, a small scale qualitative pilot was conducted with MYP students in a current IB school.

Research consistently documents that user-interface issues can interfere with the performance of students during CAA (Ashton et al., 2003; Carbó, Mor, & Minguillón, 2005; Dermo, 2009). For this reason the purposes of this pilot was to benchmark initial ideas for the design of the items and understand potential user-interface issues from students, the end-users of the platform before obtaining performance data. The results of this pilot are described below.

MYP onscreen assessment: a small scale pilot

In April 2013, the IB conducted a small scale pilot for MYP students at an IB school in the UK. In this pilot, 30 students (15 female and 15 male) took part in two mock examinations of 120 minutes each, one in the morning and one in the afternoon. Students could choose between Biology, English, History, Mathematics and the MYP Interdisciplinary task. Exams were offered onscreen within the network environment of the school, not online. Each exam was followed by a thirty minute focus group session by exam subject to gauge the students' reactions to the interface and gather their comments on the exams. The focus group questions consisted of nine questions that focused on the interface ease of use, and item levels. In addition 23 students filled in a short survey in the week following the pilot. This survey focused on their perceptions of the various subjects as well their experiences of technology use for learning activities.

As presented above, many of these students are digitally literate, using their mobile devices for school work every week. None of the students thought that doing exams on a mobile device or laptop would disadvantage them; on the contrary, a clear majority thought it would benefit them. The students thought the onscreen exam interface "looked really cool" and that the interface was easy to navigate, although exams varied in the amount of navigation required. Students saw the attraction of e-Assessment for including a greater variety of sources, e.g. excerpts from films and short video clips. But they felt that the identical text box size failed to provide a familiar cue on how much text was expected. Being used to handwriting long answer or extended responses, they had difficulty estimating when they had reached the indicated word count. It seems many students apparently overlooked the indicator included in the interface, but some did recognise it, "I liked how it told you how much to

write for each question and I liked how it gave you your word count." All thought that technical problems aside, the exams were "about the right length" but that 90 minutes would probably sufficient for the majority of MYP students.

With the exception of biology, students found the exams interesting, with questions that made them think (see figure 4: English and Interdisciplinary in particular). With reference to biology students found the items to be too easy and felt they did not address the higher level MYP learning objectives they were accustomed to and they also found the items to be less interactive despite the inclusion of simulations.

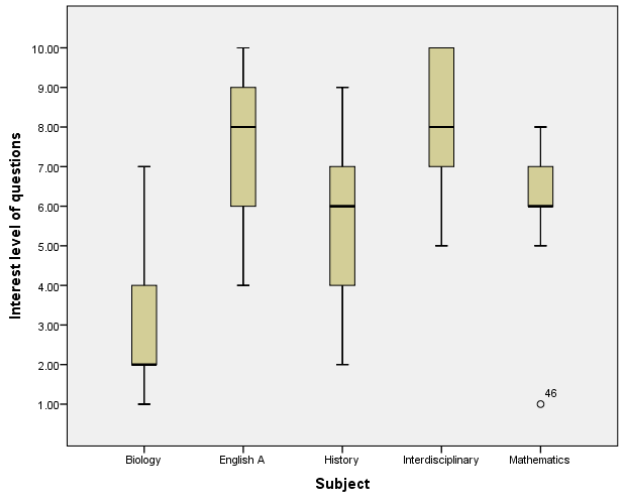

Figure 4: I thought the questions were interesting and made me think

Comparing the five exams offered, students thought that some subjects appeared to lend themselves better to e-assessment than others, with mathematics being the least popular to do onscreen, e.g. "maths ... on a computer ... is not representative of real life - quick calculations you may do in real life would probably be done by grabbing a pen and paper and jotting it down." In the mathematics exam, which offered all sources and diagrams onscreen, students missed the opportunity to think through problems using scrap paper and drawing on the sources provided, e.g. "instead of placing the graphs and pictures online you should give them to us on a hard copy so that we can scribble notes on them." They also had a number of suggestions for improvements that would help them monitor their progress through the exam. These comments were noted and are being implemented in the interface upgrade.

Conclusions

Currently students of high school age with access to digital devices would seem to prefer examination situations that match their every day learning experience. As stated above this pilot found, unsurprisingly, that participating students did a majority of their school work on a digital device and felt the quality of their work was higher if they typed. The IB's new Middle Years Programme (MYP) OEI will meet that preference. This bodes well for high user buy-in to the OEI down as it develops. As Terzis et al (2013) and Deutsch et al (2012) state high user buy in for e-assessment is an important aspect of student performance in such environments. Confounding their buy in however were interface issues that arose during the

trial. Appropriate use of testing cues such as word limits and progression indicators as well as a preference for additional navigation options were of issue. In addition students found it difficult that they didn't have a notes area or a place to show their work, particularly in maths. A closer examination of the biology items revealed that they were at too low a level for the students. Although the items did incorporate interactivity with science simulations this interactivity may have been confounded by student perceptions of the item levels.

In the next generation of the OEI under development at the time of this paper the interface has incorporated this feedback. There is now an indication of progression, multiple navigation options as well as a place for students to work out maths problems onscreen. Feedback on the item designs from this pilot was also presented to the curriculum designers, external examiners, and subject managers designing the items for the platform to better calibrate the items to MYP student levels and interactivity aspects. This 2nd generation proprietary OEI, in keeping with the design and development cycle will undergo several iterative trials building from small tasks to full-fledged exams during the 2013-2014 year. A formal pilot is expected sometime in 2015.

