The end of Maslow's hammer – or how to fit the assessment to the skill

- Stephen Miller, International Baccalaureate, stephen.miller@ibo.org
- Steve Harrington, RM Results, sharrington@rm.com

"I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail", Abraham Maslow.

This paper investigates how the International Baccalaureate (IB) develops and manages assessments, fit for the 21st Century. The IB was established to provide a challenging and comprehensive education that enables students to understand and manage the complexities of our world and provide them with skills and attitudes for taking responsible action for the future. Such an education is rooted in the belief that people who are equipped to make a more just and peaceful world need an education that crosses disciplinary, cultural, national and geographical boundaries.

We look at various types of assessment used by the IB, such as long/short responses, extended essays, onscreen assessments, collaborative projects and recorded evidence, and how these have evolved to keep pace with 21st Century skills. We demonstrate how technology has been used to facilitate marking a range of assessment types, improving marking quality and consistency, and how RM's e-marking technology has supported the IB's assessment innovation, supporting their international growth and efficiency.

We conclude with thoughts on how innovative assessment design will meet the needs of students worldwide.

Keywords: '21st century skills' 'e-marking' 'e-assessment' 'on-screen marking' 'International Baccalaureate' 'high stakes' 'marking' 'RM Results'

Background

The International Baccalaureate (IB) is a Swiss Based Education charity with registered offices in Geneva. The International Baccalaureate was formed in 1968 and issued its first IB Diplomas in 1971.

The IB is driven by its mission statement:

The International Baccalaureate aims to develop inquiring, knowledgeable and caring young people who help to create a better and more peaceful world through intercultural understanding and respect.

To this end the organization works with schools, governments and international organizations to develop challenging programmes of international education and rigorous assessment.

These programmes encourage students across the world to become active, compassionate and lifelong learners who understand that other people, with their differences, can also be right.

The IB offers four programmes of study, the Primary Years Programme (PYP) for ages 3-12 the Middle Years Programme (MYP) for ages 11-16, the Diploma Programme (DP) for ages 16-19 and the Career-related Certificate (IBCC) for ages 16-19.

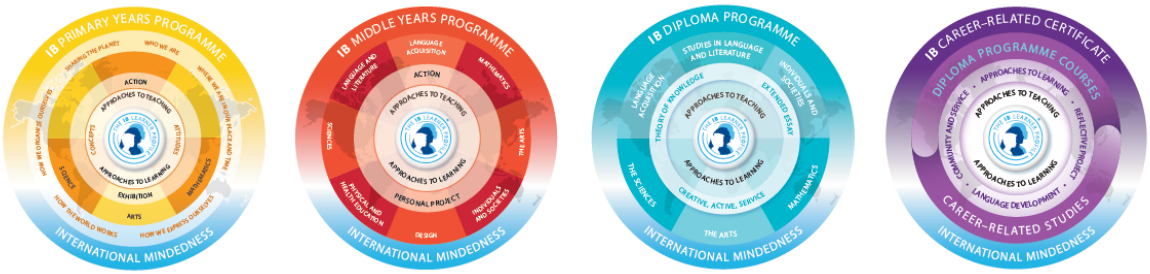

Figure 1: The IB's Primary Years Programme, Middle Years Programme, Diploma Programme and Career-related Certificate.

The organisation has global offices in Singapore (specialising in IT and finance), Bethesda, Maryland USA, (specialising in school authorisation and the IB educator network), the Hague (specialising in curriculum development and trialling for IB programmes) and the Cardiff Assessment Centre, where examination papers are produced, school and examiner moderation is conducted and examiner recruitment, training, eMarking, standardisation and grade awarding and the issuing of result is organised for the May and November Diploma Programme examination sessions.

Assessment models differ in each programme. The PYP has school based internal assessments; the current MYP has school based and optional externally moderated coursework and projects; the Diploma Programme has Internally Assessed (IA) / moderated and externally assessed components that are combined to give subjects and diploma grades; and the IBCC has a blend of Diploma subject assessments and school based vocational skills assessments.

