Assessment dilemmas when “21st century” learning approaches shift students into unfamiliar terrain

A paper presented at the 35th Annual Conference
International Association for Educational Assessment
Brisbane, Australia
13-18 September 2009

Chris Joyce, Rosemary Hipkins, NZCER

Introduction

21st century learning emphasises a need for students to not only learn “content” but also to *use* their knowledge in specific ways (see for example Gilbert, 2005). This paper discusses one small aspect of this - the ability to ask appropriate questions in the context of a science issue. Being a question *asker* is an unfamiliar role to many students and this poses challenges for assessing their questioning skills. In a traditional assessment event students are primed to be *answerers*. Asking students to develop, or even simply identify, next questions appropriate to a specific situation can be unsettling and, for many students, appears to be an unpractised skill. Students need to learn to be comfortable with this different role, but also to have access to different sorts of questions that presently seem to be unfamiliar to them and to know when to deploy them.

The types of questions discussed in this paper are not particularly difficult to pose *per se* but the evidence we will draw on certainly suggests they are unfamiliar. We discuss the sorts of questions that an inquirer might ask when attempting to understand how more generic or abstract ideas play out in a specific context. This skill, we suggest, is important for the *transfer* of science ideas to contexts where they might help resolve a dilemma or make a decision about what to do next. The whole notion of transfer is a “vexed topic” because it involves a focus on dispositions as well as knowledge and skills (Carr, 2008, p.37). Here we attempt only to tease out one tiny piece of a much larger puzzle.

As is typical of science curricula elsewhere, New Zealand’s curriculum (Ministry of Education, 2007) claims that learning science is necessary for making “informed decisions about the communication, application and implications of science as these relate to [students’] own lives and cultures” (p.28). While this might be an honourable intention, it is far from clear how learning the basic concepts of science will actually help prepare students for a more active role as users of this knowledge in real world contexts. As one curriculum commentator has noted “precious few studies of any sort have tried to connect school science with substantive, real-life outcomes” (Feinstein, 2009, p.763). Again we can only say that out paper could add one small piece to a seemingly new type of puzzle.

Our data sources

The assessment items discussed in this paper are derived from two assessment projects currently being undertaken by the New Zealand Council for Educational Research (NZCER). The two sources of data together provide evidence about what happens when students are asked to either actively generate questions or recognise relevant questions to ask.

The first project, Assessment Resources for Classroom Teachers (ARCT), is a contract NZCER has with the New Zealand Ministry of Education that includes developing assessment resources for formative purposes, and making these available to New Zealand teachers through a dedicated website, the Assessment Resource Banks (ARB). The tasks are designed for teachers to use as part of their everyday classroom activities. There are many different definitions for formative assessment, but the purpose is common to all – to improve student learning. The use of assessment for such purposes, regardless of the nature of the task, poses particular challenges for teachers. Three main actions are required – eliciting student responses, interpreting what these responses might mean in terms of next learning steps, and acting on these insights (Cowie and Bell, 1999). Writing assessment tasks for formative purposes is challenging because assessment cannot be regarded as formative *until* it results in action by the teacher and/or the student (Newton, 2007, Black, Harrison, Lee, Marshall, & Wiliam, 2002). While knowing whether students can provide the "right" answer is useful information, understanding *how* the student got to their answer provides the platform for deciding what to do next. The tasks, therefore, need to be designed to reflect this requirement. We believe these challenges are compounded when the focus is on *using* rather than having knowledge.

ARB resources are intended to support teachers to elicit, interpret and respond to students' learning challenges. Each task is accompanied by educative material to assist teachers and students with analysing student responses and making learning and teaching decisions about next steps. Pen-and-paper ARB resources are trialled nationally as sets of three or four items, often related to each other in some way. Around 200 students from a sample of New Zealand schools complete each task booklet and at the end they are given the opportunity to make a personal response about enjoyment and difficulty levels of items. (Practical tasks are treated differently and will not be further discussed in this paper.) The NZCER science education research team work together to analyse response patterns, first reading and discussing a sample of completed scripts, then building and applying a coding schedule based on the researchers' emerging understanding of the patterns of responses. Wider research projects in specific areas of interest related to formative assessment complement the more formal tasks from time to time.

