

CHAPTER 4

Eliciting Evidence of Learners' Achievement

We discovered in chapter 3 the importance of being clear about what we want students to learn. Once we've accomplished that, we need to ascertain where the students are in their learning. In many classrooms, the process of eliciting such evidence is done mainly on the fly—teachers almost always plan the instructional activities in which they will engage their students, but they rarely plan in detail how they are going to find out where the students are in their learning. This chapter emphasizes the importance of planning this process and provides guidelines on what makes a good question, as well as some alternatives to questions. The chapter also offers some practical guidance on how to use questions effectively to adjust instruction to meet students' needs.

Finding Out What Students Know

The following are two questions used in the Third International Mathematics and Science Study (TIMSS; quoted by Vinner, 1997):

Question 1

Which fraction is the smallest?

- | | |
|------------------|------------------|
| A. $\frac{1}{6}$ | B. $\frac{2}{3}$ |
| C. $\frac{1}{3}$ | D. $\frac{1}{2}$ |

Question 2

Which fraction is the largest?

- | | |
|------------------|-------------------|
| A. $\frac{4}{5}$ | B. $\frac{3}{4}$ |
| C. $\frac{5}{8}$ | D. $\frac{7}{10}$ |

These two items appear to be quite similar, although the second question is obviously somewhat more complex because of the larger numbers involved. However, the second question turns out to be *much* more difficult than the first. For middle school students in Israel, for example, the success rate on the first item was 88 percent but for the second, only 46 percent (Vinner, 1997). In other words, only about half of the students who answered the first question correctly could also answer the second.

When I ask teachers why this might be, a number of possible reasons emerge. Some suggest that the first item is easier than the second because the fractions in the first question can be easily visualized. Others suggest that apart from option B, which is obviously bigger than C, they are all unitary fractions (fractions with a numerator of 1), which makes comparison easier. A third common suggestion is that the lowest common multiple is already present in the first question but not in the second and so has to be calculated, making this a two-step problem, which is obviously more complicated—although it is interesting to note that many adults (including math teachers!) prefer to answer the second question by converting each fraction to a percentage rather than by using the method that students are typically taught in middle school.

All of these explanations are credible, but a little more detective work reveals something rather interesting. In answering the second question, 39 percent of the students chose B. Since this question was answered correctly by 46 percent of the students, it was answered incorrectly by 54 percent, but 39 percent chose the *same* incorrect answer. In other words, almost three-fourths of the students who answered incorrectly chose the *same* incorrect answer.

This is highly significant. After all, if students' errors were random, then we would expect each of the incorrect answers to be chosen as often as the other incorrect responses, so that approximately 18 percent of students would choose each of B, C, and D. Instead, option B was chosen more than twice as often as the other two incorrect responses combined. This suggests that students' errors on this item are not random but systematic.

The most plausible explanation is that the choice of option B is related to how students learn about fractions. Although confusing at first, children come to realize that one-fifth is less than one-fourth

because the whole has been divided into more pieces. From this, many students conclude that the largest fraction is the one with the smallest denominator, and the smallest fraction is the one with the largest denominator. This rule works perfectly well for unitary fractions but does not work in general.

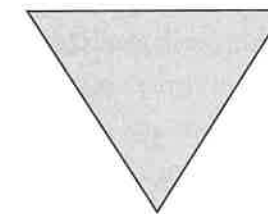
When attempting the first question, a student with this incomplete understanding of fractions will simply look for the largest denominator, find 6, and choose A—the correct answer. However, the same strategy applied to the second question leads the student to look for the smallest denominator, find 4, and choose B—an incorrect answer.

While we cannot be sure that this is the only reason for the differences in responses for the two questions, it is significant that adding the percentage of students who chose the correct answer to the second question (46 percent) to that of those who chose the answer generated by the naïve strategy (39 percent) gives us 85 percent, which is very close to the 88 percent correct figure for the first question. In other words, there is strong evidence that many students who got the first question right got it right for the wrong reason.

Why does this matter? Because when we, as teachers, ask students a question, and we get the answer we were hoping for, we tend to conclude that the students' learning is heading in the right direction. If, however, our questions are more like the first question than the second question, then there is a real danger that we will assume that our students' learning is on track when it is, in fact, heading off in a completely different direction.

Where Do Students' Ideas Come From?

Many students, when asked to describe the following shape, say it's an "upside-down triangle."



It is common to regard this as a misconception since it appears that the students do not realize that shape and orientation are independent

of each other; a triangle remains a triangle no matter which way it is oriented. However, some students describe this shape as an upside-down triangle even though they know that the orientation does not matter for the naming of the shape, because they are using vernacular, rather than mathematical, language.

A good example of this distinction is the word *square*. In the mathematics classroom, this word means a quadrilateral with four equal sides and four equal angles. In the world outside the mathematics classroom, however, the word *square* is often used to describe orientation rather than shape (for example, a picture can be "square" on a wall even when its sides are not the same length), and shapes that are, in fact, squares are described in ways based on their orientation rather than their shape (a baseball "diamond" is, after all, a square).

What seems like a misconception is often, and perhaps usually, a perfectly good conception in the wrong place. Describing a triangle as inverted is obviously an error in the mathematics classroom but a helpful aid to communication in other contexts (try Googling the phrase "inverted triangle" to see what I mean).

The same is true in other school subjects. When a child says, "I spended all my money," this could be regarded as a misconception, but it makes more sense to regard this as overuse of a general rule. Adding *d* or *ed* to a verb to form the past tense works most of the time, so the child may be guilty of nothing more than playing the odds!

Some people have argued that these unintended conceptions are the result of poor teaching. If only the teacher had phrased his explanation more carefully, had ensured that no unintended features were learned alongside the intended features, then these misconceptions would not arise. But this argument fails to acknowledge two important points. The first is that this kind of overgeneralization is a fundamental feature of human thinking. A child might initially associate the term *panda* with a particular fluffy toy but then will realize that other things are called *panda*, too. Pretty soon, the word *panda* becomes associated with a range of fluffy toys, some of which are not pandas, until the child learns to discriminate between pandas, giraffes, and lions.

The key insight here is that children are active in the construction of their own knowledge. They literally "make sense" of the things that go on around them, including what we teach, and sometimes, the sense

they make is not what we intended. Perhaps the best example of this is the answer you get when you ask children between the ages of four and seven, "What causes the wind?" "God," which is what most adults assume children will say, is a common response, but an equally popular answer is "trees." This is not likely to be something that they have been taught and misremembered. Children have observed that trees sway when the wind blows and have concluded that a correlation implies causation.

The second point is that even if we wanted to, we are unable to control the students' environments to the extent necessary for unintended conceptions not to arise. For example, many students believe that the result of multiplying 2.3 by 10 is 2.30. It is highly unlikely that they have been taught this. Rather, this belief arises as a result of observing regularities in what they see around them. They see that $7 \times 10 = 70$, $8 \times 10 = 80$, and so on, so they conclude that the result of multiplying whole numbers by ten is just to add a zero. We could make such a "misconception" less likely to arise by introducing decimals before teaching multiplying single-digit numbers by ten, but that would be ridiculous.

Teachers must acknowledge that what their students learn is not necessarily what they intended, and this is inevitable because of the unpredictability of teaching. Thus, it is essential that teachers explore students' thinking before assuming that students have understood something. However, generating questions that provide these powerful insights into students' thinking is far from straightforward.

