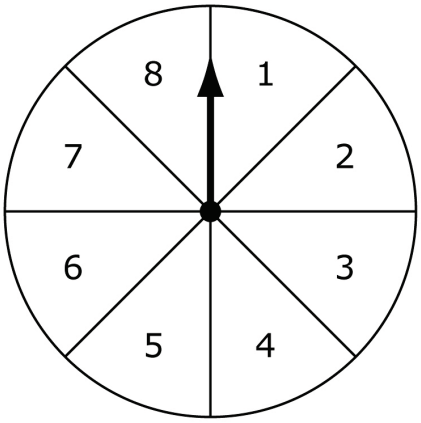
MAT.07.ER.3.000SP.E.292

Sample Item ID:	MAT.07.ER.3.000SP.E.292
Grade:	07
Primary Claim:	**Claim 3: Communicating Reasoning** Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others.
Secondary Claim(s):	Claim 1: Concepts and Procedures Students can explain and apply mathematical concepts and carry out mathematical procedures with precision and fluency.
Primary Content Domain:	Statistics and Probability
Secondary Content Domain(s):	
Assessment Target(s):	3 E: Distinguish correct logic or reasoning from that which is flawed and—if there is a flaw in the argument—explain what it is. 3 F: Base arguments on concrete referents such as objects, drawings, diagrams, and other actions. 1 I: Investigate chance processes and develop, use, and evaluate probability models.
Standard(s):	7.SP.5, 7.SP.7
Mathematical Practice(s):	1, 3
DOK:	3
Item Type:	ER
Score Points:	2
Difficulty:	M
Key:	See Sample Top-Score Response.
Stimulus/Source:	
Target-Specific Attributes (e.g., accessibility issues):	
Notes:	Part of PT set

Carl and Beneta are playing a game using this spinner.



Carl will win the game on his next spin if the arrow lands on a section labeled 6, 7, or 8.

Carl claims it is likely, but not certain, that he will win the game on his next spin.

Explain why Carl's claim is **not** correct.

Beneta will win the game on her next spin if the result of the spin satisfies event *X*.

Beneta claims it is likely, but not certain, that she will win the game on her next spin.

Describe an event *X* for which Beneta's claim is correct.

Sample Top-Score Response:

Carl's claim is incorrect. The probability that Carl will spin a 6 or higher is 0.375. This means that it is more likely that Carl will spin a number less than 6 on his next turn.

For Beneta, event *X* could be "the arrow lands on a section labeled with a number greater than 2."

Scoring Rubric:

Responses to this item will receive 0–2 points, based on the following:

2 points: The student shows thorough understanding of how to use understanding of likelihood and probability to critique the reasoning of others. The student explains why Carl's claim is incorrect and describes an appropriate event for Beneta.

1 point: The student shows partial understanding of how to use understanding of likelihood and probability to critique the reasoning of others. The student explains why Carl's claim is incorrect. **OR** The student describes an appropriate event for Beneta.

0 points: The student shows inconsistent or no understanding of how to use likelihood and probability to critique the reasoning of others. The student fails to provide a correct explanation or event.

Inside the Black Box: Raising Standards through Classroom Assessment
Author(s): Paul Black and Dylan Wiliam
Source: *The Phi Delta Kappan*, Vol. 80, No. 2 (Oct., 1998), pp. 139-144, 146-148
Published by: Phi Delta Kappa International
Stable URL: http://www.jstor.org/stable/20439383
Accessed: 17/10/2013 17:12

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Phi Delta Kappa International is collaborating with JSTOR to digitize, preserve and extend access to *The Phi Delta Kappan.*

http://www.jstor.org

This content downloaded from 128.120.194.194 on Thu, 17 Oct 2013 17:12:33 PM
All use subject to JSTOR Terms and Conditions

Inside the Black Box

Raising Standards Through Classroom Assessment

BY PAUL BLACK AND
DYLAN WILIAM

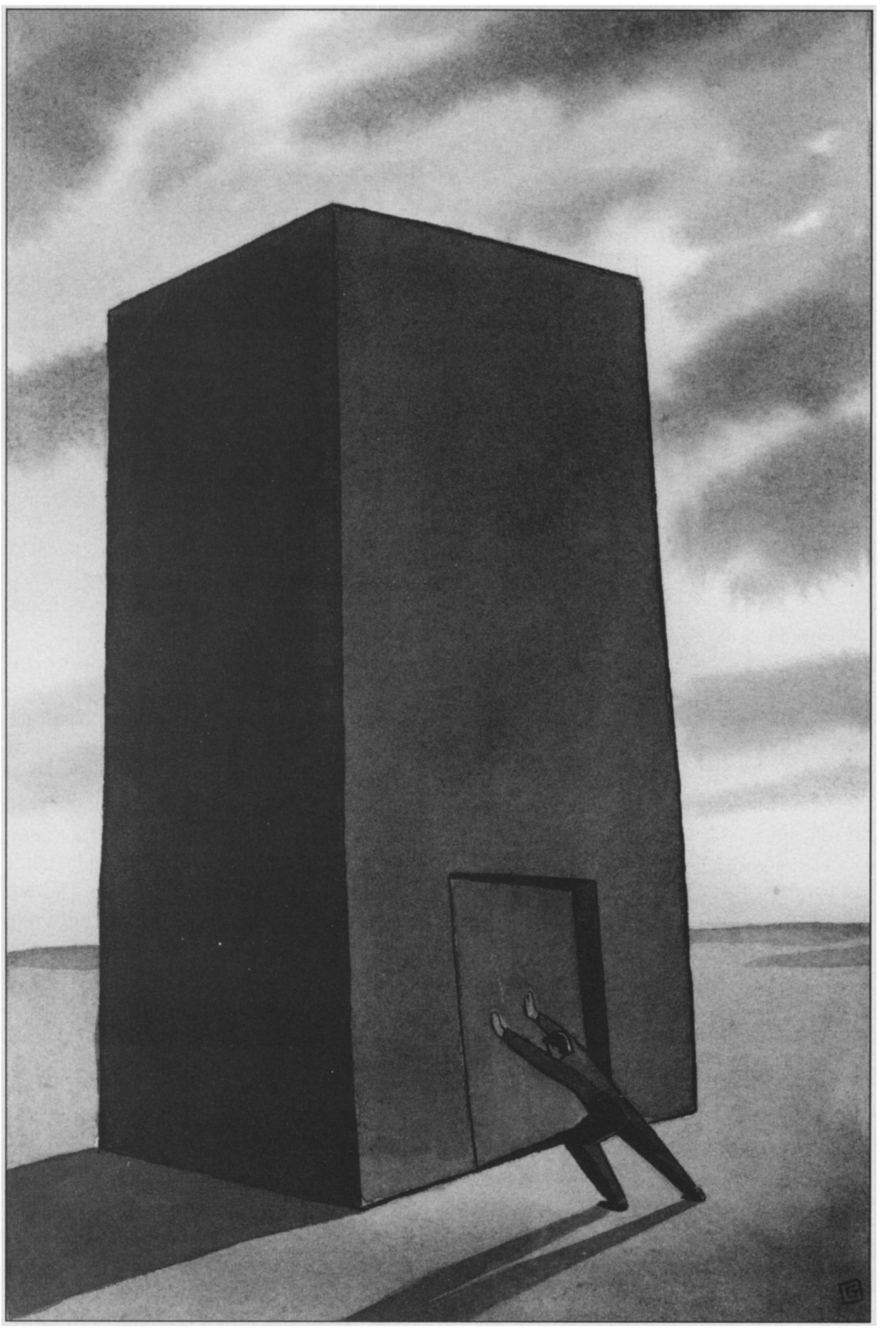


Illustration by A. J. Garces

Firm evidence shows that formative assessment is an essential component of classroom work and that its development can raise standards of achievement, Mr. Black and Mr. Wiliam point out. Indeed, they know of no other way of raising standards for which such a strong prima facie case can be made.

RAISING the standards of learning that are achieved through schooling is an important national priority. In recent years, governments throughout the world have been more and more vigorous in making changes in pursuit of this aim. National, state, and district standards; target setting; enhanced programs for the external testing of students' performance; surveys such as NAEP (National Assessment of Educational Progress) and TIMSS (Third International Mathematics and Science Study); initiatives to improve school plan-

PAUL BLACK is professor emeritus in the School of Education, King's College, London, where DYLAN WILIAM is head of school and professor of educational assessment.

