Cambridge Journal of Education

ISSN: 0305-764X (Print) 1469-3577 (Online) Journal homepage: http://www.tandfonline.com/loi/ccje20

Rational responses to high stakes testing: the case of curriculum narrowing and the harm that follows

David Berliner

To cite this article: David Berliner (2011) Rational responses to high stakes testing: the case of curriculum narrowing and the harm that follows, Cambridge Journal of Education, 41:3, 287-302, DOI: 10.1080/0305764X.2011.607151

To link to this article: http://dx.doi.org/10.1080/0305764X.2011.607151

Cambridge Journal of Education
Vol. 41, No. 3, September 2011, 287–302

Routledge
Taylor & Francis Group

Rational responses to high stakes testing: the case of curriculum narrowing and the harm that follows

David Berliner*

College of Education, Arizona State University, Tempe, USA

(*Received 22 February 2011; final version received 7 July 2011*)

The inevitable responses to high stakes testing, wherein students' test scores are highly consequential for teachers and administrators, include cheating, excessive test preparation, changes in test scoring and other forms of gaming to ensure that test scores appear high. Over the last decade this has been demonstrated convincingly in the USA, but examples in Great Britain abound. Yet the most pernicious response to high stakes testing is perhaps the most rational, namely, curriculum narrowing. In this way more of what is believed to be on the test is taught. Curriculum narrowing, however, reduces many students' chances of being thought talented in school and results in a restriction in the creative and enjoyable activities engaged in by teachers and students. The tests commonly used with narrower curricula also appear to restrict thinking skills. In addition, responses to high stakes environments can easily retard the development of achievement in later grades as a function of the restrictions on learning in earlier grades. Finally, narrowing compromises interpretations of construct validity. The dominance of testing as part of American and British school reform policies insures that many of the skills thought to be most useful in the twenty-first century will not be taught. Thus students and their national economies will suffer when nations rely too heavily on high stakes testing to improve their schools.

Keywords: assessment; testing; curriculum; thinking skills; elementary education

A plethora of negative side effects associated with high stakes testing are now well documented (Nichols & Berliner, 2007). In the USA, high stakes accountability was a key part of the No Child Left Behind Act (NCLB), legislation passed in 2001 demanding that every child in every public and charter school in the country be tested in Grades 3–8 and Grade 10. Schools that did not improve rapidly enough to have 100% of their children proficient in reading and mathematics by 2014 could have teachers and administrators fired, be reconstituted or closed. Prominently featured in NCLB were impossible goals (100% proficient) and high stakes.

The NCLB act was supposed to reduce the achievement gaps between poor and wealthy students, but data supporting that claim is in dispute (e.g., Amrein & Berliner, 2002; Braun, Chapman, & Vezzu, 2010; Chudowsky, Chudowsky, & Kober, 2009; Lee, 2008; Smith, 2007). If the gap is actually closing, it is only by the smallest of amounts. By 2008–2009, after at least five years of high stakes

*Email: berliner@asu.edu

ISSN 0305-764X print/ISSN 1469-3577 online

http://dx.doi.org/10.1080/0305764X.2011.607151
http://www.tandfonline.com

testing in all states, about one-third of all US schools failed to meet their targeted goals under the No Child Left Behind (NCLB) act (Dietz, 2010). Estimates in 2011, by the US Secretary of Education are that more than 80% of all US public schools will fail to reach their achievement targets in 2012 (Duncan, 2011), and almost every school in the nation will fail by 2014. And this widespread failure is with each state using their own testing instruments, for which they prepare students assiduously.

Further, it is quite clear that the rate of achievement gains in the US was greater before high stakes testing became national policy. And this unhappy bit of news comes from America's most respected longitudinal testing program, the National Assessment of Educational Progress (NAEP) (National Center for Education Statistics, 2009a, b). Still further discouraging news for those who advocate testing as a way to reform schools comes from PISA testing (The Program for International Student Assessment). Nations with high stakes testing have generally gone down in scores from 2000 to 2003, and then again to 2006. Finland on the other hand, which has no high stakes testing, and an accountability system that relies on teacher and school judgments much more than on tests, has shown growth over these three PISA administrations (Sahlberg, 2011). Finland is currently considered the number one high achieving nation in the world. Their enviable position in world rankings of student achievement at age 15 have occurred with a minimum of testing and homework, and a minimum of imposition on local schools by the central government (Sahlberg, 2011).

High-stakes testing has evidently slowed the growth or reduced achievement in the USA despite the pressure it induces and the attention it necessarily receives from educators. The achievement gap between blacks and whites, Hispanics and Anglos, the poor and the rich, all are hard to erase because the gap has only a little to do with what goes on in schools, and a lot to do with social and cultural factors that affect student performance (Berliner, 2006, 2009). Federal officials in the last two administrations in the USA keep looking for a magic bullet that can be fired by school leaders and teachers to effect a cure for low achievement among the poor, English language learners and among many minorities. It is, of course, mostly wasted effort if the major cause of school problems stem from social conditions outside the schools.

Responses to testing pressure

When tests of student achievement are highly consequential for teachers and administrators, the pressure induced by the tests sometimes results in cheating by some teachers and administrators, or more frequently, the breaking of standardization procedures associated with the tests. These quite common acts of resistance or despair, in response to high stakes testing (Amrein-Beardsley, Berliner, & Rideau, 2010), destroy the validity of the tests, rendering them useless for the interpretations that are desired of them. The pressure of the testing also results in teachers engaging in vast amounts of test preparation with their students, some of which also causes the validity of high stakes testing to be problematical. In addition, teachers and administrators also have moved students that they expect not to test well out of their classes and schools, or they treat them poorly hoping they will drop out, or they hold them back so they get to be a year older (and maybe a little smarter) before they have to take the tests.