Consider the following pair of equations:

$$\begin{aligned} 3a &= 24 \\ a + b &= 16 \end{aligned}$$

When asked what a and b are, many students respond that the equations can't be solved. The teacher might conclude that they need some more help with equations of this sort, but often, the reason behind the difficulties with this item is not the students' mathematical skills but their *beliefs* (Schoenfeld, 1989). When students are encouraged to talk about their difficulty, they often say things like, "I keep on getting b is eight, but it can't be because a is." The reason that many students have developed such a belief is that before they were introduced to solving equations, they practiced substituting numbers into algebraic formulas,

where each letter represented a different number. Although the students were not taught that each letter must represent a unique number, they have generalized implicit rules from their previous experience; they talk of "upside-down triangles" because we always show them triangles where the lowest side is horizontal and then compound the error when we teach them that the area of a triangle is half the base multiplied by the height (if there is no "right way up," how can a triangle have a base?).

The point here is that had the sixteen in the second equation been any other number at all, provided they had the necessary arithmetical skills, students would have solved these equations, and the teacher would, in all likelihood, assume that the class's learning was on track.

Questions that give us insight into student learning are not easy to generate and often do not look like traditional test questions. Indeed, to some teachers, they appear unfair. That is certainly the reaction of many math teachers to the following question, taken from the Chelsea Diagnostic Test for Algebra (Hart, Brown, Kerslake, Küchemann, & Ruddock, 1985):

Simplify (if possible): $2a + 5b$

The question is perceived as unfair because students "know" that in answering test questions, you have to do some work, so it must be possible to simplify this expression; otherwise, the teacher wouldn't have asked the question—after all, you don't get points in a test for doing nothing. Such a question may well fall short of standards and fairness guidelines established for items in a high-stakes test (for example, Educational Testing Service, 2002). But for finding out whether students understand a key principle in algebra, it is a useful question. If a student can be tempted to simplify $2a + 5b$, then the teacher should want to know that, because addressing this misunderstanding will be essential before the student can make progress in algebra.

Similar issues are raised by asking students which of the following two fractions is the larger:

$\frac{3}{7}$ or $\frac{3}{11}$

This item is generally not answered correctly by middle school students. It is seen by many teachers as a trick question and for that reason probably should not be used in a high-stakes test. But it is a very good

question for generating classroom discussion. The fact that this item is seen as a trick question shows how deeply ingrained into our practice is the idea that assessment should allow us to sort, rank, and grade students, rather than inform the teacher what needs to be done next.

In another example, a teacher was conducting a lesson about the molecular structure of matter. She had explained to students that water was composed of molecules, each of which had two atoms of hydrogen linked to an atom of oxygen, and asked the students to draw a sketch of what this might look like. The students in the class produced appropriate sketches, but when the teacher asked the class what was between the molecules, every single student said "water"—the students thought that the molecules were *in* the water, not that they *were* the water.

Whether to go over something one more time or to move on is a professional decision that the teacher must make, and in doing so, she needs to take a number of factors into account.⁷ For example, some aspects of the curriculum may be important but not consequential in terms of learning progressions. In other words, if a student fails to learn something at that point, he can still make progress. On the other hand, some things are absolute prerequisites for any meaningful progress. For example, there is no point trying to teach adding fractions if students are not able to generate equivalent fractions for a given fraction. It is also necessary to take into account the wider context. For example, a score of five or four (sometimes three) on one of the College Board's Advanced Placement examinations confers college credit, whereas scores of one and two do not. So a teacher might judge that moving on at a particular point is the right thing to do even though many students have not understood something, on the grounds that for a student, getting a four rather than a three represents a real difference, whereas getting a two rather than a one does not. I am not saying that this is right or wrong. It is a professional decision that each teacher needs to make for herself. What is unprofessional is to make that decision without finding out what the students in the class know. Questions that provide a window into students' thinking are not easy to generate, but they are crucially important if we are to improve the quality of students' learning.

One common objection at this point is that teachers do not have time to develop such questions, not least because they are too busy grading, but this just shows how ineffective many of our standard classroom

routines are. Every teacher has had the experience of writing the same thing on fifteen or twenty students' notebooks because the students were allowed to leave the classroom before the teacher discovered that the students had failed to understand some crucial point. So the important issue is this: does the teacher find out whether students have understood something when they are still in the class, when there is time to do something about it, or does the teacher only discover this once he looks at the students' notebooks? Viewed from this perspective, grading can be seen as the punishment given to teachers for failing to find out that they did not achieve the intended learning when the students were in front of them. Does the teacher get the class's learning back on track with the class in front of him, in one go, and when the meanings of students' responses and the teacher's questions can be negotiated, or does the teacher do it one student at a time, after they have gone away, and in writing?

There is no doubt that in the United States, most teachers spend a majority of their lesson preparation time grading student work, and almost invariably doing so alone. In other countries, much, if not the majority, of lesson preparation time is spent planning how new topics can be introduced and which contexts and examples will be used, and teachers work in groups to devise questions to find out whether their teaching has been successful. As noted previously, questions that give us this window into students' thinking are hard to generate, and teacher collaboration will help to build a stock of good questions.

Practical Techniques

Teacher-led classroom discussion is one of the most universal instructional practices. Typically, this follows a model termed "initiation-response-evaluation" (I-R-E) by researchers (for example, Mehan, 1979): the teacher asks a question, selects a student to answer the question, and then responds to the student's answer, generally with some kind of evaluation of what the student said.

Obviously, teacher talk dominates such interactions, but there are significant differences among countries in the relative proportions of teacher and student talk. American teachers seem to talk *less* than teachers in countries with higher average levels of achievement. For example, the 1999 TIMSS video study found that in US middle school

mathematics classrooms there were eight teacher words for every student word. In Japan and Hong Kong, the figures were thirteen and sixteen, respectively (Hiebert et al., 2003). So although many people assume that American teachers talk too much, they actually talk less than teachers in countries with higher performance. It would appear that how much students learn depends more on the quality than the quantity of talk.

As part of a larger study of elementary school classrooms, Ted Wragg and his colleagues analyzed one thousand teacher questions they had recorded. Over half the questions (57 percent) were managerial questions, such as "Who has finished all the questions?" or "Have you got your books?" Another one-third of the questions required only recall of previously provided information, such as, "How many legs does an insect have?" Only 8 percent of the questions asked by teachers required the students to analyze, to make inferences, or to generalize, such as, "Why is a bird not an insect?" (Brown & Wragg, 1993). Less than 10 percent of the questions that were asked by teachers in these classrooms actually caused any new learning. The others were rehearsing things the students already knew or were about classroom management. Addressing this seems like an obvious way to improve student learning.

I suggest there are only two good reasons to ask questions in class: to cause thinking and to provide information for the teacher about what to do next. As an example of the former, a sixth-grade teacher had asked a class whether a triangle could have two right angles, and students were discussing their answers in groups. In one group, one student believed that he could, by building a really long, thin triangle, because he had remembered being told by someone that parallel lines meet at infinity. Another student in the group knew that the three angles had to add up to 180° and realized that by having two right angles, he had already reached that total, and so he wondered whether it was possible to have an angle of 0° . A girl in the group was confident that she could create a triangle with three right angles, if she was allowed to place one corner at the North Pole and the other two on the equator (which the first two students dismissed, saying that it wasn't a proper triangle). Although this question looks like a closed question, answering it was a valuable activity for these students

because it caused thinking.

