Ryan, Richard M.; Sapp, Aislinn

Zum Einfluss testbasierter Reformen: High Stakes Testing (HST). Motivation und Leistung aus Sicht der Selbstbestimmungstheorie

Unterrichtswissenschaft 33 (2005) 2, S. 143-159

Quellenangabe/ Reference:
Ryan, Richard M.; Sapp, Aislinn: Zum Einfluss testbasierter Reformen: High Stakes Testing (HST). Motivation und Leistung aus Sicht der Selbstbestimmungstheorie - In: Unterrichtswissenschaft 33 (2005) 2, S. 143-159 - URN: urn:nbn:de:0111-opus-57919 - DOI: 10.25656/01:5791

https://nbn-resolving.org/urn:nbn:de:0111-opus-57919
https://doi.org/10.25656/01:5791

Unterrichtswissenschaft

Zeitschrift für Lernforschung
33. Jahrgang / 2005 / Heft 2

Thema:
Förderung der Lernmotivation

Verantwortliche Herausgeberin:
Elke Wild

Richard M. Ryan, Aislinn Sapp

Zum Einfluss testbasierter Reformen: High Stakes Testing (HST)

Motivation und Leistung aus Sicht der Selbstbestimmungstheorie

Considering the Impact of Test-Based Reforms: A Self-Determination Theory Perspective on High Stakes Testing and Student Motivation and Performance

Der Einsatz standardisierter Tests auf Länder-, Schul- und Schülerebene steht zunehmend im Zentrum des öffentlichen Interesses und liefert die Grundlage für pädagogische Reformbemühungen in der gesamten Welt. High Stakes Testing (HST) ist ein spezifischer Reformansatz, der die Vergabe von Belohnungen und Sanktionen an die Ergebnisse in solchen Vergleichstests bindet. HST ist somit eine Motivationsstrategie zur Verbesserung pädagogischer Zielvariablen. In dieser Arbeit wird das Verhältnis von HST zu behavioristisch orientierten Theorien, Theorien motivationaler Zielorientierungen und insbesondere der Selbstbestimmungstheorie (Deci & Ryan, 1985) untersucht. Es soll herausgearbeitet werden, inwiefern diese Theorien HST als Strategie zur Motivierung von Lehrern und Schülern unterstützen bzw. welche Konsequenzen HST aus Sicht der verschiedenen Theorien haben sollte. Anschließend wird ein Überblick über empirische Studien aus Amerika (wo HST zurzeit äußerst populär ist) gegeben, die die Wirkung von HST überprüft haben. Die Ergebnisse werden dann mit den theoretischen Vorhersagen verglichen. Aus diesem Vergleich werden die folgenden Schlüsse gezogen: HST ist eine stark kontrollierende, extrinsische Motivierungsstrategie, die zwar oft die gewünschten Wirkungen hat, gleichzeitig aber auch eine Reihe unerwünschter Nebenwirkungen zeigt. Hierzu zählen die Einengung der Lehrpläne, eine übertriebene Fokussierung auf das Einüben der Testinhalte, zunehmende dropout-Raten und die unzureichende Generalisierbarkeit der Effekte auf andere Lern- und Leistungsmaße. Vor diesem Hintergrund werden abschließend Möglichkeiten der Einbeziehung von Motivationstheorien – insbesondere der Selbstbestimmungstheorie – in bildungspolitische Reformbemühungen diskutiert.

Standardized tests comparing nations, schools and students have become a focus of public interest and a basis for educational reform efforts around the world. High stakes testing (HST) is a specific approach to reform based on applying rewards and sanctions contingent on attained performance on such tests. HST thus represents a motivational strategy to improve educational outcomes. Herein we discuss the relations of HST to the theoretical positions of behaviorism, achievement goal theories, and, most centrally, Self-determination Theory (Deci & Ryan, 1985). We examine the degree to which each perspective would endorse HST as an approach to motivating teachers and students, and what the theories predict in terms of consequences. We then review recent empirical studies of the effects of HST-based interventions in the United States, where that type of reform strategy is flourishing, and compare these emerging results with theoretical predictions. We conclude that HST, because it is a controlling, extrinsic form of motivating teachers and students, often raises targeted test scores while producing a number of unintended negative consequences. These include narrowing of curricula, excessive focus on test preparation, increasing dropout rates and poor generalization of test score gains to other measures of learning and achievement. We conclude by discussing how motivation theory can better inform educational policy, with an emphasis on the self-determination theory viewpoint.

