Investigating students' experiences of piloting Stride

Jeanne Marie Ryan, Research Fellow, AQA

Handan Lu, Researcher, AQA

Executive summary

Stride is AQA's new adaptive assessment, providing diagnostic tests to help students prepare for GCSE Maths.

The objective of this study was to gather feedback from students about their experiences of trialling Stride's diagnostic tests during the product's development phases. The mixed-methods research featured focus groups and surveys with students who took part in the testing periods. This study accompanied quantitative research that explored the validity and reliability of the tests.[1]

Focus groups and surveys were designed to gather feedback about student levels of engagement, their experience of how the adaptive engine tailored items to their proficiency levels, and their thoughts about how useful the results from the diagnostic tests would be for their learning and revision.

Findings from this study show that the majority of students:

- found taking part in the tests to be a positive experience

- confirmed that tests and their results would help them focus their revision

- thought that the tests were engaging and appropriately challenging due to the way the tests adapted to their proficiency levels.

AQA
Questions matter

Introduction

Many students aged 11 to 16 years struggle to make progress in maths. A student who has difficulty getting to grips with arithmetic in primary school will often go on to struggle with solving equations in secondary school.

Stride is a set of diagnostic tests that will help to quickly identify gaps in a student's understanding of fundamental maths concepts. Stride also provides students with personalised learning content based on their results. This immediate feedback enables maths teachers to provide more targeted support to their students and allows students to guide their own revision strategies more accurately.

Each test is underpinned by one of five key concepts: numbers, algebra, proportions, graphs, and shapes.

Research was carried out to evaluate how well these key concepts align with maths teachers' pedagogical practice and students' learning and understanding. Once the tests had been piloted in schools across the country, quantitative research analysed the validity and reliability of the tests, and a qualitative strand of research gathered and analysed data on user feedback relating to the Stride platform and its use in the classroom. This qualitative data, which was collected via focus groups and surveys with students, was used to inform the development of the product and any necessary modifications.

The research was conducted in two phases:

- Phase 1 – autumn/winter 2022
- Phase 2 – spring/summer and autumn/winter 2023.

This phased approach enabled iterative development by integrating user feedback at each stage.

Methods

A combination of virtual focus groups and post-test surveys were used to gather feedback from students regarding their experience of taking the diagnostic tests. Focus groups provided granular feedback about students' experiences, which allowed for improvements to be implemented throughout the course of the research; post-test surveys generated a large amount of data to ascertain whether students found the test experience positive and useful for their learning.

Focus groups

Participant recruitment and data collection

Convenience sampling was used to select schools in which virtual focus groups would be held; consequently, the sample was not representative of the entire country. Eighteen schools (including 93 students) from different regions in England were identified for participation in the focus groups; however, there was an overrepresentation of schools from the South East. Once schools had agreed to participate, maths teachers were asked to select a group of six to eight students who would represent the full range of abilities and backgrounds present in their classrooms. Prior to conducting the research, informed consent was gathered from all those involved: students, parents, teachers and head teachers.

Most of the focus groups were held within a week of students completing the test; however, some schools had to delay the focus groups by a few days due to timetabling. The focus groups were held via Teams and lasted up to 45 minutes. At the start of each focus group, students were asked to confirm that they consented to the focus group being recorded; recordings were subsequently sent to an approved third-party organisation for transcription.

Focus group questions

The focus group questions explored the following:

- students' experience of logging in and reading instructions before getting started

- the functionality of the test (eg whether students could respond easily to different question types)

- whether there was an appropriate balance of challenging and less challenging questions

- students' ideas for improvements

- students' use of the reporting tools to view their results.

Students were also asked a yes or no question about whether they thought the tests would be useful for their learning and revision.

Data analysis

Data from the focus groups was analysed using a combination of deductive and inductive approaches; NVivo software was employed for the coding process. A deductive approach was used to create codes for elements such as on-screen instructions, functionality, the adaptive engine (eg whether there was an appropriate balance between less challenging and more challenging questions) and confidence rankings. An inductive approach was used for coding of students' feedback to identify themes from their individual responses.