What occurs in high stakes environments is that individuals behave in accordance with Campbell's law. Campbell's law states that any time a social indicator takes on too much value (be it a stock price, a nations's debt, a sales report or a test score), both the indicator and the people who work with that indicator are corrupted (Campbell, 1975; Nichols & Berliner, 2007). In high pressure situations people frequently do what ever they deem necessary to achieve their goals and keep their jobs or status. That is, the value placed on the indicator thought to be a measure of a person or a system's performance corrupts individuals, and the indictor itself may quickly become invalid. For example, grades in law school are highly valued. Law students with the highest grades receive the best job offers, $100,000 per year and up, right out of law school. So many institutions, including many of the most prestigious law schools in the USA, inflate their students' grades so their law schools will look like they turn out better law students then other institutions. That is, the institution systematically corrupts its own indicators of student talent (Orlinsky, 2010). Campbell's law is ubiquitous.

While some responses to the pressures of high stakes testing are unlawful, and many more are immoral or at least questionable, one quite rational but troubling way to accommodate to the pressures to obtain ever higher test scores from students is by curriculum narrowing. This may be the most important lesson to be learned by nations using high stakes testing policies to improve student achievement.

Table 1 shows what happens when the high stakes testing is consequential for schools in just two subjects, reading (English/language arts) and mathematics. These are the two subjects tested in Grades 3–8 and Grade 10 in the USA, as required by the NCLB law. These data, and those in Table 2, come from surveys administered to a nationally representative sample of almost 500 school districts. These data refer to primary grades only.

Since these are averages, it is likely that lower performing schools (compared to the higher performing schools) added much more time to these subject matter areas in the hope of having their test scores go up. Eighty per cent of the school districts increased time in English/language arts by at least 75 minutes a week, and more than half the districts increased time in that area by 150 minutes a week or more, adding at least 30 minutes more a day. Sixty-three per cent of the districts reported that they increased mathematics time by at least 75 minutes a week and 19% added 150 minutes more each week. If a student is in a school that has done both, it is quite possible that 300 minutes of instructional time (six hours per week) have been added to his or her reading and mathematics time blocks.

Table 1. Changes since 2001–2002 in instructional time for elementary school English language arts and mathematics, in districts reporting increases (McMurrer, 2008).

Subject matter examined	Average total instructional time spent pre-NCLB (in minutes per week)	Average total instructional time spent post-NCLB (in minutes per week)	Average increase in instructional time per week (in minutes)	Average increase in instructional time as a percentage of total instructional time
English language arts	378	520	141	47%
Mathematics	264	352	89	37%

Table 2. Decreases in instructional time for various curriculum areas to accommodate increases in time for English language arts and mathematics (McMurrer, 2008).

Subject matter	Average minutes per week before NCLB	Average minutes per week after NCLB	Average decrease per week	Average decrease as a percentage of total time per week
Social studies	239	164	76	32%
Science	226	152	75	33%
Physical education	115	75	40	35%
Recess	184	144	50	28%
Art and music	154	100	57	35%

The school time has to come from somewhere. In Table 2 we see from where that time was taken.

Social studies (civics, history, law and related studies) is the curriculum area from which most time is taken. More than half of the districts reporting (53%) reduced instructional time in social studies by *at least* 75 minutes per week. But since colonial times the social studies have been deliberately designed to be a part of youth development. This curriculum area is widely believed by Americans of all political persuasions to be necessary for responsible citizenship. Americans believe that youth must understand the need to participate in democracy, for democracy is fragile, at best, without citizen involvement.

Former Supreme Court Justice Sandra Day O'Conner recognizes this new problem in curriculum:

> One unintended effect of the No Child Left Behind Act... is that it has effectively squeezed out civics education because there is no testing for that anymore and no funding for that. At least half of the states no longer make the teaching of civics and government a requirement for high school graduation. This leaves a huge gap, and we can't forget that the primary purpose of public schools in America has always been to help produce citizens who have the knowledge and the skills and the values to sustain our republic as a nation, our democratic form of government. (O'Conner, quoted in Schiesel, 2008)

Despite the widespread support for teaching this subject matter, and in direct opposition to what Americans want, instructional time in this area is down, on average, over an hour a week. Furthermore, school activities that might foster citizenship have been cut because of the need for more time in reading and mathematics. So cleaning up neighborhoods and parks, visiting nursing homes, going on field trips to the legislature, engaging in projects that examine pollution at a local level, and so forth, all have been curtailed or eliminated (Nichols & Berliner, 2007). Yet it is precisely through such projects that responsible citizenship is learned. The culprit in the demise of the social studies, of course, is fear of poor results in a high stakes testing environment.

Table 2 also reveals that instruction in science is down, on average, over an hour a week since NCLB became law. Schools can ignore a lack of growth on science tests because no sanctions are associated with that test. The payoff for schools is in increased scores on the tests of English/language arts and mathematics. Therefore science, like social studies, has been robbed of minutes to expand instructional time in curriculum areas of consequence. Thus curriculum that might help insure

America's economic competitiveness in the twenty-first century, and surely will contribute to intelligent citizenship in our science- and technology-rich future, has been sacrificed for the possibility of scoring a bit higher on a high stakes test.

