

Extended Case Study National Onscreen Adaptive Personalised Assessments of school children in Wales



The personalised assessments are innovative adaptive onscreen assessments in numeracy, and English and Welsh reading, delivered to all children aged 7 to 14. This case study reports on the first full year of live delivery onscreen.

In Wales, all children aged 7 to 14 take national assessments in procedural numeracy (number, measuring and data skills), reading and numerical reasoning (solving problems). These assessments are designed to identify areas where children and teachers need to focus their efforts in order to make progress.

In the past, these assessments were taken entirely on paper. However, the Welsh Government has decided to move the assessments to an onscreen format, starting in 2018. AlphaPlus is leading the multi-partner project team on behalf of Welsh Government, developing assessment infrastructure and content for the new assessments as they are phased in, while also delivering the legacy on-paper assessments.

The paper tests are being replaced with adaptive online personalised assessments specifically designed for use in Wales, and delivered in Welsh and English. The new adaptive assessments automatically adjust the level of questions to match the learner taking the assessment, providing a personalised assessment experience and tailoring the level of challenge.

Fact File

Number of primary schools in Wales: 1235 (of which 397 teach in Welsh)

Number of secondary schools in Wales: 187 (of which 47 teach in Welsh)

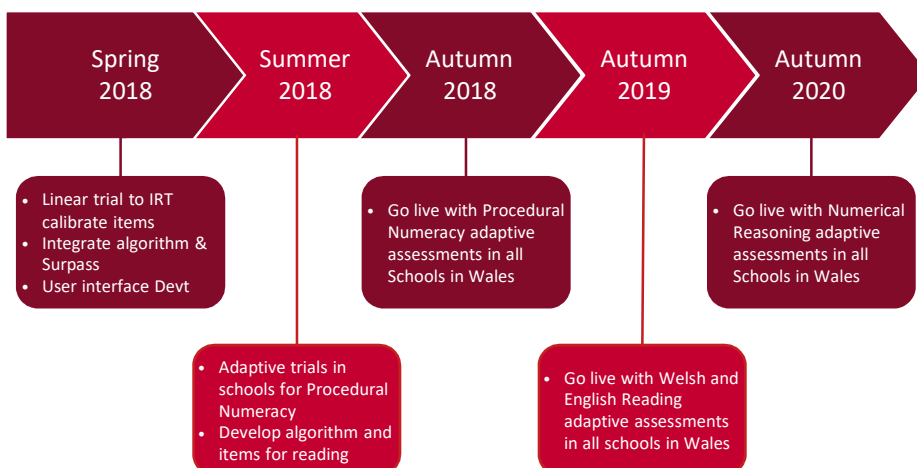
School years being assessed: - all children in state education Years 2 to 9 (aged 7 to 14) (excluding those disapplied due to special learning needs)

Number of learners taking the assessments each year: approx. 270,000

Number of personalised assessments taken to date (as of August 2019): 337,792

Project start: January 2017

Project live (first on-screen assessments): numeracy (procedural) assessment - 2018; reading assessments - 2019; numeracy (reasoning) – 2020



The numeracy (procedural) assessments have now completed their first year's live use in schools.

The reading assessments (in Welsh and English languages) go live on-screen from Autumn 2019.

The numeracy (reasoning) assessments are being developed, ready for introduction in Autumn 2020.

Turning Education & Skills
Policy into Practice

Extended Case Study

National Onscreen Adaptive Personalised Assessments of school children in Wales

2/10



How the assessment works

For an adaptive assessment to function, the difficulty level of every item in a bank must be definitively known, a process done using **item response theory**.

Modern adaptive testing relies upon **item response theory (IRT)**, in which a probabilistic model is used to estimate both how challenging items are and how able learners are. Learners' ability levels and items' difficulty parameters are computed on the same scale, allowing the difficulty level of an item to be determined independently of the ability of the learner who takes it, something that classical statistics are unable to do.

In the Welsh personalised assessments, learners are allocated an item written for their year group to begin with. Once they answer this item, an algorithm calculates an interim ability estimate, and uses this to select an appropriate item to deliver next. A learner who answers a question correctly will receive a question that is slightly more challenging; a learner who answers a question incorrectly will receive a slightly easier question.

