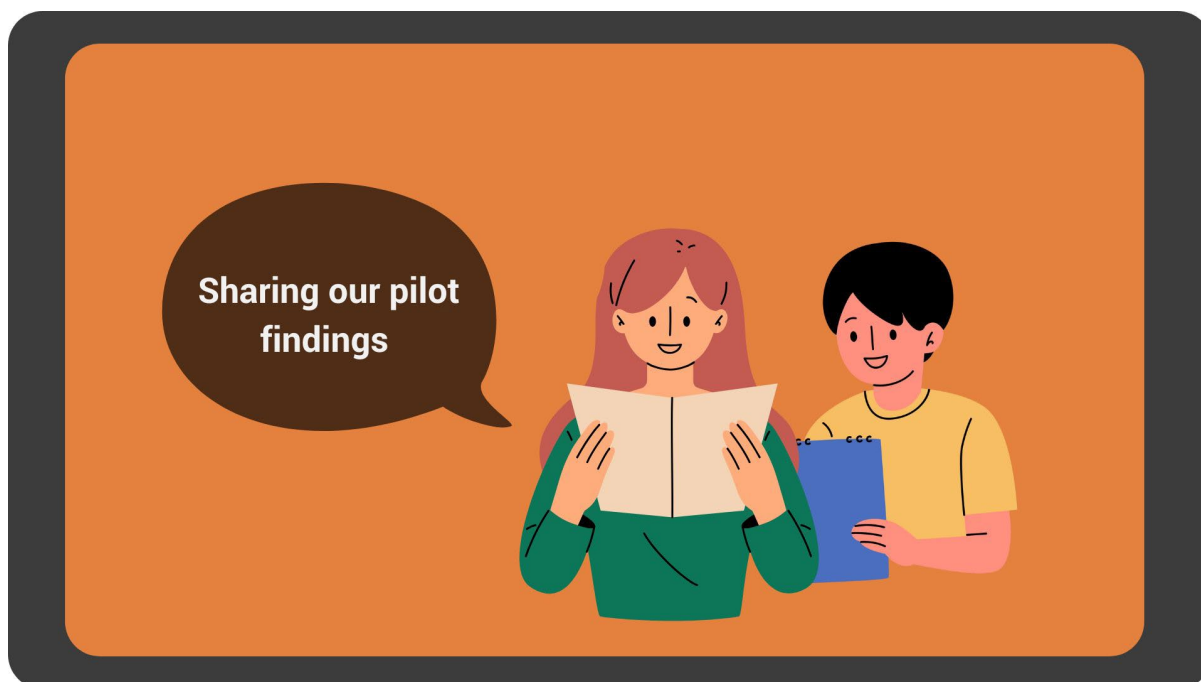


AI for assessment pilots: Five key findings

Article #11 of AI in Education Article Series: July 2025



Earlier this year, Scarlatti developed our own Artificial Intelligence (AI) agent for oral assessment to explore whether it could help improve outcomes for students who are neurodiverse, speak English as a second language, or have learning difficulties. Since then, we've undertaken two pilots, an evaluation and published our [Five key practical lessons](#) for education providers wanting to trial a similar agent. This penultimate article shares our **five key findings** for providers interested in using AI to help in assessment delivery.

This article is the **eleventh in a series titled "AI in Education"**, aimed at education providers interested in AI. The intention is for this series to act as a beginner's guide to the use of AI in education, with a particular focus on AI agents. This series has been developed as part of a project to develop an AI agent for learner oral assessment, funded by the [Food and Fibre Centre of Vocational Excellence](#). We invite you to [get in touch](#) to see a demo or explore how you could use AI agents in your own context.

Recapping our pilots

From March to May 2025, Scarlatti undertook two pilots of our AI agent for oral assessment – one with Fruition Horticulture Limited, and one with Dairy Training Limited:

	Fruition Horticulture Limited	Dairy Training Limited
Course	Hei Whanaake	Contract Milking 101
Assessment	Health and Safety Assessment (level 2)	Contract Milking Assessment (level 5)
No. of students	14	11
Student demographics	Predominantly Māori, but also Samoan and Tongan Aged ~16	Mixed ethnicities Aged ~22 to ~40
Aims of the provider	Inclusive for young Māori learners, illiterate learners + saves tutor time	Inclusive for remote and / or neurodiverse learners + saves tutor time
Number of questions	19	5
Number of learner responses	284 unique learner answers	55 unique learner answers

Below, we share our top 5 findings from these pilots. Given that our agent was a proof-of-concept and involved a small number of people, these findings should be interpreted with care.

Key finding 1: The agent appears to be highly accurate

Across both pilots, there was a **95% match between the preliminary grade given by the agent and the grade given by a human tutor**. We interpret this as high given that human graders themselves would likely have some variation between different teaching staff.

“[It] is likely better than human-to-human grading agreement” (Teaching staff).

“[The agent’s responses] were both appropriate for the question and very helpful” (Learner).

“If they say remove your jewellery, I would mark that as correct [but] there were a couple who said remove your jewellery, but they were marked wrong [by the AI as they didn’t mention ‘apart from your wedding ring’]” (Teaching staff).

Where there were grading mismatches, it was mostly when the AI had graded someone as *not yet competent*, and the tutor disagreed. At first glance, it might seem like the AI is therefore more conservative in its grading. However, these mismatches were often due to fixable inaccuracies in the training data. For example, an assessment rubric that said the learner had to mention a specific word when they in fact did not; or an exemplar learner answer that mentioned an exception, which also was not required.