Table 2 also documents that time for physical education is down, despite the fact that too many of America's youth are sedentary, overweight and show signs of Type 2 diabetes. If a link between physical activity and student physical well-being and cognition exists, then it should be a source of concern that physical education courses in the USA have dropped precipitously since the 1990s. Although acknowledged as one of the most important ways to keep a nation's medical costs down, and an activity that parents and physicians want to see promoted in the schools, physical education is sacrificed for the possibility of a few more points on state tests that are required to rise almost continuously.

Nationally, as seen in Table 2, recess was found to be down, on average, about one hour a week. Recess, in a high stakes testing environment, is wasted time. Nichols and Berliner (2007) even discovered a superintendent of schools who forbade naps for preschoolers and kindergartners. And in one school it was discovered that lunch was under 15 minutes a day so the children could have more time to study. The pressure of high stakes testing clearly results in a narrowing of the curriculum, a logical outcome of a penalty oriented program such as NCLB, where shaming and blaming of teachers and administrators for low student test scores is common. But those pressures have another pernicious effect. They seem also to turn many decent, child-centered educators, into mean spirited educators. Under pressures from high stakes testing, educators make decisions that reflect compromised ethics, if not a complete loss of their humanity.

As documented in Table 2, art and music are down an average of an hour a week, nationally. This is particularly troublesome because the USA never spent a lot of time in these subjects. A recent poll showed that 61% of California's schools did not even have one full-time-equivalent arts specialist (Woodworth, Gallagher, & Guha, 2007). With the budget problems brought on by the recession, that percentage is sure to be higher in 2010 and 2011.

The defense of the arts can be made on many grounds, but one stands out in terms of 21-first century needs, namely, that the arts are alternative ways to represent reality. Ideas expressed through the visual arts, dance and music are not presented in the verbal or mathematical symbol systems that are in everyday use. So by cutting the arts we limit the ways our students can represent the world in which our students live and about which they may choose to comment. A reduction in curricula for learning the arts, therefore, restricts our students' ways of thinking, limiting creativity. So besides defending the arts as a natural expression of our humanity, and for occasionally providing humanity with works of indescribable beauty, the arts can be justified as having economic benefits. But even this argument does not sell in a high stakes testing environment in which tests do not cover knowledge about the arts nor to they evaluate artistic performances.

A study of the arts in California, and probably generalizable to the rest of the nation, makes clear that the arts are rationed: The arts are taught primarily to the wealthy and the middle class, but not taught to the poor. These data are provided in Table 3. As can be seen from this statewide sample, close to twice as many students in schools that serve the wealthy receive instruction in the arts as do the students in schools that serve the poor.

Table 3. Percent of California students receiving instruction in various areas of the arts, by poverty level of the school they attend (Woodworth, Gallagher, & Guha, 2007).

Subject matter	Per cent studying this subject in schools serving wealthy children	Per cent studying this subject in schools serving middle class children	Per cent studying this subject in schools serving poor children
Music	45%	38%	25%
Visual arts	48%	44%	29%
Theater	17%	14%	8%
Dance	17%	14%	7%

Because of the test anxiety felt by the teachers and administrators of poor children, the study of the arts has been diminished. This provides an example of how the pressure for poor students to perform better on the tests results in an apartheid system of schooling.

Two of the nation's most visible conservative school critics, Chester Finn and Diane Ravitch (2007), also see this narrowing of the curriculum as a terrible problem for the USA. They recant their earlier statements of support for the pursuit of school reform through high stakes testing programs such as NCLB. They now say, quite elegantly:

> The liberal arts make us 'competitive' in the ways that matter most. They make us wise, thoughtful and appropriately humble. They help our human potential to bloom. And they are the foundation for a democratic civic polity, where each of us bears equal rights and responsibilities.
>
> History and literature also impart to their students healthy skepticism and doubt, the ability to question, to ask both 'why?' and 'why not?' and, perhaps most important, readiness to challenge authority, push back against conventional wisdom, and make one's own way despite pressure to conform.
>
> We're already at risk of turning US schools into test-prepping skill factories where nothing matters except exam scores on basic subjects. That's not what America needs nor is it a sufficient conception of educational accountability. We need schools that prepare our children to excel and compete not only in the global workforce but also as full participants in our society, our culture, our polity, and our economy.
>
> Abandoning the liberal arts... also risks widening social divides and deepening domestic inequities. The well-to-do who understand the value of liberal learning may be the only ones able to purchase it for their children. Top private schools and a few suburban systems will stick with education broadly defined, as will elite colleges. Rich kids will study philosophy and art, music and history, while their poor peers fill in bubbles on test sheets. The lucky few will spawn the next generation of tycoons, political leaders, inventors, authors, artists, and entrepreneurs. The less lucky masses will see narrower opportunities. Some will find no opportunities at all, which frustration will tempt them to prey upon the fortunate, who in turn will retreat into gated communities, exclusive clubs, and private this-and-that's, thereby widening domestic rifts and worsening our prospects for social cohesion and civility.

It is not just the US that has this problem. A narrowing of curriculum has been going on in England for years. England, like the USA, believed it could use high

stakes testing to improve its schools, and it too has failed to do so with that policy. The narrowing was evident a decade ago. A former chief inspector of England's schools said that the national curriculum had been 'eroded' in a 'silent revolution' that narrowed the range of subjects being taught (Baker, 2002). The inspector said that the government's focus on literacy and numeracy had led to a 'narrowing' of the primary school curriculum in a 'significant number of schools', just as it has in the USA. Geography, history, the arts and practical subjects such as design and technology were often eliminated. A head teacher of a primary school said:

> The constant focus on core curriculum areas somehow gives us teachers' guilt complexes. You always feel time pressure and feel that you must get at numeracy and at literacy, so that when you get into lovely creative areas, a bit of you thinks 'Oh my goodness, have I done the full literacy hour?' (Baker, 2002)

Another English head of school pointed out that primary schools usually teach 23 or 24 hours a week. Under the government's push for higher test scores, the schools he worked in were devoting their time as given in Table 4. If these schools actually tried to cover all these subjects a few times per week, with the usual transition times in and out of lessons, the amount of actual instructional time per non-tested subject is trivial.