The second project to which we will refer is the development of a standardised test that assesses students' ability to "think with evidence" in contexts related to issues with a science component. Information in the form of written text, photographs, diagrams, graphs, tables, or a combination of some of these is provided, and students answer multiple choice questions. Different aspects of thinking are assessed, including identifying questions relevant to the information provided. These items have been piloted with small groups, trialled with approximately 250 students nationally, and will undergo national standardisation trials in late July. One of the purposes of the standardised test is to provide direction for teachers in planning their science programme. The teacher manual that will accompany the test will potentially support teachers to use the results formatively, although the overall test is intended to provide a broader picture than individual ARB items ever could.

Asking statement-to-context questions about glaciers

In this section of the paper, we discuss a specific ARB resource, *Glaciers and global warming.* We chose this context to explore a particular set of dilemmas that confront citizens when they try to reconcile science ideas and real-life evidence in relation to complex systems. The issues that our world faces are rarely about simple cause and effect relationships and an important science (and environmental) understanding is that there are systems within systems (for example each glacier is a system in itself, but also interacts with much bigger systems such as the global water cycle and local/regional weather systems). Another important understanding is that anything that happens within a system affects to a greater or lesser degree everything else within that system. If one of science education's goals is to have students who "can participate as critical, informed, and responsible citizens in a society in which science plays a significant role" (Ministry of Education, 2007, p.17), then it is important that they learn to consider not just individual elements and direct relationships, but also the complex interactions that occur within a system. In reality, students are typically taught about individual and seemingly unconnected elements and ideas. Analytical, linear, and reductionist thinking involve understanding things by taking them apart, whereas systems-thinking takes a holistic, big picture view, and this type of thinking is more suited to making decisions about complex issues (Sterling, 2001). Changing long-established traditions in education requires, at the very least, that good exemplars of different ways of working are provided for classroom teachers.

Glaciers are cited by many scientists as being important markers of climate change. Worldwide there is a trend for glaciers to be retreating, and this general pattern is used as evidence of global warming. However, not all glaciers adhere to this pattern, including some of the glaciers on the west coast of the South Island of New Zealand. Commentators who lack a systems understanding might use this seeming contradiction to argue that global warming is not occurring, or that its effects are overstated. Therefore, glaciers provide a useful context for considering whether a theory can be refuted by any one event that doesn't immediately appear to fit the evidence used to support the theory.

The assessment item we designed gives a brief summary of the above information and provides a photograph of the Franz Josef glacier. Students are provided with some general statements about glaciers, and their task is to develop a question specific to understanding how this idea applies in the context of the Franz Josef glacier specifically (refer to Table 1). So, for example, in response to the statement "glaciers in steep, narrow valleys are less affected by small overall temperature changes than wider glaciers" students might shape a questions such as "*what is the shape of the valley in which the Franz Josef glacier is located*?" Students were provided with an example of an appropriate question related to the first statement: "There are many glaciers in the South Island and elsewhere in the world" with the accompanying question *"is Franz Josef changing in the same* *way as other glaciers?"* With hindsight we should have used an easier example but we had anticipated that the task would be too easy because, to us, it seemed so obvious what to do! We saw this first part of the task as simply setting the context for the following, more probing, questions about the status and influence of disconfirming evidence. The supervising teacher in one trial school detected the difficulty and completed a second example in discussion with his class. Even so, only some of these students went on to pose appropriate questions for the other statements provided. We had stumbled on a challenge that we now think needs much more investigation. Table 1 shows some typical examples of student responses we judged to meet the requirements of the task. In the feedback that accompanies the task on the ARB site, examples of both broad and specific questions are identified to help teachers support students by discussing what appropriate questions could look like.