The software estimates the learner's ability on a scale, and the purpose of each subsequent question is to refine the estimate, so questions are chosen to optimise those measurements. This process is repeated after every item, with ability rising and falling as learners get items right or wrong, and the algorithm selecting items at an appropriate level for a learner's current ability estimate.

"it was really good even though it was a bit hard it was still really fun"

Feedback from a Year 3 learner on the Reading adaptive assessment trial*

This process continues until the assessment system has covered the relevant aspects of the curriculum and gathered sufficient information to make a sufficiently accurate ability estimate, or until the assessment has reached its default maximum length, at which stage the assessment is terminated.

*All feedback is taken directly from forms filled in by learners & teachers

There are several additional constraints on the adaptive algorithm in these assessments:

- Curriculum category constraints ensure that learners do not solely receive items on one topic
- Minimum and maximum test length constraints ensure that learners receive an assessment of acceptable length
- Exposure controls prevent any one item from being seen by "too many" learners (the threshold of which can be adjusted)
- Item bank constraints. While learners can access items from any year if they are sufficiently able, the first few items in each assessment are aimed at that learner's own year group. Secondary school content is withheld for the youngest learners, so they cannot access content on advanced topics in the secondary curriculum not yet encountered in class.

This tailored assessment provides information to teachers and learners on strengths and weaknesses of both individual learners and whole classes, so they know where to focus their efforts to improve.

This graph maps a Year 3 learner's journey through an adaptive assessment. The red triangles are incorrectly answered items and the blue circles are correctly answered items. The y axis scale is the learner's ability estimate.

The learner answers the first few questions correctly, and quickly climbs the ability scale. Through a series of correct and incorrect answers, the range of difficulty of questions narrows, converging on the final, steady outcome. As the algorithm has reached a sufficiently accurate estimate, the assessment is terminated.



Turning Education & Skills Policy into Practice

Extended Case Study 3/10

National Onscreen Adaptive Personalised Assessments of school children in Wales



Assessment arrangements

Schools are free to schedule the assessments at any time during the academic year that they consider most appropriate to gain benefits for learning and teaching. They can be scheduled up to twice a year. For example, a school could choose to use the assessment at the beginning and end of each school year in order to see how a learner has progressed throughout the year. It is a statutory requirement for the assessment to be used once per year, with results reported to parents or carers.

Schools can choose to assess classes, small groups or individual learners according to their facilities and at a time that works for them and their learners. As the personalised assessments are different for each learner, there is no need for a whole class of learners to take the assessments at the same time or for concerns about cheating.

The personalised assessments do not have a fixed length (although there is a maximum number of questions – for example in procedural numeracy it ranges from a minimum 21 to a maximum 35 questions for learners in year 2, and from 27 to 45 for learners in year 9).

The average assessment has 28 questions and generally takes between 20 and 40 minutes to complete (younger children typically have shorter assessments). The system stops the assessment when sufficiently accurate information about the learner's ability has been obtained. Therefore, learners can take as long as they need to finish the assessments, without feeling the pressure of an imposed time constraint.

Support for schools is available through online webinars, videos and a user guide demonstrating how the assessments should be administered and how to access learner and group reports. Schools are encouraged to sign up for a webinar to gain an understanding of how the assessments work and how they can use the information from the assessments to support teaching and learning.

"I also liked how it was set out and the person sitting next to you didn't have the same question at the same time, so if you were stuck you couldn't copy them"
 Feedback from a Year 4 learner on the Reading personalised assessment trial

"For a new format, the assessment was very straight forward to set up, activate and administer. The guidelines were very helpful."
 Comment from a teacher on the Reading personalised assessment trial

The assessments are entirely online: web-based with no local server or installation of software required – and can be taken on desktop computers, laptops or tablets e.g. iPads or Chromebooks.

Sample questions from the personalised assessments

Maint esgidiau'r holl blant ym Mlwyddyn 3

Maint 2		Allwedd ☺ = 2 o blant
Maint 2½		
Maint 3		
Maint arall		

Ym Mlwyddyn 3, faint o blant sydd yna i gyd?