English teachers complain the same as do the American teachers, but in neither country are the voices of teachers often listened to by politicians. Galton and McBeath (2002) quote teachers in England. A teacher with 23 years experience says:

> Too often the subjects like art, and history and geography and the subjects that children really enjoy, and PE, are squeezed out and those children that are not academic are not getting a chance to shine. We are actually turning them off education rather than actually encouraging them to want to improve the things that they are good at because we're not actually finding out what they're good at any more.

Another teacher with nine years experience says:

> Everything else has suffered hasn't it? From my point of view I think probably the saddest thing is the arts being elbowed out. I just find that to be a subject that lots of children who don't achieve particularly academically achieve wonderfully well at the

Table 4. Approximate times for curriculum areas in primary education in one English district facing high-stakes tests (Baker, 2002).

Subject taught	Percent of time each week	Minutes per week/minutes per day
English	29	409/82
Maths	22	310/62
Science	10	141/28
Design and technology	4	56/11
Information technology	4	56/11
History	4	56/11
Geography	4	56/11
Art	4	56/11
Music	4	56/11
Physical education	6	85/17
Religious education	5	71/14

> arty crafty sorts of activities. And they're not always able to express it, because we don't have the time to dedicate.

Galton and McBeath (2002) report that many teachers noted the creative subjects were being squeezed out, with the consequence that there were fewer opportunities for children to be good at something, to succeed or to excel, and the teachers knew that this was not good for the children. In addition, it made classroom management all that more difficult. The emphasis on the core subjects, with increased focus on content, simply meant that there was less space in the school day for less structured activities, though it was in those kinds of activities that some non-academic children excelled.

The American experience is identical. In Colorado one teacher says (Taylor, Shepard, Kinner, & Rosenthal, 2003, pp. 30–31):

> We don't take as many field trips. We don't do community outreach like we used to like visiting the nursing home or cleaning up the park because we had adopted a part and that was our job was to keep it clean. Well, we don't have time for that any more.

Another Colorado teacher says:

> We only teach to the test even at second grade, and have stopped teaching science and social studies. We don't have assemblies, take few field trips, or have musical productions at grade levels. We even hesitate to ever show a video. Our second graders have no recess except for 20 minutes at lunch.

A Florida teacher says (Jones & Egley, 2004):

> Our total curriculum is focused on reading, writing, and math. There is no extra time for students to study the arts, have physical education, science, or social studies. Our curriculum is very unbalanced.

In Arizona, a teacher of English language learners talks about the test-oriented curriculum and why she might leave the field (Wright, 2007, p. 4)

> I'm going to get hired by another district to see if it's like this everywhere, because I haven't been teaching really all that long, and if it is like this everywhere, with just nothing but teach-to-the-test type stuff, and to heck with what the kids want to know, then... I'm not going to stay in the classroom because it just breaks my heart. There are things the kids just want to learn about. You teach them a little bit in these programs, but it's so structured that you don't have time to deviate from the program. I mean, we aren't allowed to have parties, they don't have recess. There is no time during the day where I am allowed to just have fun with my kids and just learn something that is just for fun. And it's really depressing.

The point about learning for fun is perhaps the most important part of this teachers' lament. We all know that students' ability to learn in areas of interest to them seems almost unlimited, as noted in their commitment to their hobbies and to acquiring skills in video games. But in this era of high stakes testing, students cannot be allowed time in school to follow their particular interests in dinosaurs, medieval armaments, the American civil war, fashion design, horses, whales, or other areas where students show intellectual passion. It is the assessment system that

defines what students should know at different grade levels, and deviation from that plan is considered dangerous because it might result in missing curriculum coverage of material reflected in some items on a high stakes accountability test.

Galton (2007) notes how this pressure has resulted in changed teacher behavior. Following up his own research of the 1970s, Galton found that teacher-centered pedagogy, characterized by interactions of a very low cognitive level, managerial in their intent, had *increased* dramatically between 1976 and 1996. Pupils had fewer opportunities to question or to explore new ideas after the tests became the primary instrument that the government used to change the schools. Assessment pressures have resulted in 42% of teachers' time being taken up with whole class teaching, compared to 18% in 1976. In primary schools in England there are now few opportunities for expressing anything that resembles creative reasoning (Galton, 2007).

More recently, the Cambridge Primary Review (Alexander, 2010) reached conclusions about English schooling that are similar to what we find in the USA. Community soundings revealed that teachers, head teachers and parents across England were concerned about the pressures that exist when high stakes tests are imposed. These professionals and citizens were also strongly against the use of league tables for judging school performance, though as in the USA they were not against assessment per se. Noteworthy for its similarity to US data is the concern about the narrowness of schooling that is inevitable when there is excess preparation for the high stakes tests, such as the English Standard Attainment Tests (SATs). Harlen (2007) drew on several surveys of time used in activities relating to assessment to conclude that pupils in Year 6 (age 11 years) spent the equivalent of 13 school days practising and taking tests, in addition to other assessment activities. Without the anxieties of high stakes testing, students and teachers could have spent several weeks in other ways, including in curricula that are far more interesting to youth and their teachers than is test preparation.

