Transitioning from paper based to computer based assessments at a national level: Challenges and opportunities

19th AEA-Europe Annual Conference, Arnhem, 7-10 November 2018

Louise Bailey, National Foundation for Educational Research, UK

l.bailey@nfer.ac.uk

Abstract

Computer-Based Assessment (CBA) is increasingly used in connection with large-scale assessments at the national or system level. These include; international assessment studies such as the PISA test (onscreen since 2015), the PIRLS test (ePIRLS from 2016) and TIMSS, (eTIMSS from 2019). Additionally, a range of national and statutory tests are switching to onscreen presentation both in Europe and beyond.

Although there are benefits to moving to CBA, such as adaptive and therefore personalised testing, automated marking and the efficient generation of a range of reports on student performance, there are also challenges, particularly in relation to large-scale assessments.

At NFER, a small-scale research project involving a literature review and a small number of case studies, of a range of countries which have made the switch to onscreen testing at a national level, were undertaken. Key government personnel with responsibility for implementing CBA were interviewed regarding the motivations and drivers for moving onscreen. The challenges that were encountered and the methods used to overcome those challenges were also probed.

Findings from this small-scale study indicated that although each country had its own motivations for transitioning to CBA, benefits and challenges were common across countries. Creative solutions to challenges were also seen in each country.

Introduction to the use of CBA for large-scale assessments

As noted by Koomen and Zoanetti, there has been a rethinking of the role that assessment plays in education, with a particular focus on the need for assessments that are flexible, efficient and which engage students, alongside a renewed emphasis placed on 'assessment-derived metrics for accountability at the teacher, school and system level' (Koomen and Zoanetti, 2016). These changes to the role that assessment plays, have led to the vast increase in use of Computer-Based Assessment (CBA), as one solution which more appropriately meets assessment needs. The potential advantages offered by CBA are listed below;

Potential advantages of CBA

- CBA assessments provide flexibility as they can be administered at an appropriate time rather than solely during a fixed time window.
- CBA assessments are also easier to modify than other assessment forms, for example font sizes can easily be changed and students can be given a choice of language in which to undertake assessments. Modifications of this type are much harder to make when dealing with paper-based tests and may require the production of more than one test paper or test version.
- Onscreen assessments are also arguably more efficient than those carried out on paper. Indepth reporting and performance analysis can be generated within a much shorter timeframe. Onscreen tests that are adaptive also require shorter testing periods.

The widespread nature of CBA

CBA has now been adopted in a range of countries including France, Italy and Australia (currently in the first year of complete transition to CBA for the national NAPLAN assessments) where it is being used for large-scale national assessment programmes. The influence of CBA is not only being seen at the national testing level but also in large-scale international assessments such as the Programme for International Student Assessment (PISA) (onscreen version available from 2015), the Progress in International Reading Literacy Study (PIRLS) (onscreen from 2016) and the TIMSS study (eTIMSS from 2019). A range of high stakes international examinations such as the International Baccalaureate (IB) Middle Years Programme (MYP) are also moving to onscreen presentation.

In this paper, I will give a brief background to the current situation regarding CBA for large-scale assessments in the UK, the driver for an NFER research study examining the transition to CBA in a range of countries internationally. The findings of the study will then be discussed in detail before the wider implications and possible developments for future assessments are considered.

Background to CBA in UK

Compared to countries such as Australia, France and Italy, where CBA is increasingly becoming embedded in large-scale national assessment programmes, the situation in the UK is more varied, with only two of the four UK countries currently pursuing CBA programmes

for statutory national testing. Scotland is in the second year of onscreen implementation and Wales began a three-year phased implementation from September 2018.

There is wide difference in the ways in which the other two UK countries: England and Northern Ireland are approaching CBA programmes. England will see the implementation of CBA for a single statutory test, the onscreen multiplication tables check for pupils in year four (aged 8-9) in 2020 whilst Northern Ireland, the first UK country to adopt statutory CBA testing in 2007, moved away from it in 2013.

