CHAPTER 20

Legislating Competence
High-Stakes Testing Policies and Their Relations with Psychological Theories and Research

RICHARD M. RYAN
KIRK W. BROWN

The development of competence in schools is an increasing focus of national concern in countries across the globe. This concern is fueled by the fact that educational outcomes, broadly considered, are linked with the health and economic well-being of nations. Beyond the obvious economic and health value of schooling to the individual person, the general expansion of education within a nation is associated with a host of outcomes, from reduced mortality and fertility to increased economic productivity and positive social change (Sen, 1999).

Because of the importance of the development of competence, governments are also increasingly attempting to legislate ways to enhance educational opportunities and outcomes. Yet much controversy exists about the appropriate ways governments can stimulate improved schools and greater academic achievement, and what kind of improvements in achievement are actually meaningful for the health and economic well-being of a nation. This issue is international and occupies headlines from Great Britain to South Korea.

In the United States, state and federal government policies aimed at obtaining greater "accountability" and "higher standards" have especially stimulated controversy. These recent policy initiatives attempt to improve school performance through high-stakes testing (HST). Specifically, high-stakes policies represent a two-pronged approach to reform. The first prong entails increased testing to gauge how students, teachers, and schools are performing relative to each other, and relative to the standards that government agencies determine all students should meet. The second prong carries the motivational component: This testing has teeth. The attainment of standards is motivated or enforced by high stakes in the form of rewards and punishments, such as
financial incentives and job security for educators, and grade retention versus promotion for students. HST reform has become, in short order, the most dominant pressure in America’s public schools and is rapidly re-shaping teaching practice and curricular contents across the nation.

What is most interesting about this approach to reform, for the purposes of this volume, is that HST policies reflect particular theories of motivation and achievement. Specifically, high-stakes reform approaches represent a view of competence promotion and teaching that reflects an operant theory of motivation (Kellaghan, Madaus, & Raczek, 1996) and a view of educational outcomes that is more closely aligned with those espousing performance goals rather than mastery or learning goals (Deci & Ryan, 2002); that is, the governmental policy is founded on the idea that making rewards and punishments more salient and contingent on test score outcomes is the most appropriate and effective way of ensuring greater student effort and learning, and more effective teaching. As such, this social policy enacts a behavioristic motivational philosophy and represents a natural experiment in the social psychology of competence. It is a policy that suggests that high-quality educational motivation is a function of external incentives, a view that at least some psychologists support (e.g., Eisenberger, Pierce, & Cameron, 1999; Hidi, 2002).

In contrast, several theories in contemporary motivational psychology predict that attempting to enhance achievement in schools through such external controls will yield some highly negative results, based on the properties of the type of motivation it incites. In particular self-determination theory (Deci & Ryan, 1985; Ryan & Deci, 2000) explicitly predicts important costs of implementing such an approach to motivating competence in public schools. Similarly, some tenets of modern goal theories (e.g., Dweck, 1991; Nicholls, 1984; Elliot, 1999) also suggest potential costs of a focus on demonstrating performance outcomes. Thus, what is scientifically engaging about the social policy debate and implementation is that results of reform should be interpretable, in accord with the varying predictions of these psychological models. What is socially engaging about the debate are the relative costs and benefits to children.

In this chapter, we examine HST reforms in the United States precisely because they illustrate the impact that social policy can have on institutional practice, and the relations (or absence of them) between empirically based research in psychology and education, and governmental policies. We highlight the nature of these test-driven reforms, the legislation surrounding them, and both the theoretically predicted impact and the current empirical data on their effectiveness and consequences. We then discuss the seeming divorce between political reforms and current empirical research in the psychology of competence and education.

To presage some conclusions, our review suggests that, to date, HST has not, in general, produced positive outcomes. Nonetheless, both the positive and negative data that have been obtained can be readily interpreted using the principles outlined in extant theories of motivation. In line with operant theory, and the general recognition of the power of contingent rewards to control behavior, high-stakes policies do indeed change behavior. They lead to increased district, school, and teacher activities intended to raise test scores. In fact, some of the behaviors that these contingencies incite are part of the problem, such as “teaching to the tests,” elimination of developmentally enriching activities that are not likely to be tested, manipulation of targeted standards, and “push-outs” of potentially low performers from the pool of test takers. In line with self-determination theory (e.g., Deci & Ryan, 2002; Ryan & LaGuardia, 1999) and some perspectives on performance-focused motivation (e.g., Midgley, Kaplan, & Middleton, 2001), these high-stakes reforms are yielding a variety of collateral or unintended negative consequences, especially in areas involving persistence and quality of learning. Among the concerns is that HST is typically “one size fits all,” requiring all students, regardless of their backgrounds, learning differences, and rates of development, to jump the same evaluative hurdles simultaneously. This approach potentially lowers the ability of schools to optimally challenge students of different talents and achievement levels, and it is of special concern regarding students with disabilities. An-
other concern is the problem of transfer. Rises in high-stakes test scores do not appear to generalize to other indices of improved achievement (e.g., other achievement measures). This poor generalizability is not necessarily due to the invalidity of the tests, but rather to the criterion contamination caused by their high-stakes implementation. The rewards and punishments that prompt an urgency to raise test scores lead to a narrowing of teaching, and therefore learning, and foster classroom dynamics that tend to decrease student motivation and engagement, as well as teacher morale and creativity. Perhaps more importantly, because HST neither provides a good basis for intrinsic motivation nor offers students optimal challenges (because the standards and methods of demonstrating performance are the same for all), reforms based on HST have been associated with increased school dropouts. These dropouts are especially salient among those already at risk, including the urban poor, students with special needs, and those for whom English is a second language—the very children whom many HST advocates have said they do not want to leave behind.

THE HIGH-STAKES TESTING MOVEMENT

There is little argument that gathering information and providing feedback about performance in educational settings is important for maintaining student and teacher motivation, and for informing educational policy (Linn, 2000; Shepard, 2000). Indeed, feedback regarding outcomes is recognized as a critical feature in improving the function of any organized system (Carver & Scheier, 1998). The function of assessment in gathering information, however, has additional impacts when the outcome data are linked with contingent rewards and punishments, as is the case in HST.