The English and US experience is replicated by Hong and Youngs (2008). They studied curriculum in Chicago and Texas, as both that district and that state responded to high stakes testing. In Chicago, the researchers found that high stakes testing seemed to narrow the curriculum and make it harder for students to acquire higher-order thinking, writing and problem-solving skills. In Texas, it was found that schooling changed in ways that emphasized rote learning, not broad intellectual skills (Hong & Youngs, 2008; McNeil, 2000). Lipman (2004) also studied the Chicago schools and reported that the accountability program insured that the more affluent students in Chicago received a much richer and more intellectually challenging curriculum than did the poor children in Chicago. Poor minority children, in particular, were required to memorize fragmented facts and information, and they were constantly taught simple test-taking techniques.

In a study of NCLB implementation in Illinois (Srikantaiah, Zhang, & Swayhoover, 2008) teachers in schools categorized as needing improvement were found to ask 16% more closed questions than was true in schools not so categorized, the schools that feel less testing pressure. And the teachers in schools not classified as in need of improvement by federal authorities were doing 100% more hands-on activities than were schools labeled as needing improvement.

Lipman is probably quite right when she says that this differential access to high quality curriculum will have significant consequences in terms of the social inequalities we will observe in the future. She joins Ravitch and Finn (cited above) in her worries about an apartheid curriculum. White students or middle class students who

possess a great deal of the cultural capital valued by schools are going to be much more likely to get to college and thus more likely to attain higher status through higher paying jobs. But low SES and minority students in communities across the nation are much more likely to end up in lower-skilled and lower-paying jobs. The decisions about curriculum and instruction in districts that are not making adequate yearly progress according to the NCLB law, results in access to rigorous curriculum for some, but not for others. This differential access to curriculum allows for the continuation of the current unequal social structures that are endemic in the USA. Apartheid education, sadly, is descriptive of the situation across districts in the US (Berliner, 2010).

Narrowness of the cognitive behavior rewarded

American schools never allowed much time for individual or group work with high cognitive demands, but now even the teachers that made some use of problem-based or project-based leaning, forms of instruction that could ignite students' interests through a more personally tailored curriculum, are not allowed to do so. One size of the curriculum is supposed to fit all students even though we are reasonably sure that the twenty-first century economy will require from our work force a broad set of skills, not a narrow one. Diversity in the knowledge possessed by students ought to be among the goals of national educational systems. Sameness in the educational outcomes of nations is likely to be unhealthy in the long run, and should be resisted.

But worse than the sameness in outcomes is the fact that the cognitive processes demanded in this era of narrow curricula offerings are themselves quite narrow. Because public educational budgets are always tight, large-scale high stakes testing is never allowed to be very expensive. The items used to assess students, therefore, are quite often multiple choice, convergent, machine-scoreable items, the cheapest items to produce for mass testing. Constructed response items requiring thoughtful, divergent, extended and creative responses by students are usually too expensive to score, requiring human raters well trained in a subject matter. Group projects are almost never used in school accountability schemes, though they permeate modern business enterprises. Sadly, the typical high stakes testing program in a US state promotes a certain impoverishment in teaching activities to produce similar and convergent ends. Preparation for answering low-level items leads easily into teaching students to memorize content and develop algorithmic skills. Thus instruction in high stakes environments is often drill oriented. Lessons that require memorization (spelling rules, punctuation, vocabulary) and mastery of rote procedures (invert and multiply, the product of two negative numbers is a positive number) are emphasized. What is sacrificed when instruction like this becomes too common is comprehension in reading, understanding in mathematics and diversity in outcomes.

Getting high scores on assessments that emphasize the mimetic and algorithmic properties of a subject matter is relatively easy with extensive test preparation. Drill and practice works well when the test items are lower level and easily inferred from the specifications for the test (usually available beforehand). But higher order thinking is sacrificed when high stakes multiple-choice testing puts pressure on teachers. This is made clear in an interesting comparison of Chinese and American college freshmen.

Chinese high school education in science is considered much more rigorous and is of much longer duration than American science education. The Chinese national curriculum in science begins in Grade 8 and continues through Grade 12, providing five years of physics, as well as instruction in other topics, culminating in a highly competitive national test for entrance into universities. This extremely high stakes test, however, requires a good deal of memory and algorithmic solutions to problems, not creative problem-solving, nor much scientific reasoning.

Freshmen in premier Chinese universities, a very exceptional group of students who had passed the rigorous exams, were tested on their knowledge of force and their knowledge of electricity and magnetism (Bao et al., 2009). Their scores were evaluated against American freshman science students who came from less prestigious US universities, institutions that admitted students with a much wider distribution of ability. The Americans had a hodge-podge of science training, with no national curriculum and sometimes even no standardized state curriculum, and no high stakes national tests of science achievement. This lack of a national high stakes test is to be noted because the comparison of freshman in the two nations revealed remarkable differences.

As Table 5 shows, in a test of concepts about force, and in a test of the concepts of electricity and magnetism, the mean Chinese scores were dramatically higher than the US scores: overwhelmingly and embarrassingly higher! The effect sizes obtained are huge. Indisputably, Chinese college freshman know a lot more science in these areas than do American college freshmen. But one other test was used in this comparison. It was of scientific reasoning. Here the elite Chinese students, compared to the less elite American students, showed no advantage whatsoever. The huge deficit of the US students on tests of simple knowledge were eliminated on what is arguably the most important of the three measures, namely, the test of scientific reasoning. There are few good explanations for this remarkable finding except that the Chinese curriculum, under the tyranny of the high stakes examination, has narrowed conceptions of what constitutes knowledge in physics. Only the approved factual knowledge is to be mastered in China, the knowledge needed to do well on the very high stakes college entrance examination. But there may have been a sacrifice of the knowledge actually needed for twenty-first century working environments.