Post-test surveys

Survey participants

All students who completed a test were asked to fill out a survey regarding their experience. Throughout the course of the spring, summer and autumn testing windows, 2,976 survey responses were received in relation to all five diagnostic tests. In many participating schools, students took part in multiple tests; the number of survey responses does not therefore reflect the total number of students who participated in the testing windows.

Survey questions

Students were asked a combination of closed and open-response questions. Closed questions, which used a Likert scale, focused on:

- to what extent students thought instructions for the test were clear
- to what extent they found questions easy to respond to
- whether they thought there was an appropriate balance of challenging and less challenging questions
- how useful they found the reporting tools showing their results
- how they would rank their experience overall, from positive to negative.

Open-response questions explored:

- students' ideas about what improvements could be made
- what they liked and disliked about the test
- how they could use the test and its results for their learning and revision.

Data analysis

Descriptive statistical analysis was used for closed responses and thematic content analysis was used for open responses.
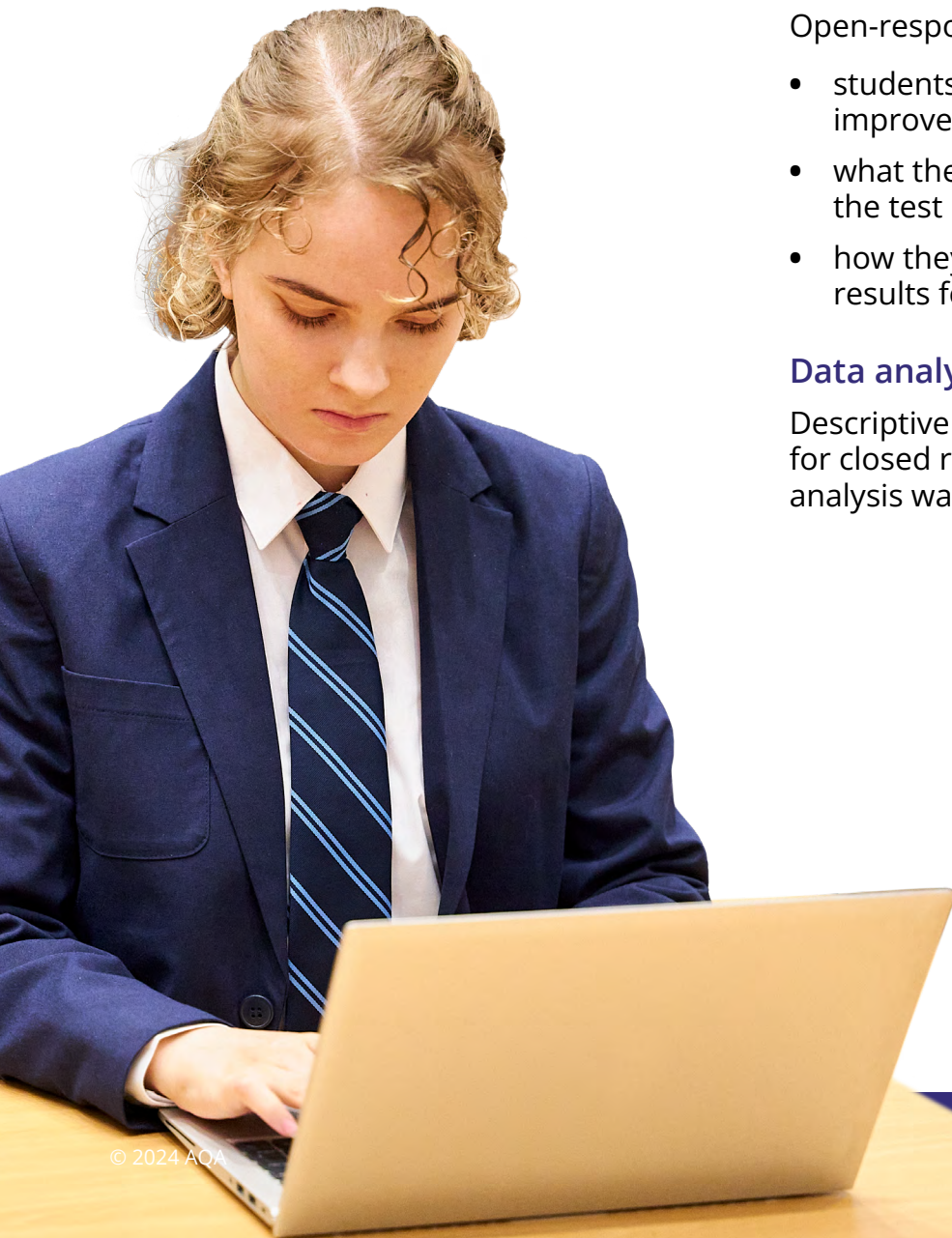
Findings