HST has been advocated as a means of motivating students and teachers alike to put in more effort, and thereby raise student achievement (Oakes, 1991; Finn, 1991). Policies instituting HST have taken on varied forms, but the common denominator in such initiatives is that state or federal governments mandate standardized testing of all students and then administer rewards or sanctions based on the results. Students, teachers, and schools that improve or do well are rewarded, and those that decline or do badly are punished. For students, HST results can be the basis for promotion versus retention, and in some states, failure on a single indicator can result in the denial of a high school diploma. Teachers in schools that perform well may get cash bonuses, while those in other schools are reprimanded or derogated. For the schools, the comparative student performance average can result in increases versus cuts in school budgets, and in some cases, poor student performance may result in administration changes or even school takeovers by the state. When the stakes get high for administrators, local officials can even add to the stakes. For example, schools have offered cash prizes, parties, exemptions from finals, scholarships, candy, and awards to high-scoring students (Keller, 2000). School superintendents have been given personal cash bonuses when scores in their districts improve. However, the principal incentive at the administrative level is the public nature of high-stakes assessments. Schools and districts are publicly compared on their test scores, with the often explicit reasoning that pride or humiliation will be attached to the differences in score attainments. Accordingly, at all levels of educational systems, raising the stakes leads to increased attention to test scores because of the consequences attached to them.

A BRIEF HISTORY OF HIGH-STAKES TESTING

The modern HST movement has roots dating back to 19th-century England. Utilitarian philosophers such as Jeremy Bentham (1748–1832) and James Mill (1733–1836) formulated principles of motivation based upon hedonic principles and associationism that provided the foundations of what would become modern behaviorism (Rachlin, 1976). In applying these principles, they suggested the systematic use of rewards and punishments to establish good learning habits in schools. The English Parliament was perhaps the first government to put HST into practice, passing numerous laws intensifying examination structures to ensure liter-
acy, including the Revised Code of Regulations (1862), which advocated a “payment by results” scheme that linked the funds awarded to schools to students’ performance on the exams. Whereas the Code promoted a wider national school system, it also prompted a rigid narrowing of curricula and an escalation of teacher-centered drill- and repetition-focused instruction. Although the Code was eventually repealed, the ideas of “streaming” or segregation of students according to ability level, evaluation by exams, and the resultant conservative methods instituted by the British system in the 19th century continued into the modern era.

In the United States, the modern instantiation of HST begins with the controversial publication of A Nation at Risk in 1983. This document, authored by the National Committee on Excellence in Education, declared that a rising tide of mediocrity was threatening the United States and its ability to compete in the world economy. (Parenthetically, one should note that despite relative stability in achievement standings since 1983, U.S. workers in 2001 were second in the world in global competitiveness according to the World Economic Forum [2002] report). Although one might assume that reform to alleviate “mediocrity” could take any number of directions, the U.S. government’s approach under President Reagan was to step up demands for a core curriculum, more homework, more discipline, and more “accountability” (e.g., performance-based pay for teachers and increased testing), not more resources for schools, in part because lawmakers sought reforms that could be easily understood and rapidly implemented. Within several years following the report, virtually all states adopted more stringent graduation requirements, and many added mandatory homework requirements. School days lengthened and extracurricular amenities shrank. Standardized testing and curricula, matched to what those tests could measure, burgeoned.

Echoing the spirit of these reforms, William Bennett, a politician and popular moralist, proclaimed that “accountability is the linchpin, the keystone, the sine qua non of the reform movement” (Toch, 1991, p. 205). The demand for accountability led quickly to a focus on tests and pressure toward better outcomes on them. Policymakers in nearly every state implemented policies to assess educational standards, and in many of these states, high-stakes consequences were attached to these outcomes, presumably as an incentive-punishment system to motivate change. High-performing schools were to be rewarded and underperformers penalized. Thus, the implementation of policy followed a behaviorist paradigm in which contingent rewards were applied to motivate (and control) teachers and students.

Although there were disappointing results from this early round of HST and many well-documented negative effects (see review by Toch, 1991), the late 1990s saw a new infusion of investment in HST policies. Politicians and business groups lobbied for still greater accountability in public schools, and states increasingly developed tests by which to rank and reward schools based on standardized test scores. Some states, such as Texas, aggressively pursued HST policies throughout the 1990s, and in so doing showed increased scores on the specific tests that were the targets of rewards and sanctions (Haney, 2000). By the first year of the new millennium, nearly all states were using HST in an attempt to foster school achievement. Nearly all states now publish school or district report cards on targeted tests, with the explicit purpose of motivating schools through public pressure or ridicule. Nearly half of all states also provide financial rewards to schools that improve on tests, and threats of administrative change or takeover for those that decline. Many states are directly paying school administrators bonus cash awards when schools under their watch improve on test scores.

Finally, states have been increasingly creating high stakes for students, as well as administrators. The most common high stake is that grade passage versus retention, and ultimately graduation, is contingent on passing a state-administered test. The high stakes of grade retention on the basis of a single examination have been applied as early as the fourth grade (e.g., in Florida). It is explicitly assumed by HST advocates that this type of contingency leads students to work harder in school (e.g., Cheney, 1991; Shanker, 1993), a point contested by critics (see Kellegan et al., 1996). At this point in time, more than half of all states have made grad-
uation from high school contingent on a standardized test performance.

A National Initiative: No Child Left Behind

In 2001, President George W. Bush succeeded in passing, with bipartisan support, landmark legislation entitled No Child Left Behind (NCLB). A stated goal of NCLB is to raise levels of achievement and close the performance gap separating middle-class from poor and underperforming minority students. The plan called for even more testing and more salient stakes for schools and students alike. Specifically, NCLB mandates annual testing in grades 3–8 in math and reading. According to the legislation, scores from such tests are to be used to determine improving and declining achievement, such that penalties and rewards can be attached to them at the level of schools and children. Schools must make steady progress every year toward raising achievement levels on these exams in each of five racial and ethnic subgroups, as well as among low-income students and those with limited English skills or learning disabilities. Failure to demonstrate improvement for any of these subgroups for 2 consecutive years results in a school being labeled low performing. According to NCLB mandates, schools deemed low performing must facilitate the transfer of students to better schools or provide private tutors for students. Schools that continue to be low performing beyond 2 years can have their administrators and staff replaced. Federal funding is made contingent on compliance with these mandates.

NCLB has many critics. Given the expectable, year-to-year deviations that occur in standardized test results, schools may frequently be categorized as low performing for what amounts to statistical issues rather than reasons of educational quality. However, such logistical concerns are not the ones most pertinent to a critique of HST as a strategy of reform. As noted, HST represents a motivational policy. Yet a number of contemporary motivational theories suggest that a host of unintended negative consequences will stem from the pressure and rewards used to externally control teaching and learning. These include narrowing of curricula, teaching to the test, less creative teaching, more superficial and nontransferable learning, more controlling behavior at all levels of power, more withdrawal of effort from at-risk students, and increased dropout rates. We turn first to these theoretical predictions, and then to a review of the accumulating empirical findings on the use of HST.