Theorists and practitioners in education have perennially been divided over the issue of how to motivate learning and achievement (Ryan & Lynch, 2003). On the one hand are those who view learning as a process that must be externally motivated (Thorndike, 1913; Skinner, 1953; Finn, 1991). They typically stress the use of control, evaluations and contingent reinforcements to foster achievement and incite engagement. On the other hand are those who view the learning process optimally stemming from internal motivations such as interest and values for learning (e.g., Deci & Ryan, 1985; Dewey, 1938; Rogers, 1969). They typically suggest minimizing the salience of external controls, and instead emphasize support for autonomy and initiative in cultivating what they see as students' natural motivation to learn. These competing ideas about how to motivate students have different implications not only for practice, but also for policy in our approach to schooling. One sees admixtures of both philosophies not only within nations, but also within schools and even within classrooms.

One aspect of educational policy where these distinct approaches to motivation are particularly relevant is the use of standardized measures of achievement. International comparisons on standardized tests such as PISA and TIMSS have amplified concerns among educators, pundits and leaders across the world with educational outcomes. In many nations this has generated pressure to raise test scores, as well as increased criticism of existing school policies and methods. Tests are being used not just to diagnose or evaluate needs for reform, but also as a means to improvement. Specifically High

Stakes Testing (HST) represents a political movement in which legislators tie external rewards and punishments to schools or students on the basis of the test scores on which they are evaluated (Clarke, Haney & Maddaus, 2000).

Calls for increased accountability have been especially shrill in both the United States and Great Britain, where recent reform strategies (the Education Reform Act in Britain and the No Child Left Behind (NCLB) legislation in the U.S.) have resulted in the widespread administration of standardized assessments to school children. Under both policies the results of examinations are used to determine student promotion and to provide a means to rank and reward "high achieving" schools and sanction those identified as "low achieving". While it is particularly conspicuous in the U.S. and Great Britain, the use of test scores as a basis for making schools and students "accountable" is not limited to these two countries. Test-based reform policies are showing wide traction across the globe. Countries as diverse as Canada, South Korea, Israel and Slovenia have experimented with HST and countries such as Germany and Japan have seen a rising political investment in test scores and pressures for reform based on them. Although we cannot review HST movements in all these nations, it is clear that HST policies are among the most formidable forces in the international landscape of education today.

Our interest in this article is to apply concepts of motivation to this debate. HST reforms represent a motivational approach because they not only put an emphasis on test scores; they also implement a strategy to raise scores through promised rewards or threatened sanctions.

Accordingly, we shall discuss the motivational implications of HST and accountability-based reforms. We shall specifically address the relations of HST policies to the theoretical positions of behaviorism, achievement goal theories, and, most centrally, Self-determination theory (Deci & Ryan, 1985). We examine the degree to which each of these perspectives would endorse or reject HST as an approach to motivating teachers and students, and what the theories predict in terms of consequences. We then review recent results of HST practices in the United States, where that type of reform is flourishing, comparing the results with theoretical predictions.

1. Theoretical Perspectives on Learning and High Stakes Testing

1.1 Behaviorism and the external view of motivation

HST by definition utilizes assessments not simply as a way of gathering information, but as a criterion for applying rewards and sanctions. In this sense, HST reflects a quasi-behaviorist view of motivation in which students and teachers are seen as primarily motivated by external rewards and punishments (Finn, 1991). This emphasis on consequences has its roots in Thorndike's (1913) law of effect, and subsequently, operant theory's

emphasis the power of reward/punishment systems in externally motivating behavior (Skinner, 1953).