The other reason to ask questions is to collect information to inform teaching, for example, when a science teacher asks a class, "Which way does light travel? From my eye to the object or from the object to my eye?" This, too, is a closed question. After all, there are two answers, and one is incorrect. But it is a valuable question to ask, because many students believe that light travels from the eye to the object being seen, rather than the other way around.

Another finding of the TIMSS video studies, which confirmed findings of earlier studies (for example, Weiss, Pasley, Smith, Banilower, & Heck, 2003; Rowan, Harrison, & Hayes, 2004) was that American classrooms were characterized by relatively low degrees of student engagement. Understanding why this is so significant requires a brief detour into some recent work on the development of human abilities.

Student Engagement

In his book *Outliers*, Malcolm Gladwell (2008b) presents data on the twenty-five members of a Junior Hockey League team in Canada called the Medicine Hat Tigers, including height, weight, date of birth, position, and so on. He then invites the reader to look at the data for unusual features. The remarkable feature of the data, which few people notice until it is pointed out to them, is that eight of the twenty-five players were born in January and another six were born in February or March. The leagues are organized by age group, with those born in the same calendar year playing together. When children begin to play hockey in competitive leagues, those born in the early part of the year are a little stronger, faster, and bigger than the rest, and so they are more likely to get picked for the team. As a result, they get more time on the ice and are also likely to get more coaching. Over time, this builds up into a significant advantage.

The same effects have been noticed in other sports, too—those born soonest after the date used to divide age groups are more likely to be playing at the highest levels. It has been termed the "Matthew effect" (Stanovich, 1986): "For to all those who have, more will be given, and they will have an abundance; but from those who have nothing, even what they have will be taken away" (Matt. 25:29, New Revised Standard Version). Why does this matter? Because we create exactly the same

kind of effect in our classrooms every day.

In almost any classroom, some students nearly dislocate their shoulders in their eagerness to show the teacher that they have an answer to the question that the teacher has just asked. In the same classroom, however, other students try to stay below the radar and avoid being called on. One teacher described his classroom thus:

I'd become dissatisfied with the closed Q&A style that my unthinking teaching had fallen into, and I would frequently be lazy in my acceptance of right answers and sometimes even tacit complicity with a class to make sure none of us had to work too hard . . . They and I knew that if the Q&A wasn't going smoothly, I'd change the question, answer it myself or only seek answers from the "brighter students." There must have been times (still are?) where an outside observer would see my lessons as a small discussion group surrounded by many sleepy onlookers. (Black et al., 2004, p. 11)

High-engagement classroom environments appear to have a significant impact on student achievement. In one study, 191 students in seven fourth-grade classrooms followed the Thinking Together program (Dawes, Mercer, & Wegerif, 2000) in which teachers were provided with twelve lesson plans that were designed to help students develop their ability to use language as a tool for thinking about science and mathematics, both individually and in collaboration with other students. The researchers found that these students outperformed controls in similar schools on both teacher-constructed measures and standardized science achievement tests (Mercer, Dawes, Wegerif, & Sams, 2004). More surprisingly, these students outperformed controls on Raven's Progressive Matrices, which is a purely spatial test of intelligence. Engaging in classroom discussion really does make you smarter.

So, when teachers allow students to choose whether to participate or not—for example, by allowing them to raise their hands to show they have an answer—they are actually making the achievement gap worse, because those who are participating are getting smarter, while those avoiding engagement are forgoing the opportunities to increase their ability.

This is why many teachers now employ a rule of "no hands up except to *ask* a question" in their classrooms (Leahy, Lyon, Thompson,

& Wiliam, 2005). The teacher poses a question and then picks on a student at random. One middle school teacher describes this technique as "pose-pause-pounce-bounce." She poses the question, pauses for at least five seconds (sometimes, to help her measure the time, she mutters, under her breath, "One, two, three, four, got to wait a little more"), pounces on one student at random for the answer, and then bounces that student's answer to another student, again at random, saying, "What do you think of that answer?"

Some teachers claim to be able to choose students at random without any help, but most teachers realize that when they are in a hurry to wrap up a discussion so that the class can move on, they are often drawn to one of the usual suspects for a good answer. This is why many teachers have found the use of some randomization device useful. Examples of these can be downloaded from the Internet, and randomization routines are often included in the software supplied with interactive whiteboards. There are now even apps for the iPhone into which student names can be entered and then drawn at random. However, a beaker of Popsicle sticks (or for science teachers, tongue depressors) on which the students' names are written does the job just as well. The Popsicle sticks are also much more flexible; one teacher has punished inattentive students by writing their name on ten additional sticks and adding them to the beaker. If there is any concern that the teacher is rigging the choice of sticks, many teachers resolve the issue by giving the beaker of sticks to a student to be the "selector" for the day.

The major advantage of Popsicle sticks can also be a disadvantage. It is essential to replace the sticks to ensure that students who have recently answered know they need to stay on task, but then the teacher cannot guarantee that all students will get a turn to answer. One way around this is to leave the selected sticks out of the beaker but then replace them when students who have already answered a question seem off task.

Selecting students at random is a radical change in the "classroom contract" (Brousseau, 1984) and will not be welcomed by many students who are accustomed to classrooms where participation is optional. Most teachers realize that being called upon at random will be a shock for students unused to participation in classrooms. However, moving to random selection can also be unpopular with students who participate regularly. For some, this is because they are no longer able to show the

teacher that they have an answer. One way to tackle this issue is to ask two students at random for responses to a question, but for the third answer, ask the remainder of the class if anyone has anything else to contribute. It is also essential to ask the question first and then pick the student or students to answer. Doing it the other way around guarantees that all but the selected student know they need not pay attention.

For other students, random questioning is unwelcome because they are unable to control when they are asked questions. In one classroom, three students who were perceived by their teachers as having their hands up in response to every question actually removed their Popsicle sticks from the beaker when the teacher wasn't looking. When this was discovered, they revealed that they only raised their hands when they had an answer, which was most of the time. However, they found the prospect of being asked to respond when they did not have an answer threatening to their image of themselves as high-performing students, and this feeling was so strong that they preferred never answering questions to the risk of being asked a question to which they did not know the answer (Barry & Hardy, 2010).

When teachers do call on students at random—what Doug Lemov (2010) terms *cold calling*—many students will still resist engagement by responding with, "I don't know." How the teacher responds to this is crucial. Some teachers accept this and move on to ask the same question of another student, but this, in effect, allows the student to opt out. By adopting a rule of "no hands up except to ask a question," the teacher has implied that the classroom should be one in which all students are engaged. However, the student has indicated that she does not wish to engage, and the teacher has, in turn, accepted this. In other words, the student has successfully resisted the teacher's attempt to change the classroom contract. A better response is for the teacher to say, "OK; I'll come back to you," and then go around the classroom and get a number of responses to the question, after which the teacher comes back to the original student and says, "Now, which of those answers do you like best?" Even when the other students have answered with a single correct response, it is still valuable to go back to the original student simply to get her to repeat the correct response, because it emphasizes that this is a classroom in which there are no opt-outs (Lemov, 2010).

Often, students will say, "I don't know," not because they do not

know, but because they cannot be bothered to think. Henry Ford once said that we know thinking is difficult because of the lengths people go to in order to avoid doing it. Although the teacher first needs to be sure that this is the reason for the student's response, in such situations, Ellin Keene suggests that teachers ask, "OK, but if you did know, what would you say?"