THEORETICAL PERSPECTIVES ON HIGH-STAKES TESTING

High-Stakes Testing as an Operant Approach

HST is based, at least implicitly, on a behaviorist view of student and teacher motivation. By putting contingent reinforcements on outcomes, the policy presumably increases efforts and behaviors associated with improvement; that is, HST advocates reason that whatever behaviors schools adopt to enhance test scores will be reinforced and selected for, whereas those associated with lower scores will be extinguished and, in the case of poor-performing schools, selected out. Not only will the behavior of teachers change, so will that of students. According to Shanker (1993), strong consequences attached to test scores will provide students with “the incentive to work hard and achieve because they know something important . . . is at stake” (p. 7).

The historical link between HST and behaviorism has deep roots. As previously noted, behaviorism emerged from a blend of British associationism and a hedonic view of human motivation, in which learned behaviors were always a function of external controls that punish or reward. It follows from this perspective that educators should utilize these external forces in regulating learning. This approach to motivation was integral to the work of perhaps the most influential of all behaviorist educators, E. L. Thorndike. The central principle of Thorndike’s theory of learning, which he called connectionism, was his law of effect, which states that if a behavior is followed by a satisfying consequence, it is more likely to occur in the future under similar conditions. Conversely, if a behavior is followed by an unsatisfying consequence, its probability of recurrence will wane. A second principle was that of
frequency: The more frequently an association is repeated, the more likely it is to recur in similar conditions. Together, these "laws" of learning underwrote educational practices focused on the use of external reinforcements, coupled with practice, drill, and repetition. Although these techniques have characterized conservative approaches to education across history (see Ryan & Lynch, 2003), connectionism gave them a specific theoretical rationale.

Thorndike was also an advocate of testing. As he stated, "Testing the results of teaching and study is . . . the sine qua non of sure progress. It is the chief means to arousing . . . the instinct for achievement" (1962, pp. 65–66). However, interestingly, Thorndike was also cautious about how such tests should be used. As he states: "Great care should be taken in deciding anything about the fate of pupils, the value of methods, the achievement of school systems and the like from scores made in a test" (p. 156).

Thorndike's behaviorism was influential in education for several decades but eventually gave way to the "radical behaviorism" of B. F. Skinner. Skinner similarly advocated the systematic application by teachers of consequences, principally positive reinforcements, to induce learning. Skinner also promoted the idea of "programmed learning," which viewed instruction not as based in relationships or interests, but rather in a well-structured and systematic application of contingent reinforcements.

Today conservative educators continue to advocate the use of rewards to control learning, both at the classroom and school system levels. Behaviorists argue that teaching is most effective when based on control through reinforcements. For example, behaviorists Cameron and Pierce (1994), in the context of reporting a now discredited meta-analysis (see Deci, Koestner, & Ryan, 1999), argued that "teachers have no reason to resist implementing incentive systems in the classroom" (p. 397). At a political level, this theme is echoed loudly. Chester Finn has argued that "the problem is that academic success yields such few rewards [sic] and indolence brings few penalties" (1991, p. 120). He, and a broad array of conservative spokespersons, have argued that putting rewards and penalties behind the test scores will effectively alter the behavior of both teachers and their students. This type of thinking has deeply influenced recent educational reforms in several nations focused on HST. In this view, instruction should be driven by measurement, and the outcomes of measurement should be the basis of rewards and sanctions for both teachers and learners (as discussed in Popham, 1983).

Our interpretation of the HST movement as reflecting an operant strategy has one very important caveat. Operant theory has always been focused on making rewards contingent on target behaviors. The twist in the HST movement is that its advocates apply contingent rewards and sanctions to performance outcomes; that is, rather than rewarding valued behaviors, such as student effort or work habits, contingencies are instead applied to test outcomes, the control over which is often questionable, especially for at-risk students. Similarly, rather than rewarding excellent teaching activities and approaches, schools are rewarded or sanctioned on their test score results. This practice is not in line with the fundamental tenets of the operant viewpoint. Indeed, we believe that the focus on performance outcomes, rather than on behaviors that students and teachers have direct control over, is one of the features of HST that lead to reinforcement of the wrong behaviors.

This focus on outcomes does find affinity from some theorists who focus on goals as motivating forces in behavior. Among those perspectives that could be aligned with HST-based reforms is the goal theory approach of Locke and Latham (1990), who argue for a high-performance model in which demanding goals are linked with both internal and external rewards to maximize organizational efficiency. Although they developed their model in application to industry, they suggest its generalizability to schooling, arguing that the high-performance model of difficult goals associated with rewards for success "should be made part of our schools as well as our work organizations" (p. 268). Advocating this linkage between measurable outcomes and performance-contingent reinforcements would seem to be fully congruent with the HST approach. A similar advocacy of applying contingent rewards to performance outcomes has also been forwarded by Hidi and Harackiewicz (2000), whose perspective on performance goals we
review in discussing theories of mastery and performance goals.

**Organismic Perspectives on Learning**

A very different view of what motivates learning and competence can be gleaned from what has sometimes been called the "liberal perspective," and sometimes the "organismic perspective," in which learning is seen as an inherent or intrinsic tendency of the person (Ryan & Lynch, 2003). In this tradition, the desire to learn is seen as a natural or basic tendency of humans. Learning is growth. However, like all growth, this inherent initiative or tendency requires support and nutriments. The result is a process (rather than outcome) focus, in which nurturance, mainly in the form of warm relationships, optimal challenges, and supports for autonomy and interest, are the most common elements.

Throughout history, educators embracing this liberal view have argued that students are not optimally motivated by external controls, but rather by support of their inherent tendency to learn. In ancient times, this view was espoused by Quintilian, who recognized that learners of different ages and types have distinct needs and interests, and held that curriculum and methods should be tailored accordingly. He deemphasized the then common use of punishment, instead stressing the importance of making learning interesting and attractive. In the Renaissance, similar views were echoed by Comenius, who focused on the strategic importance of warm student–teacher relationships and enhancing students' interest in learning. Subsequently, Enlightenment philosopher Rousseau laid the groundwork for much modern thinking in the liberal vein, emphasizing children's curiosity and natural inclination to learn under supportive conditions.