Attaching high stakes to outcomes is presumed to supply a contingency-based means of motivating both students and teachers to put in more effort and thus raise achievement (Finn, 1991; Oakes, 1991). For students the high stakes can include grade retention, and in the case of exit exams, the denial or receipt of a diploma. Some U.S. schools also offer cash prizes, parties, scholarships, candy and awards to students who score highly (Keller, 2000). At a school level, aggregated student scores have been tied to increases versus cuts in school budgets and, in cases of poor outcomes, administration changes. Some school superintendents have been given cash bonuses when their district's scores improve. However, it is the public nature of HSTs that may supply the strongest consequences. Schools are publicly compared on test scores, with the often explicit reasoning that pride and humiliation will result from success or failure.

Although HST policies have a clear connection with behaviorist methods there is one important caveat. As Ryan and Brown (2004) argued, classical operant theory focuses on the effectiveness of contingencies as applied to targeted behaviors. HST policies, in contrast, apply contingent consequences to outcomes rather than to behaviors. Ryan and Brown suggested that a danger with this outcome focus is that a wide variety of potential behaviors, both desirable (e.g., changes in instruction, improved effort, etc.) and undesirable (e.g., teaching to the test, narrowing of curriculum, cheating) can be equally "reinforced" so long as they produce the desired outcome.

Nonetheless, it is obvious that proponents of HST view their endeavor in behavioristic terms. For example, HST proponent Finn (1991) states: "the problem is that academic success yields such few rewards (sic) and indolence brings few penalties" (p. 120). He and other quasi-behaviorists believe that putting rewards and penalties behind the test scores will effectively alter the behavior of both teachers and their students.

1.2 Achievement Goal Theories: Divided Views on the Value of High Stakes

Given their focus on achievement aims and outcomes, contemporary achievement goal theories (e.g., Dweck & Leggett, 1988; Elliot & Moller, 2003; Pintrich, 2000) should have much to say about HST. Although they differ in details, these theories distinguish between goals focused on enhancing or developing one's competencies and knowledge (mastery or learning goals) and goals focused on proving or demonstrating relative ability (performance goals).

Considerable evidence has amassed demonstrating the general advantages of mastery goals relative to performance goals for many outcomes, both affective and cognitive, that are of interest to educators. Evidence suggests

that the more students focus on mastery goals the more they enjoy learning, make greater use of high level cognitive strategies, and show better integration of what is learned (Ames, 1992; Elliot & Moller, 2003; Midgley et al., 2001). By contrast, performance goals appear to foster a more superficial approach to learning. A meta-analysis by Utman (1997) showed that performance goals tend to enhance outcomes only at rote or algorithmic tasks, and typically undermine performance at more heuristic or complex tasks. Further, students with learning goals are more willing to tackle challenging or difficult material compared with those focused on performance goals (Ames, 1992; Thorkildsen & Nicholls, 1991). Finally, performance goals have been linked to greater self-handicapping (Martin et al., 2001; Urdan, Kneisel, & Mason, 1999) and greater vulnerability to helplessness (Dweck, 2002).

Although few performance goal theorists have discussed the issue directly, we submit that because HST focuses on test outcomes and uses rewards and sanctions to make those scores salient, they foster an institutional climate that emphasizes performance rather than learning goals. By making the public demonstration of scores the central issue, students, teachers and administrators alike should be more likely to adopt a performance goal orientation.

Elliot and colleagues (see Elliot & Moller, 2003) have, however, complicated this view of performance goals by differentiating two types of performance goals that they suggest differ in motivational impact. They distinguish performance-avoidance goals, in which the student is primarily motivated to avoid failure or negative outcomes, from performance-approach goals, which reflect an appetitive desire to demonstrate high performance relative to others. Much emprical data supports the view that performance-avoidance goals have many negative consequences, while performance-approach goals seem to show fewer detrimental effects, and may foster, under some circumstances (e.g., success conditions) some positive consequences (Elliot & Moller, 2003; Harackiewicz, Barron, Carter, Lehto, & Elliot, 1997).

This distinction has thus been used by some thinkers to justify the classroom use of performance goals and by extension the policies of HST reforms. For example, Hidi and Harackiewicz (2000) advocate linking performance goals with extrinsic rewards, speculating that such an approach promotes long-term motivation. Mirroring behaviorist views, Hidi (2002) states: "Why should we assume that our children will produce high level schoolwork without expecting and receiving rewards?" (p. 332).