In the USA, the eminent scholar Lauren Resnick (2010) noted that no one intended for the test-based accountability movement to end up taking us back to the minimum competency movement of the 1970s. But it did. That older and failed approach to accountability attempted to guarantee that poor and minority students

Table 5. Test scores, number in sample, and effect sizes for college freshmen in China and the US on two tests of scientific knowledge and one test of scientific reasoning (Bao et al., 2009).

Test	Chinese students' scores	US students' scores	Effect size
Force Concepts Inventory	85.9 (n = 523)	49.3 (n = 2681)	1.98
Assessment in electricity and magnetism	65.6 (n = 331)	26.6 (n = 650)	3.53
Lawson Test of Scientific Reasoning	74.7 (n = 370)	74.2 (n = 1061)	.03

would be taught the basics in the schools they were in, acknowledged then, as now, not to be very effective. Few seemed worried then that higher cognitive demands were not going to be made on those students. It was thought that the graduates of our schools, with just these minimum skills of literacy and numeracy, could find decent work in the economy. More disquieting, however, is that some were promoting this accountability model because they thought poor and minority students were incapable of learning that demanded higher levels of cognition.

To be sure memory and algorithmic learning still matters, constituting a piece of every skilled thinkers armamentarium.

> [But the] skills that are becoming valued in the twenty-first century are focused on a person's ability to participate in argumentation and discussion. Question-and-answer drills will have to be replaced by discursive processes that include productively challenging colleagues, paraphrasing, and interpreting presentations by others. And although individual performances still matter, much 'knowledge work' is 'distributed', involving collaboration with others. (Resnick, 2010, p. 188)

As Resnick and others argue, a thinking curriculum requires regular practice, of sufficient duration, in such areas of cognition best described as conceptual learning, reasoning, explaining and problem-solving. But all these cognitive activities have to be embedded in specific, challenging subject matter (Darling-Hammond et al., 2009), and these skills also need to be used across subject matter domains so that they might generalize, whenever possible. Thinking skills, even what we ordinarily think of as intelligence, are teachable skills. These skills develop in the course of reasoning about specific information and bodies of knowledge.

It is now accepted that teaching thinking skills in the absence of specific content rarely works. A person must think about some thing in particular. We also know now that drill and practice on facts and algorithms, without demands for explanation and reasoning, produces a fragile kind of knowledge, likely to help on tests close in time to the drill and practice, but likely to fade quickly after a test and unlikely to transfer to any other domains of knowledge (Darling-Hammond et al., 2009; Resnick, 2010).

Common responses to high stakes testing, therefore, are likely to be exactly what nations do *not* want for their youth. No modern democratic nation wants restriction on the cognitions youth develop as a function of teaching and learning in schools.

Narrow curriculum and cognitive development in particular domains of knowledge

Briefly, the argument here is that what we learn is heavily dependent on what we have already learned (Bransford, Brown, & Cocking, 2000; Jerald, 2006; Willingham, 2009; Darling-Hammond, et al., 2009). For example, to fully comprehend a newspaper article about confirmation hearings for a Supreme Court seat, a reader needs to know at least a little bit about what the court is and what it does. To understand the impressionist movement in art requires understanding of what came before and what the times were like in the late nineteenth century world of art and politics. To read the *Red badge of courage* or *Catcher in the rye* or *Hard times*, or to see the movie *Gone with the wind* or *Avatar*, requires enormous background knowledge fully to comprehend the texts read or fully to appreciate the

movies seen. Thus, the more narrow the curriculum is in youth, the less likely that the requisite background knowledge will be available in later grades and in the real world, after formal schooling ends.

Students who have experienced a narrow curriculum focus in the early grades are more likely to perform poorly on standardized reading and mathematics tests later because in the upper grades such assessments place more emphasis on comprehension and reasoning and less emphasis on simple decoding and algorithms. For example, Jerald (2006) reports that differences in reading scores among third graders were mainly due to how efficiently and fluently they decoded words, while variations in reading scores among tenth graders were mainly due to their vocabularies and comprehension skills. Thus, a test preparation strategy meant to improve reading scores in the early grades may actually depress reading scores in the long run.

Students who lack the background knowledge they need for the courses they must take in high school (or college, if they actually get there) are often the poor and minority children in our schools. They will struggle more than their middle class peers not because they are less intelligent, but because they have less of the requisite knowledge to learn more in a particular subject area.

Curriculum narrowing to improve the test scores of poor and minority students may, therefore, end up magnifying the achievement gaps between them and their middle class peers. Affluent students often have access to museums, go on family vacations, attend more cultural events in their community and are exposed to a much richer vocabulary. They pick up peripheral, incidental, but potentially essential knowledge, even when their schools to do a poor job of teaching about the natural world, art, music, history, geography and science. As Jerald (2006) and particularly E.D. Hirsch (2003, 2006) note, for many poor and minority children, it is the schools, and sometimes only the schools, that provide access to the background knowledge so essential to comprehending the world in which we live. When the schools narrow their curriculum offerings to gain a few points on a high stakes test, they may simultaneously be cheating their students of a richer future.