Rousseau influenced generations of subsequent educators. Outstanding among them was the Swiss educator, Pestalozzi, who viewed the aims of education not as "imposing on the child fixed doctrines and alien concepts but in helping him to develop his own constructive powers" (Silber, 1973, p. 274). His method of education entailed, first and foremost, an atmosphere of emotional security based in a warm and caring relationship between teacher and child. He advocated that knowledge be gained, when possible, through direct experience rather than through mere words passed from teacher to child. He also downplayed the utility of punishment and fear of evaluation, suggesting that if provided a secure base, the child's nature would lead to discovery and growth. Pestalozzi's philosophy was widely disseminated during the 19th century in Europe and the United States, and became a major influence on a diverse family of practitioners, including Froebel in Germany, and Montessori in Italy.

Finally, in the 20th century, Dewey (1938) emphasized the importance of cultivating interest and inquiry in crafting an education, rather than arbitrarily imposed educational tasks and goals. He stood, in this respect, in stark contrast to his behaviorist contemporary, Thorndike. In the realm of psychology, Rogers (1969) developed an influential perspective on teaching, stemming from his person-centered approach. He advocated a classroom experience that grows out of the authentic inquiry of the student. Rogers felt that the external locus of evaluation represented by traditional examinations and normative grading stifled the significant learning that grew out of a student-centered, responsive teaching environment. It was Rogers who faced off with B. F. Skinner in the 1950s and 1960s, debating the value of external control versus self-actualization in the enterprise of learning.

In summary, a long tradition of philosophy and psychology has argued against externally controlling techniques as the via regia to student learning. Instead, this tradition focuses on nurturing the natural inclination to learn, the diversity of learning abilities and styles, and the importance of students' developing their powers of self-evaluation. Importantly, the last few decades have seen the emergence of several empirically focused motivation theories that supply some support for this perspective.

*Self-determination theory (SDT; Deci & Ryan, 1985; Ryan & Deci, 2000)* is one such empirically based organismic perspective that views humans as intrinsically motivated to learn and develop competencies. However, the theory is centrally concerned with the conditions that support versus thwart these intrinsic propensities. SDT is thus particularly interested in the impact of events
such as evaluations, praise, and contingent rewards and punishments on behavior and learning.

Specifically, SDT highlights the fact that students' motivation to learn can vary in its relative autonomy, from behaviors motivated by external rewards and punishments (controlled motivations) to those that are energized by interests and values (autonomous motivations). Both evidence and theory based on SDT suggest that, to the extent that one's motivation is based on more autonomous motives, such as intrinsic motivation or well-internalized values, the more quality of learning, persistence, and affective experience are enhanced (Grolnick & Ryan, 1987; Ryan & La Guardia, 1999; Ryan, Stiller, & Lynch, 1994). On the other hand, SDT research has found that motivation based on more controlled motives, such as rewards or punishments (external regulations), or self-esteem-based pressures (e.g., ego involvement) is associated with lower quality learning, lessened persistence, and more negative emotional experience.

Because HST policies are based on the idea that rewards, punishments, and self-esteem-based pressures are effective motivators of learning, the principles of SDT apply (Deci & Ryan, 2002; Ryan & La Guardia, 1999). In what follows, we summarize the theoretical basis for those hypotheses as they relate to teacher and student motivation, and review some of the evidence supporting the validity of these hypotheses.

According to SDT, the specific effects of external events such as evaluations or feedback on human motivation depend on the psychological meaning, or functional significance, that the events have for the recipient (Deci & Ryan, 1980, 1985, 2000). The theory specifies that the functional significance of an external event, such as a test score, a tangible reward, or praise from a teacher, can be informational, controlling, or motivating. Events have informational significance when they provide effectance-relevant feedback in a noncontrolling way; that is, when an event provides individuals with specific feedback that points the way to being more effective in meeting challenges or becoming more competent, and does so without pressuring or controlling the individuals, it tends to have a positive effect on self-motivation. Events have controlling significance when they are experienced as pressure toward specified outcomes or as an attempt to control the activity and effort of the individual. According to SDT, when evaluations have controlling significance, they may produce temporary compliance, but they ultimately undermine self-motivation, investment, and commitment in the domain of activity being evaluated. Finally, events have amotivating significance when the feedback conveys incompetence to the individuals or supplies no inner or outer rationale for acting. Evaluations or reward structures based on overly challenging standards, or that are perceived to be beyond the reach of the individuals, are thus amotivating: They undermine all motivation and lead to withdrawal of effort. Teaching that does not tap into a student's interests, or that does not supply a basis for the experience of relevance or meaning, can also foster amotivation.

Both experimental and field studies have supported these predictions concerning the impact of events such as feedback and rewards on subsequent motivation. Extensive reviews are available elsewhere, but a few examples are worth detailing. In experiments with rewards, Ryan, Mims, and Koestner (1983) showed that reward structures delivered in an informational manner did not undermine intrinsic motivation, but rewards used to pressure people toward a specified outcome did. In another demonstration, Ryan (1982) showed that students who were pressured to perform by stressing that outcomes reflected ability (an ego-involving induction) were subsequently significantly less likely to engage with the target task than were students who were induced to focus on the task itself rather than task outcomes. In an experiment conducted within an elementary school, Grolnick and Ryan (1987) had students engage in a reading comprehension task under three conditions. In the first, students were told they would not be tested at all. In the second condition, they were told they would be tested, but only to determine what kinds of ideas were learned, so there were no consequences for failure or success. In a third condition, students were told they would be tested and graded, and that the grade would be delivered to their classroom teacher. This third condition represented a controlling use of
evaluations. Results showed that the controlling evaluation condition promoted not only short-term, rote memory but also produced a significantly lower level of conceptual learning and knowledge integration than the two noncontrolling conditions. Evidence from these and related studies (e.g., Benware & Deci, 1984) indicates that when tests, evaluations, and rewards are used in controlling ways, they have negative effects on students' interest, motivation, and higher level cognitive outcomes in school. Classrooms studies have added to these findings by showing that when teachers are oriented toward being controlling (e.g., using evaluations and rewards), students are less intrinsically motivated, less desirous of challenge in school, and also less confident in their skills (e.g., Deci, Schwartz, Sheinman, & Ryan, 1981; Ryan & Grolnick, 1986).