With such varied levels of CBA adoption across UK countries, it seems that many UK schools, compared to their international peers, are only just beginning to feel the effect of onscreen testing for large-scale assessments. This may have been experienced in one of two ways; firstly by participation in the onscreen version of the PISA international study (onscreen since 2015) or secondly, by involvement in the trialling of the Scottish and Welsh onscreen statutory tests.

Implications of transition to CBA

There are many advantages to the use of CBA including, but not limited to: the possibility of providing individualised assessments through the use of Computer-Adaptive Testing (CAT); automated marking; greater item interactivity and the speed at which reports on performance can be generated. However, the transition from paper-based testing to CBA for large scale and national assessments can also be complex as it involves the re-evaluation of concepts such as assessment validity (Backes and Cowan, 2018, Koomen and Zoanetti, 2016). This is due to the change in assessment mode as well as the difficulty of ensuring standardised assessment practice and conditions when there may be a range of types of computers and devices used by schools nationally. The mode effect caused by the use of computers and devices may also be differential, due to variability in the digital literacy and familiarity of students. Regarding the content of the assessment itself, there is also the question of whether CBA allows for the adequate representation of the construct that is being assessed, especially in situations where all test items are multiple choice. Where the construct is under-represented, the validity of the assessment will itself be threatened.

For large-scale national assessments to be successful, variability in teacher and practitioner support for CBA, as for paper-based testing, needs to be taken into account and addressed. However, it is also vital that adequate numbers of computers and devices, which support the chosen testing platforms, are available for use.

At the same time as UK schools are being affected by a push for onscreen testing, research indicates two specific factors that might adversely affect their onscreen testing experience. These include issues surrounding broadband connectivity and the suitability of school computers, laptops and tablets for onscreen testing. Regarding broadband, UK countries are experiencing a disparity in coverage, although England reportedly has 92% coverage, figures for the other nations vary between 85% and 89% (OFCOM, 2017).

Further to this, there is reported variability in the usefulness of the computers and devices owned by UK schools, since a study by the British Educational Suppliers Association in 2016 found that 24% of computers in primary schools and 29% of those in secondary schools in 2016 were ineffective due to age (British Educational Suppliers Association, 2016).

One potential solution to mitigate the problems caused by lack of suitable devices on which to take assessments, is to provide considerable flexibility regarding assessment timeframes, for example in ensuring that there is no set day where all assessments are required to take place.

Background to NFER interest research study and methodology

NFER interest in research surrounding the use of CBA for large-scale assessment programmes was influenced by two factors. Firstly, as a UK-based research organisation, the push for national CBA programmes, particularly seen in Scotland and Wales, and the experience of those UK countries transitioning to CBA was of interest. Secondly, the varied and ongoing nature of the transition to onscreen assessment in the UK led to an interest in the experiences of other countries that have already made a full transition to CBA. In particular, greater understanding was sought as to the different push and pull factors in addition to barriers faced and the solutions created to deal with these.

The research study had two distinct aspects: those of literature review and case study. Whilst the literature review informed understanding of the benefits and challenges of CBA and the situation regarding CBA internationally, the case studies went into greater depth by probing the experiences of three specific countries which had transitioned to CBA for large-scale assessments. Officials from all three countries who had been personally involved in the transition to CBA were interviewed.

Findings from these case study interviews form the backbone of this paper, with findings from the literature review interwoven where relevant. In particular, the literature review was helpful in guiding the selection of countries to be involved.

A range of criteria were developed to guide the selection of participant countries. These involved the types of assessments that the CBA solution was used for, the age range of pupils undertaking the assessments and the type of CBA solution utilised. Greater detail as to the final criteria can be found below in Figure 1.

Figure 1: Criteria used in country selection

Criterion 1 **Assessment type**	Criterion 2 **Pupil age range**	Criterion 3 **Type of CBA solution**
• large scale national assessments – possibly used for school accountability purposes	• within the primary and secondary education age range (5-16 years old)	• to include a range of CBA platform approaches: open source, commercial and bespoke. • also to include the use of adaptive testing.

