
Designing
Great
Hinge
Questions

Here's how teachers can get on-the-spot evidence about what students do and don't understand before moving forward with their lesson.

Dylan Wiliam

Every teacher I've ever met knows that no lesson plan survives the first contact with real students. And yet most teachers plan their lessons as though they're going to go perfectly. They plan them on the basis of assumptions they know to be false.

I'd like to suggest a small but powerful modification: Because lessons never go according to plan, teachers should build plan B into plan A. This involves designing a lesson with a "hinge" somewhere in the middle and using specific kinds of questions—what I call *hinge questions*—to quickly assess students' understanding of a concept before moving on.

The Rationale Behind the Hinge

When planning a lesson, the teacher identifies a particular concept that will be important for students to understand before moving on to other parts of the lesson. Of course, there are many such points in a lesson, but at least to start, choose one point somewhere in the middle of the lesson. At this hinge, the teacher asks a hinge question to check that the class has understood this key point of the lesson and gets a response from every single student. Depending on those responses, the teacher either moves on or goes back to review the material.

There's nothing new in this idea, but it turns out that it's rather difficult to do well. Here are some principles that teachers may find useful in designing effective hinge questions.

PRINCIPLE 1. Get a response from every student.

Teachers have always used questions to judge the level of a class's understanding before making a decision about whether to move on, but getting evidence about the whole class is difficult. For example, if a teacher wants to check whether students can recognize adverbs, he or she could ask the students to identify the adverb in a sentence like this: *The boy ran quickly across the street*. Teachers often ask students for a choral response (Hunter, 1982). However, that makes it difficult to tell who is responding correctly and who is just miming convincingly. In other words, the teacher doesn't have good evidence about who has and hasn't understood.

The teacher could, of course, ask each student in turn, but this takes a considerable amount of time, and students are likely to be influenced by the responses of other, higher-achieving students. One way to avoid this is to have all students respond at the same time by using finger voting. The teacher might ask students to identify the adverb in this sentence:

> The boy ran quickly across the street.
> (A) (B) (C) (D) (E)

Students can hold up one finger for A, two fingers for B, and so on, to indicate their choice. One of the important features of finger voting, or using ABCD cards, is that if a student hasn't made a choice, it's obvious to the teacher.

PRINCIPLE 2. Do a quick check on understanding, instead of engaging in extended discussions.

When teachers first start using hinge questions, they often find it difficult to interpret student responses; they're not sure whether a correct answer means that the students have truly understood. Although they might think that they could just ask each student to explain his or her answer, they actually never do this because, in a class of 30 students, it would simply take too long.

For the question to work as a quick check on understanding, it shouldn't take up a lot of time. Teachers should design the question so that it takes students no more than two minutes to respond, and

they should be able to collect and interpret all the student responses in 30 seconds or less. This can be achieved through using finger voting; ABCD cards; dry-erase boards; or digital technologies, such as electronic voting systems or smartphones. The technology used is far less important than the quality of the question.

PRINCIPLE 3. On the basis of student responses, decide whether to go forward or back.

Having to make such a quick decision about student understanding seems challenging at first, but that's largely because we educators are used to a process called data-driven decision making. We collect the data and then figure out what we're going to do with the information—and that can take time.

With a hinge question, the process is more like decision-driven data collection. The teacher doesn't need a lot of time to decide what to do with the data collected because the decision has already been identified. It's simple—either to go on or go back. The teacher collects just enough just-in-time data to make that decision.

If the responses indicate that most students have understood well enough the matter at hand, the teacher may go on, although he or she may arrange to talk to some students individually at a later point. If the responses indicate that few students have understood, the teacher is likely to review the material, perhaps dealing specifically with misconceptions revealed by the student responses.

When there are similar numbers of correct and incorrect responses, the teacher may either arrange for students to discuss their responses with their neighbors or conduct a whole-class discussion. Because the teacher knows who has answered correctly and incorrectly, he or she can draw certain students into the discussion at certain times, leading to better organized and more coherent classroom discussions. For example, the teacher might first ask a few of the students who thought *A* was the correct response for the reason for their choice, and then ask some of those who chose *B* for their reasons, and so on.