This content downloaded from 128.120.194.194 on Thu, 17 Oct 2013 17:12:33 PM
All use subject to JSTOR Terms and Conditions

ning and management; and more frequent and thorough inspection are all means toward the same end. But the sum of all these reforms has not added up to an effective policy because something is missing.

Learning is driven by what teachers and pupils do in classrooms. Teachers have to manage complicated and demanding situations, channeling the personal, emotional, and social pressures of a group of 30 or more youngsters in order to help them learn immediately and become better learners in the future. Standards can be raised only if teachers can tackle this task more effectively. What is missing from the efforts alluded to above is any direct help with this task. This fact was recognized in the TIMSS video study: "A focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve."[1]

In terms of systems engineering, present policies in the U.S. and in many other countries seem to treat the classroom as a black box. Certain *inputs* from the outside — pupils, teachers, other resources, management rules and requirements, parental anxieties, standards, tests with high stakes, and so on — are fed into the box. Some *outputs* are supposed to follow: pupils who are more knowledgeable and competent, better test results, teachers who are reasonably satisfied, and so on. But what is happening inside the box? How can anyone be sure that a particular set of new inputs will produce better outputs if we don't at least study what happens inside? And why is it that most of the reform initiatives mentioned in the first paragraph are not aimed at giving direct help and support to the work of teachers in classrooms?

The answer usually given is that it is up to teachers: they have to make the inside work better. This answer is not good enough, for two reasons. First, it is at least possible that some changes in the inputs may be counterproductive and make it harder for teachers to raise standards. Second, it seems strange, even unfair, to leave the most difficult piece of the standards-raising puzzle entirely to teachers. If there are ways in which policy makers and others can give direct help and support to the everyday classroom task of achieving better learning, then surely these ways ought to be pursued vigorously.

This article is about the inside of the black box. We focus on one aspect of teaching: formative assessment. But we will show that this feature is at the heart of effective teaching.

The Argument

We start from the self-evident proposition that teaching and learning must be interactive. Teachers need to know about their pupils' progress and difficulties with learning so that they can adapt their own work to meet pupils' needs — needs that are often unpredictable and that vary from one pupil to another. Teachers can find out what they need to know in a variety of ways, including observation and discussion in the classroom and the reading of pupils' written work.

We use the general term *assessment* to refer to all those activities undertaken by teachers — and by their students in assessing themselves — that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes *formative assessment* when the evidence is actually used to adapt the teaching to meet student needs.[2]

There is nothing new about any of this. All teachers make assessments in every class they teach. But there are three important questions about this process that we seek to answer:

- Is there evidence that improving formative assessment raises standards?
- Is there evidence that there is room for improvement?
- Is there evidence about how to improve formative assessment?

In setting out to answer these questions, we have conducted an extensive survey of the research literature. We have checked through many books and through the past nine years' worth of issues of more than 160 journals, and we have studied earlier reviews of research. This process yielded about 580 articles or chapters to study. We prepared a lengthy review, using material from 250 of these sources, that has been published in a special issue of the journal *Assessment in Education*, together with comments on our work by leading educational experts from Australia, Switzerland, Hong Kong, Lesotho, and the U.S.[3]

The conclusion we have reached from our research review is that the answer to each of the three questions above is clearly yes. In the three main sections below, we outline the nature and force of the evidence that justifies this conclusion. However, because we are presenting a summary here, our text will appear strong on assertions and weak on the details of their justification. We maintain that these assertions are backed by evidence and that this backing is set out in full detail in the lengthy review on which this article is founded.

We believe that the three sections below establish a strong case that *governments, their agencies, school authorities, and the teaching profession should study very carefully whether they are seriously interested in raising standards in education.* However, we also acknowledge widespread evidence that fundamental change in education can be achieved only slowly — through programs of professional development that build on existing good practice. Thus we do not conclude that formative assessment is yet another "magic bullet" for education. The issues involved are too complex and too closely linked to both the difficulties of classroom practice and the beliefs that drive public policy. In a final section, we confront this complexity and try to sketch out a strategy for acting on our evidence.

Does Improving Formative Assessment Raise Standards?

A research review published in 1986, concentrating primarily on classroom assessment work for children with mild handicaps, surveyed a large number of innovations, from which 23 were selected.[4] Those chosen satisfied the condition that quantitative evidence of learning gains was obtained, both for those involved in the innovation and for a similar group not so involved. Since then, many more papers have been published describing similarly careful quantitative experiments. Our own review has selected at least 20 more studies. (The number depends on how rigorous a set of selection criteria are applied.) All these studies show that innovations that include strengthening the practice of formative assessment produce significant and often substantial learning gains. These studies range over age groups from 5-year-olds to university undergraduates, across several school subjects, and over several countries.

For research purposes, learning gains of this type are measured by comparing the average improvements in the test scores of pupils involved in an innovation with the range of scores that are found for typical groups of pupils on these same tests.

This content downloaded from 128.120.194.194 on Thu, 17 Oct 2013 17:12:33 PM
All use subject to JSTOR Terms and Conditions

The ratio of the former divided by the latter is known as the *effect size*. Typical effect sizes of the formative assessment experiments were between 0.4 and 0.7. These effect sizes are larger than most of those found for educational interventions. The following examples illustrate some practical consequences of such large gains.

• An effect size of 0.4 would mean that the average pupil involved in an innovation would record the same achievement as a pupil in the top 35% of those not so involved.

• An effect size gain of 0.7 in the recent international comparative studies in mathematics[5] would have raised the score of a nation in the middle of the pack of 41 countries (e.g., the U.S.) to one of the top five.

Many of these studies arrive at another important conclusion: that improved formative assessment helps low achievers more than other students and so reduces the range of achievement while raising achievement overall. A notable recent example is a study devoted entirely to low-achieving students and students with learning disabilities, which shows that frequent assessment feedback helps both groups enhance their learning.[6] Any gains for such pupils could be particularly important. Furthermore, pupils who come to see themselves as unable to learn usually cease to take school seriously. Many become disruptive; others resort to truancy. Such young people are likely to be alienated from society and to become the sources and the victims of serious social problems.

Thus it seems clear that very significant learning gains lie within our grasp. The fact that such gains have been achieved by a variety of methods that have, as a common feature, enhanced formative assessment suggests that this feature accounts, at least in part, for the successes. However, it does not follow that it would be an easy matter to achieve such gains on a wide scale in normal classrooms. Many of the reports we have studied raise a number of other issues.

• All such work involves new ways to enhance feedback between those taught and the teacher, ways that will require significant changes in classroom practice.

• Underlying the various approaches are assumptions about what makes for effective learning — in particular the assumption that students have to be actively involved.

• For assessment to function formatively, the results have to be used to adjust teaching and learning; thus a significant aspect of any program will be the ways in which teachers make these adjustments.

• The ways in which assessment can affect the motivation and self-esteem of pupils and the benefits of engaging pupils in self-assessment deserve careful attention.

"The food's really not half bad, but the atmosphere leaves a lot to be desired."

Is There Room for Improvement?

A poverty of practice. There is a wealth of research evidence that the everyday practice of assessment in classrooms is beset with problems and shortcomings, as the following selected quotations indicate.

• "Marking is usually conscientious but often fails to offer guidance on how work can be improved. In a significant minority of cases, marking reinforces underachievement and underexpectation by being too generous or unfocused. Information about pupil performance received by the teacher is insufficiently used to inform subsequent work," according to a United Kingdom inspection report on secondary schools.[7]

• "Why is the extent and nature of formative assessment in science so impoverished?" asked a research study on secondary science teachers in the United Kingdom.[8]

• "Indeed they pay lip service to [formative assessment] but consider that its practice is unrealistic in the present educational context," reported a study of Canadian secondary teachers.[9]

• "The assessment practices outlined above are not common, even though these kinds of approaches are now widely promoted in the professional literature," according to a review of assessment practices in U.S. schools.[10]

The most important difficulties with assessment revolve around three issues. The first issue is *effective learning*.

• The tests used by teachers encourage rote and superficial learning even when teachers say they want to develop understanding; many teachers seem unaware of the inconsistency.

• The questions and other methods teachers use are not shared with other teachers in the same school, and they are not critically reviewed in relation to what they actually assess.

• For primary teachers particularly, there is a tendency to emphasize quantity and presentation of work and to neglect its

This content downloaded from 128.120.194.194 on Thu, 17 Oct 2013 17:12:33 PM
All use subject to JSTOR Terms and Conditions

The ultimate user of assessment information that is elicited in order to improve learning is the pupil.

quality in relation to learning.

