Text-to-Speech (Polly) Innovation Trial Evaluation and Closure Report

NCEA Online Programme

Version 1.0

Table of Contents

1. Innovation Trials and their place in the NCEA Online Programme....4

2. Trial overview....4

3. Trial objectives....5

4. Trial Summary....5

4.1 Background....5

4.2 The trial process....6

4.2.1 Phase 1: Initiation and confirmation of approach....7

4.2.2 Phase 2: Recruitment of schools and students....7

4.2.3 Phase 3: Discovery, interviews, testing and surveying....8

4.2.4 Phase 4: Analysis and reporting....9

4.3 Outcome of the trial....9

4.4 Findings....10

4.4.2 Functionality....13

4.4.3 Speed of voice....14

4.4.4 Breaks, punctuation and poetry....14

4.4.5 Familiarity....15

4.4.6 Visual tracking of voice....15

4.4.7 Availability....16

4.4.8 Speech to text....16

4.4.9 Expected uptake....16

4.4.10 Te reo Māori and Pacific Island languages....17

4.4.11 Other languages....18

4.4.12 Subjects and standards Polly can be used with....18

4.4.13 User stories....19

4.5 Evaluation of Text-to-Speech trial....19

4.5.1 Evidence gathered in support of this evaluation....19

4.5.2 Using this evaluation to inform the trial's recommendations....20

5. Evaluation of the trial process....23

5.1 Evaluation of trial processes....23

5.1.1 Was the testing adequate and capable of telling us what we need to know?....23

5.1.2 Did the trial answer our questions?....23

5.1.3 Did the trial test what we set out to test for Māori/Pacific/all learners?....23

5.1.4 Could we have better identified aspects worth testing for Māori/Pacific/all learners?....24

5.1.5 Is the evidence robust enough to support reliable decision-making?....24

5.1.6 Were those involved in the trial well enough prepared to carry out their trial roles effectively? 24

5.1.7 Did we realise the anticipated benefits? 25
5.1.8 Did the trial realise any other benefits or identify anything that could come from further work in this area? 25
5.2 Performance against plan 25
5.2.1 Review of objectives 25
5.2.2 Review of benefits 25
6. Lessons learned 26

Text-to-Speech (Polly) Innovation Trial Evaluation and Closure Report

1. Innovation Trials and their place in the NCEA Online Programme

The work of innovation trials is informed by NZQA's strategic thinking as expressed in the Digital Assessment Vision[1], Te Kōkiritanga, NZQA's Action Plan for Ākonga Māori Success, and Takiala Pasifika, NZQA's Action Plan for Pacific Learner Success. The two action plans share a common vision – that learners (ākonga Māori and Pacific learners) will 'qualify for the future world'.

In line with that vision, the NCEA Online Programme prioritises work that aims to optimise the examination experience for Māori and Pacific learners.

All data collected in innovation trials must be able to be disaggregated to enable NZQA to identify and learn about the needs or preferences of Māori, Pacific and all learners. To this end, innovation trials will collect data that enables identification of student ethnicity. Inclusion and publication of ethnicity data will be suppressed for privacy reasons where student numbers are below a level where this is appropriate.

2. Trial overview

Students who need (or benefit from) reading support in assessments, or have a preference for aural learning, are not able to listen to assessment text being read aloud in digital assessments. This includes students who:

1. are entitled to Special Assessment Conditions (SAC) and use or are entitled to use a reader in an assessment
2. are not entitled to or have not applied for SAC but require support with reading, or
3. have a preference for aural learning and are better able to decode text by listening.

Some students will prefer to use their own text-to-speech or screen reader application(s), but NZQA is not currently able to support students to use their own assistive technology in digital assessments until further investigation and testing with the digital assessment platform (Assessment Master) has been completed. It is likely some applications will not be compatible with our assessment platform. Additionally, some students will not have access to an appropriate application. By providing access to text-to-speech through our assessment platform, we can provide all students sitting NCEA Online exams with the opportunity to use text-to-speech technology. This will give students greater control over how they interact with the assessment.

Assessment Master comes with a built-in text-to-speech tool – Polly. This was evaluated in a trial run by the School Quality Assurance Liaison (SQAL) team in 2020. The findings of the trial included concerns about the speed of the reader and the reader's tone and accent. Students also noted

[1] https://www.nzqa.govt.nz/about-us/future-state/digital-assessment-vision/

issues with pausing, punctuation and decoding of some texts such as poetry, diagrams and tables. With practice, it seemed most students were able to adequately use the function.

In 2021 the Digital Assessment and Transformation team (DAT) undertook a second in-depth trial to investigate if Polly will meet student needs for text-to-speech in assessments, or if we need to investigate and trial another tool.

DAT worked closely with students, their teachers and parents to:

- understand their exam experiences
- understand what they need from text-to-speech functionality in external assessments and their experiences with text-to-speech functionality and other assistive technology in the classroom and internal assessments
- conduct user testing with Polly.

3. Trial objectives

The objectives of the 2021 text-to-speech trial were to:

- understand the requirements of students who require reading support in external assessments
- test if students' requirements for text-to-speech are met by trialling Polly or an alternative text-to speech tool
- investigate whether providing text-to-speech functionality improves the student user assessment experience of Māori, Pacific and all others.

4. Trial Summary

4.1 Background

Students who require support with reading are currently supported in NCEA external assessments with readers through SAC. While the use of readers is an effective and widely used method of support, it relies on the:

- student being entitled to SAC
- school supporting the student by making an application for SAC
- availability of readers and writers.