It was felt that since NFER interest in CBA was informed by the UK context, there should be one UK country in the final sample. Northern Ireland was chosen as the UK country that had the longest history of using CBA for statutory assessments.

Additionally, France was included as an example of a country with a both very large population size, 66.9 million in 2016, (data.worldbank.org, 2018) and a very high number of primary schools (35,000). This was seen as a possible reason behind the choice of sample-based testing, rather than full cohort testing, for the National Survey assessments.

The final list of selected countries along with their contextual information relevant to the criteria can be found in Figure 2 below.

Figure 2: Final selection of countries to take part in research

Selected country	Criterion 1 information	Criterion 2 information	Criterion 3 information	Transition to CBA
	Assessment type	Pupil age range	Type of CBA solution	
Northern Ireland	• national statutory tests • used for formative and diagnostic purpose • assessments not high stakes	• pupils in school years 4 to 7 (7-11 years old)	• adaptive tests – routing for individual pupils • commercial solution	• 2007-2012, use of CEM InCAS tests • 2012 - introduction of NINA & NILA literacy and numeracy tests • non-statutory from 2013
Australia	• national statutory tests • used for school improvement • assessments not high stakes • generation of school & student level reports	• pupils in school years 3,5,7,9 (8 – 15 years old)	• adaptive tests – routing for individual pupils • bespoke solution	• from 2018 on opt-in basis determined by territory education authorities
France	• National surveys – sample based testing • Nationwide assessments – cohort assessments		• some adaptivity • open source solution • 2 types of solution – offline & online models	• from 2016

Interviews

Prior to the interviews, broad background research was carried out on each country to include information on the education and assessment contexts in general in addition to more specific information on the use of CBA in national assessments. This research then guided the creation of interview schedules for each country which included both bespoke material in addition to the more general information that the research study sought to answer for all countries.

One interview was carried out per country, taking between 45 minutes to an hour. Permission was gained from all participants to record the interviews; recordings were used to supplement the notes taken by the researcher during the interviews and were subsequently deleted.

Research Questions

The research study aimed to answer three research questions for each of the three countries taking part in the study. These were as follows;

1. What were the key drivers for moving to CBA?
2. What have been the key successes in introducing CBA for large-scale assessment programmes?
3. What have been the key challenges / barriers to developing CBA programmes for large-scale assessments?

Overview of Findings on the Role of CBA in Participant Countries

Northern Ireland

Northern Ireland was the first UK country to transition to Computer-Based-Assessments with the adoption of the CEM InCAS tests in 2007. From 2007 to 2013, the use of CBA was statutory in primary schools with a focus on the resultant assessments being used for diagnostic purposes. Due to procurement regulations, the Council for the Curriculum, Examinations & Assessment (CCEA) were required to change assessment provider in 2012 leading to the creation of two new assessments; the NINA (Northern Ireland Numeracy Assessment) and NILA (Northern Ireland Literacy Assessment). These were taken by students in school years 4 to 7 (7 – 11 years old) during the Autumn term. The NINA and NILA tests were made non-statutory in the 2013-2014 academic year, with the last year of use being the 2016/2017 academic year.

France

The French education system has three main assessment stages: During CE1 (7 years old) students are assessed in reading, writing and mathematics; during CM2 (10 years old) students are assessed on the seven key skills learned at Primary school; and during the troisième (14 years old) students are assessed on the seven competences of learning (French language command; proficiency in modern foreign language; mathematic, science and technology knowledge; ICT proficiency; humanist culture; social and civic skills; and autnomy and initiative) (ministère education nationale, 2010).

France moved to CBA for national testing in 2016. The main driver for this was the desire to move from sample based to full cohort testing, which required a more cost effective solution. By 2018, the majority of assessments in French Primary and Secondary schools have transitioned to CBA with a few notable exceptions including the Primary Science tests. The use of adaptive tests is becoming increasingly widespread.