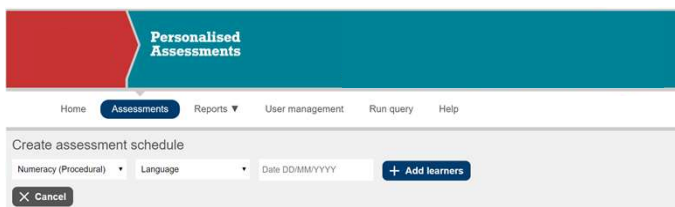
o blant



Preparing for the assessments

Access to the assessments is via the Welsh Government’s Hwb digital learning platform (<https://hwb.gov.wales>). The Hwb platform provides all maintained schools in Wales with access to a wide range of digital services and content via a Welsh Government operated national Active Directory. This means that all learners and teachers have a single user name and password which significantly reduces the authentication process for services such as personalised assessments. Before learners can take personalised assessments:

- the headteacher, or member of staff acting on their behalf, assign assessment roles for relevant staff
- staff log in to Hwb, navigate to the personalised assessments in the Hwb tools menu and schedule assessments for learners
- learners log in to Hwb and work through a familiarisation assessment so they understand the question formats.



After the assessments

Personalised feedback is made available on the assessment website one day after the assessments are completed. This is considerably faster than for the paper tests, which supports the key purpose of formative assessment, as changes to teaching can be implemented in a timely manner. Teachers have the opportunity to review individual feedback before releasing it to learners.

Language

The content of the English and Welsh versions of both the procedural numeracy assessment and the numerical reasoning assessment is identical; the questions have simply been translated from one language to the other. However, the item banks of the English and Welsh reading assessments are different, as they are based on texts from their respective languages.

The procedural personalised assessments can be taken in English or Welsh. The language of a learner’s assessment is chosen when a staff member schedules the assessment. However, a learner is able to view a copy of the question in the other language at any point during the assessment, by clicking an icon at the bottom of the screen.

All administration support, including helplines, administration website, user guides and FAQs are provided in English and Welsh.

Multi-disciplinary team

This project is a national rollout of high profile assessments for children, and there are demanding educational and technical requirements. The specialist capabilities required mean that no single organisation could deliver the project. AlphaPlus is leading the multi-disciplinary multi-partner team.

The key members of the project team are listed below.

Organisation	Role
AlphaPlus (prime contractor)	<ul style="list-style-type: none"> • Project leadership • Project management • Assessment & psychometric design • CAT algorithm design • English reading assessment development • Data collection & analysis
BTL	<ul style="list-style-type: none"> • Online assessment platform
WJEC	<ul style="list-style-type: none"> • School support, helpdesk & training • Welsh reading assessment development • Welsh language support
CITO	<ul style="list-style-type: none"> • Open source adaptive algorithm
Acumina	<ul style="list-style-type: none"> • Numeracy assessment development (reasoning & procedural) • Numeracy assessment design
Pia	<ul style="list-style-type: none"> • Test item accessibility
Method 4	<ul style="list-style-type: none"> • User interface development

Turning Education & Skills Policy into Practice

"the assignment was made [in] a very good way like when I forgot to choose two it said are you sure you want to proceed when I only [chose] one."

Feedback from a Year 5 learner on the Reading adaptive assessment trial

Benefits

The personalised assessments are specifically designed for formative benefit - so teachers have accurate and timely information about the reading and numeracy skills and subskills of their learners and an understanding of strengths and areas for improvement in these skills.

The assessments are focused on understanding learner progress, and not school performance or accountability. The assessments offer teachers a performance measure against a national benchmark (a standardised score that compares performance to all other learners in Wales), but the outcomes are purposefully not used as an accountability measure.

Adaptive assessments enable learners to move up or down the ability range so that all learners get approximately 50% of the questions right and 50% wrong. Learners may therefore move up and down the bank to access content from higher or lower years – challenging the more able while not demotivating those who are struggling. Allowing learners from different years to "roam" through a single item bank means that the assessments truly test what the learners can do, making no judgement of what was expected to be taught in a given year.