Statement	Question
Glaciers in steep, narrow valleys are less affected by small overall temperature changes than wider glaciers.	**Broad question:** What shape is the Franz Josef glacier? **More specific question:** Does the Franz Josef glacier flow down a wide or a narrow valley?
Some glaciers have a layer of rubble on top. It makes them look dirty but also acts like a cooling layer of insulation that traps the cold.	**Specific question:** Does the Franz Josef glacier have a layer of rubble on top?
Changes to glaciers are related to weather that happened several years ago.	**Broad question:** What sorts of weather patterns are common in the Franz Josef area? **More specific question:** Was the weather unusually cold in the Franz Josef area several years ago?
Glaciers have always gone through periods of retreat (shrinking) and advance (growing).	**Broad question:** How has the Franz Josef glacier changed over time? What is Franz Josef's history of shrinking and growth? Is Franz Josef just in its growth phase? **More specific question:** Is the overall trend for the Franz Josef glacier to be gaining or losing its ice volume?

Table 1 **Examples of satisfactory questions asked by students**

The glacier resource was aimed at Years 9 and 10 (ages 13-15), and was trialled nationally by 244 Year 10 students. Trial students worked independently (although we recommend that teachers take a more collaborative approach when they come to use this task formatively), and there was no pre-teaching, i.e. the students completed the assessment outside the context of their classroom programme. Most statements were rated “difficult” for the trial students to convert into questions (between 20% and 39% of students gave a satisfactory response to any one statement). One exception was a statement about the manner in which a layer of rubble on the top of a glacier can act as insulation. This layer was visible in the provided photograph and the response was rated “moderately difficult” (between 40% and 59% of trial students posed a satisfactory question).

A few students shaped particularly interesting pattern-seeking questions that captured the essence of the systems challenge and extrapolated from Franz Josef to glaciers in general: *“Does the age and global position of Franz Josef simply mean that it is in advance while others are in retreat?” “Are other glaciers exposed to same/similar weather conditions as well?”* These responses went beyond what we expected, so what might have made this task so difficult for other students? When analysing patterns of student responses to formative tasks for the ARBs, we pay careful attention to patterns in incorrect or inappropriate responses, as discussed next.

Assessment dilemmas when assessing for science-related competencies

The “nature of science” achievement objectives in the New Zealand Curriculum (NZC) were written in a manner that would allow the more generic “key competencies” for the overall curriculum framework to be developed during science learning (Barker, Hipkins, and Bartholomew, 2004). As defined in NZC, key competencies are “capabilities for living and lifelong learning” and “draw on knowledge, attitudes and values in ways that lead to action” (Ministry of Education, 2007, p.12). That is, they have important *dispositional* components. Carr (2008) identifies three dimensions to the disposition to demonstrate competency in a specific task: students have to be ready, willing and able. The patterns of missing or inappropriate responses we discuss next suggest that some students were not “willing” but others were not “ready” (knowing when and how to respond) or perhaps “able” (having the right intellectual tools for the job) to respond in the manner the task required.

The glacier task put students in an unfamiliar role and some of them told us forthrightly that they did not like this! We were particularly interested in the responses of some students who, judging by their answers across the whole trial set, were high achievers who were very capable of thinking differently. Most were girls, and they did answer the glacier question as intended but appeared to be quite unsettled by the task. Some, for example, complained that it made them "feel dumb". We interpreted this to mean they were used to knowing the answers to give and being asked to ask questions moved them to a place where they could not be so sure they were “right”. Supporting this interpretation, students who were interviewed when the item was being piloted made the point that they were not used to asking their own questions; when they did an investigation in class the teacher gave them the question. Unfamiliarity with the role of being a question asker also seems likely to be the reason that a greater number of students than usual either did not attempt this item, or only partly completed it. (On average, a quarter of the sample did not attempt a question for any one statement.)