The second issue is *negative impact*.

• The giving of marks and the grading function are overemphasized, while the giving of useful advice and the learning function are underemphasized.

• Approaches are used in which pupils are compared with one another, the prime purpose of which seems to them to be competition rather than personal improvement; in consequence, assessment feedback teaches low-achieving pupils that they lack "ability," causing them to come to believe that they are not able to learn.

The third issue is the *managerial role* of assessments.

• Teachers' feedback to pupils seems to serve social and managerial functions, often at the expense of the learning function.

• Teachers are often able to predict pupils' results on external tests because their own tests imitate them, but at the same time teachers know too little about their pupils' learning needs.

• The collection of marks to fill in records is given higher priority than the analysis of pupils' work to discern learning needs; furthermore, some teachers pay no attention to the assessment records of their pupils' previous teachers.

Of course, not all these descriptions apply to all classrooms. Indeed, there are many schools and classrooms to which they do not apply at all. Nevertheless, these general conclusions have been drawn by researchers who have collected evidence — through observation, interviews, and questionnaires — from schools in several countries, including the U.S.

An empty commitment. The development of national assessment policy in England and Wales over the last decade illustrates the obstacles that stand in the way of developing policy support for formative assessment. The recommendations of a government task force in 1988[11] and all subsequent statements of government policy have emphasized the importance of formative assessment by teachers. However, the body charged with carrying out government policy on assessment had no strategy either to study or to develop the formative assessment of teachers and did no more than devote a tiny fraction of its resources to such work.[12] Most of the available resources and most of the public and political attention were focused on national external tests. While teachers' contributions to these "summative assessments" have been given some formal status, hardly any attention has been paid to their contributions through formative assessment. Moreover, the problems of the relationship between teachers' formative and summative roles have received no attention.

It is possible that many of the commitments were stated in the belief that formative assessment was not problematic, that it already happened all the time and needed no more than formal acknowledgment of its existence. However, it is also clear that the political commitment to external testing in order to promote competition had a central priority, while the commitment to formative assessment was marginal. As researchers the world over have found, high-stakes external tests always dominate teaching and assessment. However, they give teachers poor models for formative assessment because of their limited function of providing overall summaries of achievement rather than helpful diagnosis. Given this fact, it is hardly surprising that numerous research studies of the implementation of the education reforms in the United Kingdom have found that formative assessment is "seriously in need of development."[13] With hindsight, we can see that the failure to perceive the need for substantial support for formative assessment and to take responsibility for developing such support was a serious error.

In the U.S. similar pressures have been felt from political movements characterized by a distrust of teachers and a belief that external testing will, on its own, improve learning. Such fractured relationships between policy makers and the teaching profession are not inevitable — indeed, many countries with enviable educational achievements seem to manage well with policies that show greater respect and support for teachers. While the situation in the U.S. is far more diverse than that in England and Wales, the effects of high-stakes state-mandated testing are very similar to those of the external tests in the United Kingdom. Moreover, the traditional reliance on multiple-choice testing in the U.S. — not shared in the United Kingdom — has exacerbated the negative effects of such policies on the quality of classroom learning.

How Can We Improve Formative Assessment?

The self-esteem of pupils. A report of schools in Switzerland states that "a number of pupils . . . are content to 'get by.' . . . Every teacher who wants to practice formative assessment must reconstruct the teaching contracts so as to counteract the habits acquired by his pupils."[14]

The ultimate user of assessment information that is elicited in order to improve learning is the pupil. There are negative and positive aspects of this fact. The negative aspect is illustrated by the preceding quotation. When the classroom culture focuses on rewards, "gold stars," grades, or class ranking, then pupils look for ways to obtain the best marks rather than to improve their learning. One reported consequence is that, when they have any choice, pupils avoid difficult tasks. They also spend time and energy looking for clues to the "right answer." Indeed, many become reluctant to ask questions out of a fear of failure. Pupils who encounter difficulties are led to believe that they lack ability, and this belief leads them to attribute their difficulties to a defect in themselves about which they cannot do a great deal. Thus they avoid investing effort in learning that can lead only to disappointment, and they try to build up their self-esteem in other ways.

The positive aspect of students' being the primary users of the information gleaned from formative assessments is that negative outcomes — such as an obsessive focus on competition and the attendant fear of failure on the part of low achievers — are not inevitable. What is needed is a culture of success, backed by a belief that all pupils can achieve. In this regard, formative assessment can be a powerful weapon if it is communicated in the right way. While formative assessment can help all

This content downloaded from 128.120.194.194 on Thu, 17 Oct 2013 17:12:33 PM
All use subject to JSTOR Terms and Conditions

pupils, it yields particularly good results with low achievers by concentrating on specific problems with their work and giving them a clear understanding of what is wrong and how to put it right. Pupils can accept and work with such messages, provided that they are not clouded by overtones about ability, competition, and comparison with others. In summary, the message can be stated as follows: *feedback to any pupil should be about the particular qualities of his or her work, with advice on what he or she can do to improve, and should avoid comparisons with other pupils.*

Self-assessment by pupils. Many successful innovations have developed self- and peer-assessment by pupils as ways of enhancing formative assessment, and such work has achieved some success with pupils from age 5 upward. This link of formative assessment to self-assessment is not an accident; indeed, it is inevitable.

To explain this last statement, we should first note that the main problem that those who are developing self-assessments encounter is not a problem of reliability and trustworthiness. Pupils are generally honest and reliable in assessing both themselves and one another; they can even be too hard on themselves. The main problem is that pupils can assess themselves only when they have a sufficiently clear picture of the targets that their learning is meant to attain. Surprisingly, and sadly, many pupils do not have such a picture, and they appear to have become accustomed to receiving classroom teaching as an arbitrary sequence of exercises with no overarching rationale. To overcome this pattern of passive reception requires hard and sustained work. When pupils do acquire such an overview, they then become more committed and more effective as learners. Moreover, their own assessments become an object of discussion with their teachers and with one another, and this discussion further promotes the reflection on one's own thinking that is essential to good learning.

Thus self-assessment by pupils, far from being a luxury, is in fact *an essential component of formative assessment.* When anyone is trying to learn, feedback about the effort has three elements: recognition of the *desired goal*, evidence about *present position*, and some understanding of a *way to close the gap* between the two.[15] All three must be understood to some degree by anyone before he or she can take action to improve learning.

Such an argument is consistent with more general ideas established by research into the way people learn. New understandings are not simply swallowed and stored in isolation; they have to be assimilated in relation to preexisting ideas. The new and the old may be inconsistent or even in conflict, and the disparities must be resolved by thoughtful actions on the part of the learner. Realizing that there are new goals for the learning is an essential part of this process of assimilation. Thus we conclude: *if formative assessment is to be productive, pupils should be trained in self-assessment so that they can understand the main purposes of their learning and thereby grasp what they need to do to achieve.*

Dialogue with the teacher provides the opportunity for the teacher to respond to and reorient a pupil's thinking.

The evolution of effective teaching. The research studies referred to above show very clearly that effective programs of formative assessment involve far more than the addition of a few observations and tests to an existing program. They require careful scrutiny of all the main components of a teaching plan. Indeed, it is clear that instruction and formative assessment are indivisible.

To begin at the beginning, the choice of tasks for classroom work and homework is important. Tasks have to be justified in terms of the learning aims that they serve, and they can work well only if opportunities for pupils to communicate their evolving understanding are built into the planning. Discussion, observation of activities, and marking of written work can all be used to provide those opportunities, but it is then important to look at or listen carefully to the talk, the writing, and the actions through which pupils develop and display the state of their understanding. Thus we maintain that *opportunities for pupils to express their understanding should be designed into any piece of teaching, for this will initiate the interaction through which formative assessment aids learning.*

Discussions in which pupils are led to talk about their understanding in their own ways are important aids to increasing knowledge and improving understanding. Dialogue with the teacher provides the opportunity for the teacher to respond to and reorient a pupil's thinking. However, there are clearly recorded examples of such discussions in which teachers have, quite unconsciously, responded in ways that would inhibit the future learning of a pupil. What the examples have in common is that the teacher is looking for a particular response and lacks the flexibility or the confidence to deal with the unexpected. So the teacher tries to direct the pupil toward giving the expected answer. In manipulating the dialogue in this way, the teacher seals off any unusual, often thoughtful but unorthodox, attempts by pupils to work out their own answers. Over time the pupils get the message: they are not required to think out their own answers. The object of the exercise is to work out — or guess — what answer the teacher expects to see or hear.