"it was a fun test and I liked it because it was challenging me a bit"

Feedback from a Year 4 learner on the Reading adaptive assessment trial

After taking the reading personalised assessment trial, of learners responding to our questionnaire:

- 74% found the questions to be just right (not too easy or too difficult)
- 70% said that they enjoyed the assessment

One of the key challenges of national testing is sustaining learners' engagement. The national personalised assessments have two key benefits here:

1. Because each learner attempts questions at or about their level, the tests are engaging for all. Paper tests, which are the same for all learners in a given school year, typically have significant floor and ceiling effects (too easy for some learners, too difficult for others), meaning that not all learners experience a test at an appropriate level (the range of abilities within a year is much wider than the difference between years).
2. The adaptive assessment typically takes less time than an equivalent linear test to reach an equivalent measurement accuracy. Shorter tests or assessments can mean better learner engagement.

"simple format and good length of time"

Feedback from a Year 8 learner on the Reading adaptive assessment trial

Aside from the engagement benefits, there are operational benefits for schools:

1. The assessments are personalised, so there is no need for whole year groups to take the assessment at the same time.
2. The assessments are marked automatically, whereas the current paper-based tests are marked in schools.
3. The assessments offer high-quality feedback, providing schools with information on the skills of individual learners and whole classes.
4. As the assessments are automatically marked and calibrated, this feedback is available the day after the assessment which supports the key purpose of formative assessment, as changes to teaching can be implemented in a timely manner.

"more fun than doing it on paper."

Feedback from a Year 5 learner on the Reading adaptive assessment trial

Turning Education & Skills
Policy into Practice

Accessibility

From the project onset, the Welsh national personalised assessments have been designed to allow children with different access needs to fully demonstrate what they know and can do. Once a national test moves to onscreen, all the assessments are delivered onscreen – there are no paper versions of the onscreen assessments. Every learner that is cognitively able to take the assessment takes it onscreen, with access support where needed.

Accessibility was built into all technology and content development from the outset, for example:

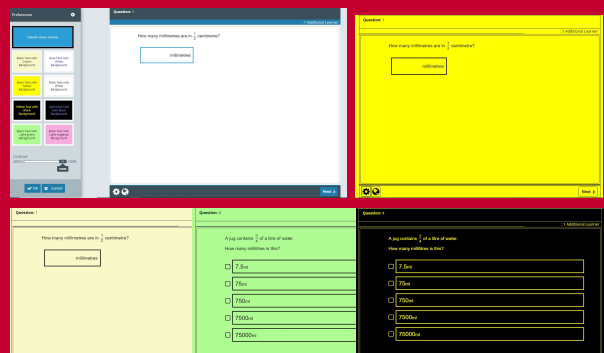
- The personalised assessments are WCAG2.0 AA compliant and accredited by the [Digital Accessibility Centre](#) and developed with contributions from specialist access organisations. We conducted surveys with special schools and representative bodies to understand the full range of disabilities to be accommodated and of favoured tools and approaches. We also ensured that special schools were included during trialling, working with learners with various physical and behavioural disabilities to ensure fair access for all.
- As well as following best practice for accessible e-assessment, the assessments were developed with general accessibility in mind. For example, all items in the procedural numeracy assessment have been written in the active voice, using short, straightforward sentences, and the assessment interface is simple, uncluttered and can be personalised.
- Modified assessments are available to support learners with visual impairment or learners who normally use large print as part of their normal classroom practice. There is also guidance to support those administering the modified assessments to learners with a hearing impairment or who use sign language.

"I found it easier doing the test online. Much better than working on the enlarged booklet I usually need to have. I could change the colour and size of text for myself."

Feedback from a Year 4 learner on the Reading adaptive assessment trial

- In Welsh, where possible, we have ensured that the language used is accessible e.g. avoiding difficult concise forms of irregular verbs, ensuring that the impersonal form of the verb is not overused.
- The assessments use the [BTL Surpass](#) platform, a platform with years of user experience refinement for many of the world's largest e-testing providers, including Prometric, AQA, WJEC and City & Guilds. Surpass provides the following accessibility support for learners in test delivery:
 - Colour preference options for dyslexic and visually impaired users
 - Alternative text descriptions for images
 - Additional time for learners with special requirements
 - Keyboard only navigation
 - Text-to-speech support through JAWS screen reader
 - Screen magnification support.