Key Finding 1: the majority of students had a positive experience taking the tests

Analysis of data from both the focus groups and post-test surveys showed that the majority of students had a positive experience taking part in the tests. From the 2,976 survey responses received across all five diagnostic tests, an average of **77% of students** reported that they had an overall positive experience (Figure 1).
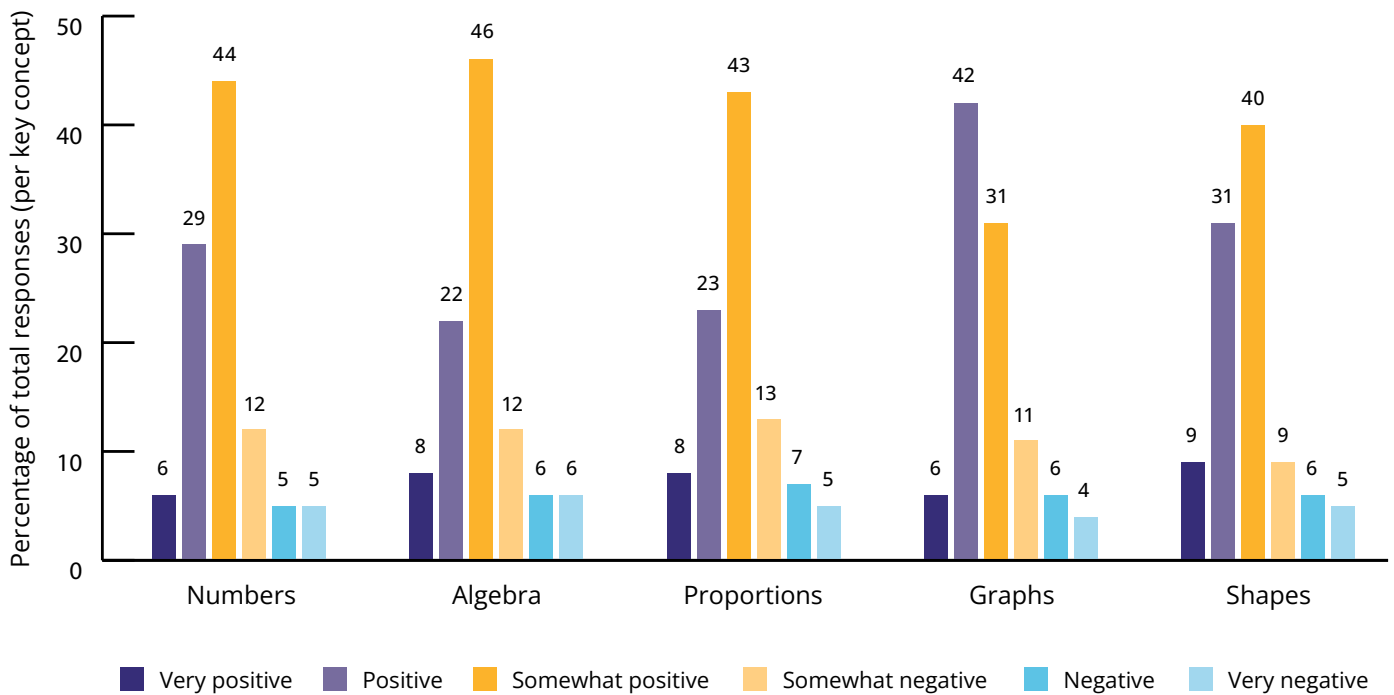