Students who use assistive technology, particularly text-to-speech or screen readers in the classroom and in internal assessments are unable to use this technology in external assessments as it is not currently supported by the digital exam platform. Some students who use technology for support in the classroom may not be entitled to SAC due to having a preference for aural learning rather than a learning or physical disability or having a learning difference not considered significant enough to be entitled to SAC.

Giving students the opportunity to use text-to-speech in an assessment will:

- increase the accessibility of NCEA Online
- support students to meaningfully engage with the assessment by removing the burden of having to read in a high-pressure environment

- provide them with independence and more control over their experience.

Assessment Master, the digital exam platform used by NZQA to deliver NCEA Online, has a built-in text-to-speech function 'Polly'. Polly is an Amazon Web Services (AWS) service widely used internationally for providing synthesised speech. When Polly is turned on in an assessment, an option to read the text aloud is displayed when text is highlighted on the screen. This can be text that is part of the assessment such as instructions, resources, questions or text the student has entered as part of their response. NZQA currently has Polly turned off pending the results of this investigation into student experience of testing Polly.

In 2020 SQAL asked schools with SAC provision for reader and / or writer to test Polly and complete a short survey. The findings were that:

- 78 percent of the students found the facility useful or very useful
- only 25% (7) of these 28 students said they would use it for digital exams, with 61% (17) of these students using a mix of the facility and a human reader
- most reported finding the voice difficult to listen to and were not happy with the speed the text was read at
- a lack of pauses to reflect punctuation or line breaks (in poems) was also reported.

A second trial was initiated in 2021 with DAT working closely with students and SENCOs to test Polly and explore the student experience of using Polly. This second trial allowed NZQA to build on the findings from the first trial by in-depth investigation that included observing students testing Polly and interviews with students.

DAT initially observed students testing Polly during school visits. Empathy interviews were used to find out as much as possible about students' experience of using Polly, sitting exams and learning. Group discussions were also used to test and discuss Polly.

The Covid outbreak in August of 2021 meant this approach was no longer feasible and school visits were cancelled. The approach changed at this point with students testing Polly remotely and completing the student interview online.

Findings of the trial indicated that changes need to be made to Polly for it to meet the requirements of most students. Most students reported issues with the voice and speed of Polly. NZQA currently has control over whether Polly is available for an exam, and what exams Polly is available for, but does not have control over the voice or the speed of the voice. Any changes to voice or speed will need to be made by RM (Aus), the Assessment Master vendor. RM (Aus) does not have capacity to make these changes until later in 2022 and may not be able to complete some of the changes requiring significant development effort until 2023. Improvements to Polly currently scheduled to be delivered in time for end of year exams in 2022 are the:

- ability for the NZQA administrator to turn Polly on and off by student
- ability for the NZQA administrator to change the Polly voice
- visual tracking of the text as it is ready aloud by Polly.

4.2 The trial process

The trial had four phases:

- Phase 1: Initiation and confirmation of approach
- Phase 2: Recruitment of schools and students

- Phase 3: Discovery, interviews, testing and surveying
- Phase 4: Analysis and reporting.

*Due to the 2021 Delta outbreak and lockdowns, phase 2 was repeated after school visits were cancelled.

4.2.1 Phase 1: Initiation and confirmation of approach

In December 2020 members of DAT, SQAL and the editing team (Secondary Exams) met to discuss the results of the 2020 trial and the next steps for trialling the Assessment Master text-to-speech function (Polly). An innovation trial proposal was approved by the Project Steering Committee (PSC) in early 2021. It was agreed to use a customer-centric approach and:

- work directly with students to test Polly
- discuss student experience of text-to-speech in the classroom to gain insights into their exam experience
- identify a preferred voice.

The results of the trial would inform a decision about whether Polly could be offered to students in their external assessments with or without improvements to it.

This approach was combined with the work on the assistive technology research and trial to understand what assistive technology is being used, and if and how students would like to use it, in exams.

4.2.2 Phase 2: Recruitment of schools and students

Polly testing and student interviews were part of a wider programme of school visits organised by the DAT Customer Experience team (CX team) designed to hear student voice across several initiatives and changes planned by NZQA.

To identify schools that may have wanted to participate in school visits, the DAT CX team used:

- data from digital assessments
- SAC data
- previous trial participation data
- expressions of interest from schools that had participated in practice exams
- feedback from SQAL.

Invitations to participate in school visits were sent to schools. Visits, including interviews with SAC students and Special Education Needs Coordinators (SENCOs), were organised with schools that responded positively to the invitation.

One school visit was completed in early August 2021. All other school visits that had been arranged were cancelled when the Level 4 lockdown was announced on the 17th of August. Following the cancellation of school visits, the PSC accepted a proposal to use existing contacts to identify and recruit students to participate in the trial. Contacts included contacts at:

- the Ministry of Education
- secondary schools
- the Dyslexia Foundation
- NZQA.

Students would test Polly remotely, and then would be interviewed online or by phone to discuss their feedback on Polly and exam and learning experiences. To avoid undue pressure on students sitting exams and to gain insights from students who would be sitting exams in the future, students not sitting exams including students in years 9 and 10 were invited to complete testing.

Students were recruited using the following channels:

- emails to Resource Teachers – Learning and Behaviour (RTLB) Cluster Managers asking if any of the schools in their area would be interested in participating in the trial
- direct contact with existing contacts such as SENCOs and teachers
- the Dyslexia Foundation offered to post a request from NZQA on Facebook. Students or their parents contacted NZQA and expressed interest in participating
- NZQA employees with contacts eligible to participate in the trial asked those contacts if they would be interested in participating.

Using this approach 63 potential contacts were identified. 21 of the 63 contacts completed the trial and provided NZQA with feedback. Many of the contacts initially expressed interest in the trial, but found they were unable to participate due to end of year demands on their time.