How Performance Standards Affect Teachers

The finding that when teachers use controlling strategies and performance pressure to motivate students, the students become less self-motivated and less engaged in school, raises an interesting issue. What factors lead teachers to be controlling? One answer is that they may become controlling when they themselves are pressured to get children to perform. An experiment performed by Deci, Spiegel, Ryan, Koestner, and Kauffman (1982) addressed this issue. Participants simulated teachers with the task of helping students learn a cognitive-perceptual task. The teachers all had the same set of problems to work with and were given the same preparation. However, before entering the teaching session, one group was explicitly told that it was their job to make sure their students performed "up to high standards," whereas another group received no such pressure. The sessions were recorded and rated for differences in teaching strategies. The results showed that the participants who were explicitly pressured to produce high student achievement were more controlling and less supportive of students' autonomy. Specifically, teachers in the performance standards condition engaged in more lecturing, criticizing, praising, and directing—all techniques that have been shown to have a negative impact on students' interest in learning and their willingness to undertake greater academic challenges. Flink, Boggiano, and Barrett (1990) followed up on this reasoning by examining a newly introduced school-based curriculum for elementary students across several schools. They showed that, as predicted, teachers pressed toward higher standards were more likely to engage in controlling instructional behaviors. In line with SDT, the more they did so, the more their students actually performed more poorly on objective test-score outcomes. This is consistent with a wide body of literature linking evaluative pressure with poorer performance in schools (Kohn, 1996; Ryan & Stiller, 1991), as well as dropout rates (Hardre & Reeve, 2003).

From the SDT perspective, creating a test-driven evaluative focus not only leads teachers to be more controlling but also leads students to be more externally regulated and/or ego involved in their motivational orientation. According to SDT, ego involvement is potentiated whenever a person's esteem is linked with attainment of specific outcomes (deCharms, 1968; Plant & Ryan, 1985; Ryan, 1982). Accordingly, ego involvement can motivate effort, just as rewards can. However, like most performance-contingent rewards, ego involvement is a controlling form of extrinsic motivation, and it runs the risk of undermining internal motivations based in value or interest. Furthermore, unless one is ensured of success when applying effort, ego involvement can have deleterious immediate effects. The more ego involving a context, the more many students, particularly the less confident ones, withdraw effort in order to reduce the diagnosticity of tests and thus protect their self-esteem (Martin, Marsh, & Debus, 2001). Additionally, even for students who try to do well, such evaluation-based motivations tend to foster more superficial and less integrative learning processes, thus debilitating long-term knowledge retention and growth (Golan & Graham, 1990; Grolnick & Ryan, 1987).

Beyond this, the evidence suggests that focusing parents' concerns on performance outcomes will lead them, like teachers, to use pressuring motivational strategies that will backfire, leading to lower achievement over the long term (Ginsburg & Bronstein, 1993; Grolnick, 2003; Grolnick, Gurland, Decourcey, & Jacob, 2002; Grolnick &
Ryan, 1989). In short, pressure (whether it be through rewards or esteem-related threats) to meet externally dictated or controlled standards usually translates into lower quality teaching and less effective motivational practices, unwittingly undermining high-quality performance, as well as the interest and task involvement that facilitate it.

It should also be mentioned that use of uniform evaluative standards for all students regardless of their starting points or resources, which is a invariant feature of HST policies, violates another tenet of SDT’s approach to motivation. According to the theory, people are most intrinsically motivated when they are optimally challenged—when the tasks set by or for them are within reach. Tasks that are overly challenging have amotivational significance, and thus undermine motivation altogether, leading to lower effort withdrawal, helplessness, and lower confidence and self-esteem (Deci & Ryan, 1985; Ryan & La Guardia, 1999; Vallerand & Reid, 1984). The evidence is clear: If the bar appears to be too high, many students will experience futility and withdraw their effort. People are simply not motivated by the prospect of failure.

Moreover, test-based reforms seem to ignore the diversity of ways in which students both learn and demonstrate learning. As Gardner (1991) has argued, even a well-constructed test may be a nonoptimal challenge for some children, and may present a distorted picture of how well that student has mastered or understood material. Because the hallmark of HST is a single criterion, it favors those who are most apt within its format.

Together, these tenets of SDT would suggest that HST will have a number of negative effects, many of which are undoubtedly unintended (see Ryan & LaGuardia, 1999). The controlling reward structure behind HST should, according to SDT, externally regulate the behavior of teachers. They are thus predicted to engage in those behaviors instrumentally tied to test scores, regardless of their inherent value or worth. One should thus see a narrowing of curricula, more teaching to the test, more controlling motivational techniques used in classrooms, and less positive experience on the part of students and teachers alike. Because of the motivational dynamics set in motion in the classroom, SDT also predicts greater dropout rates among students, especially those at risk for failure or alienation, since withdrawal of effort is a common fallout of controlling and nonoptimal pressures, and uninspiring classroom practices. Systems such as state and district administrations will, because of the high stakes, be driven to “fuzzy accounting methods” (e.g., wavering standards), pushing out students who might bring down scores, and using other devices to maximize the target outcome, regardless of other costs of such behaviors. Yet, because there is pressure on narrowly defined test-score outcomes, scores on targeted tests should increase, but such increases will not necessarily generalize to other indices of achievement, because these increases were obtained through methods that do not incite more self-motivation, interest, and value for learning.

Achievement Goal Theories: Divided Views on the Value of Performance Goals and High Stakes

Another family of theories that has relevance to HST initiatives is those that concern performance versus mastery goals in the achievement domain, and the conditions that inspire them (e.g., Dweck & Leggett, 1988; Elliot, 1999; Nicholls, 1984; Pintrich, 2000). Although the theories differ in some details, the critically important distinction is between goals that are focused on increasing or developing one’s competence or knowledge (called mastery or learning goals) and those focused on proving or demonstrating one’s competence or ability (often called performance goals). HST, by focusing on the demonstration of specific test scores and using rewards to make that demonstration salient, represents an institutional climate that one might expect to catalyze performance goals; that is, by making the demonstration of competence the most salient issue, students, teachers, and administrators alike would be likely to adopt a performance goal orientation.