Key finding 2: The agent appears to be easy to use

The **majority of learners (83%) found the AI agent “easy” or “very easy”** to use. They attributed this to the flexibility of completing the assessment when it suited them, and to it being easier than writing down their answers. This is a positive result, given that this was the first time learners had used the AI agent for assessment.

“Very clear and understands you well” (Teaching staff).

“It was a lot different than what I am used to, but [it] was easy to navigate” (Learner).

“It had me repeating a couple questions that I thought that I had answered pretty well or to the standard it was setting” (Learner).

Where there were challenges, it was sometimes due to learners using the agent for the first time or technical issues (which could likely be fixed in a production version). Other times, it was due to specific features we had piloted but could change in future (e.g., having a follow-up question at the end of the conversation rather than immediately after a learner’s answer).

Key finding 3: The agent appears to be quite enjoyable

A reasonable number of learners (62%) reported that using the agent was “enjoyable” or “very enjoyable”. These learners appeared to find it enjoyable because the agent provided instant feedback and enabled them to ask questions. We see this as a strong result, given that assessments would not typically be enjoyable.

“A few [learners said to me] I don't understand it, and I said, ‘just ask it to repeat the question in a different way’... and it did, and you just watched them go ‘oh, yeah now I get it’” (Teaching staff).

“The instant feedback was great” (Learner).

“We need to talk to people... not computers” (Learner).

Where learners gave feedback that it wasn’t very enjoyable, this was often due to fixable technical challenges or a preference for tutor-run assessments. However, we note that tutor-run oral assessments are often not feasible for providers due to the cost.

Key finding 4: The agent can empower learners

Teaching staff also **observed how AI-run assessments could help empower learners**, particularly in the Fruition pilot. This sense of empowerment was seen amongst young Māori and Pasifika learners. Staff attributed this to the agent making the assessment feel easier and faster to complete, while also allowing learners to ask questions and not feel judged. Together, this helped learners build their confidence in their ability to do well.

“Some of the students seemed upbeat after they’d done it... They were like, oh, this was really easy... It gave them confidence to not be afraid” (Teaching staff).

“Most of our students have been let down by the education system and refuse to ask questions... But AI isn’t going to judge them. It empowers them” (Teaching staff).

“I have a lisp, and [it] wasn’t picking up my words, but after a couple of goes it got it right” (Learner).

However, while our sample is small, there was some evidence to suggest that learners with speech impediments or limited vocabulary may be more likely to be marked incorrectly. If true, this would be an example of bias. This should be investigated further, with mitigations built into any agent.

Key finding 5: The agent will likely save time

Teaching staff also suggested that the agent could save time in administering and grading assessments. For example, within the Fruition pilot, staff calculated the agent to have **saved learners approximately 31 hours**. If it was streamlined into their existing LMS, they estimate that it could have **saved the tutor approximately 20 hours**.¹ These varied depending on the previous format of the assessment and the number of learners in the course. Teaching staff noted that these savings could enable them to spend more time providing learners with one-on-one support.

“Tutors [spend] an incredible amount of time supporting the learners to engage with the paper tasks... they're having to make decisions all the time is to prioritise who their attention goes to and for how long they can do and often learners miss out. [It would be different with an AI agent available to help]” (Teaching staff).

“It was so easy that it repeats and simplifies, I was able to do the assessment in 10 minutes [rather than the allocated 2 hours]” (Learner).

“If the backend [could be developed to be] easy, so I didn’t utilise a lot of time, I [would be] 100% for AI assessment / oral assessment...” (Teaching staff).

However, to maximise time savings, the agent would need to be integrated with the organisation’s learner management system in a way that enables streamlined grading and reporting processes.

Learners' and pilot organisations interest going forward

The majority of learners (88%) **were interested in seeing more AI agents in assessments or were open to using them again after improvements**. This uptake is likely to increase with further demonstrations and technical fixes.

“[Using the agent] is a lot easier and won't make my fingers tired” (Learner).

“The agent is good if you don’t know how to write down the ideas in your head” (Learner).

Both pilot organisations also expressed that they would like to continue using the AI agent to assess their learners.

¹ These estimates are based on using the AI agent for this one 15-minute oral assessment in one course with 1 tutor and 14 learners, and comparing the time used to the time needed when doing this same assessment on paper.

“I would like to see it across all of our assessments this year” (Operations Manager).

“I think that AI will have a massive impact on that kind of equitable resource or support to engage with assessment... The bureaucracy around assessment demands and reporting and documentation is massive, and I can see AI cutting down on that hugely” (Academic Manager).

“I've been thinking lots and lots about [how to incorporate it more] ... If you could combine [AI and something that collects photos or other evidence out on the field] ... so that [the tutor] is out there on the orchard and he just needs to speak to the AI and the AI prompts them to... whack this picture in and now whack this in here and did all of that... That would be a game changer” (Academic Manager).

However, some learners still prefer written assessments, due to personal reasons or discomfort with using AI. This suggests that having an option for learners to write the assessment is still important.

“I prefer writing, just personal preference” (Learner).

Scarlatti's take

These findings show largely positive results overall. This suggests that our AI agent for oral assessment has strong potential and that only minor fixes are needed.

It also makes us wonder how else AI agents could be used in assessment, especially in lower risk use cases. See our next and final article for more on these possibilities.

As we wrap up this project, we invite you to:

- **Reflect** on how AI could support you or your learners
- **Get in touch** with Scarlatti for a demo or an obligation-free chat on AI agents.