There are still two main categories of assessments in French primary and secondary schools; the National Survey (sample-based) tests and the Nationwide full cohort assessments. These are marked by a combination of machine scoring and human marking; the sample based national assessments contain both constructed (human marked) and selected response (machine scored) items whilst the full cohort nationwide assessments solely contain multiple choice items and are thus all machine scored.

Australia

Australia is in the first year of transition to onscreen assessment, with online versions of the National Assessment Program – Literacy and Numeracy (NAPLAN) assessments ready for use. Decisions regarding the timeline for transition to the onscreen NAPLAN assessments have been devolved to individual state and territory education authorities, leading to a phased introductory period. NAPLAN assessments are taken every May by students in school years 3 (8-9 years old), 5 (10-11 years old), 7 (12-13 years old) and 9 (14-15 years old).

The online NAPLAN tests are full cohort and make use of multistage testing, where items of similar difficulty are grouped into testlets. Progression through the assessment depends on performance and ability as students are routed to testlets containing easier or more challenging items dependent on their performance on the previous testlet. The Australian

Curriculum, Assessment and Reporting Authority (ACARA), use a bespoke item authorinig system due to political demands that item development and delivery should be kept separate.

Overview of Key Findings

CBA systems and programmes differed in each country, likely due to the effect of differing cultures, political demands and levels of enthusiasm for the adoption of CBA.

For example, in Northern Ireland, the need to adhere to a five-year procurement cycle meant that after this period, CCEA moved away from the CEM InCAS tests, instead developing two new assessments; the NILA and NINA through contracts with external suppliers. Although these tests were used for diagnostic purposes to inform teaching and learning rather than for high stakes accountability purposes, the need to change test provider presented CCEA with additional challenges. Notably, the need to maintain staff and stakeholder engagement for two new national assessments directly after the establishment of the InCAS tests.

In Australia, one key driver towards CBA not seen in the other two countries was that of the diverse nature of the student population and the desire to provide assessment content that was suitable for a wide range of student groups including: disadvantaged students; students from rural areas; and indigenous students. It was felt that the use of CBA, in combination with CAT, was the only way to ensure students accessed content that was appropriate to them through the provision of individualised tests. Although the creation of nuanced and tailored assessments was a benefit widely reported on by participants, emphasis tended to be given to the greatly increased size of the item banks used to create these assessments, rather than on the social and cultural reasons behind the need for individualised assessments.

Key Drivers for the Transition to CBA

Research by Binkley, suggests that there are commonly two driving forces behind the desire to innovate assessment through the use of technology: the perceived generation of efficiency and the desire to use technology to transform the educational landscape (Binkley et al, 2012). In the main, drivers for the transition to CBA uncovered by this small-scale study can be categorised under these two overarching categories.

Firstly, regarding the desire to generate efficiency, drivers falling under this category included: the speed at which performance reports and diagnostic feedback can automatically be generated; the cost effectiveness and efficiency of using CBA rather than paper-based testing for full cohort assessments and the positive impact of CBA on teacher workload.

Importantly, cost effectiveness should be considered a long term rather than short term benefit due to the high initial cost incurred by the development of CBA solutions. It may be years before the cost benefit of CBA can be realised.

Regarding the speed at which test feedback can be generated, the Australian case study interview stated that the use of CBA will eventually reduce the timeframe for feedback reaching students to days rather than, as currently the case, months.

Concerning the second category of forces governing the transition to CBA, that of educational transformation, the main driver in this category was the possibility of providing a tailored or nuanced assessment for students, through the use of CAT. This centred on the ability to expand assessment item banks, giving a total item pool that could be up to three times bigger than those used for paper based tests. Participants emphasised the positive impact this could have on both curriculum coverage and the range of item difficulties – ensuring that all students would be guided to items that were suited to their individual ability levels and would highlight what they could achieve rather than what they could not.