A large body of evidence suggests that very different behaviors and quality of learning typically follow from performance versus learning and mastery goals. This evidence suggests that the more students are
focused on learning or mastery goals, the more extensively they enjoy learning, make greater use of higher level cognitive strategies, experience greater efficacy, and show better integration of what is learned (Ames & Archer, 1987; Elliot, McGregor, & Thrash, 2002; Midgley, Anderman, & Hicks, 1995; Midgley et al., 2001). Performance goals, by contrast, appear to foster a more superficial approach to learning, because the motivation is to demonstrate rather than attain competence. For example, a meta-analysis by Utman (1997) suggests that performance-focused goals can produce enhanced performance at rote or algorithmic tasks but tend to undermine performance at more heuristic or complex tasks. Furthermore, students with learning goals are often more willing to tackle challenges and difficult material, whereas those with performance goals are often more interested in demonstrating competencies already attained (Ames, 1992; Thorkildsen & Nicholls, 1991). Finally, performance goals have been linked to greater self-handicapping (Martin et al., 2001; Urdan, Kneisel, & Mason, 1999) and may leave students more vulnerable to helplessness when failure occurs (Dweck, 2002).

However, despite the numerous advantages of mastery goals in learning contexts, Elliot and his colleagues (see Elliot & Thrash, 2002) introduced an important distinction within goal theories between performance-avoidance and performance-approach goals. *Performance-avoidance* goals concern situations in which the student is primarily motivated to avoid failure or negative outcomes in the demonstration of performance. *Performance-approach* goals refer to a more appetitive desire to positively demonstrate high performance. Much empirical literature supports the view that the adoption of performance-avoidance goals has many negative consequences. By contrast, performance-approach goals seem to show fewer detrimental effects and can inspire some positive consequences (Elliot & Moller, in press).

It is important to realize that current HST systems do not, at least strategically, aim differentially to foster performance-approach rather than performance-avoidance goals. Indeed, the rhetoric of HST suggests that advocates expect that both desire to attain success and fear of failing at these demonstrations are engendered. Indeed, they may activate both to different degrees, both across and within individuals (Elliot & Moller, in press; Midgley et al., 2001).

Nonetheless, among the achievement motivation theorists focused on the performance versus the mastery goal distinction, opinions are divided as to the implications of the findings. Some theorists seem quite positive about having performance goals coupled with rewards be a central focus in classrooms. For example, Harackiewicz, Barron, Carter, Lehto, and Elliot (1997) argued that performance-approach goals are "adaptive" in settings where achievement is competitively defined or based on normative comparisons, because those whose adopted goal is to demonstrate high performance are more likely to do so. Hidi and Harackiewicz (2000) further advocate linking performance goals with extrinsic rewards. They speculated that performance goals linked with reward contingencies may be effective in promoting long-term interest and intrinsic motivation, especially among unmotivated and at-risk students. As Hidi (2002, p. 332) puts it: "Why should we assume that our children will produce high level schoolwork without expecting and receiving rewards?" Such thinking clearly mirrors the philosophy of HST advocates such as Bennett and Finn.

In contrast, other researchers in this domain hold that a focus on promoting performance demonstrations rather than mastery development in real-world classrooms will yield few positive and many negative motivational outcomes. Midgely et al. (2001), for example, highlight the fact that an emphasis on performance goals at best supports and rewards only highly achievement-oriented students who are certain about their abilities, and even for many of them, it leads to an extrinsic and superficial focus, and to vulnerability, if academic setbacks occur. In a context that emphasizes performance goals, they further suggest that many students, especially those with lower or uncertain abilities, will show increased self-protective strategies such as self-handicapping and withdrawal of effort. Thus, performance-focused classrooms may lead some students to be more extrinsically motivated to perform well, but, at the same time, it will lead to lessened intrinsic motivation and
withdrawal of effort among those at risk for failure, a prediction in opposition to the view of Hidi and Harackiewicz (2000).

Between these views, Elliot and Moller (in press), even while highlighting the clear benefits of students adopting performance-approach goals, suggest that institutional policies should still be directed toward a mastery focus. For them, performance-approach goals, when they arise, are a natural expression of competence urges (Elliot et al., 2002). However, in their view, policies aimed at performance put many students at risk for undermining effects, because many will adopt an avoidance focus under such a circumstance.

Thus, performance-mastery goal theories lack consensus regarding the effects of establishing performance goals as a modus operandi in schools and, by implication, on the effects of HST reforms. Some in this tradition suggest a positive influence of performance goals linked with contingent rewards on promoting interest and achievement efforts, whereas others suggest that a performance goal focus backed by high stakes will lead to numerous deleterious results, especially for at-risk students. Still others suggest the need to develop strategies that could foster performance-approach orientations, without simultaneously generating performance-avoidance concerns in the same setting, although ways to do that have not yet been explicated.

THE RESULTS OF HIGH-STAKES TESTING

Given the clear, yet opposing predictions from theories of motivation on the impact of HST, it is interesting to look at what the accumulating evidence actually shows. It is important to note that full-fledged HST programs are still being phased in within most states; thus, the full impact of HST has not yet been felt. In addition, although anecdotes abound, only a few credible empirical studies are available. Nonetheless, there is a growing body of evidence associated with these initiatives, and we review the most extensive studies to date.

Moon, Callahan, and Tomlinson (2003) surveyed a nationally stratified random sample of teachers on the effects of state HST programs on their classroom practices. Results indicated that classroom practices were strongly affected, especially in schools serving students in the lowest socioeconomic strata. Teacher reports suggested that HST was indeed salient, and that increases in test scores are not necessarily a result of student academic attainment, but are more due to test preparation. Test preparation associated with HST was reported to drive out other instructional activities, because much time was taken in the classroom to review and practice for state testing. Test preparation was especially intense in poorer districts. The authors speculated that one result of HST is a narrowing of the curriculum and the implementation of practices that may actually run counter to effective instruction, student self-direction and autonomy, and opportunities for interaction between students. Indeed, the authors suggested that the very salience of HST in the minds of teachers may be restricting educational opportunities, particularly among those from the most impoverished areas. Moon et al. further suggested that when teachers specifically teach to the test, the scores may no longer represent the broader domain of knowledge for which they are supposed to be an indicator, especially in schools serving disadvantaged students, where the test preparation was reported to be more intensive.

A study by McNeil and Valenzuela (2000) of Texas teachers arrived at similar conclusions. They found that teachers were encouraged or required to reallocate time away from core subjects not tested on the state examinations, and to eliminate or curtail special projects, experiments, library research, extensive writing, or oral assignments. This was especially true in schools that might be lower in absolute performance levels (i.e., those serving less affluent students). Much time was also reported being spent specifically on test-taking strategies rather than substantive issues.