This approach provided a range of students across different demographic groups.

Students interviewed were:

- predominantly year 9 and 10 but also years' 11, 12 and 13
- from co-ed, single sex, state and state-integrated schools
- from a range of schools between decile 2 and 10
- from a range of ethnicities including Māori and Pacific peoples.

Students with Dyslexia, low vision, Attention Deficit Hyperactivity Disorder (ADHD), Autism Spectrum Disorder (ASD), anxiety and Dyspraxia participated in the trial.

4.2.3 Phase 3: Discovery, interviews, testing and surveying

Instructions for testing were sent to SENCOs and other teachers or school staff and students or their parents. The instructions included a link to the Assessment Master testing environment and access to an AWS voice generator where they could listen to the different voice options available with Polly. When students had completed testing, they or a parent or teacher contacted us to let us know they were ready to be interviewed.

Interviews were conducted with students, parents and teachers. Some students felt uncomfortable discussing their feedback directly with NZQA so it was more appropriate to discuss the feedback directly with a person they trusted and were familiar with and who had observed the testing, such as the parent or teacher.

As we were using empathy interviews, the interviews involved using broad and open-ended questions to prompt full and open feedback. Students were asked the following questions as a starting point (questions were not limited to the following):

- What they generally thought about using Polly. What they liked and disliked.
- What we could do to improve Polly.
- What aspects of the exam Polly worked well in, and where it didn't work well.
- If they would use Polly in an exam, and how they would like to use it.
- What we would need to do before they would use Polly in an exam.
- Whether they would like to use Polly with or without a reader / writer.

- What voices they liked and why they liked or didn't like them. They were also asked to give each voice a score out of five.
- What assistive technology they used in the classroom and how they found using it. Would they like to use that assistive technology in an assessment?
- If they had sat exams, what sort of experience they had sitting the exam.
- If they used a reader/writer, what sort of experience they had with the reader/writer.

Students and parents who were unable to complete testing or organise a time for an interview before Christmas were asked to complete a survey. In total, three students completed the survey.

It became evident before testing was complete that most students required some improvements be made to Polly to optimise it for their use. Improvements were, for the most part, related to being able to control the speed of the voice, changing the voice to one students preferred, and being able to visually track where Polly was in the text.

4.2.3.1 Personas and journey maps

Prior to interviews with students, we created a series of personas and journey maps to represent students who would benefit from using Polly. These personas were verified with SQAL, and then with students once interviews had taken place. Personas and journey maps help us understand the current journey of students and to identify opportunities where we can make improvements that add value for those students.

4.2.4 Phase 4: Analysis and reporting

As empathy interviews were used to elicit feedback on the testing, most of the results from the trial are based on observations and are qualitative rather than quantitative. Numbers of students or an indication of the general response is given where possible. Quantitative data is provided for the feedback on Polly voices as students were asked to give each voice a ranking out of five.

A breakdown of students by demographic has not been provided because the low numbers interviewed could lead to students being identified by their demographic information.

A total of 21 students were interviewed, not including students in group discussions. Two focus groups in total tested and discussed Polly during school visits but other students provided casual feedback on text-to-speech during unstructured discussions.

Scores from the trialling of the Polly voices were calculated. See the results of voice testing for more information.

Key findings relating to potential improvements to Polly have been written as user stories for development by RM (Aus) expected to be completed in the second half of 2022.

4.3 Outcome of the trial

This trial provided NZQA with further opportunity to understand the experience students have with Polly and whether it could be used to support them to sit digital exams, and what improvements need to be made, if any. We were able to build on the findings from the 2020 trial. We were also able to explore in more depth the areas identified in the 2020 trial.

The trial clearly identified key areas for improvement to Polly and the findings provide evidence that the PSC can use to decide the next steps for piloting, improving and implementing Polly.

While it will never be possible to provide a perfect text-to-speech tool that suits the preferences of all students (via our digital exam platform), most students (approximately 70%) indicated that they probably wouldn't use Polly in an exam in its current state. A smaller number of students (approximately 30%) said they would use Polly in an exam in its current state.

There are other factors that will influence whether a student chooses to use Polly or not, such as the teacher or SENCO view and the increase in willingness to use Polly as it becomes more familiar over time.

4.4 Findings

Findings presented are the results of empathy interviews with students or their parents and teachers. As open-ended questions were used in the interviews, findings are qualitative rather than quantitative data. The survey provides some quantitative data, but this is limited as only three students completed the survey.

Where possible, numbers of students have been given, but for the most part, general statements about the response to Polly or a particular aspect of Polly has been given. The only exception to this is the results of voice testing where students were asked to give each voice a score out of five.

Voice
Students were asked to listen to the following voices available within Polly and give each voice a score out of 5:

- English accent female (Amy)
- English accent male (Brian)
- US accent female (Salli)
- US accent male (Matthew)
- Australian accent female (Nicole)
- Australian accent male (Russel) – current default setting.

Overall, students preferred Brian, with Amy a close second. Since completing testing of voices with students a New Zealand voice 'Aria' has been released by Amazon Web Services. Aria includes several common te reo Māori phrases and kupu. Aria was not tested as we were not aware of her availability when testing commenced.