Future steps

As the MYP OEI develops it will continue to seek to fulfil the promise of CAA through the use of media-rich onscreen assessment items that get at higher order skills such as problem solving, critical thinking and inquiry by providing transfer problems derived from the MYP curriculum. Currently teams of IB curriculum managers, subject managers, MYP staff and examiners are refining and working on items that leverage the potential of ICT for assessment. Future trials will continue to examine student user interface issues as well as performance, item reliability and validity issues to ensure that IB standards for external assessment are maintained. Future trials will also incorporate an examination of a system that will ensure that accessibility for students with assessment access requirements is met. Future work will also be devoted to linking the OEI to the existing IB infrastructure for results reporting and e-marking by external examiners. In keeping with the IB model of involving stake holders in a collaborative approach teachers and students will be utilized to inform the process so that there is a minimum negative backwash on teaching by its implementation. The new MYP promises a forward-looking assessment model that is rigorous, new, and fit for future learners. In the longer term MYP eAssessment will enable a more reliable, accessible and valid large scale external assessment for this age group.

Reference list

Adler, Jerry (2013, April). 1993. *Wired Magazine* . n. pag Retrieved 5 September 2013: http://www.wired.com/magazine/2013/04/genwired/

Ashton, H. S., Schofield, D., & Woodger, S. C. (2003). Piloting Summative Web Assessment in Secondary Education. In *Proceedings for 7th CAA Conference*. Loughborough, UK. Retrieved from http://caaconference.co.uk/pastConferences/2003/procedings/index.html

International Baccalaureate (2013). IB Facts & Figures. *IB World*, *September*(68), 20–21.

Carbó, J. M., Mor, E., & Minguillón, J. (2005). User navigational behavior in e-learning virtual environments. In *IEEE/WIC/ACM International Conference on Web Intelligence*. Compiégne, France.

Conole, G., & Warburton, B. (2005). A review of computer-assisted assessment. *Alt-J Research in Learning Technology, 13*(1), 17–31. doi:10.1080/0968776042000339772

Connelly , J., E. Dockrell & J. Barnett (2005). The slow handwriting of undergraduate students constrains overall performance in exam essays, *Educational Psychology: An International Journal of Experimental Educational Psychology, 25:1*, 99-107

Connelly, V., S. Campbell, M. MacLean & J. Barnes (2006). Contribution of lower-order letter and word fluency skills to written composition of college students with or without dyslexia. *Developmental Neuropsychology, 29:1*, 175-196. DOI:10.1207/s15326942dn2901_9

Cranmer, S. (2006) Children and young people's uses of the Internet for homework, *Learning, Media and Technology, 31:3*, 301-315, DOI:10.1080/17439880600893358

Dermo, J. (2009). e-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology, 40*(2), 203–214. doi:10.1111/j.1467-8535.2008.00915.x

Deutsch, T., Herrmann, K., Frese, T., & Sandholzer, H. (2012). Implementing computer-based assessment – A web-based mock examination changes attitudes. *Computers & Education, 58*(4), 1068–1075. doi:10.1016/j.compedu.2011.11.013

Eurostat (2013). http://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do, statistics extracted 6 March 2013.

Graham S., N. Weintraub & V.W. Berninger (1998). The Relationship Between Handwriting Style and Speed and Legibility, *The Journal of Educational Research, 91:5*, 290-297, DOI: 10.1080/00220679809597556

Internet World Stats (2013). http://www.internetworldstats.com/stats.htm, statistics extracted 10 March 2013

Mogey, N & J. Hartley (2013): To write or to type? The effects of handwriting and word-processing on the written style of examination essays, *Innovations in Education and Teaching International, 50:1*, 85-93

Peverly, S.T. (2006). The importance of handwriting speed in adult writing. *Developmental Neuropsychology, 29:1*.

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching, 49*(3), 363–393.

Quellmalz, E. S., Davenport, J. L., Timms, M. J., DeBoer, G. E., Jordan, K. a., Huang, C.-W., & Buckley, B. C. (2013). Next-Generation Environments for Assessing and Promoting Complex Science Learning. *Journal of Educational Psychology*.

Ras, E., Maquil, V., Foulonneau, M., & Latour, T. (2013). Empirical studies on a tangible user interface for technology-based assessment : Insights and emerging challenges. *International Journal of e-Assessment*, *3*(1), 1–19.

Ricketts, C., & Wilks, S. (2001). IS COMPUTER BASED ASSESSMENT GOOD FOR STUDENTS ? In *Proceedings of the Fifth International Computer Assisted Assessment Conference*. Loughborough, UK: Loughborough University. [Retrieved from http://caaconference.co.uk/pastConferences/2003/procedings/index.html]

Terzis, V., Moridis, C. N., & Economides, A. a. (2013). Continuance acceptance of computer based assessment through the integration of user's expectations and perceptions. *Computers & Education*, *62*, 50–61.

Warburton, B. (2013). CAA – Whither and Whence ? The last decade and the next decade. In *Proceedings for 17th CAA Conference* (pp. 1–13). Loughborough, UK: Loughborough University.

Zakrzewski, S., & Steven, C. (2000). A Model for Computer-based Assessment: The catherine wheel principle. *Assessment & Evaluation in Higher Education*, *25*(2), 201–215.