With over 45 years of stability in programme design and the learner profile at the core of an IB education, the IB has always focussed on communication, international understanding, breadth of study, critical thinking and, through constructivist approaches, responsibility for learning.

21st century skills

Assessment of Teaching of 21st Century Skills (ATC21S) has, through research, categorised 21st century skills under the following headings (ATC21S, 2014):

- **Ways of thinking.** Creativity, critical thinking, problem-solving, decision-making and learning
- **Ways of working.** Communication and collaboration
- **Tools for working**. Information and communications technology (ICT) and information literacy
- **Skills for living in the world.** Citizenship, life and career, and personal and social responsibility

All IB programmes place the learner at the centre, fostering responsibility for learning and providing the meta-language to describe t approaches to learning. The IB has had a defined learner profile since 1997. Figure 2 shows how the IB Learner profile concurs with common statements of 21st century skills (highlighted in yellow) and demonstrates that they are not unique to the 21st century.

Figure 2: The IB learner profile showing the links to the 21st Century skills in yellow.

The IB Learner Profile translates the IB mission into skills that are developed through an IB education and evaluated through IB Assessments. The Learner Profile represents attributes valued by IB World Schools and by the IB and cumulatively represent international mindedness. The Learner Profile has evolved since it was first stated in 1997, undergoing a number of small changes, the most recent to shift focus from learners referred to as "they" to a more inclusive "we". The development of these skills and attributes are achieved through the expertise of IB school teachers and educators supported by the IB educator network, IB curriculum content and curriculum frameworks and assessed in the DP and IBCC through IB developed summative assessments.

The IB is uniquely positioned, being independent of national government influences, to ensure consistency of approach, content and assessment over many years. A review of recent Diploma results shows that assessment standards are highly conserved over time with pass rate variation of less than 1% (IB, 2013):

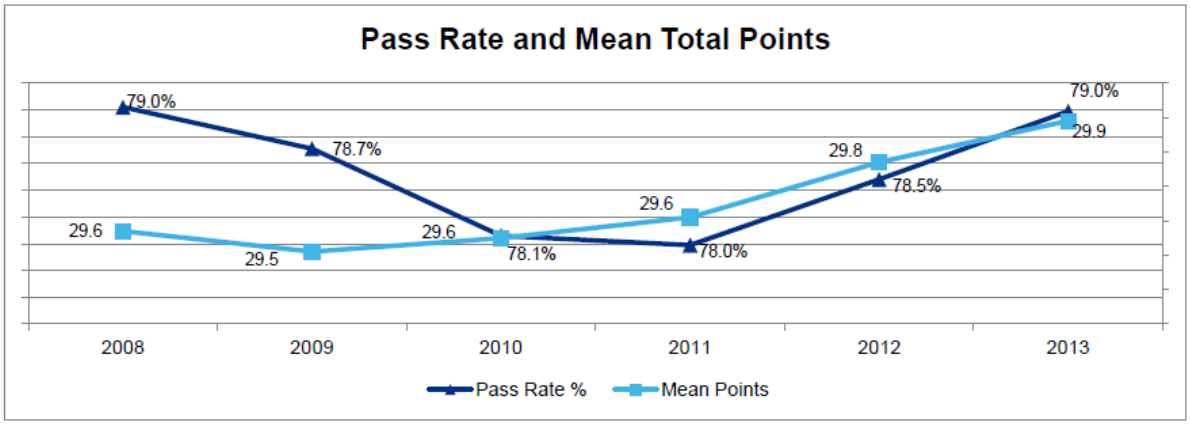

Figure 3: This graph shows very small variations of pass rate (<=1%) and mean points scores (<0.3 points) over a six year period.

