

Frequently Asked Questions

Science: Thinking with evidence

In May 2010 we presented our science test, *Science: Thinking with evidence* to teachers around the country. Here we answer in more detail some of the frequent questions. First, two general points to bear in mind:

- i The tests in *Science: Thinking with evidence* (STwE) assess just one aspect of learning in science. Of all the knowledge, skills and dispositional attributes students could develop through science, STwE focuses on “thinking with evidence”. We chose this as an activity that brings together the key competency “thinking” and the “nature of science” strand of the science learning area. Science is an appropriate context in which to develop this competency, and science contexts can develop dimensions of thinking with evidence that might not be as readily developed through other learning areas. STwE was never designed to test all possible learning outcomes in science. The results of the test should not be reported as if it did.
- i The purpose of the tests is primarily formative. The questions we address below mostly point to its use for summative reporting purposes. Because there is a common scale across all four tests, you can measure progress over time, but STwE is primarily designed to identify areas of existing strength and areas where students might need scaffolded support to develop their thinking competencies. How you use the results to determine next learning steps is the main focus of the teacher manual.

Reporting achievement using STwE

Q: Can we link the test results to the curriculum levels? (Our school requires us to report to parents against these.)

A: This question is often asked and not just about science! The practice of reporting against curriculum levels assumes that the achievement objectives describe aspects of progress in sufficient detail that differences between students’ performances can be determined - both across time for each individual (i.e. “progress” made) and between individuals at any one time (i.e. comparative achievement). Recent policy advice to the Ministry of Education, *Directions for Assessment in New Zealand*, had this to say about that assumption:

Some of the assumptions underpinning the national curriculum, particularly those that relate to levels and achievement objectives, continue to be contestable. Analyses show that, by

and large, the objectives do not provide a sufficiently clear basis for discriminating levels of achievement or judging learning progress. If schools feel they must accumulate assessment records and data as evidence of curriculum coverage and student learning, loose criteria and surface coverage can undermine the quality of assessment and the validity of the interpretations and decisions that follow (Absolum, Flockton, Hattie, Hipkins, and Reid, 2009, p.12).

Adding to the general challenge described here, the Nature of Science strand of the curriculum is new and “double-levelled” so there is even less detail than in AOs for other areas. You *can* report each student’s personal progress in thinking with evidence if you use different STwE tests at a suitable time interval because all the tests share a common scale. They are broadly linked to chronological age, which in turn is broadly linked to curriculum levels. But it would be misleading to say that STwE results are indicative of overall science progress at the various curriculum levels. No such tool exists, nor probably could exist.

Q: We are reporting achievement in junior science at achieve, merit and excellence levels so students are familiar with these before they reach NCEA. Can we use the test results to do that?

A: The three levels of any one achievement standard (A/M/E) differentiate between qualitatively different demonstrations of achievement of the same task (at least in theory). Students all do the same thing but some do it much better than others. The nature of this “better” achievement is described in the standard. In STwE students all do the same task (i.e. answer the same test). Some mostly get the “easy” questions correct, some can answer easy and moderately difficult questions correctly, and some can display the mix of competencies needed to get even the more difficult questions correct. The teacher manual describes the characteristics of each item that *in combination* determine how easy or hard it will be. The *position* of each item on the scale reflects this combination of competencies. The STwE scale is continuous whereas achievement standards “chunk” performance into four levels, but the scale does differentiate qualitative performance as just described. To the extent that you wish to make the distinction, you can say that students with lower-scale scores have achieved at a qualitatively lower level of competence and those with the highest scale scores have shown “excellence” in this type of thinking. Be careful that people do understand that this is not just about getting more “right” answers. A/M/E is about the *types* of items you can answer correctly. Also bear in mind that the STwE tests assess just one aspect of learning in science.

Q: Multi-choice questions are not like the NCEA focus on written descriptions (A), explanations (M) or discussions (E). Is this a shift away from NZQA assessment requirements?

A: See the above answer and the summary of the purpose of STwE. NZQA had no part in the development of STwE and it is not intended to signal anything at all about how

NCEA should be managed. Differentiating between a description, an explanation and a discussion is never entirely straightforward even though we might recognise clear examples of each when we see them. A lot of professional judgment goes into recognising the *qualitative* difference in performance when this is determined by the use of a descriptive taxonomy such as describe/explain/discuss. STwE determines difficulty using a *quantitative* process of item response analysis. This is then built into the test scale so that teachers are free to focus on *why* students achieved as they did, without needing to first work through complex judgements about what their level of achievement actually is.

Q. Te Kotahitanga uses the results of asTTle tests to analyse student achievement levels and/or improvement, but there is no “science asTTle”. Can we use STwE for this purpose?

A: Refer to the answer to the question on reporting against curriculum levels. Like asTTle, STwE has a common scale so each student’s improvement over time can be monitored if you keep careful records. It may be possible to link broad patterns of achievement gains to an initiative such as Te Kotahitanga but care is needed in how you interpret such data shifts. Correlations do not necessarily indicate causes. For example, over time we might expect at least some of the gains students show to be linked to our own developing understandings of what we might teach differently now that we have a nature of science strand. The STwE manual supports this aim by providing guidance about how to do just that.

Q: Is a student able to use a reader to complete the STwE tests?

A. You really need to think about what you are going to use the data for. If it is for formative purposes, you obviously already know this student has problems with reading. However, by providing this support you may get a better idea of how well they can use the information.

However, if you want to fairly compare the students’ results with the national data, then being able to interact with the ways that evidence is presented is an important component of this test. In this case it wouldn't be appropriate to have a reader.

Be careful, if you are not “comparing apples with apples” (i.e., a student is supported with a reader while others are not), that this is really clear when you are reporting on the test results.

Practical questions about using STwE

Q. Can we photocopy the student answer sheets and test booklets?

A. No. Copying booklets and answer sheets is a breach of copyright. The test booklets are reusable so you will only need to purchase full sets the first time you use the test, but you will need to continue purchasing packs of answer sheets. Without this ongoing support it would not be viable for us to develop tests such as this.

Q. What support is there for in-depth analysis of each student's response?

A: The manual describes how you can look for patterns in each student's responses and in overall class responses. This is currently something you have to do "by hand". (If support for the test allows us to offer a marking service in the future, this could be done automatically. See below.) In depth analysis does take time so you might prefer to take a group of students who are worrying or interesting rather than doing this for your whole class. The manual shows how to work out "next steps" for your students, which means you can proactively address any learning needs the analysis shows up.

Q: Can our students take the STwE test online?

A: STwE is currently offered electronically on the Marking Service
www.nzcermarking.org.nz

Q. Is a marking service available for STwE?

A: NZCER is a small organisation, so time and resources for these kind of developments need to fit into our current work plan and the priorities we have. If there is enough demand we will consider providing this option in the future. Please email marking@nzcer.org.nz if this would be of benefit to your school.

Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, I. (2009). *Directions for Assessment in New Zealand (DANZ)*. Wellington: Ministry of Education.