General feedback about the voices included comments about the robotic nature of the synthesised voices with some students saying they didn't like to listen to electronic or synthesised voices:

> *I prefer voices that don't sound too electronic and can be understood.*
>
> *Both students are on the spectrum (ASD) and not normally good with interpreting emotion but both commented that the voices lacked emotions and were hard to engage with.*
>
> *All voices monotonous with not much variation in pitch and pace.*
>
> *Both students didn't like the computer-generated voice so an audio recording of someone reading aloud the text that they could stop and start would be better.*
>
> *The voice was too monotonous.*

Some students who had experience using other text-to-speech solutions or screenreaders remarked that the Polly voices were not robotic:

> *Polly voices are quite natural compared to the voices used by JAWS which is much more synthesised.*
>
> *I was surprisingly impressed. I like that it is not robotic.*

Students commented on the accents of the Polly voices with several stating they preferred the British voices over the Australian and US voices:

> *British accents are far easier to understand.*
>
> *Found English voices better, read the punctuation better. Easier to listen to at that speed.*
>
> *Enunciation of words much clearer with English accents.*
>
> *All the voices sound like they have accents.*
>
> *There is a risk to misinterpret some things because of accent.*

Familiarity with the accent was noted as a factor influencing which voices students preferred, with students or their parents or teachers commenting that they are used to US and British accents from watching YouTube and other content channels and some students listen to audio books which sets their expectations in terms of what accents are acceptable.

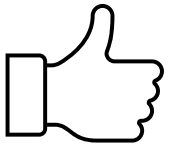
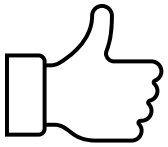
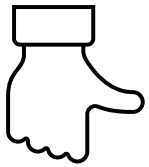
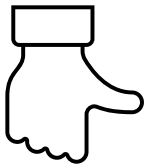
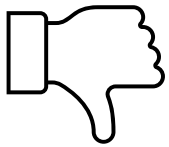
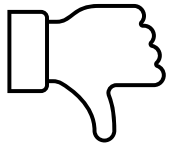
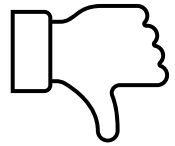
4.4.1.1 Voice testing results

Please see the diagram on the following page for an overview of testing results.

Feedback was mixed across all voices with even the top scoring voices receiving a mix of positive and negative scores and comments.

Brian, a male British accent, was the highest scoring voice. Amy, a female British accent, was the second highest scoring voice. The third highest scoring voices were Matthew with a United States male accent and Salli with a United States female accent. The feedback about the accents indicates that Brian and Amy were liked for their familiarity but also because the British accents provided clearer enunciation of words.

4.4.1.2 *Voice testing results overview*

Brian	Amy	Matthew	Salli	Joanna	Nicole	Russell
British Male	British Female	US Male	US Female	US Female	Australian Female	Australian Male
						
3.1	**3.0**	**2.9**	**2.9**	**2.3**	**2.3**	**2.0**
Not as robotic	Nice and clear and maybe slower.	Sounds too robotic and mean	**Very funny**	Too fast and robotic.	Way too robot	Sounds a little weird
Very mono Sounds like he works with Thomas the Tank Engine	Wheezy	Aggressive	Nice and clear and slower.	Quite fast and sounds like a computerised voice.	Nice, younger	Nice and clear and slower.
Like a robot off YouTube	I like listening to the British accent.	Quite natural and easy to listen to	Better but a bit airy. More breaks, easier to listen to	Great accent, nice tone, ok speed	Clearer, more gaps between the words. Doesn't even sound Australian	Very jarring to listen too, and robotic. It would annoy me.
Didn't like accent. Sounds a bit pompous!	Could be hard to listen to for long periods of time but more fun, entertaining and engaging	Hard to understand	Voice is too soft and dainty	Sounds more human	Sounds like a news reporter.	Sounds quite natural and is the clearest voice
Sounds like Peppa Pig's Dad	Didn't like the accent. Sounds a bit posh.	Unusual pauses and a little fast	Clear to listen to and less American sounding	Accent is annoying	Unusual pronunciation	Very Australian.
I like him he sounds like the Google guy.	Pleasant to listen to	He reads fast and has a harsh voice	Weird and fairly squeaky and bouncy (up and down) -annoying.	Great accent, nice tone, ok speed	She says it syllable by syllable	Robotic and monotone. Hard to understand

4.4.2 Functionality

When asked, students were generally positive about Polly's functionality:

> *It was cool.*
>
> *There was nothing I didn't like about it.*

Polly was considered easy to use by some students who liked being able to highlight the text and click on play:

> *Highlighting text was easy, simple to use.*
>
> *I liked how easy it is to get it to use text.*
>
> *Was easy to go back and listen again if needed.*
>
> *Found it easy to access and intuitive to use.*
>
> *Liked the highlighting and having control.*
>
> *Handy to be able to read aloud text she'd written. She was able to find mistakes.*
>
> *Liked being able to highlight and play.*
>
> *It helps me in the exam as does a reader/writer does and might let me access parts I want re-read quicker.*

Some students found Polly difficult to use.

> *Had to make sure didn't highlight too much at once. Highlighted everything and it read out words he didn't want to listen to.*
>
> *Found it hard to get it to work, couldn't always see the TTS bar when highlighting. (maybe a browser issue?).*
>
> *Biggest issue was scrolling up instead of scrolling down. The toolbar in the middle of the text was very irritating.*
>
> *Using Microsoft Edge and a mouse pad. I couldn't always get the box to pop up when highlighting the text. Sometimes it took three attempts.*

Some students weren't clear how Polly worked with the highlighter and notes tool:

> *Can you use highlight and notes and Polly at the same time?*

There were comments about the floating toolbar:

> *Toolbar quite often gets in the way of the text that is being read out – should sit off to the side so the user can follow along. Many of us scroll up as opposed to down and if you select text by scrolling up the toolbar will consistently sit on top of the text being read.*
>
> *The Stop and Close buttons on the toolbar ultimately do the same thing – would be good if it continued to read if you closed the toolbar.*

4.4.3 Speed of voice

Most students were not comfortable with the default speed the text was read at.