In contrast to these students who were not willing, a number of students did not appear to be ready or able. They did attempt this part of the task but accessed inappropriate types of question, presumably because they perceived these might fit the task, given the science context. Interestingly some of these questions made the task seem more difficult than we had intended it to be. We coded them as three main types:

Questions to find information or explanations: Some students wrote questions of the type that is familiar to them from traditional science learning. They posed questions for which they might find the answer to in books or on the internet: *"Why have some glaciers got rubble on top?" "How long does it take for the weather to affect the glaciers?"*

Investigative questions: Some students drew on their familiarity with typical "fair testing" investigations in science to pose questions that could be answered by an empirical investigation: *"By how much does a rubble layer slow down melting?"*

Adversarial questions: A few students wrote "rebuttal" type questions of the sort that might be used in a debate or a court of law, seeing this as a combative exercise rather than one of attempting to establish an overall picture of the relationship between parts and events in a complex system: *"Does appearance not matter? Rubble can be the preserver of our glaciers." "Is it scientifically proven that rubble insulates glaciers? Has it been measured?"*

Finally, despite the written and pictorial information provided with the task, it became evident that lack of contextual knowledge about glaciers was an obstacle to meaningful task completion for a small group of students. For example one student commented: *"if you wanted to move a glacier to a cooler place to keep it colder it might say in another fact that they can't survive when they are moved"*. In their feedback, other students told us they were reluctant to do the task because they didn't *"know anything about glaciers"*. While it is no surprise that lack of contextual knowledge seems to be a barrier for students who are used to thinking about assessment as being about showing what they already know, we think there is a deeper, more difficult issue to address here and we will return to the role of context in the demonstration of competency in the final sections of the paper.

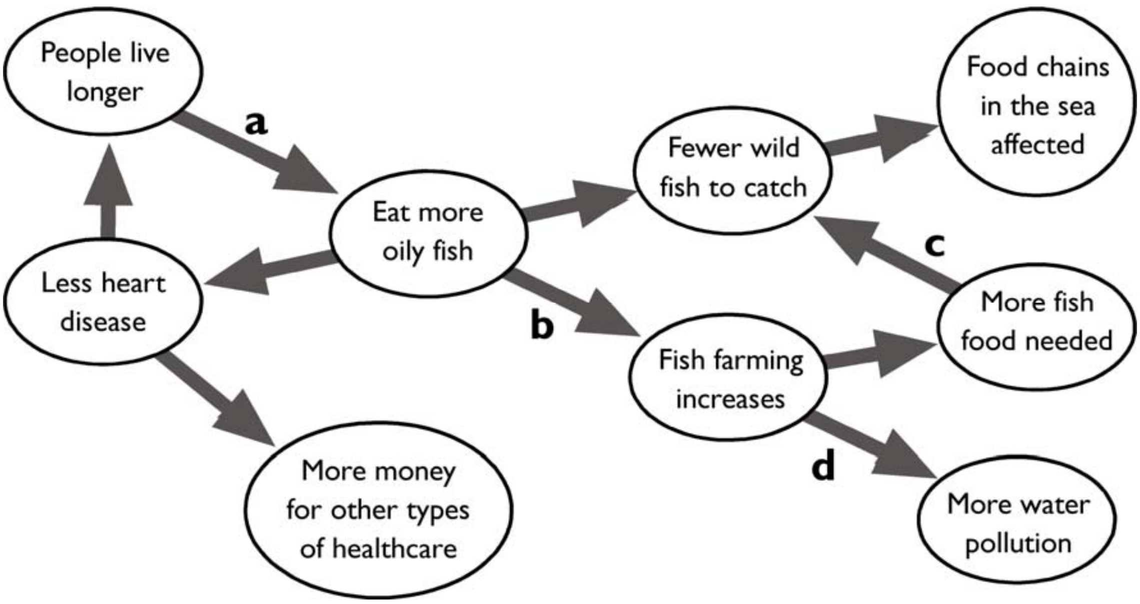
Recognising the relevant question to ask