Key Benefits and Successes of CBA Programmes

Interestingly, the main benefits reported for CBA programmes were very similar across all participants. As suggested by Backes and Cowan, these focused on large item banks, fast generation of performance reports and flexible test design (Backes & Cowan, 2018). In many cases the reported advantages of CBA programmes matched the original drivers for the transition to CBA in each of the countries. This was seen most clearly through the emphasis given to the speed and efficiency with which performance reports could be generated. Additionally however, participants focused on another aspect of the performance reports as a benefit for using CBA; that of the increased range of options available for reporting. For example, participants highlighted both the ease of generating reports for a wide audience including teachers, parents and children, in addition to a range of assessment purposes including those used for formative, summative or diagnostic purposes.

The use of Technology Enhanced Items (TEIs), defined by the Smarter Balanced Assessment Consortium as "a computer-delivered item which includes specialised interactions for collecting response data" (Smarter Balanced Assessment Consortium, 2012) was another benefit remarked on, particularly during the French case study interview. The advantage of using the log file data associated with TEIs when analysing performance data in order to gain

greater insight into individual student performance was also touched upon. This will be discussed in greater detail in the section on the impact of CBA on assessments of the future.

Key Challenges to the Establishment of National CBA Programmes

The main challenge to the establishment of national CBA programmes across all three countries appeared to be the preparedness of schools to test in this manner. In particular, and as seen in the UK, schools in all three countries faced difficulties with ICT resourcing and access to high-speed broadband connections as well as the challenge of ensuring staff engagement and support for CBA programmes. However, the interviews also uncovered a range of creative solutions used to deal with these challenges.

Firstly regarding the suitability of ICT devices, solutions created to deal with this differed across all three countries depending on the social and political contexts of each country. For example, in both France and Northern Ireland, government grants provided additional ICT equipment, whereas in Australia Bring Your Own Device (BYOD) policies supported students using their own devices for assessment purposes, alongside the use of school laptops and devices. The use of BYOD resulted in the creation of further solutions, such as the use of lock-down browsers to ensure students could not access the internet or their own desktops whilst taking assessments. Additionally the use of a technical readiness test in Australia enabled schools to ensure that computers, laptops and devices were technically prepared to run the testing platform prior to the start of the testing period.

In terms of ICT resourcing, subtle differences were also seen in the ways that countries carried out policies that were seemingly very similar. For example although additional ICT resources were made available through government grants in both France and Northern Ireland, the supply of the resultant equipment differed. In France, sets of tablet computers to be taken by test administrators into primary schools across the country were purchased, whereas in Northern Ireland, the equipment was supplied to individual schools.

Secondly, regarding access to adequate broadband connections, solutions again varied. The Australian Curriculum Assessment and Reporting Authority (ACARA) stated a preference for the technical readiness test accompanied by an online test model rather than the use of an offline solution. This was due to perceived issues with security and data corruption. In contrast, the French Directorate of Evaluation; Forecasting and Performance (DEPP) pursued two different assessment models, an online and offline model. The offline model, where test administrators upload test responses after assessments have taken place, is used in Primary

schools and the online model, where assessment responses continuously and instantaneously upload, is used in Secondary schools. This is to fulfil the very different broadband needs found in French Primary and Secondary schools.

The issue of confidence in, and engagement with, the use of CBA for nationalised assessments, was seen as a challenge by both the CCEA in Northern Ireland and ACARA. In Australia, this arose from knowledge of the varied confidence levels of teaching staff with the use of an online assessment system. In Northern Ireland, the political context of changing assessment provider due to procurement cycle regulations, made ensuring stakeholder support for the NILA and NINA assessments crucial. Solutions to this challenge involved the rollout of exercises to increase teacher familiarity with online assessment processes by the platform provider (Australia) and the use of stakeholder events and consultations (Northern Ireland) to inform development of the NILA and NINA assessments. Ensuring sufficient time to adequately pilot CBA assessments was also seen as essential though could be difficult to achieve particularly in cases where assessment reform is linked to political agendas.

Looking to the future – impact of CBA

It has been argued that technology-based assessments provide a dual level of interactivity: that of how the test-taker interacts with the assessment but also, crucially, the way in which assessments can now adapt based on the responses of the test-taker (Wirth & Klieme, 2003). Whilst all participants cited personalised assessments as a driving force behind transitioning to CBA, the use of log-file data, to ensure test accessibility whilst gaining clarity on how students approach certain test items, was discussed as a future assessment innovation. Additionally, the use of AI to enable certain subject aspects, such as reading fluency, to be assessed and marked online was also considered as a potential future direction for CBA assessments.