Standard assessment format and alternative colour options, suitable for dyslexic and visually impaired learners



- A small number of questions require the use of tactile diagrams for learners with visual impairment. These ensure curriculum coverage for learners who are using screen readers or having the assessments read by an adult. The learner only needs to use the booklet if one of the questions requiring a diagram is selected by the system.
- Braille text and question booklets are used to support the reading assessments, as these cannot be used with screen readers (other than for instructions). The Braille assessments are produced in both grade 1 and grade 2 Braille.

Turning Education & Skills Policy into Practice

Extended Case Study

National Onscreen Adaptive Personalised Assessments of school children in Wales

7/10



The three assessments

Procedural Numeracy	Reading	Numerical Reasoning
Live in schools from the academic year 2018/2019.	Live in schools from the academic year 2019/2020.	Live in schools from the academic year 2020/2021.
Constructed from a large bank of approx. 1,800 items, covering the National Numeracy Framework .	Constructed from a large bank of approx. 200 texts and 900 items, covering the reading elements of the National Literacy Framework .	Constructed from a large bank of approx. 1,250 items, covering the National Numeracy Framework .
Parallel banks for English and Welsh assessments.	Independently developed banks for English and Welsh assessments.	Parallel banks for English and Welsh assessments.
The assessment is not divided into sections.	<p>The assessment comprises items of different lengths, ranging from 1 mark items to 8 mark question sets. It is divided into 'sections': learners begin the assessment with short texts (150 – 200 words and 3 items) or sentence completion, and then move onto medium (200 – 350 words and 5 items) and long texts (350 – 450 words and 8 items).</p> <p>There are curriculum and source coverage constraints on the algorithm which means that learners receive a variety of different text types and questions assessing a range of reading skills.</p>	<p>The format of the assessments currently under development comprises 3 item types of varying complexity:</p> <ul style="list-style-type: none"> • Category A – discrete single mark items • Category B – groups of 2-4 linked single mark items • Category C – single items with 2-4 marks • Category D – stimulus questions – made up of Category A and C items, grouped around a common stimulus (approximately 8 marks in total). <p>The assessment is divided into 'sections', starting with Category A and Category B items, then moving on to include category C items, and ending with a Category D stimulus question.</p>
The assessments adapt at the item level, i.e. the system estimates the ability after each item and chooses an appropriate next item.	The assessments adapt at the text level, i.e. the system selects an appropriate text, and the learner answers all the questions associated with that text; the system then estimates the learner ability and chooses an appropriate next text.	The assessments adapt at the section level, i.e. the system selects an appropriate section, and the learner answers all questions associated with that section; the system then estimates the learner's ability and chooses an appropriate next section.