Figure 1: Students' ratings of their test experience (survey data)

In alignment with the survey data, students who participated in focus groups all confirmed that they had a positive experience taking the tests. Focus group students also reported that they enjoyed taking part in the tests because each test helped them identify their strengths and weaknesses.

> "I liked it, it was a good opportunity to check what I know and what I need to work on."

Student, School J

> "Yeah, I like that this could be used to test what you know and what you're good at and what you need to improve on."

Student, School D

Key Finding 2: the majority of students stated that the tests would help them revise

Findings showed that, on average, **79% of students** who completed a survey felt the tests would be helpful for their revision (Figure 2) and for their learning more generally. In particular, students appreciated being able to view their results immediately after finishing the test.
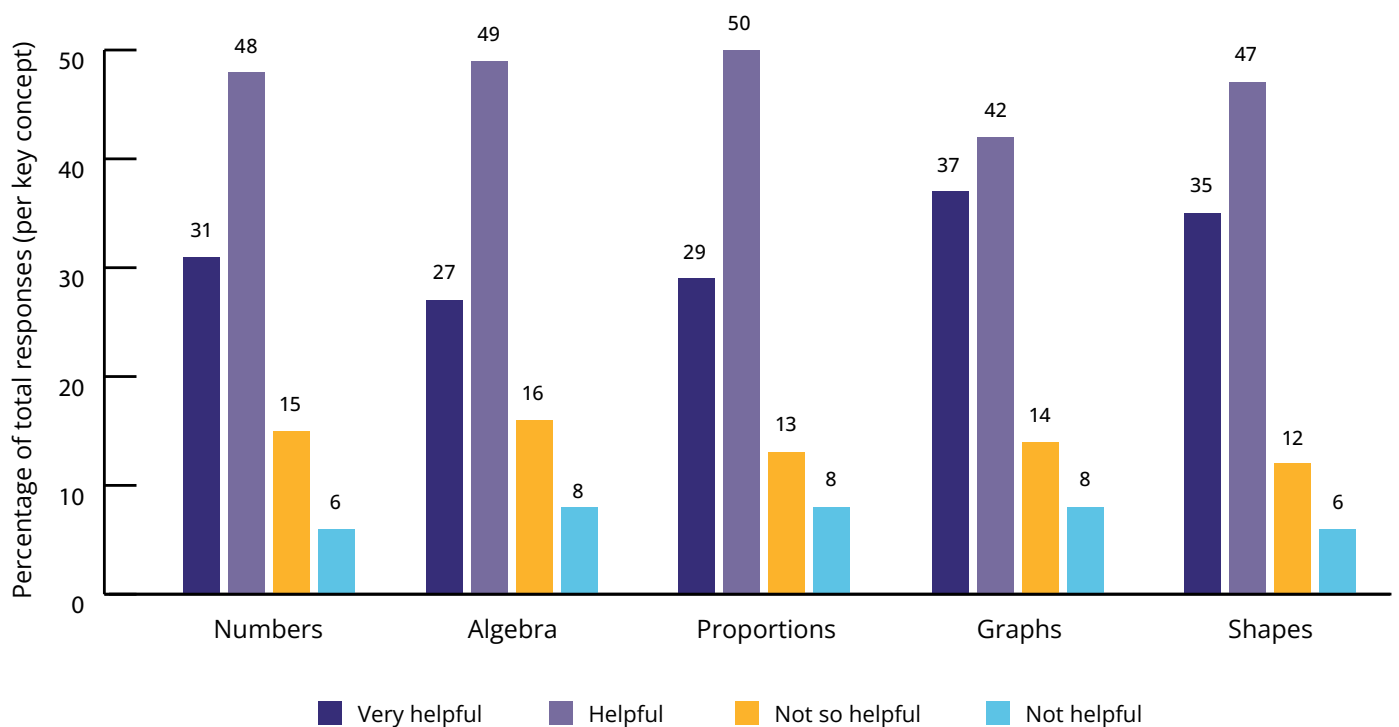