A particular feature of the talk between teacher and pupils is the asking of questions by the teacher. This natural and direct way of checking on learning is often unproductive. One common problem is that, following a question, teachers do not wait long enough to allow pupils to think out their answers. When a teacher answers his or her own question after only two or three seconds and when a minute of silence is not tolerable, there is no possibility that a pupil can think out what to say.

There are then two consequences. One is that, because the only questions that can produce answers in such a short time are questions of fact, these predominate. The other is that pupils don't even try to think out a response. Because they know that the answer, followed by another question, will come along in a few seconds, there is no point in trying. It is also generally the case that only a few pupils in a class answer the teacher's questions. The rest then leave it to these few, knowing that they cannot respond as quickly and being unwilling to risk making mistakes in public. So the teacher, by lowering the level

This content downloaded from 128.120.194.194 on Thu, 17 Oct 2013 17:12:33 PM
All use subject to JSTOR Terms and Conditions

Tests given in class and tests and other exercises assigned for homework are also important means of promoting feedback.

of questions and by accepting answers from a few, can keep the lesson going but is actually out of touch with the understanding of most of the class. The question/answer dialogue becomes a ritual, one in which thoughtful involvement suffers.

There are several ways to break this particular cycle. They involve giving pupils time to respond; asking them to discuss their thinking in pairs or in small groups, so that a respondent is speaking on behalf of others; giving pupils a choice between different possible answers and asking them to vote on the options; asking all of them to write down an answer and then reading out a selected few; and so on. What is essential is that any dialogue should evoke thoughtful reflection in which all pupils can be encouraged to take part, for only then can the formative process start to work. In short, the dialogue between pupils and a teacher should be *thoughtful, reflective, focused to evoke and explore understanding, and conducted so that all pupils have an opportunity to think and to express their ideas.*

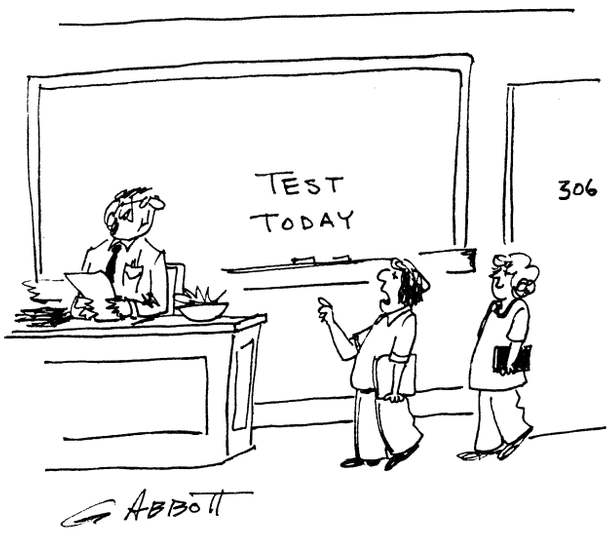
Tests given in class and tests and other exercises assigned for homework are also important means of promoting feedback. A good test can be an occasion for learning. It is better to have frequent short tests than infrequent long ones. Any new learning should first be tested within about a week of a first encounter, but more frequent tests are counterproductive. The quality of the test items — that is, their relevance to the main learning aims and their clear communication to the pupil — requires scrutiny as well. Good questions are hard to generate, and teachers should collaborate and draw on outside sources to collect such questions.

Given questions of good quality, it is essential to ensure the quality of the feedback. Research studies have shown that, if pupils are given only marks or grades, they do not benefit from the feedback. The worst scenario is one in which some pupils who get low marks this time also got low marks last time and come to expect to get low marks next time. This cycle of repeated failure becomes part of a shared belief between such students and their teacher. Feedback has been shown to improve learning when it gives each pupil specific guidance on strengths and weaknesses, preferably without any overall marks. Thus the way in which test results are reported to pupils so that they can identify their own strengths and weaknesses is critical. Pupils must be given the means and opportunities to work with evidence of their difficulties. For formative purposes, a test at the end of a unit or teaching module is pointless; it is too late to work with the results. We conclude that *the feedback on tests, seatwork, and homework should give each pupil guidance on how to improve, and each pupil must be given help and an opportunity to work on the improvement.*

All these points make clear that there is no one simple way to improve formative assessment. What is common to them is that a teacher's approach should start by being realistic and confronting the question "Do I really know enough about the understanding of my pupils to be able to help each of them?"

Much of the work teachers must do to make good use of formative assessment can give rise to difficulties. Some pupils will resist attempts to change accustomed routines, for any such change is uncomfortable, and emphasis on the challenge to think for yourself (and not just to work harder) can be threatening to many. Pupils cannot be expected to believe in the value of changes for their learning before they have experienced the benefits of such changes. Moreover, many of the initiatives that are needed take more class time, particularly when a central purpose is to change the outlook on learning and the working methods of pupils. Thus teachers have to take risks in the belief that such investment of time will yield rewards in the future, while "delivery" and "coverage" with poor understanding are pointless and can even be harmful.

Teachers must deal with two basic issues that are the source of many of the problems associated with changing to a system of formative assessment. The first is *the nature of each teacher's beliefs about learning*. If the teacher assumes that knowledge is to be transmitted and learned, that understanding will develop later, and that clarity of exposition accompanied by rewards for patient reception are the essentials of good teaching, then formative assessment is hardly necessary. However, most teachers accept the wealth of evi-



"It has been said that a fool can ask more questions than a wise man can answer."

This content downloaded from 128.120.194.194 on Thu, 17 Oct 2013 17:12:33 PM
All use subject to JSTOR Terms and Conditions

dence that this transmission model does not work, even when judged by its own criteria, and so are willing to make a commitment to teaching through interaction. Formative assessment is an essential component of such instruction. We do not mean to imply that individualized, one-on-one teaching is the only solution; rather we mean that what is needed is a classroom culture of questioning and deep thinking, in which pupils learn from shared discussions with teachers and peers. What emerges very clearly here is the indivisibility of instruction and formative assessment practices.

The other issue that can create problems for teachers who wish to adopt an interactive model of teaching and learning relates to *the beliefs teachers hold about the potential of all their pupils for learning*. To sharpen the contrast by overstating it, there is on the one hand the "fixed I.Q." view — a belief that each pupil has a fixed, inherited intelligence that cannot be altered much by schooling. On the other hand, there is the "untapped potential" view — a belief that starts from the assumption that so-called ability is a complex of skills that can be learned. Here, we argue for the underlying belief that all pupils can learn more effectively if one can clear away, by sensitive handling, the obstacles to learning, be they cognitive failures never diagnosed or damage to personal confidence or a combination of the two. Clearly the truth lies between these two extremes, but the evidence is that *ways of managing formative assessment that work with the assumptions of "untapped potential" do help all pupils to learn and can give particular help to those who have previously struggled.*

Policy and Practice

Changing the policy perspective. The assumptions that drive national and state policies for assessment have to be called into question. The promotion of testing as an important component for establishing a competitive market in education can be very harmful. The more recent shifting of emphasis toward setting targets for all, with assessment providing a touchstone to help check pupils' attainments, is a more mature position. However, we would argue that *there is a need now to move further, to focus on the inside of the "black box" and so to explore the potential of assessment to raise standards directly as an integral part of each pupil's learning work.*

It follows from this view that several changes are needed. First, policy ought to start with a recognition that the prime locus for raising standards is the classroom, so that the overarching priority has to be the promotion and support of change within the classroom. Attempts to raise standards by reforming the inputs to and measuring the outputs from the black box of the classroom can be helpful, but they are not adequate on their own. Indeed, their helpfulness can be judged only in light of their effects in classrooms.

The evidence we have presented here establishes that a clearly productive way to start implementing a classroom-focused policy would be to improve formative assessment. This same evidence also establishes that in doing so we would not be concentrating on some minor aspect of the business of teaching and learning. Rather, we would be concentrating on several essential elements: the quality of teacher/pupil interactions, the stimulus and help for pupils to take active responsibility for their own learning, the particular help needed to move pupils out of the trap of "low achievement," and the development of the habits necessary for all students to become lifelong learners. Improvements in formative assessment, which are within the reach of all teachers, can contribute substantially to raising standards in all these ways.

