

# Jumping to conclusions: Making too much of one response

Charles Darr

These days we are keen to make the most of our test data, particularly when it comes to informing next steps for teaching and learning. Sometimes we may examine how students have responded to particular questions on a test. This is good practice. It can be very enlightening to see what students found difficult (or easy) within a particular assessment and to think about the responses they made. It is also always important to know what was actually tested and how this matches our learning intentions. However, we must take care when dealing with responses to single questions. Making judgements as to what an individual knows (or doesn't know) on the basis of a response to a single question is problematic and can lead to poor decision making.

How a student engages with and responds to a question will often depend on the way the question is worded, the stimulus or context in which it is embedded, and even whether or not it contains a picture or diagram. In the case of a multiple-choice question, changing the options that are presented as alternative answers can also significantly alter how students respond to the question. On top of this, students can have good and bad days; a correct answer one day might represent a moment of inspiration, just a lucky guess, a momentary lapse, or the impact of some outside information.

This uncertainty at the question level makes it difficult to unpack what a student really knows about a concept or concepts being tested by a particular question. It also makes it unwise to use a response to a single question to reach a firm conclusion as to what they must be taught next.

Recently, the New Zealand Council for Educational Research (NZCER) had an opportunity to look at how unstable responses to individual questions can be. A group of students was asked by their teachers to retake a multiple choice test after the answer sheets from the original administration of the test went missing. After the students had repeated the test the original answer sheets were found. Later when the two sets of answer sheets were shared with NZCER, we were able to examine how students responded to the same question twice.

Table 1 shows that, overall, the test results were consistent. As can be seen, the average total scores at each year level were similar from one testing occasion to the next (the difference perhaps indicating a slight practice effect). In addition, the correlations are high, suggesting students have maintained similar rankings from one test administration to the next.

TABLE 1 AVERAGE TEST RESULTS FOR  
A REPEATED TEST

	Mean score Time 1	Mean score Time 2	Correlation
Year 7 (201 students)	16.55	18.36	0.85
Year 8 (186 students)	18.25	19.26	0.80

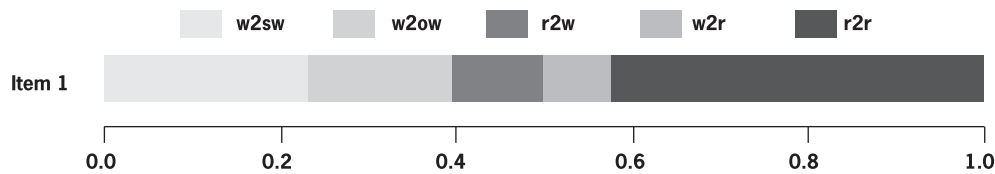
At the question level, however, the way the students responded is much less consistent from one testing occasion to the next. Many of the students were surprisingly willing to change their answers.

To investigate this more closely we examined the five valid ways students can respond to a question when given the opportunity to answer it twice (remember the questions are multiple choice). They are (coded as in Figure 1):

- Choose the same Wrong option both times (w2sw)
- Choose a Wrong option the first time and a different Wrong option the second time (w2ow)
- Choose the Right option the first time and a Wrong option the second time (r2w)
- Choose a Wrong option the first time and the Right option the second time (w2r)
- Choose the Right option both times (r2r)

Figure 1 uses a graphical representation to show the way the students responded to the first question in the test over the two test administrations. At first glance we can see that a sizeable proportion of students have changed their answer on

FIGURE 1 RESPONSE PATTERNS FOR QUESTION 1



the second administration. About 20 percent of the students who chose the correct option on the first testing occasion have changed their answer to a wrong option on the second. A similar percentage of the students, whose first response was wrong, have chosen the correct option on their second attempt. Interestingly, another fairly large group who chose a wrong option the first time have chosen a different wrong option the second time.

These kinds of response patterns are evident for most of the questions in the test. They suggest that it would have been unwise after the first administration to use a response to a question to definitively diagnose whether an individual student did (or did not) understand the concepts involved. As demonstrated, a sizeable proportion of the students would go on to do something different given a second opportunity to respond.

This variability of response to a single question is not restricted to multiple-choice questions. When a test is well targeted to students' achievement levels, many of the questions will test the limit of what they really know and some of their answers will be tentative.

### So, are responses to individual questions any use?

Examining how students respond to individual questions can still be very useful. We need to move, however, from what could be called a diagnostic frame of reference to a more formative frame. That is, instead of making a definitive diagnosis of what a student can or cannot do, we want to become involved in assessment behaviours that will allow us to work with learners to provide useful and valid feedback. So what can we do with responses at the question level?

### Use student responses as a starting point for further investigation

Responses—particularly incorrect responses—can give us clues as to what could be investigated further. A response to a question can be a great way to begin a learning conversation with a student or group of students, or help us select material for a follow-up assessment (perhaps using a focused Assessment Resource Bank item) to get a more in-depth view of what students know.

### Look at group responses

When we see an incorrect response repeated across a group of students we can be confident that, overall, the students are having issues with what is being assessed. We should think about what the question is asking and investigate with the students what might be going on. Notice here that we are not pinpointing exactly who knows what but are recognising that, within the group as a whole, some understanding does not line up with the expectations of the question. For a multiple-choice question analysing which distractors are being chosen can be very useful here.

### Look at patterns of responses across similar questions

Once a pattern of responses has emerged across several related questions we can begin to make judgements with more confidence. However, any conclusions we come to should still be tentative; making judgements about an area such as understanding of fractions will require a large number of questions.

### Take other information into account

When examining a response to a question or group of questions we need to think about what we know about the students and how that can inform our interpretations of the data. When a student gets a question wrong (or right) in a test we need to think about how that lines up with what we know from other sources of data, including our own day-to-day interactions with the students. If in doubt we should investigate further.

The response a student makes to a single test question should not lead to definitive judgements as to what they know (or don't know) and what they must learn next. However, a response to a question can be used as a catalyst for further investigation and engagement with learners and add to our toolbox of formative techniques.

**Charles Darr** is a senior researcher and manager of the assessment design and reporting team at the New Zealand Council for Educational Research.

**Email:** [charles.darr@nzcer.org.nz](mailto:charles.darr@nzcer.org.nz)