Conclusion

Curriculum narrowing is the most rational of the many responses that occur to high stakes testing. While the narrowing, and the test preparation that accompanies the narrowing, is likely to result in higher scores on the tests that are so consequential for teachers and administrators, there are many side effects of this response to the high stakes testing policies. Most notable is the clear evidence that a great deal of the curriculum deemed desirable for our schools by a broad spectrum of citizens is, instead, curtailed in high stakes environments. A great deal of time for learning is added to those subjects that are tested, and a great deal of time is subtracted from those subjects not tested. In addition, many of the instructional activities in the curriculum areas tested are of a low level in terms of the cognitive processes that are called for by students. Drill in memorization and practice of algorithms predominates. The tests themselves are also not demanding of higher cognitive processes. Demands for higher cognitive processes, what is ordinarily called thinking skills, are not taught frequently enough in schools that are heavily pressured to improve achievement.

Limiting the curriculum to be learned and the cognitions used to think about subject matter restricts the growth of our students' understanding as they progress through school. Less often discussed is that the constructs to be measured – reading comprehension and mathematical understanding – are likely not to be well measured in high stakes testing environments that have limited the conception of the construct and the cognitions required to demonstrate competency in the domain. Thus, the construct validity of the test whose scores determine so much is compromised in ways that are impossible to estimate, but are likely to be substantial.

The system of high stakes testing in the USA and England has produced many negative side effects. These negative effects are magnified among those who are in minority groups and are poor. It is past time to abandon high stakes testing policies and to experiment with alternative systems to monitor, evaluate and improve our schools.

Note on contributor

David C. Berliner is Regents' Professor of Education Emeritus at Arizona State University and has taught at universities throughout the world. He is a member of the National Academy of Education, the International Academy of Education, and a past president of both the American Educational Research Association (AERA) and the Division of Educational Psychology of the American Psychological Association (APA). He is the winner of numerous awards including the AERA award for distinguished contributions to education and the E.L. Thorndike award from the APA for lifetime achievements. He is co-author of the bestseller *The manufactured crisis*, as well as *Collateral damage: How high stakes testing corrupts American education* and the textbook *Educational psychology*, now in its sixth edition. He co-edited the first *Handbook of educational psychology*, *Talks to teachers* and *Perspectives on instructional time*. Professor Berliner has authored more than 200 published articles, technical reports and book chapters.

References

Alexander, R.J. (2010). *Children, their world, their education*. London: Routledge.

Amrein, A.L., & Berliner, D.C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved June 24, 2010, from http://epaa.asu.edu/epaa/v10n18/

Amrein-Beardsley, A., Berliner, D.C., & Rideau, S. (2010). Breaking professional law: Degrees of cheating on high stakes tests. *Education Policy Analysis Archives, 18*(14). Retrieved June 24, 2010, from http://epaa.asu.edu/ojs/article/view/714

Baker, M. (2002). 'Silent erosion' of school curriculum. Retrieved June 24, 2010, from http://news.bbc.co.uk/2/hi/uk_news/education/1991816.stm

Bao, L., Cai, T., Koenig, L., Fang, K., Han, J., Wang, J., Liu, Q., Ding, L., Cui, L., Luo, Y., Wang, Y., Li, L., & Wu, N. (2009). Physics: Learning and scientific reasoning. *Science, 323*(5914), 586–587.

Berliner, D.C. (2006). Our impoverished view of educational reform. *Teachers College Record, 108*(6), 949–995. Retrieved June 24, 2010, from http://www.tcrecord.org/content.asp?contentid=12106

Berliner, D.C. (2009). Poverty and potential. Out-of-school factors and school success. Boulder, CO and Tempe, AZ: Education and the Public Interest Center, University of Colorado/ Education Policy Research Unit, Arizona State University. Retrieved June 10, 2009, from http://epicpolicy.org/publication/poverty-and-potential

Berliner, D.C. (2010). New analysis of achievement gap: ½ x ½ = 1½. *Washington Post*. Retrieved June 29, 2010, from http://voices.washingtonpost.com/answer-sheet/guest-bloggers/new-analysis-of-achievement-ga.html?wprss=answer-sheet

Bransford, J.D., Brown, A.L., & Cocking, R.R. (2000). *How people learn: Brain, mind, experience, and school. Committee on developments in the science of learning*. Washington, DC: National Academy Press.

Braun, H., Chapman, L., & Vezzu, S. (2010). The Black–White achievement gap revisited. *Education Policy Analysis Archive, 18*(21). Retrieved February 14, 2011, from http://epaa.asu.edu/ojs/article/view/772

Campbell, D.T. (1975). Assessing the impact of planned social change. In G. Lyons (Ed.), *Social research and public policies: The Dartmouth/OECD Conference* (pp. 3–45). Hanover, NH: Dartmouth College, The Public Affairs Center.

Chudowsky, N., Chudowsky, V., & Kober, N. (2009). *State test score trends through 2007–2008: Are achievement gaps closing and is achievement rising for all?* Washington, DC: Center on Education Policy. Retrieved June 30, 2010, from http://www.cep-dc.org/index.cfm?fuseaction=document_ext.showDocumentByID&nodeID=1&DocumentID=292

Darling-Hammond, L., Barron, B., Pearson, D., Schoenfeld, A.H., Stage, E.K., Zimmerman, T.D., Cervetti, G.N., & Tilson, J.L. (2009). *Powerful learning: What we know about teaching for understanding*. San Francisco: Jossey-Bass.