The Glacier ARB task, imperfect as it was, raised a number of interesting questions about students' disposition to be question-askers. How much of their reluctance related to the unfamiliar role? Would those who appeared to lack knowledge of the types of questions that were appropriate, nevertheless be able to recognise such questions if they did not have to generate them? How much of a barrier was the perceived unfamiliarity of the context? As outlined above, patterns of answers for the ARB task pointed to all three possibilities, but gave us no way of determining their *relative* influence. The development of items for the *Science: Thinking with evidence* test provided an opportunity to revisit students' abilities as question askers. In a multiple-choice format, question-asking becomes a more passive "question-recognising". While this project is still very much a work in progress, some emerging patterns suggest that the lack of familiarity with the active role of being a question asker might be the greater barrier. As we outline next, many students can recognise an appropriate "systems" question when they see it.

Figure 1 shows an illustrative item from the *Thinking with Evidence* draft test at Year 9 (the first year of secondary schooling). The item is positioned as part of a *student* investigation that requires them to pose a question (some similar items asked about *scientists'* questions). Like the glacier question, this context is likely to be relatively unfamiliar to many students, and requires thinking at a systems level. All students in the initial trial attempted the question, so the “willingness” challenge of the ARB item completion was not apparent. Furthermore, initial data showed that students found it relatively easy to recognise the appropriate question, with 77% correctly selecting "What could happen if everyone ate more oily fish?" The strongest distracter was "Should we farm tuna and salmon for our food?" Most of the other "question-recognising" items in this test produced similar results. These results suggest that the greater challenge for students in *posing* pattern-seeking questions is unfamiliarity with this type of task.

Omega 3

Tuna and salmon are oily fish. They eat smaller fish in the sea. Oily fish have a chemical called Omega 3 which is good for us. For example, Omega 3 can protect people from getting heart disease. This is why doctors say people should eat more of these fish. Deanne drew this diagram to explore possible *consequences* of the doctors’ suggestion.



6 **What question do you think Deanne was exploring?**

A Should we farm tuna and salmon for our food?
B What sort of food do tuna and salmon eat?
C What could happen if everyone ate more oily fish?
D Are there food chains in the sea?

Figure 1 **Question-recognising item from Science: Thinking with Evidence test**

The challenge of transferring learning to new contexts

In the case of the part of the glacier task we have discussed, the focus of the assessment is not on what students know about glaciers, nor even really about whether they can ask questions that help them to decide the relevance of the patterns of the Franz Josef glacier to global warming. What we really want to know is whether students can and will ask questions that relate general science principles to specific contexts. If the concepts taught in school science are expected to inform critical decision-making as an adult (which is an assumption in NZC and, we suspect, many other national curricula) then there appears to be a corresponding assumption that, provided the concepts or facts have been correctly understood, the question of transfer to the issue in question will take care of itself. However, whether and how this actually does happen is the subject of some debate (Carr, 2008), and the glacier task suggests that, at the very least, there could be some specific "deep learning" (Egan & Madej, 2009) needed to underpin any such attempts in practice.

This third example is drawn from a small research project that preceded the development of a suite of ARB items. The research investigated the ways in which students from three schools, all aged from eight to twelve years, understood relationships in a stream ecosystem. The classroom activities we designed were intended to foster their systems thinking. The assessment task took the form of adding details to a simple outline sketch that represented several familiar elements of a stream and its surroundings (as discussed by Assaraf & Orion, 2005). The children also wrote answers to three short questions about relationships they could see in their drawings. While most students only considered direct relationships, some were able to identify a more sophisticated dynamic chain of related events. One boy (Matt) who appeared to have a good grasp of thinking at a systems level particularly interested us. As this piece of text shows, he did not have the contextual knowledge to identify plausible consequences from the chain of events he imagined:

> Companies by water drop oil/waste into the waterways, therefore killing the trout and other fish. The reeds will overgrow, algae would spread and this will cause blockage of the drains.