Challenges

The IB continues to face a number of challenges in how it assesses programmes through external examinations and through moderation of coursework. In the Diploma Programme, all courses are taught in one of three core languages, English, Spanish and French and all assessments are also delivered in these three languages, with mother tongue entitlement leading to a further 80 languages being assessed in Studies in Language and Literature (Group 1). Diploma programme assessments comprise two or three sat papers per subject, depending on whether it is Standard or Higher Level a number of internally assessed and externally moderated coursework elements, externally assessed Theory of Knowledge, written tasks, written assignments and Extended Essays plus performance based material in dance, film, music, theatre and visual arts and evidence of participation in Community Action and Service. There are few multiple choice style assessments compared to the volume of assessment material and these can only be found in The Sciences (Group 4). External Assessment includes short and long answer exam questions, extended written work, visual and audio records of performances and exhibitions and research and presentations. Internal Assessment (moderated by the IB) includes Language Orals, Science practical work, Maths portfolios and short and extended writing.

All assessment in the DP draws on the founding principles that they are "valid....really assess the whole endowment and personality of the pupil ...[and] at the same time [are] sufficiently reliable to assure pupils ...that justice has been done" (Peterson, 2003). Maintaining reliability and validity over this breadth of assessments across three languages with world regions separated by time-zoned papers in many instances is both a challenge of design and of logistics. With over 40 years of organising and administering examination sessions, however, the 150 staff at the Cardiff Assessment centre produce, deliver and organise the marking of Diploma Programme internal and external assessments twice a year, not necessarily with ease, but certainly with confidence and enthusiasm.

In a typical May Assessment session, the IB assesses over 1600 discreet components consisting of over 2 million separate submissions. The challenge of this scale coupled with the logistics of providing assessment to Diploma schools in 147 countries and managing 10,000 examiners spread across 147 countries is enormous. In addition, the number of registered diploma candidate has doubled in number from 2007 to 2014.

Meeting the IB challenge starts with assessment design. This is typically a collaborative activity spread amongst question paper authors managed by the Examination Paper Production team. Examination paper production is managed through a workflow in Microsoft Sharepoint where roles and handovers are controlled and managed in collaboration with Assessment Subject Managers. Initial design is in English and when the paper is finalised the translations are managed by the IB's dedicated translation department.

Questions in Diploma Papers are designed to reflect the learner profile and take the student beyond recall and into analyses, critical thinking, problem solving and effective communications. These examples from recent papers convey some of the nature of IB DP assessments:

- Theory of Knowledge: "*It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts*" (Arthur Conan Doyle). Consider the extent to which this statement is true in two or more areas of knowledge.
- Language and Literature: How was your understanding of cultural or contextual considerations of the work developed in the interactive oral?
- Biology: Discuss the ethical arguments concerning the eating of animal products.

- Philosophy rubric: In your response you are expected
    - to develop a philosophical response in an organized way
    - draw upon, and show a **holistic appreciation of, the skills, material and ideas** developed throughout the course
- Mathematical Exploration: The specific purposes of the exploration are to
    - Develop students' personal insight into the nature of mathematics and to develop their ability to ask their own questions about mathematics
    - Provide students with the opportunity to experience for themselves the beauty, power and usefulness of mathematics

Using eMarking to meet the challenges

To meet the challenge of marking both internally and externally assessed materials reliably across languages, time zones and increasing candidature, the IB has transitioned from paper-based marking to eMarking.

To cope with the increase in candidates taking diploma exams each year the IB has partnered with RM Education to introduce onscreen eMarking of 100% of sat papers over the last four years. At the start of a session, candidate details are fed into RM Assessor and scripts (making a single journey) are sent by schools to a scanning centre in either the UK or in the US. Scanned images are matched to candidate data and anonymised whole scripts are presented to examiners to mark by whole paper.

RM Assessor is a multi-platform web based marking system with an interface translated with IB defined terminology in English, Spanish and French. At the start of a marking session, the standardisation team definitively marks a number of qualification and seed scripts that demand a range of marks across the mark structure. At the same time the standardisation team will edit and finalise the mark scheme that examiners will use to aid their marking. When marking starts, examiners first qualify to mark by marking one or two standardisation sets and once qualified, they maintain their marking quality by blind marking definitively marked seeds presented during their eMarking.