Evidence that HST leads to “teaching to the test,” which in turn crowds out the teaching of skills not on the tests and the provision of enriched experiences that might better engage students’ interest in additional knowledge seeking, may underlie the concern with the generalizability of score gains. This issue can be partly addressed by examining transfer, or the extent to which gains
on HST are reflected in evidence of improved achievement on other, nontargeted measures. Little research exists on the validity of test-score increases on HST, despite the fact that it is a crucial bone of contention between HST advocates and their opponents.

Perhaps the most comprehensive look at this issue was an 18-state study by Amrein and Berliner (2002). To test the transfer of score increases on high-stakes examinations, they obtained scores on non-HST that overlap with HST in their assessment of achievement domains. These were the ACT (established by the American College Testing Program), Scholastic Aptitude Test (SAT), National Assessment of Educational Progress (NAEP), and Advanced Placement (AP) tests. Their evidence suggested, contrary to that of HST advocates, that when transfer is considered, level of learning in those states with salient HST policies remains level or falls below previous levels once HST is implemented. In contrast, states without high-stakes graduation tests were more likely than states that had imposed them to show improvements on these outside tests. Indeed more than two-thirds of states posted decreases on ACT performance after high-stakes graduation exams were implemented.

Neil and Gaylor (2001), using the NAEP as a metric, similarly showed that states without HST were more likely to show score improvements than states with them; that is, NAEP scores were not improved by HST initiatives, and they also had many other potentially negative consequences. They specifically suggested that HST may widen educational outcome inequities between the rich and the poor rather than ameliorate them.

With so much attention paid to test scores, an equally important gauge of school performance is high school dropout rate. Although dropouts are hard to track and are often systematically misreported (Orfield, Iosen, Wald, & Swanson, 2004), available data show that both dropouts and students leaving high schools for equivalency diplomas are on the rise, with notable escalation in the past few years as HST policies have intensified. Indeed, Reardon and Galindo (2002), for example, studying students between 8th- and 10th-grade in districts with and without HST policies, estimated that the imposition of HST increased the odds of dropout by 39%.

Although accounts differ, one possibility is that as states required students to pass tests for promotion, more pupils were held back. In turn, convincing data suggests that the mere fact of retention dramatically increases the probability of dropout (Natriello, 1998). In addition, if one assumes that HST imposes even modestly more difficult standards, that, too, could lead to a motivation and discouragement among students already at risk for failure.

A related issue is the concern that HST may lead many students to seek a general equivalency diplomas (GED). Studies comparing high school graduates to young people who received equivalency diplomas show that even among those with similar academic scores, those who complete high school have higher earnings, secure better employment, and commit fewer crimes. One reasonable account of this is that the confidence, self-esteem, and work habits of young adults is greater if they graduate from high school than if they drop out to earn a GED, and that confidence translates into better adult outcomes. In other words, if HST drives students out of school, this has costs, most of which will be borne by children from lower income families.

Jacob (2001) examined the effects of high-stakes high school examinations on student retention, especially among low achievers, who, some have argued, would most benefit from a performance-based focus (e.g., Hidi & Harackiewicz, 2000). His findings, based on analysis of data from 15 states, showed that students in the bottom 20th percentile of achievement who faced such requirements were 25% more likely to drop out in states with tests. He also found, however, that use of the tests had no significant effect on subsequent academic achievement for the population considered as a whole.

Another way to examine the impact of HST policies is to examine the results in Texas, where the most widely cited and lauded HST program has been in place since the early 1990s. HST policies in Texas have been described in the press as the “Texas Miracle,” and have become a model for other reform efforts, including the federal NCLB program. This enthusiasm was partially based on the fact that scores on the
Texas State Achievement Tests (the TAAS) had shown large gains under the high-stakes regimen; TAAS scores provided evidence of a decreasing gap between minority and white students. An independent report by Grissmer, Flanagan, Kawata, and Williamson (2000) of the RAND Corporation initially suggested that the high-stakes policies themselves might have facilitated this positive trend. However, a subsequent report by RAND investigators (Klein, Hamilton, Mccaffrey, & Stecher, 2000) found that such gains in TAAS scores did not match trends on other measures, raising serious questions about the meaning of these achievement gains, or their transfer, and about the validity of the score gains. With regard to the achievement gap, results from other tests besides the TAAS also suggested that the gap might have slightly widened in Texas, over the same period that TAAS scores suggested it was closing.

At the same time, evidence of higher grade retention and dropout rates in Texas has accumulated (Haney, 2000), and outright cheating on results has been documented (Hoff, 2000; Johnston, 1999). Haney (2000) found that increased dropout rates in Texas were especially high among Latino and African American students. Haney linked these dropouts with aggregate score gains, arguing that Texas students’ gains in NAEP scores were directly related to exclusion rates. Haney concluded that the apparent rise in scores was illusory. Tracking these dropouts, Haney found that approximately one-third of students leave school before graduation, often as a direct result of being retained in grade 9 by schools focused on obtaining good HST scores.

Moreover, evidence from Texas points to considerable teaching to the test, again, especially intensively in low-performing schools serving pockets of poverty and minority students. Such teaching to the test can give the appearance of “closing the gap” when that is not occurring, because of the criterion contamination this behavior causes (Carnoy, Loeb, & Smith, 2000; McNeil & Valenzuela, 2000). For such reasons, Popham (1999) concludes that judgments about school quality based on changes in HST scores are not likely to be valid.

Despite the limitations of the empirical studies thus far conducted, it is not unreasonable to suggest that the evidence points to the very kinds of changes predicted by some of the motivational theories we reviewed. Under HST, outcome-focused behavior change does indeed occur, no doubt due to the power of rewards and sanctions. Yet these changes are often a “monkey’s paw,” representing deleterious classroom and institutional processes that hurt especially the most vulnerable populations. This in turn suggests that the HST policies may be exacerbating the problem they are designed to correct. Nonetheless, these negative results should not be taken as a definitive summary or as the final chapter. We reiterate that the results of HST policies are still unfolding. At the same time, there are clearly problems with the impact of HST, which predictably motivates counterproductive processes in both classroom and school administration arenas. It is ultimately the economically disadvantaged students, as well as the frontline teachers who serve them, that appear to suffer the most serious costs.