Other possibilities that support the students without letting them off the hook are allowing students to "phone a friend," or for multiple-choice items, they can "ask the audience" or ask to go "fifty-fifty," where two incorrect responses are removed. All these strategies derive their power from the fact that classroom participation is not optional, and even when the student resists, the teacher looks for ways to maintain the student's engagement.

Wait Time

How much time a teacher allows a student to respond before evaluating the response is also important. Teachers do not allow students much time to answer questions (Rowe, 1974), and if they do not receive a response quickly, they will often "help" the student by providing a clue, weakening the question in some way, or moving on to another student. However, the amount of time between the student's answer and the teacher's evaluation of that answer is just as, if not more, important. Of course, when the question is a simple matter of factual recall, then allowing a student time to reflect and expand upon the answer is unlikely to help much. If you don't know the capital of Connecticut, extra time isn't going to help. But when the question requires thought, increasing the time between the end of the student's answer and the teacher's evaluation from the average wait time of less than a second to three seconds produces measurable increases in learning, although according to Kenneth Tobin (1987), increases beyond three seconds have little effect and may cause lessons to lose pace.

Alternatives to Questions

Asking questions may not be the best way to generate good classroom discussions. Asking a class which country was most to blame for the outbreak of World War I invites students to plump for one country or another. If, instead, the teacher makes a statement, such as, "Russia

was most to blame for the outbreak of World War I," students seem to respond more thoughtfully because they realize that just agreeing or dissenting is not enough—reasons have to be given (Dillon, 1988). Similarly, asking, "Are all squares rectangles?" is likely to yield a less-thoughtful discussion than framing what is, in effect, exactly the same question as a statement: "All squares are rectangles."

The quality of discussion is usually enhanced further when students are given the opportunity to discuss their responses in pairs or small groups before responding (a technique often called "think-pair-share").

Evaluative and Interpretive Listening

John Wooden was one of the greatest college basketball coaches of all time—some have suggested that he was the greatest coach of any sport (Serwer, 2010). He was once asked why other coaches were not as successful, and he said, "They don't listen. Listening is the best way to learn. You have to listen to those who you are supervising" (Serwer, 2010). But of course, what matters is *how* you listen.

When teachers listen to student responses, many focus more on the correctness of the answers than what they can learn about the student's understanding (Even & Tirosh, 1995; Heid, Blume, Zbiek, & Edwards, 1999). It is easy to identify such teachers because when they get incorrect answers from students, they respond by saying things like, "Almost," "Close," or "Nearly; try again." What the teacher is really saying is, "Give me the correct answer so that I can get on with the rest of my script for the lesson." Brent Davis (1997) called such teacher behavior "evaluative listening."

Teachers who listen evaluatively to their students' answers learn only whether their students know what they want them to know. If the students cannot answer correctly, then the teachers learn only that the students didn't get it and that they need to teach the material again, only, presumably, better. However, when teachers realize that there is often information about *how* to teach something better in what students say—and thus how to adjust the instruction to better meet students' needs—they listen *interpretively*. What such teachers seek to learn from the students' responses is not, "Did they get it?" but rather, "What can I learn about the students' thinking by attending carefully to what they say?" The shift from evaluative to interpretive listening was perhaps

most eloquently summarized by a girl in a seventh-grade classroom, who, when asked if she had noticed any change in her teacher over recent months, said, “When Miss used to ask a question, she used to be interested in the right answer. Now she’s interested in what we think” (Hodgen & Wiliam, 2006, p. 16).

Question Shells

There are a number of general structures that can help frame questions in ways that are more likely to reveal students’ thinking. One is the general structure of “Why is _____ an example of _____?” We saw one of these earlier (“Why is a bird not an insect?”). Similarly, rather than asking, “Is magnesium a metal?” more thoughtful and reasoned responses are likely to be generated by asking, “Why is magnesium a metal?” Other examples are shown in table 4.1.

Table 4.1: Examples of the Use of Question Shells

Original	Reframed
Is a square a trapezoid?	Why is a square a trapezoid?
Is carbon a metal?	Why is carbon not a metal?
Is <i>etre</i> a regular verb?	Why is <i>etre</i> an irregular verb?
Is this a sentence or a clause?	Why is this a clause rather than a sentence?
Is slate a metamorphic rock?	Why is slate a metamorphic rock?
Is <i>The Merchant of Venice</i> a comedy [or a tragedy]?	Why is <i>The Merchant of Venice</i> a comedy [or a tragedy]?
Is 23 prime?	Why is 23 prime?
Is photosynthesis an endothermic reaction?	Why is photosynthesis an endothermic reaction?

Another technique is to present students with a contrast and then ask them to explain the contrast, as shown in table 4.2.

Table 4.2: Examples of Questions Reframed in Terms of Contrasts

Original	Reframed
What is a prime number?	Why is 17 prime and 15 not?
What was life under apartheid like?	How were the lives of blacks and whites different under apartheid?
Is a bat a mammal?	Why is a bat a mammal and a penguin not?

Hot-Seat Questioning

Another useful technique for deepening classroom discussion is “hot-seat questioning.” In the typical classroom, the teacher scatters questions around the class, which can create student engagement but tends to lead to a rather flat discussion in which there is little development of the subject matter. In hot-seat questioning, the teacher asks a student a question and then a series of follow-up questions to probe the student’s ideas in depth. Other students in the class pay close attention because they know that at any minute, the teacher can turn from the student in the hot seat to anyone else in the class (chosen with a Popsicle stick!) and say, “OK, summarize for me what Jamilla just said.”

All these techniques are very effective ways of creating classroom engagement, but they are not necessarily good sources of evidence for teachers’ instructional decision making. Probably the most common instructional decision made by teachers every day is, “Do I need to go over this one more time, or is it OK to move on?” In most classrooms, teachers generate the evidence for this decision by making up a question on the spur of the moment, asking the class, and selecting one of the students who raise their hands to answer. If that student answers correctly, the teacher generally says, “Good,” and moves on. Asking a random student rather than a volunteer would be better, but the real problem is that the teacher is making a decision about the readiness of the whole class on the basis of the performance of one or two students. If teachers are to harness the power of high-quality questioning to inform their instructional decisions, they need to use all-student response systems routinely.

All-Student Response Systems

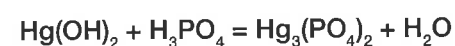
The idea of an all-student response system is very simple: the teacher asks a question in such a way that allows him to get a response from every student in real time. Some teachers employ class polls—going around the class and asking each student for his or her view on a question, such as, “Global warming: natural or anthropogenic?” or, “Macbeth: mad or bad?” but these work well only when every student expresses a view, and getting around the whole class can cause a considerable loss of momentum. It is best to collect the information from every student at the same time.

Techniques for All-Student Response

Many teachers collect information from the whole class simultaneously with techniques like “thinking thumbs” or “fist to five,” where students indicate their confidence in their understanding either by the position of the thumb (pointing up: confident; horizontally: not sure; down: still confused) or with a number from zero to five by holding up a clenched fist (zero) or the appropriate number of fingers. The problem with such techniques is that they are self-reports, and, as we know from literally thousands of research studies, self-reports are unreliable.