Four steps to implementation. If we accept the argument outlined above, what needs to be done? The proposals outlined below do not follow directly from our analysis of assessment research. They are consistent with its main findings, but they also call on more general sources for guidance.[16]

At one extreme, one might call for more research to find out how best to carry out such work; at the other, one might call for an immediate and large-scale program, with new guidelines that all teachers should put into practice. Neither of these alternatives is sensible: while the first is unnecessary because enough is known from the results of research, the second would be unjustified because not enough is known about classroom practicalities in the context of any one country's schools.

Thus the improvement of formative assessment cannot be a simple matter. There is no quick fix that can alter existing practice by promising rapid rewards. On the contrary, if the substantial rewards promised by the research evidence are to be secured, each teacher must find his or her own ways of incorporating the lessons and ideas set out above into his or her own patterns of classroom work and into the cultural norms and expectations of a particular school community.[17] This process is a relatively slow one and takes place through sustained programs of professional development and support. This fact does not weaken the message here; indeed, it should be seen as a sign of its authenticity, for lasting and fundamental improvements in teaching and learning must take place in this way. A recent international study of innovation and change in education, encompassing 23 projects in 13 member countries of the Organisation for Economic Co-operation and Development, has arrived at exactly the same conclusion with regard to effective policies for change.[18] Such arguments lead us to propose a four-point scheme for teacher development.

1. *Learning from development.* Teachers will not take up ideas that sound attractive, no matter how extensive the research base, if the ideas are presented as general principles that leave the task of translating them into everyday practice entirely up to the teachers. Their classroom lives are too busy and too fragile for all but an outstanding few to undertake such work. What teachers need is a variety of living examples of implementation, as practiced by teachers with whom they can identify and from whom they can derive the confidence that they can do better. They need to see examples of what doing better means in practice.

So changing teachers' practice cannot begin with an extensive program of training for all; that could be justified only if it could be claimed that we have enough "trainers" who know what to do, which is certainly not the case. The essential first step is to set up a small number of local groups of schools — some primary, some secondary, some inner-city, some from outer suburbs, some rural — with each school committed both to a school-based development of formative assessment and to collaboration with other schools in its local group. In such a process, the teachers in their classrooms will be working out the answers to many of the practical questions that the evidence presented here cannot answer. They will be reformulating the issues, perhaps in relation to fundamental insights and certainly in terms that make sense to their peers in other class-

This content downloaded from 128.120.194.194 on Thu, 17 Oct 2013 17:12:33 PM
All use subject to JSTOR Terms and Conditions

rooms. It is also essential to carry out such development in a range of subject areas, for the research in mathematics education is significantly different from that in language, which is different again from that in the creative arts.

The schools involved would need extra support in order to give their teachers time to plan the initiative in light of existing evidence, to reflect on their experience as it develops, and to offer advice about training others in the future. In addition, there would be a need for external evaluators to help the teachers with their development work and to collect evidence of its effectiveness. Video studies of classroom work would be essential for disseminating findings to others.

2. *Dissemination.* This dimension of the implementation would be in low gear at the outset — offering schools no more than general encouragement and explanation of some of the relevant evidence that they might consider in light of their existing practices. Dissemination efforts would become more active as results and resources became available from the development program. Then strategies for wider dissemination — for example, earmarking funds for inservice training programs — would have to be pursued.

We must emphasize that this process will inevitably be a slow one. To repeat what we said above, *if the substantial rewards promised by the evidence are to be secured, each teacher must find his or her own ways of incorporating the lessons and ideas that are set out above into his or her own patterns of classroom work.* Even with optimum training and support, such a process will take time.

3. *Reducing obstacles.* All features in the education system that actually obstruct the development of effective formative assessment should be examined to see how their negative effects can be reduced. Consider the conclusions from a study of teachers of English in U.S. secondary schools.

> Most of the teachers in this study were caught in conflicts among belief systems and institutional structures, agendas, and values. The point of friction among these conflicts was assessment, which was associated with very powerful feelings of being overwhelmed, and of insecurity, guilt, frustration, and anger. . . . This study suggests that assessment, as it occurs in schools, is far from a merely technical problem. Rather, it is deeply social and personal.[19]

The chief negative influence here is that of short external tests. Such tests can dominate teachers' work, and, insofar as they encourage drilling to produce right answers to short, out-of-context questions, they can lead teachers to act against their own better judgment about the best ways to develop the learning of their pupils. This is not to argue that all such tests are unhelpful. Indeed, they have an important role to play in securing public confidence in the accountability of schools. For the immediate future, what is needed in any development program for formative assessment is to study the interactions between these external tests and formative assessments to see how the models of assessment that external tests can provide could be made more helpful.

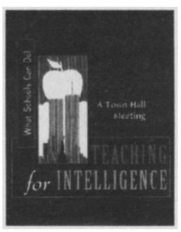
All teachers have to undertake some summative assessment. They must report to parents and produce end-of-year reports as classes are due to move on to new teachers. However, the task of assessing

Now Available on Videotape

THE FOURTH INTERNATIONAL

TEACHING FOR INTELLIGENCE CONFERENCE

In April 1998, the world's educational leaders on teaching and instruction gathered to discuss student achievement. More than 2,200 educators from across the nation came to New York City to focus on strategies and practices that promote teaching for intelligence.



Town Hall Meeting
Live Satellite Broadcast
Item #1683
$299.00 with facilitator's guide

HOWARD GARDNER, JAMES COMER, LINDA DARLING-HAMMOND and 20 more education experts shared their insight on the nature of intelligence and how to transfer this knowledge to the classroom.

SkyLight
Training and Publishing Inc.

Meet the Experts Interviews

Kay Burke
Performance Assessment
#1689 (12 minutes)

Arthur Costa
Habits of Minds/Dispositions of Thinking
#1696 (14 minutes)

Elliot Eisner
The Arts and the Aesthetic Intelligence
#1692 (12 minutes)

Robin Fogarty
The Brain-Compatible Classroom
#1698 (14 minutes)

Herb Ginsburg
Early Childhood Education
#1694 (10 minutes)

John Goodlad
School Reform and Renewal
#1690 (10 minutes)

Donna Ogle
Reading Strategies
#1695 (12 minutes)

Theodore Sizer
Reshaping High Schools
#1693 (16 minutes)

Robert Sternberg
Successful Intelligence
#1691 (13 minutes)

Uri Triesman
School Improvement
#1697 (12 minutes)

$19.95 for each videotape or $179 for 10–tape package

CALL 800-348-4474 OR 847-290-6600 TO ORDER.

This content downloaded from 128.120.194.194 on Thu, 17 Oct 2013 17:12:33 PM
All use subject to JSTOR Terms and Conditions

pupils summatively for external purposes is clearly different from the task of assessing ongoing work to monitor and improve progress. Some argue that these two roles are so different that they should be kept apart. We do not see how this can be done, given that teachers must have some share of responsibility for the former and must take the leading responsibility for the latter.[20] However, teachers clearly face difficult problems in reconciling their formative and summative roles, and confusion in teachers' minds between these roles can impede the improvement of practice.

The arguments here could be taken much further to make the case that teachers should play a far greater role in contributing to summative assessments for accountability. One strong reason for giving teachers a greater role is that they have access to the performance of their pupils in a variety of contexts and over extended periods of time.

This is an important advantage because sampling pupils' achievement by means of short exercises taken under the conditions of formal testing is fraught with dangers. It is now clear that performance in any task varies with the context in which it is presented. Thus some pupils who seem incompetent in tackling a problem under test conditions can look quite different in the more realistic conditions of an everyday encounter with an equivalent problem. Indeed, the conditions under which formal tests are taken threaten validity because they are quite unlike those of everyday performance. An outstanding example here is that collaborative work is very important in everyday life but is forbidden by current norms of formal testing.[21] These points open up wider arguments about assessment systems as a whole — arguments that are beyond the scope of this article.

4. *Research.* It is not difficult to set out a list of questions that would justify further research in this area. Although there are many and varied reports of successful innovations, they generally fail to give clear accounts of one or another of the important details. For example, they are often silent about the actual classroom methods used, the motivation and experience of the teachers, the nature of the tests used as measures of success, or the outlooks and expectations of the pupils involved.

However, while there is ample justification for proceeding with carefully formulated projects, we do not suggest that everyone else should wait for their conclusions. Enough is known to provide a basis for active development work, and some of the most important questions can be answered only through a program of practical implementation.