> *Found Polly too slow and she struggled to comprehend what was being read.*
>
> *Pace the biggest issue but was better with some voices than others.*
>
> *If you increase the pace it sounds a bit more normal and less robotic.*

Some students said the pace the text was read at was too fast and others found it too slow. Two students reported they were happy with the pace:

> *Was at a good speed, slower than I read.*
>
> *Pace was good. I could follow it.*

Several students commented they would like to control the speed:

> *Ability to speed up and slow down would be helpful.*
>
> *Options to slow down and speed up the voice rate would be good.*
>
> *Would like to choose the speed.*
>
> *Couldn't change pace which threw me off.*

Some students said they wouldn't use Polly until the issue with the speed was resolved. As there was a mixed response to the pace, with some students saying it was too fast or too slow, a solution could be to give students control of the speed the text is read at.

4.4.4 Breaks, punctuation and poetry

Many students commented on Polly's inability to add breaks and pauses in the correct places in the text. This was particularly an issue for poetry where Polly did not recognise the line breaks. It should be noted that editors have the ability to check Polly's behaviour and may be able to modify the set up of an exam so that Polly reads the text more naturally.

This was an area where students commented that a human reader had a distinct advantage over a digital solution:

> *Readers can't give answers but can read text as it's meant to be read, putting breaks and pauses at the right points.*
>
> *Didn't recognise poetry format, just ran through it.*
>
> *Raced through poetry, no punctuation, monotone, no expression. Needed to have something to give the feel of the poem. Didn't have the flow.*
>
> *Particular instances where the voice seems very fast as there is no full stop between the heading and the question, so it doesn't take a pause. The inflection and pauses at commas and full stops is really good though.*
> *Being able to add pauses between words is even more useful than slowing down and speeding up. Readers are trained to pause between words so student can process what has been said.*

Other students commented that the way Polly read the text was very natural:

It was fluent with punctuation. The meaning was clear and it was as if you had read it in your own head.

4.4.5 Familiarity

Having familiarity with Polly before using it in an assessment was key to using it for many students:

Familiarity is the biggest issue with Polly. It would be easier to use once he is familiar with it.

People would need to practice with it.

Familiarity is the key. If he is using the technology at school then it will be easy to use in an exam.

The most important thing is to be familiar with new tools before an exam so it's not confusing and I don't have to wade through instructions during the exam.

Needed to know what to do or would get stuck. Wouldn't want to go into an exam for the first time without using it before hand.

It will be very helpful in exams if I'm familiar with it before going into the exams.

Those students who had used other text-to-speech tools or screenreaders generally found Polly easier to use.

4.4.5.1 Similarity

Students were asked if they were using any other assistive technology, particularly text-to-speech tools or screen readers to support them in the classroom or internal assessments. Many students did not use or have access to text-to-speech tools or screenreaders. Some who were year 9 or 10 were waiting for an assessment at the beginning of year 11 (2022) and depending on the results of that assessment would be provided with appropriate assistive technology to support their learning. Of those who were currently using some sort of assistive technology, most agreed that Polly's functionality had some limitations but was like tools they were already using.

Polly is similar functionality to what they are used to. (SENCO).

if the pause and speed functions were added it would be close to Word Q in functionality.

it's not as good as my screenreader, probably has about a 80% recognition rate.

4.4.6 Visual tracking of voice

In its current set up in Assessment Master, Polly does not provide any visual tracking on the screen as it reads the highlighted text aloud. This is a feature most other text-to-speech tools provide. Highlighting each word as it is read aloud means the student knows where they are in the text. This is particularly useful if they stop the text-to-speech. For students who need support with reading, identifying where the text-to-speech tool was in the text when it was stopped can be very difficult, and with the pressure of any exam, also stressful.

It is likely a developer can add this functionality to Polly.

Several students commented that not being able to visually identify where in the text Polly was, was an issue:

> *Would be good if each word was highlighted as it was read – much easier to follow along.*
>
> *Having the word highlighted as it is read out is important. It lets you know where you are in the text.*
>
> *Couldn't tell where she was in the text as there is no tracking. Very hard to find the place when stopping the text-to-speech.*

4.4.7 Availability

Groups of students who did not require support with reading and weren't likely to use Polly commented on its availability in exams. The general consensus was that Polly should only be available for those who want to use it as it can't be turned on and off by the user and will display every time text is highlighted (unless the highlighter and notes function has been turned on).

> *Don't turn on Polly for everything. People will mess with it. Also I highlight to read.*
>
> *Don't want to put it on by accident.*

4.4.8 Speech to text

About half the students interviewed noted they really wanted access to speech-to-text rather than text-to-speech. Dependent on the outcome of testing, students may be able to bring their own speech to text tools into exams in late 2022.

> *I would definitely use text to speech as an option. Even better would be speech to text cos I can't spell.*

4.4.9 Expected uptake

Most students reported they would use Polly in an assessment but only if improvements were made to it. The main improvements were being able to control the speed of the voice and replacement of the current voice with one of the preferred options.

> *Would much rather have this in an exam that not have the option.*
>
> Some students said using Polly would just be a matter of getting used to it
>
> *Found using it was trial and error, not complicated once got used to it.*

One SENCO reported that after testing with a group of students with Autism Spectrum Disorder, that if anything went wrong they would quickly give up and the ECM would need to fix any issues very quickly.