Figure 2: Students' ratings of how useful the tests would be for their revision (survey data)

Students were presented with their results via an interactive screen that allowed them to view their areas of strength and weakness using a RAG (red, amber, green) system. In agreement with the survey findings, all students participating in the focus groups confirmed that they thought these reporting tools would be very useful when planning their revision and identifying areas to focus on.

Students who ranked the tests as being potentially less helpful were asked how they could be improved to make them more useful. The students responded that they would want to have targeted sets of questions addressing their weaknesses as identified through the RAG status reporting.

In response to this feedback, AQA has developed 'Learn' modules. After a student takes part in a test, the Stride platform identifies the learning objectives where the student requires additional practice and generates individualised practice sets (Learn modules) that they can work through to strengthen these areas of weakness. Each time a student completes a Learn module, Stride updates to reflect their improvement and generates additional Learn modules until a student has successfully strengthened all their learning objectives.

Key Finding 3: students found the tests engaging as they adapted to their proficiency levels

In the post-test surveys, students were asked whether they thought the tests were presenting them with the correct balance of challenging and less challenging questions. The majority of students agreed that the balance was appropriate for them and that the different types and levels of questions held their attention throughout the assessment.

Student 3: Seeing our results was really **useful**.

Student 4: Yeah, it was very useful. And it showed what I needed to revise on and what I needed to do next and how to improve.

Students, School N

The focus groups conducted during each testing window were used to explore this issue further. In the first round of focus groups, students flagged that the adaptive engine would sometimes give them a group of challenging questions without a break, which became frustrating for them; they therefore asked for a better mix of challenging and less challenging questions.

At the same time, other students said they received a string of less challenging questions, which negatively impacted their engagement; they requested that more challenging questions be mixed in. In response to this feedback, the adaptive engine was modified so that students would receive a better mix of questions that would be more or less challenging depending on their proficiency levels.

Following these modifications to the adaptive engine, students who participated in subsequent focus groups reported that the balance of questions was correct. Perhaps more importantly, the way in which the questions adapted to their proficiency levels made them feel adequately challenged and, as they progressed, positively rewarded.

> "Yeah, I feel like this test helps people more because then not everything's like easy. They're not just getting 100% every time. It helps them, like, actually learn things."
>
> **Student, School P**

> "I felt that when I did the AQA test, when I answered the question correct, it felt like kind of rewarding. It's interesting because it's measuring you as you're going along. You feel more of like a pressure and commitment to get questions right."
>
> **Student, School G**

Conclusions

The main purpose of this research was to ensure that students would have a positive and engaging experience when taking the diagnostic tests and, importantly, that these tests would be helpful for their learning and revision. Recruitment for the study was challenging due to time and resource constraints in schools. However, any issues were mitigated through the study design.

Analysis of the survey and focus group data revealed three important findings:

- The majority of students (on average, 77% of survey respondents) enjoyed taking part in the tests and found them to be engaging.

- Most students (on average, 79% of survey respondents) reported that the results from the tests, ie the RAG status of strengths and weaknesses, would be helpful for them to focus their revision.

- The iterative improvements that have been made to the tests have resulted in students finding that there was a good balance of questions adapted to their proficiency levels, ranging from more to less challenging.

𝕏 **@AQA_Research**

aqa.org.uk/about-us/our-research

research@aqa.org.uk

G02587_01b

AQA
Questions matter