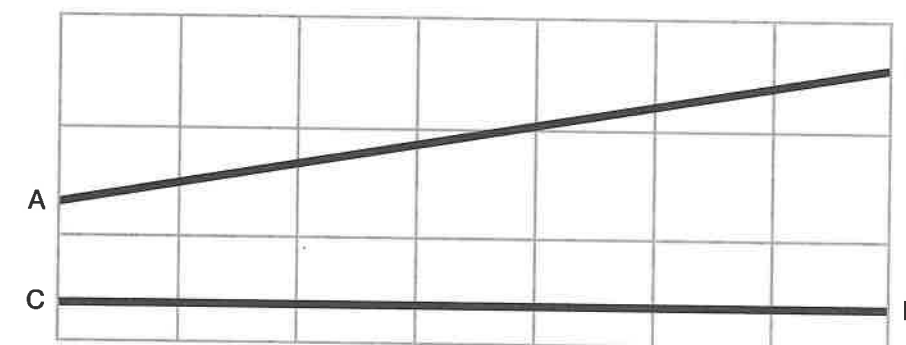
However, a very small change can transform useless self-reports into a very powerful tool. Simply make sure that the question being asked is cognitive rather than affective—in other words, that it is asking about thinking, not a feeling. For example, an elementary school teacher was teaching her students when the word *its* needed an apostrophe and when it didn't. On the board, she wrote the sentence, “Its on its way” and invited students to come up to the board and add any necessary punctuation. One student came up and added a period; a second added an apostrophe to the first *its*; and a third student added an apostrophe to the second *its*. The teacher then asked, “Is that now correct?” and every student had to respond by holding out a thumb pointing up or down. Just this small change in the way the question is posed creates a situation in which there is no place to hide for the students. If they signal that it is correct when it is not, then they reveal that they do not understand. On the other hand, if they signal that it is incorrect, then the teacher is likely to ask them to go to the front of the class to correct the error.

The same basic idea can be used in every school subject. In a high school chemistry class, for example, students were learning how to balance chemical equations. On the board, the teacher wrote the basic unbalanced equation for the reaction of mercury hydroxide with phosphoric acid to produce mercury phosphate and water:



Students were invited to go up to the board to suggest changes to help balance the equation, and after there were no more volunteers, the teacher asked the students to indicate, by thumbs up or down, whether the equation was now balanced.

In an elementary school math class, the teacher wanted to check on students' understanding of length and asked the students to indicate, by holding up one, two, or three fingers, which of the statements were true of the lines shown in the following problem.



Which of the following statements is true?

1. AB is longer than CD.
2. AB is shorter than CD.
3. AB and CD are the same length.

A middle school science teacher had been teaching students to distinguish between different kinds of levers. After explaining that the key principle of the classification of levers concerns the relative arrangement of the load, the effort, and the fulcrum, she illustrated the principle with three examples: a seesaw (type 1), a wheelbarrow (type 2), and a deep-sea fishing rod (type 3). To check the students' understanding, she asked the class how a pair of tweezers would be classified, requesting that each student hold up one, two, or three fingers to indicate his or her response. She was surprised when most of the students indicated that they thought the tweezers were a type 2 lever. When she asked them why, the students replied that there were two arms to the tweezers. She realized that she needed to get the students to understand that it is the relative distribution of the effort, load, and fulcrum that is important, not the number of components, by introducing more examples, such as a pair of scissors and a nutcracker.

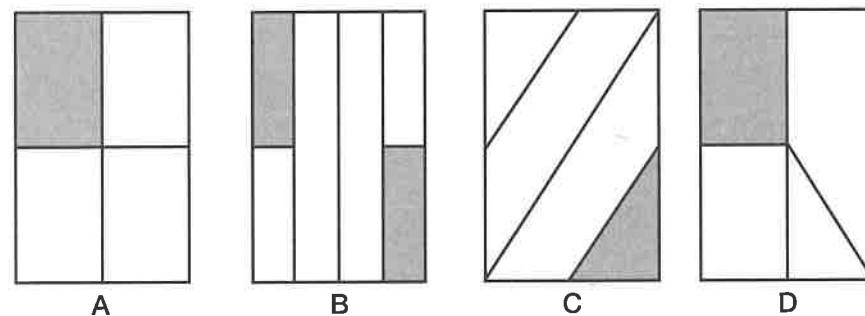
In each of these four examples, the teacher was able to ensure both student engagement—after all, it is very easy to tell if a student has not voted—and high-quality evidence to help decide what to do next. The problem with using fingers and thumbs is that they work best with

questions for which a single response is required (a drawback of most electronic clickers, too). That is why many teachers have adopted ABCD cards.

ABCD Cards

Each student has a number of cards, each of which bears a single letter. Some teachers just use A, B, C, and D, while others use sets with nine cards: A, B, C, D, E, F, G, H, and T (for true/false questions). These can be used just like fingers, but they also can be used with questions for which there is more than one correct answer, as in the following problem.

In which of the following diagrams is one-fourth of the area shaded?



Most students should recognize that A and B represent one-fourth, and hopefully, most students will also realize that in diagram C, the fraction shaded is not one-fourth. Many teachers tell students that if one out of four regions in a diagram is shaded, the four regions have to have the same area for this to represent one-fourth, and as a result, many students come to believe that the fraction shaded in D is not one-fourth, because the regions are not the same area.

The use of multiple correct answers allows the teacher to incorporate items that support differentiation, by including some responses that all students should be able to identify correctly but also others that only the ablest students will be able to answer correctly. Such differentiation also helps keep the highest-achieving students challenged and, therefore, engaged.

ABCD cards can also be used when there are no right and wrong answers but different views. A teacher was discussing with a class the Heysel Stadium disaster of 1985, during which thirty-nine people died and over six hundred people were injured in rioting before a soccer game between Liverpool and Juventus. To help organize the discussion of

this emotive issue, the teacher asked the students to indicate who they thought was most to blame:

- A. The Liverpool fans
- B. The Juventus fans
- C. The police
- D. The match organizers
- E. The builders of the stadium

During the discussion, students displayed the cards indicating their views on their desks, and the teacher brought them in to the discussion at the appropriate time (students who changed their minds during the discussion were encouraged to change the letters they were displaying). At the end of the discussion, the teacher asked the students to vote again, and she was pleased to see that the class now had a much more complex view of the tragedy, with all of them displaying at least two cards.

One elementary school teacher takes the idea of cards one step further with what she calls "letter corners." She uses multiple-choice questions with four options (A, B, C, and D), and if each option has at least three students supporting it, she sends the students to the four corners of the room, which she has labeled A, B, C, and D. The task of the students in each corner is to work out how to convince the students in the other corners that their choice is the best. Occasionally, students sneak from one corner to another. They move surreptitiously because they worry that the teacher will think that they are cheating, whereas, of course, the teacher celebrates such changes of mind as indicating that the students have learned something through discussion with peers.

ABCD cards can also be used to bridge two lessons. A middle school math class was learning to solve equations with a single unknown, and five minutes before the end of the lesson, the teacher asked the students to work out the answers to six questions. Four of the questions related to what the class had done that lesson. As the students held up cards indicating their responses to these questions, the teacher was pleased to see that most of the students answered correctly. The last two questions related to what she planned to do for the following lesson, where the unknown quantity appeared on both sides of the equation. Very few students answered either of these two questions correctly, so she could see that

what she had planned to do for the next lesson was, in fact, appropriate.

A major difficulty with ABCD cards is that they generally require teachers to have planned questions carefully ahead of time, and so they are less useful for spontaneous discussion, which is where mini whiteboards come in.