MOTIVATION THEORIES AND EDUCATIONAL REFORM

One conclusion we reach from reviewing this material concerns the relevance of debates between theories of motivation to policies attempting to legislate competence in schools. We have underscored how policymakers have, at both state and federal levels, enacted policies driven by a naive behaviorism in their attempts to motivate improvements in school performance. Unlike behaviorists, however, they have applied rewards and sanctions contingently upon performance outcomes (test scores) rather than desired behaviors, and they have also not appreciated the well-documented deleterious effects that even a well-structured contingency management approach can yield in domains such as learning and education. At the same time, results bespeak the power of such contingencies to change behavior, if not necessarily for the better.

The specific deleterious effects of such high-stakes policies have been predictable, and sometimes explicitly predicted by some motivational perspectives, whereas others have not addressed these “collateral” conse-
quences. Most notably, self-determination theory has specifically argued that these reforms would foster teaching to the test, narrowing of the learning experience, relatively poor transfer of knowledge, and increased dropouts among those most disadvantaged (Ryan & LaGuardia, 1999; Ryan & Stiller, 1991). All of these predictions have come home to roost in states that have used HST. Similar deleterious effects may have been predictable from some goal theories as well, particularly the perspectives of Dweck (2002) and Midgeley et al. (2001). These views stand in contrast to the views of those who have advocated greater emphasis on performance goals in classrooms linked with high stakes. Rather than facilitating achievement in at-risk students, such motivational interventions seem especially harmful to vulnerable groups. If nothing else, one lesson we should learn from this is that our theoretical and empirical differences are far from merely "academic."

SOME POLICY IMPLICATIONS

Empirical research is critical to informed policy in education, yet the gulf between the types of reforms suggested by educational research and those being implemented by policymakers appears vast. In part, this stems from the fact that policymakers want clear-cut actions, an urge that the implementation of high-stakes and standardized tests appears to satisfy. At the same time, as the effects of this "natural experiment" unfold, we should make sense of the results and outcomes, learning from the implementation (Hamilton, Stecher, & Klien, 2002). To do so we use the lens of SDT, which has specifically predicted many of these effects.

The SDT perspective suggests that tests can have both informational and controlling effects, and the high-stakes approach has largely undermined the informational value of standardized testing. Policymakers might first remember the purpose of testing: To gain information that can be used to advocate for those assessed. The informational use of tests would be represented by using tests to help identify students who may be most disadvantaged and in need of resources, and perhaps to identify curricular issues or problems with teaching methods. Informational use of tests would also require that they be useful to teachers—that they would not simply be a scorecard at the end of a year, but a useful indicator of gaps in knowledge, while there is still time to redress the situation. The current practice in most HST states is year-end testing, with individual score reports often not going to the teacher who taught the subject matter until the following year, which is of little educational benefit to the participating students.

More importantly, the positive effects that can come from the informational function of tests are undermined when policymakers place high stakes behind test outcomes. The implementation of high-stakes contingencies based upon test performance, which are intended as "motivators," actually do have a strong impact. They lead to practices that distort the validity of the outcomes, and that instigate deleterious institutional behaviors. They narrow curricula, decrease individualized approaches, and make even more vulnerable those students who are at risk for retention and dropout. Taking the stakes out of the heart of testing policies would make the testing more informationally valuable. Whereas high stakes contaminate the criterion, removing the stakes might make standardized testing all the more useful, and less engendering of damaging processes.

A further important issue concerns the fact that any standardized paper-and-pencil measure may be a poor fit with the learning and performance styles of some learners, making it inappropriate as a sole criterion for attaining credentials. "One size fits all" as a model of outcomes is a regressive step in schools, where for years educators have been developing approaches to address more effectively diversity in learning styles, interests, and skills. Moreover, basing high-stakes decisions on a single indicator is unfair to students, and even unethical, given the lack of validity of most of the tests for this purpose (American Educational Research Association, 2000). Accountability does not need to be actualized by only a single, uniform test. Instead, schools that use alternative approaches and curricula could develop and justify alternative assessments. This would in fact lead toward greater innovation rather than drying up choice and diversity, which has been the trend under HST.

In a context where testing was used for in-
formational rather than controlling purposes, educational experiments might actually permit better judgement on their effectiveness, and indeed catalyze more innovation and progress. For instance, there appears to be growing evidence that high schools organized into small schools or learning communities, where personalized attention is available, are effective in promoting achievement (e.g., Howley & Bickel, 2000; Meier, 1998; National Research Council, 2004). Effective non-high-stakes testing could both verify and extend such data, and be a basis for justifying such structural reforms to policymakers and taxpayers. Similarly, an innovative and highly successful experiment in redesigning urban high schools was the creation of the New York Performance Standards Consortium (NYPSC) schools. These schools had served as models and were recognized for their high educational standards, high attendance, and low dropout and college success rates (Darling- Hammond, Ancess, & Ort, 2002). However, NYPSC schools were built around a portfolio-based assessment system that was deemed integral to the form of instruction, which itself was highly individualized rather than standardized. These successful schools are being forced under New York’s rigidly enacted high-stakes regimen to change their practices and teach to the tests. In a non-high-stakes atmosphere, standardized tests might have been one among several useful indices affirming their efficacy, but in a high-stakes atmosphere, the curriculum will be bent to the shape of tests, and a successful innovation stifled.

An important take-home point is that the introduction of high stakes behind test scores distorts the validity of tests as an indicator of true excellence in the classroom, or of school quality. Amrein and Berliner (2002) described this distortion effect by evoking the Heisenberg Uncertainty Principle. According to the principle, the more important any quantitative indicator becomes in decision making, the more likely it will distort and corrupt the process it is intended to monitor. Because high-stakes policies attach reward and punishment contingent on test scores, they especially have such distorting and corrupting consequences. They make the meaning of test score changes questionable, and they make inferences from score changes problematic. Combined with the fact that most states use percentage-passing rates on tests that are not equivalent from year to year, many of the inferences concerning the outcomes of reform are without a sound scientific basis.

While the massive educational experiment called HST is still in progress, it is clear that what is driving national and state education policy is not sound educational theory or research, but a blend of political expediency and naive faith in the efficacy of rewards and punishments. Research that has accumulated points to complex, and often negative, effects that may not be willingly received by politicians who, in many instances, may “have already decided” that HST is an effective approach (Hamilton et al., 2002). On a more positive note, we suggest that current work in the field of motivational psychology is highly relevant to, and capable of, meaningfully informing the process of education reform. The question is, who might be listening?

REFERENCES


the accountability rhetoric. Cambridge, MA: Harvard University Civil Rights Project.


Popham, W. J. (1999). Educators’ indifference to the misuse of standardized tests is having calamitous consequences. *Education Week*, 18, 48–32.