The use of seeded scripts, particularly for longer pieces of writing is not seen as conforming to a mark but more as achieving a standard. Examiners who mark outside tolerance are coached by team leaders and given feedback via secure communication tools within the system. Examiners who do not qualify to mark or who do not mark seeds within IB determined tolerances are stopped from marking scripts. Unlike IB's paper based sampling model, raw scores become final scores.

The introduction of eMarking has brought a large number of benefits to the IB and ultimately to candidates and to schools. These benefits permeate all stages of the marking process.

Schools send scripts directly to one of the two scanning centres, meaning scripts now take fewer journeys around the world. US public schools, they have reported savings on courier fees. And the IB has reduced its carbon footprint.

The standardisation team members, who used to travel from around the world to meet in person now work virtually, viewing and provisionally marking the same scripts in RM Assessor and then engaging online to select and mark examples of scripts for standardisation and seeding. The principal examiner is able to add more scripts to the seeding pool as the session progresses. During standardisation setup, examiners can practise with RM Assessor in a familiarisation mode or view live scripts and practise marking them in simulation mode. Once the mark scheme is released and standardisation set up is complete, examiners can begin to qualify and start marking.

With examiners working all around the world in different time zones, RM have developed the capability for a pool of team leaders to be set up in RM Assessor who can support examiners when they need assistance during qualification or after marking seeds outside tolerance. This pool means that examiners in a different time zone to their team leader's time zone do not need to wait for feedback.

The introduction of eMarking has also allowed the IB to move away from examiners having an allocation of scripts to mark and instead they examiners have a target for their marking, drawing down scripts from a script store at RM. Adding additional scripts to a target or removing an examiner and their open scripts is a simple activity and more efficient than packing and retrieving scripts in the conventional process. Anonymising scripts removes any unintentional bias in marking and breaks the conventional school to examiner relationship. RM Assessor calculates all paper totals from the marks entered and examiners cannot award too many marks. The need for clerical checking disappears with eMarking as does the need for the physical storage of scripts.

By stopping examiners who cannot mark within tolerance, the IB has significantly reduced the number of scripts that need to be remarked at the end of a session immediately before results issue.

Scripts can be selected at the click of a button for grade awarding instead of the time consuming physical retrieval process. The system also retains archived boundary scripts for use at grade award meetings and presents current session boundary scripts against selected mark points. What used to be retained in a warehouse is now archived on a portable hard drive.

To ensure the IB can assess the range of materials required to demonstrate the learner profile attributes, the IB has also developed systems for candidates and schools to upload essays, oral examinations and visual arts materials. Theory of Knowledge essays have been uploaded since 2009 and transferred into RM Assessor for eMarking since 2010. IB schools upload Group 1 and Group 2: Language Acquisition audio materials and Visual Arts portfolios for marking. In 2015, IB plans to use RM Assessor to mark these materials.

Uploading materials has meant that the assessment material from schools is no longer blocked by customs or damaged in transit.

Applying the eMarking quality model to all assessment types

Development for uploading rich media files and portfolios will continue into 2015 to allow the upload of all externally assessed materials and internally assessed samples. These materials will be pushed into RM Assessor for eMarking allowing the IB to use a single quality assurance model across all assessment in the Diploma Programme.

Other developments around eMarking are to introduce question item based marking. This enables examiners to mark single questions in large volumes whilst focusing on a small part of the mark scheme. This focus reduces the cognitive load, facilitates marking to tighter tolerances and improves the overall marking reliability.

Marking by Question Item Group (QIG) supports specialisms in marking and enables greater Examiner retention. For example, examiners can mark particular genres in literature and examiners who would not qualify to mark a whole paper will qualify to mark at least one question. Tracking item level marking will also allow enhanced reporting of candidate performance against objectives and against schools and worldwide averages.

Further plans include the structuring of unstructured papers to allow generic answer booklets to be marked by QIG.