Dietz, S. (2010). *How many schools have not made adequate yearly progress under the No Child Left Behind Act?* Washington, DC: Center for Educational Policy. Retrieved June 24, 2010, from http://www.cep-dc.org/index.cfm?fuseaction=document_ext.showDocumentByID&node ID=1&DocumentID=303

Duncan, A. (2011). Winning the future with education: Responsibility, reform and results. Oral testimony of US Secretary of Education Arne Duncan given to the US Congress, March 9, 2011. Retrieved June 20, 2011, from http://www.ed.gov/news/speeches/winning-future-education-responsibility-reform-and-results

Finn, C.E., & Ravitch, D. (2007). Not by geeks alone. *The Education Gadfly, 7*, 30. Retrieved June 15, 2010, from http://www.edexcellence.net/gadfly/index.cfm?issue=302/1a3539 See also: *The Wall Street Journal*, Commentary, August 8, 2007, page A13.

Galton, M. (2007). *Learning and teaching in the primary classroom*. London: Paul Chapman Publishing.

Galton, M., & MacBeath, J. (2002). *A life in teaching? The impact of change on primary teachers' working lives*. London: National Union of Teachers.

Harlen, W. (2007). *Assessment of learning*. London: Sage.

Hirsch, E.D. Jr. (2003). Reading comprehension requires knowledge – of words and the world. *American Educator, 27*(1), 17.

Hirsch, E.D. Jr. (2006). *The knowledge deficit*. Boston: Houghton Mifflin Company.

Hong, W.-P., & Youngs, P. (2008). Does high stakes testing increase cultural capital among low-income and racial minority students? *Education Policy Analysis Archives, 16*(6). Retrieved January 14, 2009, from http://epaa.asu.edu/epaa/v16n6/

Jerald, C.D. (2006). *The hidden costs of curriculum narrowing*. Washington, DC: Learning Point Associates Issue Brief. The Center for Comprehensive School Reform and Improvement. Retrieved June 24, 2010, from http://www.centerforcsri.org/files/CenterIssueBriefAug06.pdf

Jones, B., & Egley, R. (2004). Voices from the frontlines: Teachers' perceptions of high stakes testing, *Education Policy Analysis Archives, 12*(39). Retrieved June 15, 2009, from http://epaa.asu.edu/epaa/v12n39

Lee, J. (2008). Is test driven accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies. *Review of Educational Research, 78* (30), 608–644.

Lipman, P. (2004). *High-stakes education: Inequality, globalization, and urban school reform*. New York: Routledge Palmer.

McMurrer, J. (2008). *Instructional time in elementary school subjects. A closer look at changes for specific subjects*. Washington, DC: Center on Education Policy. Retrieved May 15, 2010, from http://www.cep-dc.org/index.cfm?fuseaction=document.showDocumentByID&nodeID=1&DocumentID=234

McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

National Center for Education Statistics. (2009a). *The nation's report card: Mathematics 2009 (NCES 2010–451)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education. Retreived June 24, 2010, from http://nces.ed.gov/nationsreportcard/pubs/main2009/2010451.asp

National Center for Education Statistics. (2009b). *The nation's report card: Reading 2009 (NCES 2010–458)*. Washington, DC: Institute of Education Sciences, US Department of Education. Retrieved June 24, 2010, from http://nces.ed.gov/nationsreportcard/pdf/main2009/2010458.pdf

Nichols, S.N., & Berliner, D.C. (2007). *Collateral damage: The effects of high stakes testing on America's schools*. Cambridge, MA: Harvard Education Press.

Orlinsky, K. (2010). In law schools, grades go up, just like that. *New York Times*. Retrieved June 24, 2010, from http://www.nytimes.com/2010/06/22/business/22law.html?th&

Resnick, L.B. (2010). Nested learning systems for the thinking curriculum. *Educational Researcher, 39*(3), 183–197.

Sahlberg, P. (2011). *Finnish lessons: What can the world learn from educational change in Finland*. New York: Teachers College Press.

Schiesel, S. (2009). Former Justice promotes web-based civics lessons. *New York Times*. Retrieved June 24, 2010, from http://www.nytimes.com/2008/06/09/arts/09sand.html

Smith, M.S. (2007). NAEP 2007-What about NCLB? Powerpoint presentation made by Marshall S. Smith at Berkeley, CA, October 10, 2007, cited with permission of the author.

Srikantaiah, D., Zhang, Y., & Swayhoover, L. (2008). *Summary: Lessons from the classroom level: Federal and state accountability in Illinois*. Washington, DC: Center for Educational Policy. Retrieved May 20, 2010, from http://www.cep-dc.org/index.cfm?fuseaction=document_ext.showDocumentByID&nodeID=1&DocumentID=250

Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2003). *A survey of teachers' perspectives on high stakes testing in Colorado: What gets taught, what gets lost*. CSE Technical Report 588. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Willingham, D. (2009). Reading is not a skill – and why this is a problem for the draft national standards. *Washington Post*. Retrieved June 24, 2010, from http://voices.washingtonpost.com/answer-sheet/daniel-willingham/willingham-reading-is-not-a-sk.html

Woodworth, K.R., Gallagher, H.A., & Guha, R. (2007). *An unfinished canvas. Arts education in California: Taking stock of policies and practices*. Summary Report. Menlo Park, CA: SRI International. Retrieved January 14, 2009, from http://policyweb.sri.com/cep/publications/AnUnfinishedCanvasSummaryReport.pdf

Wright, W.E. (2007, April). *Impact of high stakes testing policies on English language learners in Arizona*. Paper presented at the meetings of the American Education Research Association, Chicago, IL.