PRINCIPLE 4. Design hinge questions that elicit the right response for the right reason.

Hinge Questions with Distractors

When teachers make a judgment about whether students have understood something, they can make either of two mistakes: They can conclude that students haven't understood something when, in fact, they have, or they can conclude that students *have* understood something when, in fact, they haven't. These two kinds of mistakes differ greatly in their ramifications. Assuming students do know something when they actually don't (a *false positive* in researchers' jargon) is far more serious than assuming they don't know something when, in fact, they do (a *false negative*).

Assuming students do know something when they actually don't **is far more serious than assuming they don't** know something when, in fact, they do.

Some people point out that you can't conclude much from a single question. Those who construct tests note that you typically need at least 30 questions to get a reasonably reliable score for a student. This is true, and it would be a valid criticism of the idea of a single hinge question—*if* teachers were going to make high-stakes decisions on the basis of students' responses.

But the decision here is not high-stakes. More important, the teacher is trying to make a decision about the needs of the whole group, not of individual students. If the response of a student to a 30-item test provides a reasonable basis for drawing conclusions about that student, then the responses of 30 students to a single question probably provide a reasonable basis for drawing conclusions about that class.

We have to take great care in designing hinge questions so students don't get the correct answer for the wrong reason. This is perhaps the most important property of a good question. After all, if students with

PHOTOS L-R:SERGE MIKHAILOV, TOM GOWANLOCK, AMV_80, TOMASZ MAZON, MOVIT/SHUTTERSTOCK

the right thinking and those with the wrong thinking can answer a question in the same way, it's not very valuable as a diagnostic tool. The aim is always to have students with the right thinking and students with the wrong thinking give different answers.

The following question about measurement in science, from Stanford University science education professor Jonathan Osborne (2011), illustrates how we can do this:

> Janet was asked to do an experiment to find how long it takes for some sugar to dissolve in water. What advice would you give Janet to tell her how many repeated measurements to take?
>
> A. Two or three measurements are always enough.
> B. She should take five measurements.
> C. If she is accurate, she only needs to measure once.
> D. She should go on taking measurements until she knows how much they vary.
> E. She should go on taking measurements until she gets two or more the same.

This question has a single correct response (it's *D*), and so students have a 20 percent chance of answering the question correctly. Obviously, in a class of 30, six students would likely guess the correct response. But this question is so well designed that students who don't understand the main point are highly unlikely to get the correct answer. The incorrect responses—what test developers call *distractors*—are so plausible that they're attractive to students with incomplete understandings.

One of the best ways to develop such questions is to start from the partial or incomplete understandings that students have and generate questions that help identify which students have which misconceptions. The aim is to create what Philip Sadler (1998) calls *distractor-driven assessments*.

We have to **take great care in designing hinge questions** so students don't get the correct answer for the wrong reason.

People with deep knowledge of a subject, but little knowledge of teaching it to school-age students, often find it difficult to generate good distractors. They know the subject, but they don't know the difficulties that students have with it. This latter kind of knowledge—what Lee Shulman (1986) called *pedagogical content knowledge* and Deborah Ball and colleagues call *knowledge for teaching* (Hill, Rowan, & Ball, 2005)—is developed only through sustained experience working with students.

I suggested earlier that we could check whether students know what an adverb is by asking them to identify the adverb in the following sentence: *The boy ran quickly across the street.* However, students with an incorrect understanding can still answer the question correctly. Many students believe that an adverb usually follows the verb, so they may identify *quickly* as an adverb simply because it follows the verb *ran*. Offering an alternative version of the sentence, with the adverb moved to the end, will likely be more effective:

> The boy ran across the street quickly.
> (A) (B) (C) (D) (E)

Students who believe that an adverb typically comes immediately after the verb will now give a different response than those with a correct understanding.