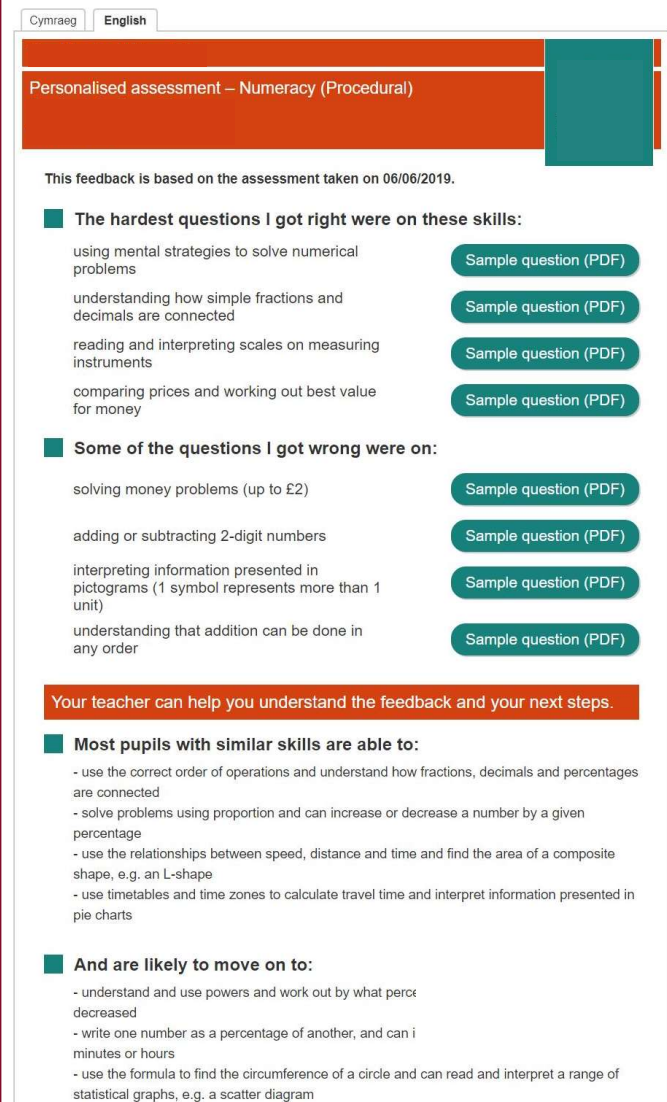
Turning Education & Skills Policy into Practice

Feedback

The purpose of the personalised assessments is to provide teachers and learners with information on skills so that they can understand areas to work on and their plan next steps. Ensuring teachers and learners receive clear, targeted feedback is crucial to the success of the assessments.

A range of reports is provided about the performance of both individual learners and groups/classes. The two most important for individual learners are the learner feedback and the individual learner progress reports.

Learner feedback



The screenshot shows a web interface for a personalised assessment report. At the top, there are tabs for 'Cymraeg' and 'English'. The main heading is 'Personalised assessment – Numeracy (Procedural)'. Below this, it states 'This feedback is based on the assessment taken on 06/06/2019.' The report is divided into several sections:

- The hardest questions I got right were on these skills:**
 - using mental strategies to solve numerical problems (Sample question (PDF))
 - understanding how simple fractions and decimals are connected (Sample question (PDF))
 - reading and interpreting scales on measuring instruments (Sample question (PDF))
 - comparing prices and working out best value for money (Sample question (PDF))
- Some of the questions I got wrong were on:**
 - solving money problems (up to £2) (Sample question (PDF))
 - adding or subtracting 2-digit numbers (Sample question (PDF))
 - interpreting information presented in pictograms (1 symbol represents more than 1 unit) (Sample question (PDF))
 - understanding that addition can be done in any order (Sample question (PDF))
- Your teacher can help you understand the feedback and your next steps.**
- Most pupils with similar skills are able to:**
 - use the correct order of operations and understand how fractions, decimals and percentages are connected
 - solve problems using proportion and can increase or decrease a number by a given percentage
 - use the relationships between speed, distance and time and find the area of a composite shape, e.g. an L-shape
 - use timetables and time zones to calculate travel time and interpret information presented in pie charts
- And are likely to move on to:**
 - understand and use powers and work out by what per cent decreased
 - write one number as a percentage of another, and can i minutes or hours
 - use the formula to find the circumference of a circle and can read and interpret a range of statistical graphs, e.g. a scatter diagram

Principles of feedback

The priority of the assessment is to support teaching and learning, not as a school accountability measure, and the feedback reports have been designed to reflect this. To ensure the final reports were useful while still maintaining their formative purpose, Welsh Government held and hosted focus groups with teachers and school leaders, as well as with parents and learners, and had discussions with Regional Education Consortia, the Expert Advisory Group and the Project Board.

- The learner feedback from the personalised assessment does not include a score. This is because the focus of this feedback is on understanding what a learner can do, what they may need to work on, and where they are going next.
- The Learner Feedback Report highlights the learner's strengths and areas for further development, based on the skills in the National Literacy and Numeracy Framework, as measured in the assessment.
- Scores are not the main focus on the group reports, because the main focus of the assessments is identifying learners' skills and next steps to inform future learning and teaching.