The importance of stakeholder engagement, previously discussed as a key challenge in the transition to onscreen assessment, to future assessment developments can not be overstated. At the time of this research project, the Australian Education Council announced that the use of automated essay scoring in the ACARA NAPLAN tests will no longer go ahead. Protests from teaching unions and concerns that automated essay marking would decrease both the validity and reliability of assessments may well have played a vital role in the termination of this policy (Robinson, 2018).

Conclusion

Although this was a small-scale research study, it did lead to interesting findings. Notably, although the drivers, benefits and challenges of transitioning to CBA appeared similar across participants, the solutions created to overcome challenges differed depending on the social and political contexts of the countries themselves. This study specifically focused on countries using CBA for Primary and Secondary level assessments, however it was notable that all large-scale national assessment programs discussed in this study were classified as low-stakes assessment programmes. Furthermore that they did not incorporate the entire primary to secondary age range but focussed instead on the 7-15 age range. Whether high stakes national tests and qualifications will also move to onscreen administration and over what timeframe will be interesting to see.

References

Backes, B & Cowan, J. (2018). Is the Pen Mightier Than the Keyboard? The Effect of Online Testing on Measured Student Achievement. Working paper 190. National Center for Analysis of Longitudinal Data in Education Research.

Binkley, M., Erstad, O. Herman, J. Raizen, S. Ripley, M. Miller-Ricci, M. & Rumble, M. (2012) Defining twenty-first century skills. Cited in Koomen, M. & Zoanetti, N. (2018). Strategic planning tools for large-scale technology-based assessments. *Assessment in Education: Principles, Policy & Practice,* 25:2 [online]. Available: https://www.tandfonline.com/doi/pdf/10.1080/0969594X.2016.1173013?needAccess=true [26th September 2018].

British Educational Suppliers Association, (2016). *Information and Communication Technology in UK State Schools.* Issue 19: 2016: Full Report.

Data.worldbank.org. (2018). *France | Data.* [online]. Available: https://data.worldbank.org/country/france [27th September, 2018].

Koomen, M. & Zoanetti, N. (2018). Strategic planning tools for large-scale technology-based assessments. *Assessment in Education: Principles, Policy & Practice,* 25:2 [online]. Available: https://www.tandfonline.com/doi/pdf/10.1080/0969594X.2016.1173013?needAccess=true [26th September 2018].

OFCOM. (2017). *Connected Nations 2017: Data Analyis*. [online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0016/108511/connected-nations-2017.pdf [27th September 2018].

Ministère Education Nationale. (2010). Files on School Education: School Education in France. [online]. Available: http://media.eduscol.education.fr/file/dossiers/50/3/enseignement_scolaire_VA_135503.pdf [27th September 2018].

Robinson, N. (2018) 'NAPLAN: Robot marking of school tests scrapped by education ministers", *The e-Assessment Association* [online]. Available: http://www.e-assessment.com/news/naplan-robot-marking-school-tests-scrapped-education-ministers/ [26th September 2018]

Smarter Balanced Assessment Consortium. (2012). *Technology-Enhanced Items Guidelines*. [online]. Available: https://www.measuredprogress.org/wp-content/uploads/2015/08/SBAC-Technology-Enhanced-Items-Guidelines.pdf [27th September 2018].

Wirth, J. and Klieme, E. (2003) Computer-based assessment of problem solving competence. *Assessment in Education: Principles, Policy & Practice,* (10), 329-345. Cited in Koomen, M. & Zoanetti, N. (2018). Strategic planning tools for large-scale technology-based assessments. *Assessment in Education: Principles, Policy & Practice,* 25:2 [online]. Available: https://www.tandfonline.com/doi/pdf/10.1080/0969594X.2016.1173013?needAccess=true [26th September 2018].