Directions for future research could include a study of the ways in which teachers understand and deal with the relationship between their formative and summative roles or a comparative study of the predictive validity of teachers' summative assessments versus external test results. Many more questions could be formulated, and it is important for future development that some of these problems be tackled by basic research. At the same time, experienced researchers would also have a vital role to play in the evaluation of the development programs we have proposed.

Are We Serious About Raising Standards?

The findings summarized above and the program we have outlined have implications for a variety of responsible agencies. However, it is the responsibility of governments to take the lead. It would be premature and out of order for us to try to consider the relative roles in such an effort, although success would clearly depend on cooperation among government agencies, academic researchers, and school-based educators.

The main plank of our argument is that standards can be raised only by changes that are put into direct effect by teachers and pupils in classrooms. There is a body of firm evidence that formative assessment is an essential component of classroom work and that its development can raise standards of achievement. We know of no other way of raising standards for which such a strong prima facie case can be made. Our plea is that national and state policy makers will grasp this opportunity and take the lead in this direction.

---

1. James W. Stigler and James Hiebert, "Understanding and Improving Classroom Mathematics Instruction: An Overview of the TIMSS Video Study," *Phi Delta Kappan*, September 1997, pp. 19-20.

2. There is no internationally agreed-upon term here. "Classroom evaluation," "classroom assessment," "internal assessment," "instructional assessment," and "student assessment" have been used by different authors, and some of these terms have different meanings in different texts.

3. Paul Black and Dylan Wiliam, "Assessment and Classroom Learning," *Assessment in Education*, March 1998, pp. 7-74.

4. Lynn S. Fuchs and Douglas Fuchs, "Effects of Systematic Formative Evaluation: A Meta-Analysis," *Exceptional Children*, vol. 53, 1986, pp. 199-208.

5. See Albert E. Beaton et al., *Mathematics Achievement in the Middle School Years* (Boston: Boston College, 1996).

6. Lynn S. Fuchs et al., "Effects of Task-Focused Goals on Low-Achieving Students with and Without Learning Disabilities," *American Educational Research Journal*, vol. 34, 1997, pp. 513-43.

7. OFSTED (Office for Standards in Education), *Subjects and Standards: Issues for School Development Arising from OFSTED Inspection Findings 1994-5: Key Stages 3 and 4 and Post-16* (London: Her Majesty's Stationery Office, 1996), p. 40.

8. Nicholas Daws and Birendra Singh, "Formative Assessment: To What Extent Is Its Potential to Enhance Pupils' Science Being Realized?," *School Science Review*, vol. 77, 1996, p. 99.

9. Clement Dassa, Jesús Vazquez-Abad, and Djavid Ajar, "Formative Assessment in a Classroom Setting: From Practice to Computer Innovations," *Alberta Journal of Educational Research*, vol. 39, 1993, p. 116.

10. D. Monty Neill, "Transforming Student Assessment," *Phi Delta Kappan*, September 1997, pp. 35-36.

11. *Task Group on Assessment and Testing: A Report* (London: Department of Education and Science and the Welsh Office, 1988).

12. Richard Daugherty, *National Curriculum Assessment: A Review of Policy, 1987-1994* (London: Falmer Press, 1995).

13. Terry A. Russell, Anne Qualter, and Linda McGuigan, "Reflections on the Implementation of National Curriculum Science Policy for the 5-14 Age Range: Findings and Interpretations from a National Evaluation Study in England," *International Journal of Science Education*, vol. 17, 1995, pp. 481-92.

14. Phillipe Perrenoud, "Towards a Pragmatic Approach to Formative Evaluation," in Penelope Weston, ed., *Assessment of Pupils' Achievement: Motivation and School Success* (Amsterdam: Swets and Zeitlinger, 1991), p. 92.

15. D. Royce Sadler, "Formative Assessment and the Design of Instructional Systems," *Instructional Science*, vol. 18, 1989, pp. 119-44.

16. Paul J. Black and J. Myron Atkin, *Changing the Subject: Innovations in Science, Mathematics, and Technology Education* (London: Routledge for the Organisation for Economic Co-operation and Development, 1996); and Michael G. Fullan, with Suzanne Stiegelbauer, *The New Meaning of Educational Change* (London: Cassell, 1991).

17. See Stigler and Hiebert, pp. 19-20.

18. Black and Atkin, op. cit.

19. Peter Johnston et al., "Assessment of Teaching and Learning in Literature-Based Classrooms," *Teaching and Teacher Education*, vol. 11, 1995, p. 359.

→ Dylan Wiliam and Paul Black, "Meanings and Consequences: A Basis for Distinguishing Formative and Summative Functions of Assessment," *British Educational Research Journal*, vol. 22, 1996, pp. 537-48.

21. These points are developed in some detail in Sam Wineburg, "T. S. Eliot, Collaboration, and the Quandaries of Assessment in a Rapidly Changing World," *Phi Delta Kappan*, September 1997, pp. 59-65.

This content downloaded from 128.120.194.194 on Thu, 17 Oct 2013 17:12:33 PM
All use subject to JSTOR Terms and Conditions

How Students Should be Taught Mathematics:

Reflections from Research and Practice.

Jo Boaler, Professor of Mathematics Education, Stanford University.

Mathematics classrooms should be places where students:

Develop an inquiry relationship with mathematics, approaching math with curiosity, courage, confidence & intuition.

Talk to each other and the teacher about ideas – Why did I choose this method? Does it work in other cases? How is the method similar or different to methods other people used?

Work on mathematics tasks that can be solved in different ways and/or with different solutions.

Work on mathematics tasks with a low entry point but a very high ceiling – so that students are constantly challenged and working at the highest and most appropriate level for them.

Work on mathematics tasks that are complex, involve more than one method or area of mathematics, and that often, but not always, represent real world problems and applications.

Are given growth mindset messages at all times, through the ways they are grouped together, the tasks they work on, the messages they hear, and the assessment and grading.

Are assessed formatively – to inform learning - not summatively to give a rank with their peers. Students should regularly receive diagnostic feedback on their work, instead of grades or scores. Summative assessments are best used at the end of courses.

Mathematics classrooms should be places where students believe:

Everyone can do well in math.

Mathematics problems can be solved with many different insights and methods.

Mistakes are valuable, they encourage brain growth and learning.

Mathematics will help them in their lives, not because they will see the same types of problems in the real world but because they are learning to think quantitatively and abstractly and developing an inquiry relationship with math.

MAT.08.ER.3.0000F.E.205 Claim 3

Sample Item ID:	MAT.08.ER.3.0000F.E.205
Grade:	08
Claim(s):	**Claim 3: Communicating Reasoning** Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others.
Secondary Claim(s):	Claim 1: Concepts and Procedures Students can explain and apply mathematical concepts and carry out mathematical procedures with precision and fluency.
Primary Content Domain:	Functions
Secondary Content Domain(s):	
Assessment Target(s):	3 E: Distinguish correct logic or reasoning from that which is flawed, and—if there is a flaw in the argument—explain what it is. 1 E: Define, evaluate, and compare functions. 1 A (Gr 7): Analyze proportional relationships and use them to solve real-world and mathematical problems. 3 F: Base arguments on concrete referents such as objects, drawings, diagrams, and actions.
Standard(s):	8.F.3, 7.RP.3
Mathematical Practice(s):	1, 2, 4, 7
DOK:	3
Item Type:	ER
Score Points:	2
Difficulty:	M
Key:	See Sample Top-Score Response.
Stimulus/Source:	
Target-Specific Attributes (e.g., accessibility issues):	Calculators may be used on this item.
Notes:	Part of PT set.

Samir was assigned to write an example of a linear functional relationship. He wrote this example for the assignment.

> The relationship between the year and the population of a county when the population increases by 10% each year

Part A

Complete the table below to create an example of the population of a certain county that is increasing by 10% each year.

Year	Population of a Certain County
0	
1	
2	
3	
4	

Part B

State whether Samir's example represents a linear functional relationship. Explain your reasoning.

Sample Top-Score Response:

Part A

Year	Population of a Certain County
0	100,000
1	110,000
2	121,000
3	133,100
4	146,410

Part B

Samir's example is not a linear functional relationship. The population does not increase by the same amount each year, so the relationship is not linear.