Others liked it because it improved accessibility in assessments:

> *Text-to-speech evens out the playing field.*

4.4.9.1 Use with or without reader

When asked if they would want to use Polly with or without a reader, most responded they would still like to have their reader in their assessment, especially as many also required a writer. Some said they would use Polly for reading and the writer for writing.

> *I'd like to use it in an exam if all the glitches were fixed. I would still use a writer for writing.*
>
> *Yes would like to use a reader and writer to write for him in exams but would like to use Polly to read.*
>
> *I would use it in an exam. I would use a writer if no speech to text tool but would use the TTS instead of the reader.*
> *Students prefer to use a computer over a reader - more tech savvy.*

Others said that over time they expected they'd get used to Polly and not need the reader.

Some said that if given the choice, they would opt for the reader:

> *If they had the choice of a reader / writer or TTS they would choose the reader / writer. With the Reader / writer you get variation in tone.*
>
> *I'd opt for a person not a computer generated voice.*
>
> Several students said they preferred the idea of using Polly as it meant they were not singled out and could sit the exam in the main exam centre with other students:
>
> *He feels a bit weird about being different and singled out. Even with the c=pen which he likes using he will probably stop using it because of the stigma of using it.*

Others stated they would prefer Polly over a reader

> *I would use it in an exam. I would use a writer if no speech to text tool but would use the TTS instead of the reader.*
>
> *Much prefer to use this software over a reader/writer.*

SENCOs noted issues with the availability of readers and logistical issues with organising them:

> *Finding readers and writers is hard.*
>
> *Easier to make SAC computer use application than a reader application.*
>
> *It is hard to provide students with a reader/signer who hasn't supported them in the classroom.*

4.4.10 Te reo Māori and Pacific Island languages

Only two of the 21 students who tested Polly were studying te reo Māori. Both commented that pronunciation of te reo Māori kupu was very poor.

Other comments about the pronunciation of kupu from te reo Māori and Pacific Island languages were:

It needs to pronounce words correctly in te reo Māori.

Voices couldn't handle Māori words.

One student pointed out that she couldn't read well enough to identify any mispronounced words in the text so would miss those words if the word could not be identified via Polly:

Pronunciation of Māori and Pacific words is a big issue. They are unrecognisable.

While we can recommend that Polly is not offered for any te reo Māori assessment, te reo Māori kupu are frequently used as part of everyday New Zealand English speech and are prevalent in assessment texts. Students using Polly will encounter these kupu and will be reliant on recognising them by either interpreting the Polly pronunciation or being able to read the word in the text. Polly includes the capability for the NZQA editor to add a phonetic pronunciation guide for non-English words including te reo Māori kupu. These individual pronunciation guides can be collated into a lexicon of common words.

However, there are currently no products on the market that can consistently pronounce te reo Māori correctly. Te Hiku Media have been engaged to develop a text-to-speech solution that supports te reo Māori by 30 June 2022. This will then be integrated with Assessment Master.

Any kupu that are part of the exam text will need to be added to Polly by the editors, but we will not be able to add every kupu that may be used by a student as part of their exam response.

Words from Pacific Island languages have some prevalence in New Zealand English, but prevalence is not as high as te reo Māori. Similarly, NZQA editors can add words from Pacific Island languages to a common lexicon but will not be able to account for every word that may be used by a student in their exam response.

4.4.11 Other languages

Polly supports other languages. None of the students interviewed were studying other languages with the exception of te reo Māori.

Making Polly available to support other languages is an option but would require using voices specific to that language for the exam (for example French). This has not been identified as a need but if it is in the future NZQA could work with RM (Aus) to modify Polly to manage this scenario.

4.4.12 Subjects and standards Polly can be used with

Polly was tested with a Level 1 unfamiliar text English standard, 90851. This standard was selected as it provides a large section of text that can be tested. Students were encouraged to add their own text to the response boxes and test Polly reading different sections of text aloud.

The general conclusion is that Polly would be effective on any text-based assessment. When subjects that use graphs, numbers, formula and other text devices come are made available in the future, Polly will need to be revaluated for its appropriateness for these subjects.

4.4.13 User stories

User stories have been created for improvements to Polly that were identified by a significant number of students. The Assessment Master vendor RM (Aus) will work with NZQA to prioritise and scope the improvements for development. Development is expected in the second half of 2022 with any changes tested and released for students to use in the end of year assessments. Priority of the user stories is based on feedback from students.

Improvements include changing the voice to a voice preferred by students, giving students the ability to control speed and providing the NZQA administrator with more control over Polly, including the ability to change the default voice and speed and turn Polly on and off at the student level.

4.5 Evaluation of Text-to-Speech trial

4.5.1 Evidence gathered in support of this evaluation

Evaluation question	Data collection Method	Source of data
1. Would students be willing to replace their SAC-provided reader with text-to-speech technology use?	Trial survey disaggregated by Māori and Pacific Student interviews	Participating students
2. Would students be willing to use the trialled TTS technology instead of the technology they currently use?	Trial survey disaggregated by Māori and Pacific Student interviews	Participating students
3. What are the limitations of text-to-speech in an NCEA Online setting?	Trial survey disaggregated by Māori and Pacific Student interviews	Participating students
4. What was the experience of Māori and Pacific students using text-to-speech for assessment activities?	Trial survey disaggregated by Māori and Pacific Student Interviews	Participating Māori and Pacific students
5. What are the benefits of using text-to-speech in an NCEA Online assessment activity?	Trial survey disaggregated by Māori and Pacific Student Interviews	Participating students
6. Which group of students realised the most benefits from using text-to-speech in an NCEA Online assessment activity?	Trial survey disaggregated by Māori and Pacific Student Interviews	Participating students

7. What is the impact of text-to-speech on the work of the editor?	Student interviews Editor assessment of text-to-speech trial	Participating students Participating editors
8. Does text-to-speech use improve the student assessment experience?	Trial survey disaggregated by Māori and Pacific Student Interviews	Participating students
9. Does use of text-to-speech improve assessment access and/or outcomes for ESOL students?	Trial survey disaggregated by Māori and Pacific Student Interviews	Participating ESOL students
10. Which assessments is it appropriate to use Polly in?	Trial survey Student interviews Feedback from SRMs and Secondary Exams	Participating students Participating SRMs and Secondary Exams

4.5.2 Using this evaluation to inform the trial's recommendations

The following table examines the project's success in addressing the evaluation questions asked, the criteria or indicators applied and the standards achieved.