The impact of technology availability on assessment design

In 2013, there were 511m sales of tablets and Personal Computers (Gartner, 2014a and 2014b). There are an increasing number of low and lower cost devices: Google Chrome based laptops, Datawind's 7 inch Aakash 2 Android tablet - £26 per unit (BBC, 2012) to name two. Sales of smart phones hit over 0.9 billion in 2013 (Gartner, 2014c). This means that globally students' access to technology through PCs, tablets and smart phones will increase. The proportion of internet users (International Telecommunication Union, 2013) within the developing and developed world are increasing but access is significantly lower in the developing countries (31 users per 100 compared to 77 users per 100). This means that a digital divide may exist between students in developed and developing counties.

Massive Open Online Courses (MOOCs) will encourage the use of online course materials and associated assessments. This development may act to increase internet use across developed and developing countries.

New communication technologies may enable developing countries to leap frog the technologies used in developed countries. For example, developing countries could use 4G mobile, satellite, microwave or laser links instead of installing physical cable.

With the difference in internet access between developing and developed countries likely to persist, a mixed delivery model is likely to exist for some years. Internet delivery should be achievable to devices in developed countries and for developing countries internet distribution to schools with onward distribution to devices via school networks should be achievable.

Near Field Communications (NFC) and Bluetooth communications provide low power wireless communication between devices. For example, NFC could be used to set-up wireless settings on tablet devices when the student enters the examination centre.

So, we should ask ourselves how this technological environment will influence assessment design

Firstly, increasing numbers of students will use technology devices and the internet regularly. Students in the developed world will increasingly use technology in their education. To fully understand a student's ability more assessments may need to make use of technology.

Secondly, writing high stakes assessments is a lengthy process. Curriculum frameworks and assessment designs need to be developed, quality assured and signed off. This process can take several years and hence it is mostly today's technology that will impact assessments in five years' time.

In the 21st Century, solving problems using a range of sources is common place. The IB are developing on-screen tasks for their Middle Years Programme focussed on assessing these attributes using an on-screen environment. The tasks will include text, audio, video and interactive stimulus materials. In addition to analysing information from range of sources and reaching a conclusion, the students will create a reflective journal of their learning at the end of their final assessment.

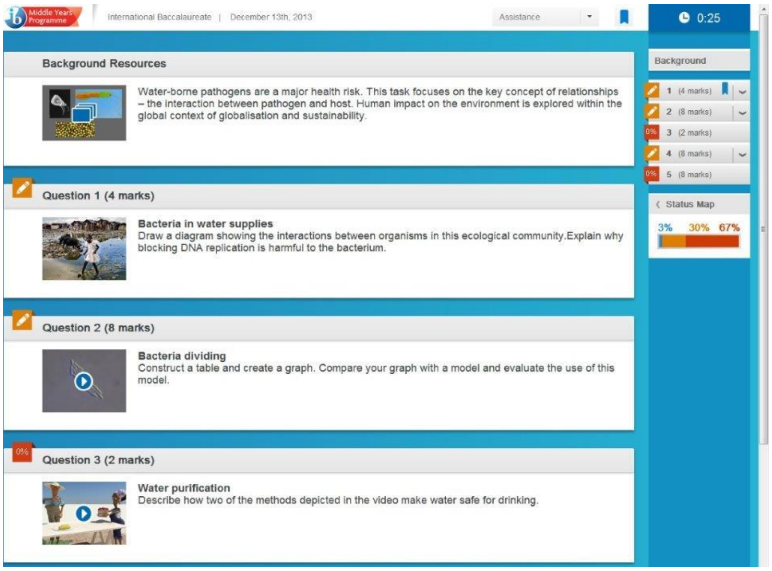

Figure 2: A sample Middle Year Programme Task illustrating the range of sources and interactive nature of the tasks.

Currently, the MYP has no summative assessment and so introducing eAssessment leapfrogs the paper assessment model. The IB will pilot the MYP onscreen assessments available in schools in 2015 and offer the full range of disciplines from 2016.

The IB expects to retain portfolios of work and group project work as a constituent part of its assessments as described above.

Technology to support Awarding Bodies

When candidate responses are entered on-screen, a range of technology assisted marking approaches become cost effective.