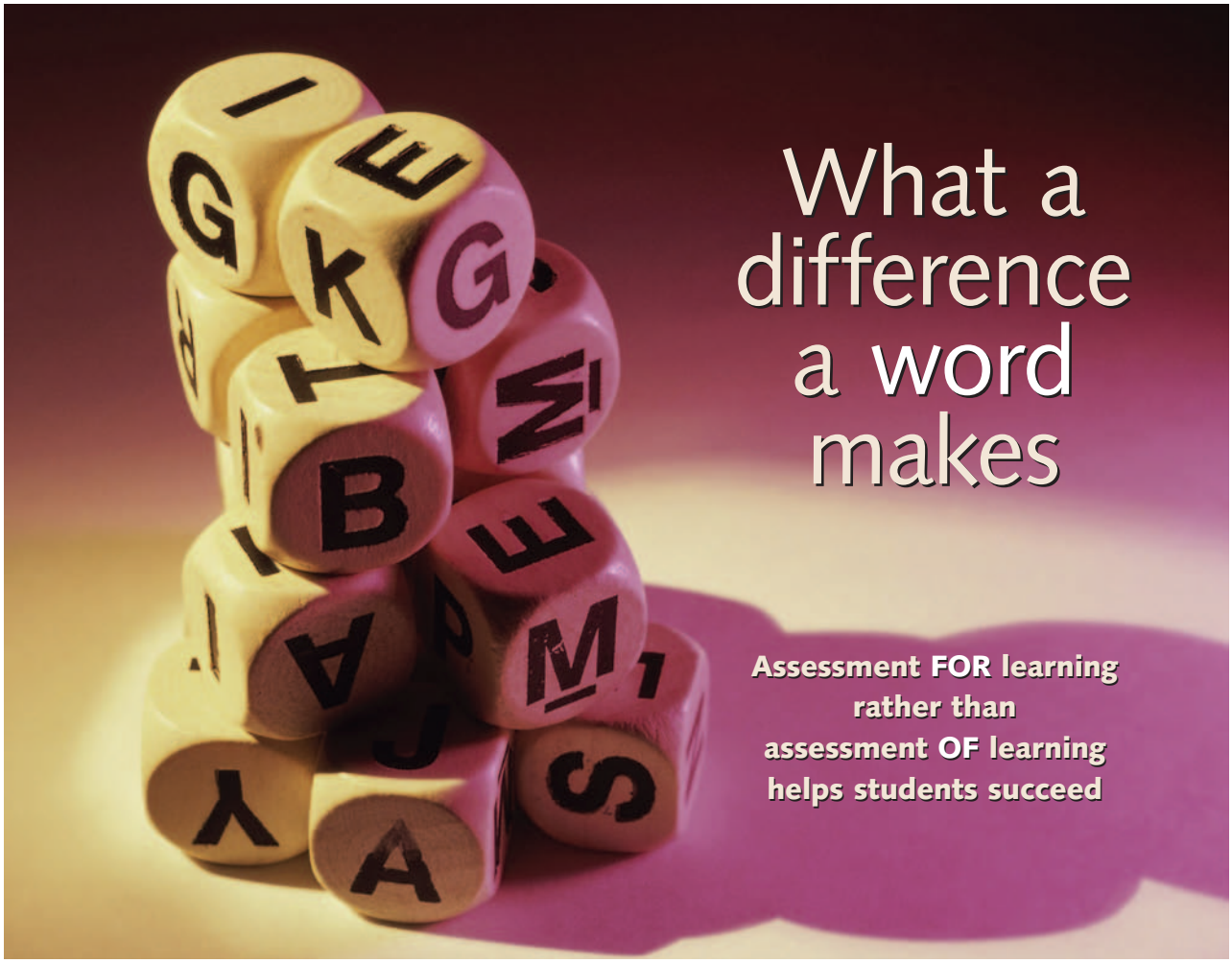
Scoring Rubric:

Responses to this item will receive 0-2 points, based on the following:

2 points: The student shows a thorough understanding of how to distinguish between linear and nonlinear relationships. The student correctly completes the table showing a 10% increase in population each year, states that the relationship is not linear, and provides a clear and complete explanation of why the relationship is not linear. Rounding to the nearest whole number for the population is permitted.

1 point: The student shows a partial understanding of how to distinguish between linear and nonlinear relationships. The student makes calculation errors in the table but gives a response in *Part B* that corresponds with the numbers in the table OR the student completes the table correctly but provides an incorrect explanation in *Part A*.

0 points: The student shows inconsistent or no understanding of how to distinguish between linear and nonlinear relationships. The student makes major errors when completing the table and gives a response in *Part B* that does not correspond to the values entered in the table.



BY RICK STIGGINS AND JAN CHAPPUIS

Without question, assessment remains among the very hottest topics in school improvement. High-stakes state accountability assessments and adequate yearly progress continue to represent the driving forces of school improvement these days. But, as accountability systems evolve, attention to this topic has turned in an interesting direction. Educators have concluded that testing once a year does not provide sufficient evidence to inform many crucial, more frequently made instructional decisions, which has generated renewed interest in *formative assessment*.

Traditionally, the term has referred to assessments used to support learning. But, in the current environment, formative assessment as defined by the test publishers has taken on a narrow meaning. In this context, it refers to a system of more frequent summative assessments administered at regular intervals (often quarterly) to determine which students have not yet met state standards — an early warning system, if you will.

We both applaud and, at the same time, challenge this thinking. On the

one hand, it helps us identify students who need help when we still have time to help them. On the other hand, while this very expensive assessment process helps us identify the problem, it doesn't help those students find greater success. For that, we must expand our definition. Enter *assessment for learning*.

Assessment *for* learning happens in the classroom and involves students in every aspect of their own assessment to build their confidence and maximize their achievement. It rests on the understanding that students, not just adults, are data-driven instructional decision makers. Several key features differentiate assessment *for* learning from formative assessment as currently being sold by test publishers: To begin with, state standards are deconstructed into classroom-level learning targets, which we translate into language our students understand so they know what they are responsible for learning. In addition, we turn those classroom-level targets into dependably accurate classroom assessments, aspects of which we integrate into daily instruction. In short, everyone understands the definition of success from the outset and we generate an ongoing flow of descriptive feedback that permits students to watch themselves grow. In this case, students and their teachers become partners in the classroom assessment process, relying on student-involved assessment, record keeping, and communication to help students understand what success looks like, see where they are now, and learn to close the gap between the two.

The good news is that research has shown for years that consistently applying principles of assessment *for* learning has yielded remarkable, if not unprecedented, gains in student achievement, especially for low achievers (Black & Wiliam, 1998). Results verify positive impacts across grade levels and school subjects.

However, the troubling news is that we weren't given the opportunity to learn to apply principles of assessment *for* learning during our preparation to teach. It remains the case that colleges of education often fail to include this kind of assessment training in their programs. And lest we believe that teachers can turn to their principals for assistance in this regard, be advised that assessment training of any sort remains virtually nonexistent in leadership training programs across the nation.

We know what teachers need to know and understand to apply principles of assessment *for* learning effectively in their classrooms. We know what will happen to their students' confidence, motivation, and achievement if they learn those lessons. We know how to deliver these tools to their hands in an efficient and effective manner.

Research has shown that consistently applying principles of assessment *for* learning has yielded remarkable, if not unprecedented, gains in student achievement.

RICK STIGGINS is the founder of Assessment Training Institute and JAN CHAPPUIS is a professional development associate at ATI. You can contact them at the Assessment Training Institute, 317 S.W. Alder St., Suite 1200, Portland, OR 97204, phone (800) 480-3060, fax (503) 228-3014, e-mail: ati@assessmentinst.com.