Mini Whiteboards

The mini whiteboard is sometimes touted as a modern invention but is really just the latest reincarnation of the slates used in 19th century classrooms. Whiteboards are powerful tools in that the teacher can quickly frame a question and get an answer from the whole class, whether asking elementary school students to write down a four-letter word with a short “i” sound or, as we saw in chapter 2, asking students in an AP calculus class to sketch the graph of

$$y = \frac{1}{(1+x)^2}$$

One teacher wanted to use whiteboards, but there was insufficient money in the school’s budget to acquire these, so instead, she placed sheets of letter-sized white card stock inside page protectors to provide a low-cost alternative. She realized that this was actually far more flexible than the whiteboards, because she was able to preprint different images on the inserts for specific lessons. When she was teaching math, the insert could be a sheet of graph paper. When she was teaching geography, it could be a map of the United States.

Exit Passes

When questions require longer responses, teachers can use the exit passes described in chapter 2. Examples of exit pass questions are:

- Why can’t you have a probability greater than 1?
- What is the difference between mass and weight?
- Why are historians concerned with bias when analyzing historical sources?

Exit pass questions work best when there is a natural break in the instruction; the teacher then has time to read through the students’ responses and decide what to do next. In the example in chapter 2, the teacher discarded the exit passes, but if students have written their names on the back of the card, the teacher can use the exit passes as

place settings for the next period with that class. The teacher can either create homogenous groups, so that she can work with the students who are having the greatest difficulty, or create heterogeneous groups in which there is at least one student who provided a good answer. Of course, since the work hasn’t been assessed, the students with the good answers don’t know who they are, which leads to more open discussion.

All these techniques create student engagement while providing the teacher with evidence about the extent of each student’s learning so that the teacher is able to adjust the instruction to better meet the students’ learning needs. Of course, the quality of the evidence—and, therefore, the quality of the instructional adjustments—depends on the quality of the questions asked, and so the next section of this chapter looks in more detail at what kinds of questions work best.

Discussion Questions and Diagnostic Questions

The following is a question that might be used in a middle school mathematics classroom.

Look at the following sequence:

3, 7, 11, 15, 19, . . .

Which is the best rule to describe the sequence?

- A. $n + 4$
- B. $3 + n$
- C. $4n - 1$
- D. $4n + 3$

Many students will select A; they see that each number in the sequence is four more than the number preceding it, so the rule must be “keep adding 4.” Others choose B, because although rule A tells you to keep adding 4, it does not tell you where to start, whereas B indicates that the sequence starts with 3, and numbers are added to 3 to generate the remaining numbers in the sequence. Interestingly, students who choose either A or B often justify their choices by saying that their teacher told them that n could be any number! The disadvantage with both A and B is that generating, say, the hundredth term requires first generating the previous ninety-nine. (These kinds of rules are sometimes called “term-to-term rules” since they allow one to generate the next term from the previous term.) Rules C and D are based on the observation that the

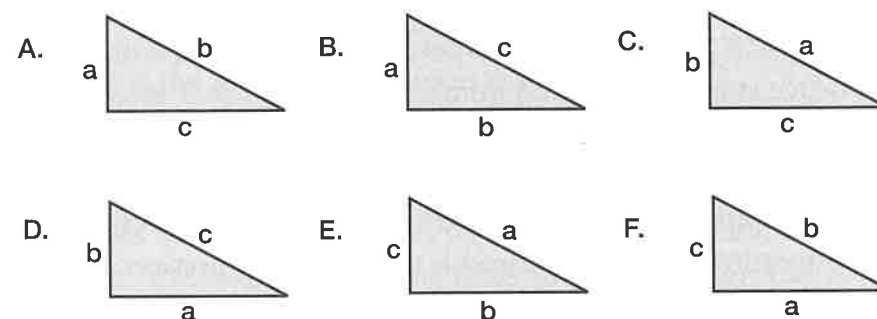
numbers in the sequence can be viewed as one fewer or three greater than the four times table. (These kinds of rules are sometimes called "position-to-term rules" since they allow one to generate each term simply by knowing its position.) Rule C allows one to find the hundredth term by multiplying 100 by 4 and subtracting 1 to get 399. Rule D works in the same way, except that one has to remember that the hundredth term is generated by setting n to 99, since the first term corresponds to setting n to 0, rather than 1, and then adding 3.

This question can lead to a valuable discussion in the mathematics classroom, since it allows the teacher to challenge the idea that mathematics is a right-or-wrong subject. For example, rules A and B can be thought of as correct even though they do not use the standard mathematical notation, but they are less powerful than rules C and D. Rules C and D are equivalent, but one can make an argument that C is more elegant than D because to find the hundredth term, one sets n to 100 with rule C, but one has to remember to set it to 99 with rule D.

Just because someone chooses rule C, that does not mean that he understands why rule C is superior to rule A, for example. A student may have chosen C because it looks more "mathematical." The teacher learns little just by seeing which of these alternatives students choose. She has to hear the reasons for the choices. That is why this question is a valuable *discussion* question, but it is not a good *diagnostic* question. It can lead to fruitful discussion in the class, but there is no point in asking this question unless you are going to have the discussion. If you don't have twenty minutes for the class discussion, it's probably not worth asking the question.

Consider, as an alternative, the following question.

In which of these right triangles does $a^2 + b^2 = c^2$?



In some ways, this is a sneaky question, because there are two correct—and four incorrect—responses. In total, there are sixty-four possible choices of response to this question (for each of the six triangles, the student needs to decide whether the equation is true, so there are $2 \times 2 \times 2 \times 2 \times 2 \times 2$ possible combinations). Therefore, if students select B and D, and none of the others (provided they haven't copied these from other students), the teacher can be reasonably confident that they know how to apply Pythagoras' rule and can move on. Of course, if he has time, the teacher could decide to ask the students for the reasons for their choices, but the point is that, with this question, the teacher gains concrete evidence about the students' learning *without* having to have the classroom discussion.

The crucial feature of such *diagnostic* questions is based on a fundamental asymmetry in teaching; in general, it is better to assume that students do not know something when they do than it is to assume they do know something when they don't. What makes a question useful as a diagnostic question, therefore, is that it must be very unlikely that the student gets the correct answer for the wrong reason.

In science, a teacher might ask the following question to prompt discussion.

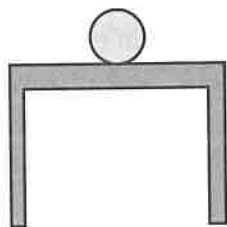
Ice cubes are added to a glass of water. What happens to the level of the water as the ice cubes melt?

- A. The level of the water drops.
- B. The level of the water stays the same.
- C. The level of the water rises.
- D. More information is needed to be sure.

A teacher who has been focusing on Archimedes' principle hopes that the students choose B, but there are valid reasons for choosing alternatives. If evaporation is a significant factor, response A would be a better choice. The question does not specify that the ice cubes are floating. If so many cubes are added that they are resting on the bottom of the glass, then the water level will rise as the ice melts. Very able students may realize that B is not, in fact, correct—whatever the science teacher thinks. When ice melts, it draws heat from the surrounding water—it actually takes an appreciable amount of energy to convert ice at 0°C to water at 0°C . So, as the ice melts, it cools the surrounding water by

more than just the effect of the ice being colder than the water. If the water is above 4°C , then as it cools, it will contract, so A would then be correct (although the amount may be too small to measure). On the other hand, if the water is below 4°C , then cooling it will cause it to expand, so C would be correct. And since all these answers depend on the assumptions made, it is fair to claim that one needs more information to be sure, so D is correct, too. As with the previous question on sequences, the teacher learns little about the quality of his students' thinking just by seeing which response was chosen; the teacher needs to hear the reasons for the choices, so while this may be a good *discussion* question, it is a poor *diagnostic* question.