Evaluation question	Criteria or indicator	Findings
1. Would students be willing to replace their SAC-provided reader with text-to-speech technology use?	Data reveals students' level of confidence as to whether a SAC-provided reader could be replaced with text-to-speech.	The students who had sat exams and responded and had used a reader were able to state whether they would use text-to-speech independently in a digital exam, use it with the reader or would not use the text-to-speech.
2. Would students be willing to use the trialled text-to-speech technology instead of the technology they currently use or are familiar with?	Data reveals students' level of confidence in trialled text-to-speech technology.	At least half the respondents had used assistive technology in the classroom and/or internal assessments and were able to compare their assistive technology with text-to-speech and discuss what they would prefer to use in a digital assessment.

3. What are the limitations of text-to-speech in an NCEA Online setting?	Data reveals number and seriousness of any limitations of text-to-speech functionality as used in NCEA Online.	All students provided feedback on limitations of the current functionality and indicated whether those limitations would deter them from using Polly. Based on the frequency limitations were reported and the seriousness of the limitations we were able to prioritise improvements to address these limitations for development.
4. What was the experience of Māori and Pacific students using text-to-speech for assessment activities?	Disaggregated data reveals details of the user experience of Māori and Pacific students with trialled text-to-speech technology.	Just over 10% of responses were from students who identified as Māori and there was one response from a student who identified as Pacific Peoples. Responses from Māori and Pacific students were compared with all others. These students noted that the text-to-speech does not pronounce te reo Māori kupu correctly. Students from the category 'others' also noted that pronunciation of te reo Māori was poor. There was otherwise no notable differences in the experience reported by Māori and Pacific students.
5. What are the benefits of using text-to-speech in an NCEA Online assessment activity?	Disaggregated data reveals any benefits identified by Māori, Pacific and all other students from using text-to-speech in NCEA Online assessment activity.	There was no notable differences in the experience reported by Māori and Pacific students.

6. Which group of students realised the most benefits from using text-to-speech in an NCEA Online assessment activity?	Disaggregated data reveals benefits experienced by groups of students when using text-to-speech in an NCEA Online assessment activity	Students from the following groups participated in the trial: Students with dyslexia, students with low vision, students with dyspraxia, students with ASD, students with ADHD, students with anxiety, Māori students, Pacific students. Responses were compared between groups and there were notable differences between students with dyslexia and those with ASD. Students with dyslexia were more likely to use and benefit from Polly as they had less concerns with synthesised voice. We did not interview any students with English as a second language.
7. What are the requirements for editing assessment content when used with text-to-speech?	Any impacts of text-to-speech on assessment content are identified through analysis of student and or editor feedback.	Student feedback identified content and features that did not render well with text-to-speech such as poetry or any other text with line breaks. The Editors have indicated they will be able to make editorial changes to improve the rendering of text-to-speech with poetry.
8. Does text-to-speech use improve the student assessment experience?	Disaggregated data reveals whether students believe using text-to-speech improved their assessment experience.	Students including Māori and Pacific students were able to discuss whether they believed using text-to-speech would improve their assessment experience, providing evidence for recommendations for next steps for this work.
9. Does use of text-to-speech improve NCEA assessment access and/or outcomes for ESOL students?	Data reveals whether ESOL students believe using text-to-speech can improve their access to and outcomes of NCEA assessment.	ESOL students did not participate in the trial so we are not able to evaluate whether ESOL students would benefit from using text-to-speech in an assessment.

10.Which assessments is it appropriate to use Polly in?	Data reveals which assessments and standards students would like to use Polly in. Data reveals which assessments and standards SRMs and Secondary Exams want to make Polly available for.	Students discussed the subjects they prefer to use text-to-speech with. They also provided feedback on their experience of using text-to-speech with different types of text such as poetry vs prose.

5. Evaluation of the trial process

5.1 Evaluation of trial processes

5.1.1 Was the testing adequate and capable of telling us what we need to know?

Despite having to change the approach to recruitment of students due to the Covid-related lockdown and cancellation of school visits, we still gathered adequate evidence to inform a decision about use of text-to-speech in digital assessments.

Two group discussions and four interviews were completed before the lockdown.

Of 63 potential contacts interviewed after the lockdown, 21 completed the trial and provided feedback. Each was interviewed in depth or provided feedback via a survey.

In many cases, parents, teachers or SENCOs were able to observe the testing and provide us with their observations as well as the student's feedback.

Feedback clearly identified aspects of the text-to-speech that were found to be inadequate and require improvement as well as those aspects that students generally liked.

Interviews gave us a clear indication of whether students would like to use Polly in assessments and what recommendations could be made to support this.

5.1.2 Did the trial answer our questions?

The trial sufficiently answered our questions.

5.1.3 Did the trial test what we set out to test for Māori/Pacific/all learners?

Yes, Māori and Pacific learners were included in the trial and students were asked about their use of other languages including te reo Māori. Several students observed that pronunciation of te reo Māori kupu and words from Pacific languages was poor.