- Responses for the questions can be collated and unique student responses can be presented to be marked by one or more examiners. For example, Figure 3 shows 1306 candidates provided the answer Magnesium. The Examiner has marked this as correct and this mark is applied to all of the 1306 scripts containing this response.

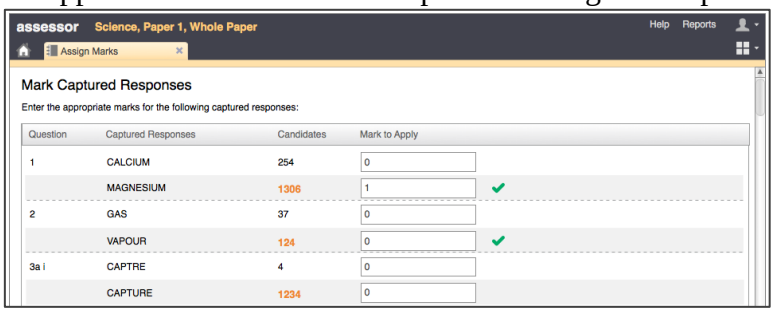

Figure 3: Mock-up showing how collated short response questions could be marked.

- Short response scoring algorithms can be created by the examining team and applied to all student responses.
- Short response and essay scoring technologies have existed for several years but only become applicable when candidate responses are received electronically. One model is to use essay scoring technology in conjunction with human marking and where there is a significant difference, the student response is marked by a different human marker. This method will result in improved marking reliability.

- With candidates submitting electronic responses, plagiarism detection software can be used.
- Automatic marking of student actions as well as evaluating the student's actual response are possible. The PISA 2015 tasks will employ this technique (PISA, 2013).

Conclusion

The IB will continue to reflect 21st Century skills through its learner profile and extend its suite of assessments to include authentic tasks in on-screen assessments. By doing so, the IB will meet the valid expectations of students who predominantly use computers instead of pen and paper. The use of RM Assessor will be extended to enable portfolios of work, multi-media files and e-assessment output to be eMarked.

References

Peterson, A.D.C., 2003: School Across Frontiers – The Story of the International Baccalaureate and United World Colleges.

IB, 2013: The IB Diploma Statistical Bulletin: May 2013 Examination Session: Available online at https://www.ibo.org/facts/statbulletin/dpstats/ (Accessed 17th March 2014)

ATC21S, 2014: Available at http://atc21s.org/index.php/about/what-are-21st-century-skills/ (Accessed 18th March 2014).

IB, 2014: IB Learner Profile, Available at http://www.ibo.org/programmes/profile/ (Accessed 17th March 2014).

Gartner, 2014a: Gartner Says Worldwide Tablet Sales Grew 68 Percent in 2013, With Android Capturing 62 Percent of the Market, Available online at http://www.gartner.com/newsroom/id/2674215 (Accessed 10th March 2014).

Gartner, 2014b: Gartner Says Worldwide PC Shipments Declined 6.9 Percent in Fourth Quarter of 2013, Available online at http://www.gartner.com/newsroom/id/2647517 (Accessed 10th March 2014).

BBC, 2012: India backs Aakash 2 tablet-based national education project, Available online at http://www.bbc.co.uk/news/technology-20297872 (Accessed 10th March 2014).

BBC, 2013: Baked in Britain, the millionth Raspberry Pi, Available online at

http://www.bbc.co.uk/news/technology-24435809 (Accessed 10th March 2014).

Gartner, 2014c: Gartner Says Annual Smartphone Sales Surpassed Sales of Feature Phones for the First Time in 2013, Available online at http://www.gartner.com/newsroom/id/2665715 (Accessed 10th March 2014).

International Telecommunication Union, 2013: ITU_Key_2005-2013_ICT_data.xlsx. Available online at http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx (Accessed 10th March 2014).

PISA, 2013: Draft Collaborative Problem Solving Framework, Available at http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf (Accessed 10th March 2014).

BBC, 2014: Student visa system fraud exposed in BBC investigation, Available at http://www.bbc.co.uk/news/uk-26024375, (Accessed 10th March 2014).