Feedback on individuals is available on the assessment website one day after the assessments are completed, and teachers have the opportunity to review reports before releasing them to learners. Up to four statements are presented in the learner feedback report along with a sample question for each statement (not exactly the same question that the learner received).

The descriptors at the bottom of the report take the learner's final result from the assessment and map this against the complete bank of questions. The first descriptor summarises the questions in the bank that relate to the learner's current attainment level. The second descriptor summarises the questions in the bank that are a bit more difficult than those at the learner's current attainment level.

Extended Case Study

National Onscreen Adaptive Personalised Assessments of school children in Wales

9/10



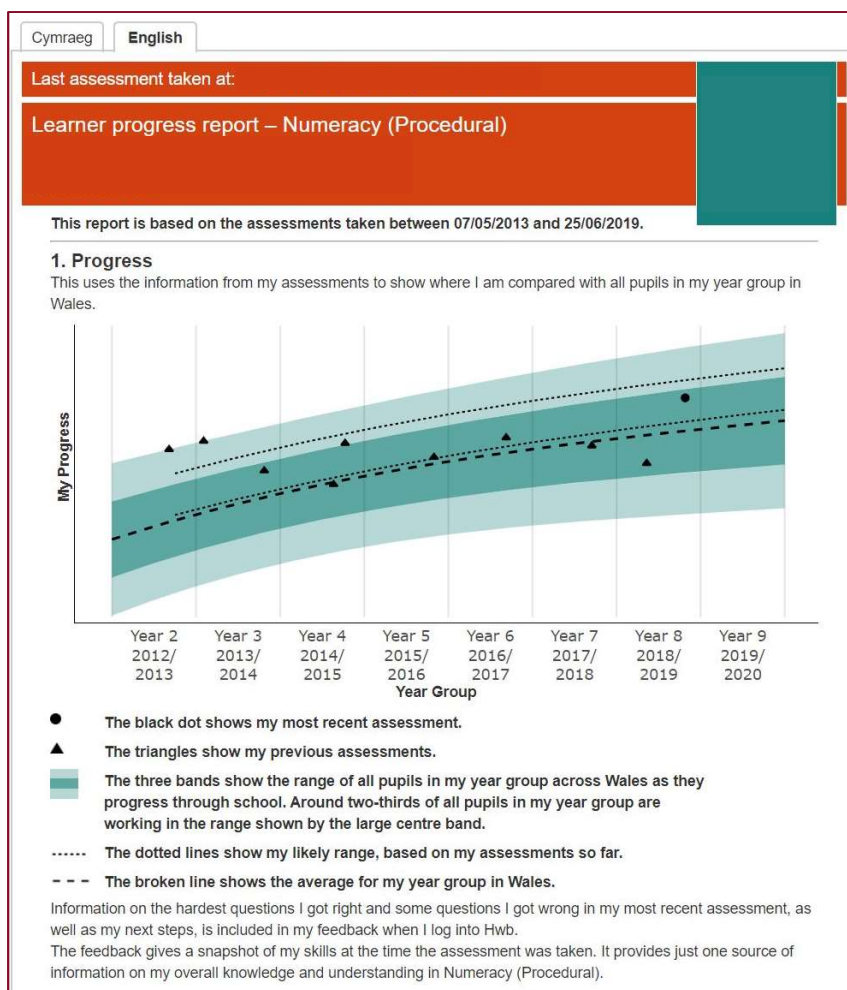
Individual learner progress report

Individual learner reports, issued at the end of the first year of the personalised assessments, show a learner's progress over the different assessments they take, based on a standardised score

The first section of the report (right) shows the learner's progress. The assessments taken by the learner are shown as black dots on the graph. Using the information from their assessments, the graph shows where a learner is in comparison to all other learners in their year group across Wales. As the learner takes more assessments, their progress will be mapped from one year to the next.

The second section of the report shows the learner's age-standardised score for their most recent assessment. This shows how well a learner has done compared with other learners in Wales born in the same year and month and accounts for variations of life experience within a cohort.

The score range is 70–130 and the average is 100.