Matt made several contextual errors in his diagram completion and written text: neither trout nor eels are herbivores and trout, at least, cannot live in brackish water where reeds grow. Although Matt could obviously think about systems at a generic/conceptual level, we now wonder if he might have thought to check the biological facts he gave, given more time than the task allowed and some hint that such a step would be wise. Because we did not ask the students who did this task what questions they would need to ask to check the validity of their response, we have no way of knowing how well he might have done. It is easy to see the missed opportunity in hindsight, but it is only as these examples have started to accumulate and build an overall pattern that we have been sensitised to the specific concept-to-context transfer challenge. An irony is that the types of questions Matt needed to ask to check his response would have been much simpler for him than the thinking he did successfully demonstrate (for a more detailed discussion see Hipkins, Bull, and Joyce, 2008).

Implications for formative assessment

Assessment is about making generalisations from the data we collect (Wiliam, 2008). When using assessment formatively, the purpose of finding out what students can do is to adjust classroom teaching to improve their ability to do so across a wide range of contexts. One class that trialled the glacier task, on the whole, did noticeably better than any other group, and this pattern held for other items in the trial set. We suspected these students had experienced rather different learning opportunities and that these had strengthened their dispositions to think differently, if not their confidence that this was the correct way to proceed. If teaching can make a difference, what might other teachers learn from using tasks such as those discussed here and what further challenges might we anticipate?

The glacier example illustrates that making teaching decisions is complex, even given a carefully elaborated example to use. Several factors that may affect the way students complete the task, and the interactions between these factors may mean that it is difficult to decide what needs to be attended to first. Teachers and students need to establish next steps together, and the support material can only alert the teacher to the possibilities. There is still an inquiry to undergo: do the students need to learn more about the context? Are there some important scientific understandings they need to understand? Are they able to take a big picture view when considering an event? Do they need more practice in considering what questions to ask when confronted with a claim and its supporting "evidence"? Most importantly, students need to meet this type of task in a range of familiar and unfamiliar contexts before they can be judged on their ability to be question-askers. This is also a challenge for the designers of assessment tasks.

References

Assaraf, O., & Orion, N. (2005). Development of system thinking skills in the context of Earth system education. *Journal of Research in Science Teaching, 42* (5), 518-560.

Barker, M., Hipkins, R., & Bartholomew, R. (2004). *Reframing the essential skills: Implications for and from the science curriculum*. Wellington: Ministry of Education.

Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2002). *Working inside the black box.* London: nferNelson.

Carr, M. (2008). Can assessment unlock and open the doors to resourcefulness and agency? In Swaffield, S (Ed.). *Unlocking assessment: Understanding for reflection and application.* London and New York: Routledge.

Cowie, B. and Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Policy and Practice,* 6(1), 110-116.

Egan, K. and Madej, K. (2009). Learning in Depth: Students as Experts. *Education Canada,* 49(2), 18-23.

Feinstein, N. (2009). Prepared for what? Why teaching "everyday science" makes sense *Phi Delta Kappan* (June), 762-766.

Gilbert, J. (2005). *Catching the Knowledge Wave? The Knowledge Society and the future of education.* Wellington: NZCER Press.

Hipkins, R., Bull. A., and. Joyce, C. (2008). The interplay of context and concepts in primary school children's systems thinking. *Journal of Biological Education 42,* 2, 73-77.

Ministry of Education (2007). *The New Zealand Curriculum.* Wellington: Learning Media.

Newton, P., (2007). Clarifying the purposes of educational assessment. *Assessment in Education: principles, policy & practice,* 14(2), 149-170.

Sterling, S. (2001). *Sustainable education: Revisioning learning and change.* Devon: Green Books.

Wiliam, D. (2008). Quality in assessment. In Swaffield, S (Ed.). *Unlocking assessment understanding for reflection and application.* Oxon: Routledge.