Consider the following question, developed by Mark Wilson and his colleagues at the University of California, Berkeley (Wilson & Draney, 2004, p. 149).



The ball sitting on the table is not moving. It is not moving because:

- A. No forces are pushing or pulling on the ball.
- B. Gravity pulls the ball down but the table is in the way.
- C. The table pushes up with the same force that gravity pulls down.
- D. Gravity is holding it on the table.
- E. There is a force inside the ball preventing it from rolling off the table.

The first response, A, is typically selected by students with the naïve conception that if there is no movement, then there are no forces in play. The second, B, on the other hand, looks quite reasonable. Is gravity pulling the ball down? Well, we might quibble about the language, but the ball is certainly being attracted by the gravitational pull of the earth, and the table is definitely in the way. So response B will be attractive to many students.

Response C is what the science teacher is hoping for. She has been teaching that when there is no movement, this could either mean that there are no forces acting on the object or, much more commonly, that

the forces acting on the object are in equilibrium. Response D describes the situation in a similar way to response B, and response E is selected by students who, instead of seeing inertia as a property of matter, see it as a force (presumably at least in part because of the way the word *inertia* is used in everyday language).

The first and last responses, therefore, are obviously incorrect and related to well-known naïve conceptions that students have about the physical world. Response C is clearly the best answer, but what about B and D? They are correct, in an intuitive sense at least, so should the teacher accept either B or D as correct? In the science classroom, the answer is no. Responses B and D might be acceptable naïve ways of looking at the situation, but what the teacher is trying to establish is whether her students can think about this situation like a physicist. Physics is an unnatural way of thinking about the world—if it were natural, it wouldn't be so hard for students to learn. For example, when one ventures out in cold weather, it certainly feels as if the cold is coming in through one's clothes, but physicists have learned that it is more productive to think about the situation as one in which heat is escaping.

This question is not really asking, "What's happening here?" It is asking, "Can you think about this situation like a physicist?" And from that perspective, B and D are not acceptable. The only acceptable response is C. However, it is the plausibility of B and D that makes this such a valuable question. Students who are not secure in their understanding of forces will be tempted to opt for B or D, and so only students with a good understanding of the idea of forces in equilibrium will choose C. If a teacher asks this question, and students select C, then the teacher can be reasonably confident that they have understood the key idea about forces in equilibrium. She could, of course, ask students to explain the reasons for their choices and have a whole-class discussion, but this is not necessary for the teacher to gain valuable evidence about the students' understanding.

Discussion and diagnostic questions work in other school subjects, too. A history teacher might ask the following question.

In which year did World War II begin?

- A. 1937
- B. 1938
- C. 1939

D. 1940

E. 1941

Most European historians would probably select 1939, while Americans might regard the war as having started when the United States joined the conflict in 1941. However, interesting arguments can (and have!) also been made for 1937 (second Sino-Japanese war), 1938 (annexation of the Sudetenland), and 1940 (the Tripartite Pact). One particularly interesting feature of this question is that it focuses attention on the issue of what the phrase *world war* actually means. In 1939, Germany was at war with Britain and France, and Japan was involved in conflicts in Manchuria, but at that time, there was no formal alliance between Germany and Japan. However, even in 1939, the conflict was global in the sense that armed forces from British colonies such as India, Canada, Australia, and New Zealand were involved, and German colonies in Africa bordered those of Britain and France.

Again, the point is that the teacher learns little about the quality of student thinking from hearing which answer a student chooses; the teacher needs to hear reasons for the choice, and that means hearing from every student in the classroom. That's great if you have time to hear from every student, but this kind of item is useless as a quick check on student understanding.

For a good example of a diagnostic question in history, consider the following question, which was mentioned as a possible exit pass question earlier.

Why are historians concerned with bias when analyzing sources?

- A. People can never be trusted to tell the truth.
- B. People deliberately leave out important details.
- C. People are only able to provide meaningful information if they experienced an event firsthand.
- D. People interpret the same event in different ways, according to their experience.
- E. People are unaware of the motivations for their actions.
- F. People get confused about sequences of events.

This multiple-choice question was generated by using the most interesting incorrect responses from the exit passes. Most history teachers would agree that a class in which every student selects D as the best

answer is probably ready to move on. Again, a discussion might be helpful, but it is not necessary to help the teacher judge whether the students have understood the main point of this instructional sequence. It is the quality of the distractors that is crucial here—it is only because the distractors are so plausible that the teacher can reasonably conclude that the students who choose D have done so for the right reason.

The following item was generated by a group of language arts teachers from Cherry Hill, New Jersey, who were, at the time, teaching persuasive writing to their students.

Which of these is the best thesis statement?

- A. The typical TV show has nine violent incidents.
- B. The essay I am going to write is about violence on TV.
- C. There is a lot of violence on TV.
- D. The amount of violence on TV should be reduced.
- E. Some programs are more violent than others.
- F. Violence is included in programs to boost ratings.
- G. Violence on TV is interesting.
- H. I don't like the violence on TV.

Most language arts teachers would agree that D represents the best thesis statement *within the genre of persuasive writing*. If the writing were a factual reporting of a study on the amount of violence on TV, then response A might be a better thesis statement (analogous to an abstract for a research article) and F represents a perfectly reasonable opinion. Several of the other responses represent views that students hold, and, of course, the one that drives language arts teachers crazy is B—when students confuse a thesis statement with an introductory statement.

The following is an example from the teaching of modern foreign languages developed by a group of teachers of Spanish from Chico Unified School District, California.

Which of the following is the correct translation for "I give the book to him"?

- A. Yo lo doy el libro.
- B. Yo doy le el libro.
- C. Yo le doy el libro.
- D. Yo doy lo el libro.

E. Yo doy el libro le.

F. Yo doy el libro lo.

Anyone who has taught Spanish to native English speakers knows that they have two particular difficulties in learning about pronouns: which pronoun to choose and where it should go. This item is ingenious because in one of the responses, the correct pronoun is in the wrong place; in another, the placement is correct, but the pronoun is incorrect; and in others, there are multiple errors. This question took quite a while to generate—certainly much more time than teachers would normally take to write a single question—but this question will still be a good question in twenty years' time, because native English speakers will still be having the same difficulty with pronoun selection and placement when learning Spanish.

Diagnostic questions can be used in a variety of ways. They can be used as "range-finding" questions to find out what students already know about a topic before beginning the instruction. A university teacher teaching an introductory course on botany poses the question "What proportion of the water taken in by the roots of a corn plant is lost through transpiration?" and asks the students to hold up one, two, three, four, or five fingers according to whether they think the answer is 10 percent, 30 percent, 50 percent, 70 percent, or 90 percent. In teaching this course over many years, she has found that this simple question gives her a good indication of students' understanding of how plants work. Students who know little about plants assume that because plants die if they cannot get enough water in through the roots, the plants would seek to minimize the amount of transpiration in order to minimize water loss, and thus, they tend to select 10 percent or 30 percent. Students who understand the role played by water in transporting nutrients within the plant realize that most of the water taken in through the roots is lost through transpiration.

While using questions for range-finding at the beginning of a sequence of instruction is important, for many teachers, diagnostic questions are most useful in the middle of an instructional sequence to check whether students have understood something before moving on. The central idea here is that the teacher designs each lesson with at least one "hinge" in the instructional sequence. The hinge is a point at which the teacher checks whether the class is ready to move on through the use

of a diagnostic question. How the lesson proceeds depends on the level of understanding shown by the students, so the direction of the lesson hinges at this point.