An age-standardised score converts a learner's raw score into a standardised score which takes into account the learner's age in years and months at time of assessment and indicates how they are performing relative to learners of the same age. For example, if two learners, one who is 6 years and 1 month at the time of the assessment and one who is 6 years and 9 months, receive the same raw score on the personalised assessments, the younger learner will receive a higher age-standardised score.

2. Age-standardised score

This is worked out by comparing my assessment score with the scores of other pupils in my year group who were born in the same year and month as me. The score range is 70–130 and the average is 100. If I have a very high age-standardised score it may be shown as 'more than 130' (130+). If my skills are developing more slowly I may have a score of 'less than 70' (<70). A score from a personalised assessment is just one piece of information about how my skills are developing.

In my most recent assessment my age-standardised score was: 70

hwb.gov.wales/personalised-assessments

Group reports

A number of reports for groups of learners, summarising the performance of the whole group, are also available for staff to request and use.

Group members and scores

Learners	Hwb ID	Date of last assessment	Age-standardised score	Progress score
		10/12/2018	83	966

Turning Education & Skills Policy into Practice

Extended Case Study

National Onscreen Adaptive Personalised Assessments of school children in Wales

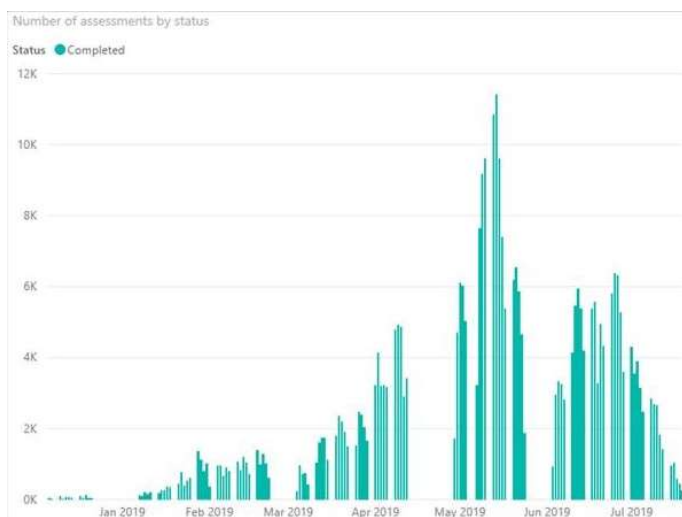
10/10



Usage

The table to the right details the number of assessments taken in their first year of delivery. In the first year of use, 96% of learners in years 2 to 9 in Wales completed at least one personalised assessment.

As seen in the graph below, the majority of assessments were taken in the first half of May, which is the testing window for the paper-based tests that the online assessments replace. However, this is likely to change in future years as the personalised assessments can be used at any point during the academic year. Schools can choose to use the assessments at whatever time they consider to be of most benefit to learning and teaching.



Live assessments (August 2019)	Unique learners completing assessments
Numeracy (Procedural) MLP user	357
Numeracy (Procedural) Braille user	11
Numeracy (Procedural)	268,375
Total	268,743

Response

Although there have been one or two small glitches, as would be expected of a technical development of this size, we have been hearing positive feedback from both teachers and learners.

Following the trial of the Procedural Numeracy adaptive assessment, of teachers responding to our questionnaire:

- 78% thought learners were engaged with the assessments
- 83% thought the assessments were about the right length for the majority of learners
- 75% found it easy or very easy to use the Administration System to access the feedback
- 65% found the individual learner report to be useful or very useful
- 63% found the feedback to be useful or very useful

Following the trial of the Reading adaptive assessment, of teachers responding to our questionnaire:

- 71% found the content of the assessment to be as they expected
- 74% thought that learners were engaged

"The feedback is detailed and thorough. It will be very useful for the new class teachers in September."

Comment from a teacher on the Procedural Numeracy personalised assessment

"I love this assessment I probably think it was my favourite assessment to be honest."

"Prefer it to paper assessments."

"It was really fun and it helped me so I wonder if you could make more of them"

"I tried my best and it was better on the computer because on the paper it would be hard to read my handwriting"

Feedback from learners on the personalised assessments

Turning Education & Skills Policy into Practice