Hinge Questions with Multiple Correct Responses

Another way of making it less likely that students get the right response to a question for the wrong reason is by offering multiple correct options. When a question has multiple correct answers, the chances of a student getting the correct response by guessing drop markedly.

For example, the teacher might ask a class to identify all adverbs in the following sentence:

> Fred ran the race well, but unsuccessfully.
> (A) (B) (C) (D) (E)

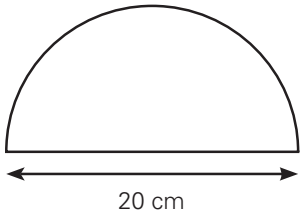
If students aren't told how many of the options are correct, then there are 32 possible responses. (Students have to make a separate choice about the correctness of each of the five options, so there are 2 x 2 x 2 x 2 x 2 combinations.) Students are far less likely to guess the correct answer when faced with 32 possible choices as opposed to the one-in-five chance that the single correct response offered.

Of course, no matter how carefully designed the question, it may not always work. I posted the question above to a group of students, and one

student who seemed to understand what adverbs were thought the sentence had only one—*unsuccessfully*. I asked him whether *well* was an adverb, and he said it wasn't because "Fred might have been sick yesterday."

No matter how well we design our questions, students will sometimes get them right for the wrong reasons and get them wrong for the right reasons. It's important to be sensitive to such possibilities.

Multiple correct answers also enable teachers to build a degree of challenge into their questions to stretch higher-achieving students. For example, a mathematics teacher asked a class of 7th graders to find the area of the following semicircle:



The students were given the following options from which to choose:

A) $\frac{\pi \times 20}{2}$

B) $\frac{\pi \times 20 \times 20}{2}$

C) 50π

D) $\frac{\pi}{2}\left(\frac{20}{2}\right)^2$

E) $\frac{\pi \times 10 \times 10}{2}$

The first two options correspond to well-known student errors, whereas the last three are all correct. However, it's far more challenging for students to recognize that *D* is correct than to recognize that options *C* and *E* are correct. As long as the easiest options represent the minimum level of achievement needed to move on, then a single question can give the teacher the information he or she needs to make that decision as well as keep the highest-achieving students on their toes.

The **teacher collects just enough just-in-time data** to decide whether to go on or go back.

Well Worth the Time

Kirschner, Sweller, and Clark (2006) have pointed out, "The aim of all instruction is to alter long-term memory. If nothing has changed in long-term memory, nothing has been learned" (p. 77). Now, the fact that students know something today doesn't mean they'll know it next week, but if they *don't* know it today, it's highly unlikely they'll know it next week. That's why checking for understanding in a planned way, by using hinge questions, is so valuable.

Every day, teachers typically make dozens of decisions about what to do next in group instruction, and it's simply not possible to plan a hinge question for each decision. But by planning at least a few questions carefully, teachers can get better-quality evidence about what their students can and can't do in time to do something about it.

Designing good hinge questions is usually harder than teachers imagine. I've found that groups of teachers typically take more than an hour to design one good question. But the benefits of doing so are huge. It means that you can find out what's going wrong with students' learning when they're right in front of you and that you can put the whole class's learning back on track right away. If you don't have this opportunity, then you'll have to wait until you grade their work. And then, long after the students have left the classroom, you'll have to try to get their learning back on track, in writing, one student at a time.

References

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.

Hunter, M. C. (1982). *Mastery teaching.* El Segundo, CA: Tip Publications.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.

Osborne, J. (2011, February 11). *Why assessment matters.* Paper presented at the annual conference of SCORE (Science Community Representing Education), London, UK. Retrieved from www.score-education.org/media/6606/purposejo.pdf

Sadler, P. M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching, 35*(3), 265–296.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(1), 4–14.

Dylan Wiliam (dylanwiliam@mac.com) is emeritus professor of educational assessment at University College London. His most recent book is *Embedding Formative Assessment: Practical Techniques for K–12 Classrooms* (Learning Sciences International, 2015).