The te reo Māori text-to-speech trial will address these issues and Polly should not be turned on for any assessment that uses te reo Māori for large sections of text.

Improving Polly so it tracks the text on the screen as it is read aloud will help students find any mispronounced te reo Māori kupu in the text.

Māori and Pacific students did not make any distinctive observations about the text-to-speech and feedback largely reflected that of the group as a whole.

5.1.4 Could we have better identified aspects worth testing for Māori/Pacific/all learners?

Modifying Polly to handle te reo Māori and Pacific languages is out of scope for this trial and is part of the scope of the te reo Māori text-to-speech trial.

Polly was initially trialled with a group of Māori students at the only school visit that was completed before lockdown. As this was a group discussion, these students are not included in the total number of students interviewed. These students did not have learning needs that meant they would use Polly and many of them stated they would not use Polly but believed that other students would use Polly. They also gave feedback on the voice.

Students were also asked about use of text-to-speech during a visit to a Kura. Feedback was that all the students in the focus group would use text-to-speech in both English and te reo Māori but they would find speech to text even more useful.

Due to changing the approach to recruiting students it was not possible to specifically recruit Māori and Pacific students and/or it was not always appropriate to ask students to identify their ethnicity. However, several students identified themselves as Māori or as Pacific peoples during the interview or this was already known to the interviewer.

5.1.5 Is the evidence robust enough to support reliable decision-making?

The findings from the 2021 text-to-speech trial provides a sufficient body of evidence to support reliable decision making.

The students that participated represented a broad cross section of New Zealand students who require support with reading and included Māori and Pacific and other ethnicities. More males participated than females (13 out of 21) but this may be reflective of the tendency that males are more likely than females to be diagnosed with a learning difference such as dyslexia.[2]

Students or their parents were able to articulate their experience of testing Polly, name the voices they preferred to listen to and provide additional information about assistive technology they used in the classroom and internal assessments and their exam experience (if relevant).

The findings clearly demonstrate the improvements that need to be made to Polly to increase the number of students willing to use it in an NCEA Online assessment.

5.1.6 Were those involved in the trial well enough prepared to carry out their trial roles effectively?

Yes. A customer experience approach was used for this trial and this was supported by the CX team. School visits were organised by the CX team. Questions about Polly were included in any school visit completed by the CX team.

[2] https://www.psychologytoday.com/us/blog/raising-readers-writers-and-spellers/201404/are-more-boys-girls-dyslexic

SQAL have supported the trial including with the selection of schools and advice.

5.1.7 Did we realise the anticipated benefits?

The key benefits expected from this trial were, an understanding of:

- whether using text-to-speech would contribute to students' experience of equitable access and outcomes and relevant credentialed learning sustains their ability to succeed in higher education and the workplace
- whether text-to-speech could contribute to sustained confidence in the assessment system as demands of it change.

Students who were part of the trial largely agreed that having text-to-speech available in exam would improve their exam experience and identified changes to the tool that would increase uptake of Polly if it was available in digital exams. Some students noted that having text-to-speech available increased their autonomy during the assessment and ability to fully participate in the exam.

Māori and Pacific students responded positively to the opportunity to have text-to-speech available in assessments. Polly does not pronounce kupu from te reo Māori and Pacific Island languages correctly, however provision of a text-to-speech tool that supports te reo Māori is being managed by another project.

Assistive technology is being regularly used in classrooms and internal assessments. This was validated through discussions with students, teachers and SENCOs. Making Polly available in NCEA online assessments will mean NZQA is keeping pace with the needs of students and sustaining confidence that the assessment system can continue to meet the changing needs of its users. By providing support for students who do not have access to assistive technology, availability of TTS in NCEA exams would contribute to students' experience of equitable access and outcomes and relevant credentialed learning to sustain their ability to succeed in higher education and the workplace.

5.1.8 Did the trial realise any other benefits or identify anything that could come from further work in this area?

No additional benefits were identified.

5.2 Performance against plan

5.2.1 Review of objectives

The objectives of the 2021 text-to-speech trial were satisfied. We were able to work closely with students to test Polly and understand if it was a suitable text-to-speech tool to make available in assessments.

5.2.2 Review of benefits

5.2.2.1 NCEA Online benefits

NCEA Online benefits	This trial's contribution to benefit(s)
Students' experience of equitable access and outcomes and relevant credentialed learning sustains their ability to succeed in higher education and the workplace.	Students confirmed that they were likely to use text-to-speech tools in their assessments and this would increase their independence and ability to participate in the exam. Accessibility is improved for students who do not have access to assistive technology.
Sustained confidence in the assessment system as demands of it change.	Assistive technology is commonly used in the classroom, workforce and at home. Making text-to-speech available in exams means NZQA is offering students an experience that reflects the way they use technology in other environments.
Greater efficiency and enhanced quality assurance.	N/A

5.2.2.2 Other benefits identified for (potential) realisation

No other benefits were identified during the trial.

6. Lessons learned

The following lessons were learned during the lifecycle of the project:

- Some students may find it difficult to articulate or may be reluctant to articulate their experiences. Parents and teachers are sometimes able to advocate for the student, share observations and provide insights into the experience.
- Phone calls are often a less intrusive channel for interviews that students may find more comfortable than face to face on Teams or other video conferencing software.
- While many students or their parents expressed interest in participating in the trial, only a third of that number were able to complete the trial. Ensuring any process to register interest in a trial is broad and captures as many potential contacts as possible is required to support an adequate number of students completing the trial and reliable trial results.