While the design of diagnostic questions to be used at hinge points in lessons—*hinge-point questions* for short—is much more craft than science, work with teachers suggests that the following two principles are useful guidelines. First, it should take no longer than two minutes, and ideally less than one minute, for all students to respond to the question; the idea is that the hinge-point question is a quick check on understanding, rather than a new piece of work in itself. Second, it must be possible for the teacher to view and interpret the responses from the class in thirty seconds (and ideally half that time). If the teacher takes any longer than this to make sense of the students' responses, there is a risk of students getting off task and consequent disruptive behavior.

The main requirement of good hinge-point questions is nicely illustrated by the following item, used as a sample item in the National Assessment of Educational Progress (NAEP, 2005) program for fourth-grade mathematics:

There are two flights per day from Newtown to Oldtown. The first flight leaves Newtown each day at 9:20 and arrives in Oldtown at 10:55. The second flight from Newtown leaves at 2:15. At what time does the second flight arrive in Oldtown?

When I ask teachers how this item might be improved, they highlight the lack of information about whether these times are a.m. or p.m., the fact that the linguistic demands may be much more taxing than the mathematical demands for some students, the fact that one has to assume that both flights take exactly the same amount of time, and so on. These are all valid criticisms, but there is a much more fundamental problem with this item that, in my experience (and I have asked this of literally thousands of teachers), not one teacher in a hundred notices; students who think there are a hundred minutes in an hour and those who know there are just sixty get the same answer.

Earlier in this chapter, I proposed that there are two good reasons to ask questions in classrooms: to cause thinking and to provide the teacher with information that assists instructional decision making. Whether questions satisfy the first of these is relatively easy to evaluate. For the second, a key requirement of questions to assist instructional decision

making is that students with the right idea about whatever it is we want them to know, understand, or be able to do should get different answers from students who do not have the right idea. Or, to put it another way, ideally it would be impossible for students to get the right answer for the wrong reason. One way to improve hinge-point questions is to involve groups of teachers. One group of teachers is given the task of developing a question, which is then handed to a second group of teachers whose task it is to try to identify a way in which a student might get the correct answer with incorrect reasoning. If they can, then the question is handed back to the first group to be refined.

In the case of the question about flights, one immediate improvement is simply to change the start time of the first flight from 9:20 to 9:05 (and presumably also to add a.m. or p.m.). With this modification, students who think there are one hundred minutes in an hour will get the answer 3:65, while those who know there are sixty minutes in an hour will answer 4:05. No question will ever be perfect, but by constantly seeking to understand the meaning behind students' responses to our questions, we can continue to refine and polish our questions and prompts.

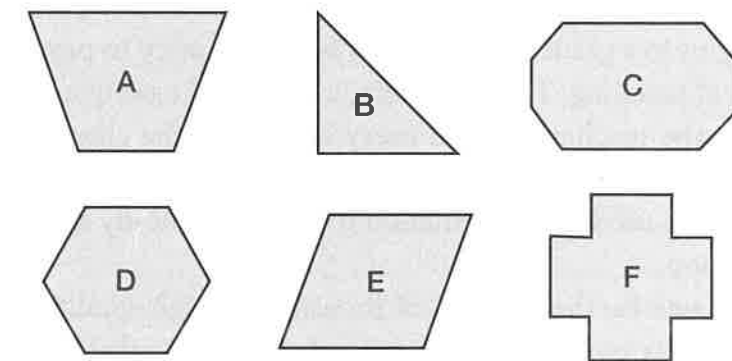
A second requirement of questions to assist instructional decision making—much less important than the first but still useful to bear in mind when designing questions—is that the incorrect answers should be *interpretable*. That is, if students choose a particular incorrect response, the teacher knows (or at least has a pretty good guess) why they have done so. A multiple-choice question for which the incorrect responses relate to well-known naïve conceptions, such as the question on pronouns in Spanish, is one example.

While some people argue that multiple-choice questions have no place in educational assessment, there are good reasons for preferring the multiple-choice format for hinge-point questions. Multiple-choice questions are often criticized because they assess only low-order skills such as factual recall or application of a standard algorithm, although as the previous examples show, they can, if carefully designed, address higher-order skills. Another problem with multiple-choice questions used in tests is that there is no opportunity for the student to negotiate the meaning of the question with the question-setter or the machine that does the scoring.

In the classroom, these are much less important considerations, and

multiple-choice questions have one great feature in their favor: the number of possible student responses is limited. When a teacher is faced with a sea of thirty whiteboards, each with a different response, it can be bewildering. Using multiple-choice questions provides a means for sorting all the students' responses ahead of time, so that precious classroom time is not spent trying to make sense of the students' answers.

Sometimes it makes sense to administer the question as a series of simple questions. For example, an elementary school teacher has been teaching her students about lines of symmetry in two-dimensional shapes, and toward the end of the lesson, she displays the following images.



She calls out "A," and each student in the class has to hold up a number of fingers to show how many lines of symmetry shape A has. She then does the same for B, C, D, E, and F. She doesn't try to remember how each student responds. Instead, she focuses on just two things:

1. Are there any items that a significant proportion of the class answers incorrectly and will need to be retaught with the whole class?
2. Which two or three students would benefit from individualized instruction?

Many students indicate that they think that shape D has just three lines of symmetry and that shape E has no lines of symmetry even though it is a rhombus. She points out to the class that D has six lines of symmetry, three through the corners and three through the midpoints of the sides, and reminds the class to look for lines of symmetry that are not horizontal, vertical, or at 45°. She realizes that the incorrect answers

may not necessarily indicate poor mathematical understanding. After all, in the case of shape D, holding up just three fingers could be a result of students thinking that only one hand would be necessary, and as a result of the distortion of the data projector, the image of shape E might not actually be the rhombus shape intended. Nevertheless, she is well aware that in teaching, as noted previously, it is better to assume students do not know something when they do than to assume that they do know something when they don't.

In this episode, the teacher administered, graded, and took remedial action regarding a whole-class quiz in a matter of minutes, without giving herself a pile of grading to do. She does not have a grade for each student to put in a gradebook, but this is a small price to pay for such an agile piece of teaching. This is, in fact, a canonical example of formative assessment: the teacher engaged every student in the class by requiring them to indicate the number of lines of symmetry for each of the six shapes, and she used the information to make on-the-fly adjustments to her instruction.

There is one further benefit of focusing on high-quality questions, and that is their portability. Most teachers find worksheets or lesson plans developed by other teachers to be of limited usefulness. However, high-quality questions seem to work across different schools, districts, states, cultures, and even languages. Indeed, sharing high-quality questions may be the most significant thing we can do to improve the quality of student learning.

Conclusion

Years ago, David Ausubel (1968) argued that the most important factor influencing learning is what the learner already knows and that the job of the teacher is to ascertain this and to teach accordingly. Students' conceptions are not random aberrations but the results of sophisticated and creative attempts to make sense of their experiences. Within a typical classroom, there is clearly not enough time for the teacher to treat each student as an individual, but with careful planning and the thoughtful application of the techniques presented in this chapter, the teacher can make the classroom a much more engaging place for students and one in which the teacher is able to make rapid and effective

instructional adjustments to meet the learning needs of all students. Once the teacher knows where learners are in their learning, she is in a position to provide feedback to the learners about what to do next—and this is the subject of the next chapter.