Competence in assessment *for* learning

The chart on p. 12 details five keys to classroom assessment quality, with each broken down into specific competencies teachers need to master

Sound classroom assessment practice

1. Clear purposes Assessment processes and results serve clear and appropriate purposes.	a. Teachers understand who uses classroom assessment information and know their information needs. b. Teachers understand the relationship between assessment and student motivation and craft assessment experiences to maximize motivation. c. Teachers use classroom assessment processes and results formatively (assessment *for* learning). d. Teachers use classroom assessment results summatively (assessment *of* learning) to inform someone beyond the classroom about students' achievement at a particular point in time. e. Teachers have a comprehensive plan over time for integrating assessment *for* and *of* learning in the classroom.
2. Clear targets Assessments reflect clear and valued student learning targets.	a. Teachers have clear learning targets for students; they know how to turn broad statements of content standards into classroom-level learning targets. b. Teachers understand the various types of learning targets they hold for students. c. Teachers select learning targets focused on the most important things students need to know and be able to do. d. Teachers have a comprehensive plan over time for assessing learning targets.
3. Sound design Learning targets are translated into assessments that yield accurate results.	a. Teachers understand the various assessment methods. b. Teachers choose assessment methods that match intended learning targets. c. Teachers design assessments that serve intended purposes. d. Teachers sample learning appropriately in their assessments. e. Teachers write assessment questions of all types well. f. Teachers avoid sources of mismeasurement that bias results.
4. Effective communication Assessment results are managed well and communicated effectively.	a. Teachers record assessment information accurately, keep it confidential, and appropriately combine and summarize it for reporting (including grades). Such summary accurately reflects current level of student learning. b. Teachers select the best reporting option (grades, narratives, portfolios, conferences) for each context (learning targets and users). c. Teachers interpret and use standardized test results correctly. d. Teachers effectively communicate assessment results to students. e. Teachers effectively communicate assessment results to a variety of audiences outside the classroom, including parents, colleagues, and other stakeholders.
5. Student involvement Students are involved in their own assessment.	a. Teachers make learning targets clear to students. b. Teachers involve students in assessing, tracking, and setting goals for their own learning. c. Teachers involve students in communicating about their own learning.

SOURCE: *Classroom Assessment for Student Learning: Doing it Right—Using it Well* by Richard Stiggins, Judy Arter, Jan Chappuis, and Steve Chappuis. (Portland, OR: Assessment Training Institute, 2004). Reprinted with permission.

to tap the full potential of assessment *for* learning (Stiggins, Arter, Chappuis, & Chappuis, 2004).

First, we need to know why we're assessing. If assessment is the process of gathering evidence to inform instructional decisions, teachers must begin the assessment process by asking:

- What decisions?
- Who's making the decisions?
- What kind of information will be helpful?

The assessment must produce that information, and it must take into account the needs of the student as a crucial decision maker.

Second, quality assessments can arise only from a clear vision of the achievement to be mastered. We cannot dependably assess targets we have not completely defined and mastered ourselves. Neither can we communicate them clearly to students.

Third, we develop and use assessments in a manner that yields accurate results. We select proper assessment methods, high-quality items and scoring guides, and plan for careful sampling of achievement. And we minimize distortion in results due to bias.

Fourth, results must feed into

effective communication systems that deliver accurate information into the hands of the intended user(s) in a timely and understandable manner. For students, this includes receiving descriptive feedback while there is still time to use it to improve.

And finally, students must be taught the skills they need to be in control of their own ultimate academic success: self-assessment and goal setting, reflection, keeping track of and sharing their learning.

Becoming competent in assessment *for* learning — what won't work and why

No Child Left Behind has lit an assessment fire in our nation: All things assessment-related sell fast. But we can't buy assessments that will circumvent teachers' need for deeper assessment expertise. Off-the-shelf assessments may be marketed as "formative assessments," but they don't help teachers understand or apply the strategies that have been proven to increase student learning. They do not show teachers how to make learning targets clear to students, or how to help students differentiate between strong and weak work. They do not help teachers understand what kinds of feedback are most effective or how to find the time to provide that feedback. They do not help teachers show students how to assess their own strengths and weaknesses, nor do they emphasize the motivational power of having students track and share their learning. They cannot substitute for the professional development needed to cause changes in assessment practice in the classroom.

Neither can we "workshop" our way to assessment competence. A professional development model designed to provide a quick workshop fix or to economize on time at the expense of deep understanding will fail. Developing assessment expertise goes beyond teaching people how to create a test. It goes beyond showing how to convert rubric scores to grades or how to develop a standards-based report card. It examines well-established assessment practices that are harmful to students and their learning, like factoring practice work (such as homework) in the final grade, giving tests without first understanding what specific learning each item addresses, and keeping students in the dark about the learning for which they are responsible.

If teachers assign lower grades to late work, give zeros for cheating, or factor attendance into grades, a workshop on grading is unlikely to change such unsound practice. It takes an ongoing investment of cognitive effort for teachers to think and come to embrace arguments for not doing these things, to discuss reasons for wanting to continue those grading practices, and to work out acceptable substitutes that both hold students accountable for developing good work habits and communicate effectively about those work habits.

Students must be taught the skills they need to be in control of their own ultimate academic success: self-assessment and goal setting, reflection, keeping track of and sharing their learning.

Changing habits is not easy. It takes work in and out of class to build assessment *for* learning environments that meet the student's information needs along with the teacher's. Increasing descriptive feedback while reducing evaluative feedback means

that the teacher must figure out ways to comment on the quality of student work and then schedule time for students to act on that feedback before being graded. Teaching students to assess their own work takes class time as well as practice. It is difficult to delete content coverage in order to accommodate these activities on a regular basis — there is already more to teach than there is time.

Developing assessment competencies requires that people rethink both what they do now and what beliefs led them to adopt those practices. It requires that they make decisions about what to give up and what to retool. The workshop model of professional development cannot offer the support needed for such changes.

What will work? Learning teams

In the learning team approach to professional development, participants engage in a combination of independent study and ongoing small-group collaboration with a commitment to helping all group members develop classroom assessment expertise. The process begins with an infusion of new ideas that can come from several sources: attending workshops, reading books and articles, watching videos, and observing other teachers at work. It continues with ongoing opportunities to discuss and work through the cognitive consonance and dissonance that arise when practice and beliefs conflict. But most importantly, it requires that each team member transform new assessment ideas into actual classroom practices with which they experiment. In this way, they and their students learn valuable lessons about what works for them and why.

Few teachers currently use the words "assessment" and "joy" in the same sentence.

When the experiences of such hands-on learning are shared among teammates in regular team meetings, all members benefit from the lessons of each partner. When teams commit to shaping the ideas into new classroom practice, reflecting on the results, and sharing the benefits with each other, professional growth skyrockets. Teams reach their ultimate goal of changing classroom assessment practices in specific ways that benefit students.

This is challenging work and can be even painful at times; few teachers currently use the words "assessment" and "joy" in the same sentence. Yet if we don't begin this dialogue, this study of assessment *for* learning, we are relegating assessment to its accountability role and passing up its potential benefits to students. Let us fundamentally rethink how assessment is used in our classrooms, eliminate its negative effects on students, and act collaboratively to ensure that our classroom assessment practices maximize, not just measure, our students' achievement.

Resources in assessment *for* learning

Research on assessment's impact on student achievement:

- "Creating a system of accountability: The impact of instructional assessment on elementary children's achievement scores," by Samuel J. Meisels, Sally Atkins-Burnett, Yange Xue, Donna DiPrima Bickel, and Seung-Hee Son. (2003). *Educational Policy Analysis Archives, 11*(9), 19. Retrieved from http://epaa.asu.edu/eapp/v11n9/
- "The impact of classroom evaluation on students," by Terence J. Crooks. (1988). *Review of Educational Research, 58*(4), 438-481.
- "The role of classroom assessment in student performance on TIMSS," by Michael C. Rodriguez. (2004). *Applied Measurement in Education, 17*(1), 1-24.
- "The search for methods of group instruction as effective as one-to-one tutoring," by Benjamin Bloom. (1984). *Educational Leadership, 41*(8), 4-17.

Valuable professional development materials:

- *Assessment FOR Learning: An Action Guide for School Leaders,* by Steve Chappuis, Richard J. Stiggins, Judith Arter, and Jan Chappuis. Portland, OR: Assessment Training Institute, 2005.
- *Capturing All of the Reader Through the Reading Assessment System,* by Rachel Billmeyer. Omaha, NE: Dayspring Printing, 2004.
- *Creating Writers,* by Vicki Spandel. New York: Addison Wesley Longman, 2001.
- *How to Grade for Learning,* by Ken O'Connor. Thousand Oaks, CA: Corwin Press, 2002.
- *Scoring Rubrics in the Classroom*, by Judith Arter and Jay McTighe. Thousand Oaks, CA: Corwin Press, 2001.

REFERENCES

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Educational assessment: Principles, policy and practice, 5*(1), 7-74. Also summarized in "Inside the black box: Raising standards through classroom assessment," *Phi Delta Kappan, 80*(2), 139-148.

Stiggins, R., Arter, J., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right—using it well.* Portland, OR: Assessment Training Institute.