In contrast, other goal theorists hold that a focus on performance goals will, outside rarefied laboratory demonstrations, yield few positive, and many negative motivational outcomes. Midgley et al. (2001), for example, argue that an emphasis on performance goals at best rewards only those highly

achievement oriented students who are certain about their abilities, and even for many of them it can lead to an extrinsic and superficial focus to learning. Further, they suggest that students with lower or uncertain abilities will show increased self-protective strategies like self-handicapping and withdrawal of effort, and lessened intrinsic motivation. Elliot and Moller (2003), while highlighting the benefits of approach versus avoidance performance goals, suggest that institutional policies should still be directed towards a mastery or learning goal focus. They reason that policies aimed at performance may put students at risk for negative effects, as many will adopt an avoidance focus under such circumstances. We not only concur, but further suggest that HST inevitably fosters both performance-avoidance and performance approach goals in real world classroom settings given the normative nature and salience of high stakes assessments. This will be true whether or not practitioners explicitly intend to incite approach orientations.

In sum, achievement goal theories lack consensus regarding the effects of HST as a modus operandi in schools. Some goal theorists suggest that linking performance goals with rewards will have a positive influence, whereas others suggest that this will yield deleterious results. Still others suggest the need to foster performance-approach goals without at the same time generating performance avoidance concerns, although realistic ways to do that in real world (i.e., non-laboratory) settings have not been explicated.

1.3 The Self-determination view of motivation

Self-determination theory (SDT) is an empirically based theory of motivation that is primarily concerned with promoting in students an interest in learning and a valuing of the educational process. From this perspective people are viewed has having an innate tendency to learn and to develop competencies and SDT attempts to delineate the conditions that support versus thwart this intrinsic propensity. Therefore strategies such as the use of evaluations, contingent rewards and performance pressures are of particular interest within SDT.

In SDT motivation is seen as a complex construct, as many different motives can underlie a given behavior. SDT argues that these different motives can be differentially associated with an individual's performance, well-being and subsequent motivation. The most general distinction is between intrinsic motivation and extrinsic motivation. Intrinsic motivation concerns the doing of an activity for its inherent satisfactions, whereas extrinsic motivation concerns doing a task or activity for its instrumental value (Ryan & Deci, 2000). Within SDT extrinsic motives are further differentiated into those that are externally or heteronomously regulated or controlled versus those that are more internally regulated or autonomous. SDT-based research has consistently demonstrated that more internalized and autonomous motivation is associated with a host of positive outcomes from greater well-

being to increased persistence. Delineating the factors which facilitate internalization and support autonomy is thus of central importance in SDT.

Some principle tenets of SDT suggest that intrinsic and autonomous extrinsic motives emerge when the learning climate facilitates the student's experience of being: a) volitional rather than controlled; b) competent and optimally challenged rather than under or over-challenged; and c) belongingness rather than alienation or detachment. That is, SDT argues that learners are most fully functioning when their basic needs for autonomy, competence and relatedness are satisfied.

More specifically, the motivational effects of an external event like a test score, feedback, a reward or a punishment contingency depends upon its functional significance – that is, the psychological meaning – that the event has for an individual's basic needs (Deci & Ryan, 2000). The functional significance of any event can be either informational, controlling, or amotivating.

With respect to testing, assessments are experienced as informational when they provide feedback that students (or teachers) can utilize in becoming more competent or effective, in a context that supports volition. Informational feedback is also referred to as effectence relevant feedback, insofar as informational events facilitate feelings of competence, rather that being experienced as pressure toward specified outcomes. According to SDT to the extent any event is experienced as informational, it tends to have a positive impact on self-motivation.

Events have a controlling functional significance when they are experienced as pressure toward a specified outcome, or as an attempt to control behavior. Rewards, for example, are often experienced as controlling because the recipient often sees him/herself as being controlled by the rewarder (Deci, Ryan & Koestner, 1999). Tests too, especially when connected with rewards and sanctions, can be experienced as controlling (Ryan & Brown, 2004). Although controlling events, if potent enough, may initially prompt compliance, people tend to comply in the least effortful way possible, and controlling regulations fail to inspire ongoing self-motivation in those subjected to them. In schools controlling motivational strategies have been shown to foster more "surface" forms of learning (e.g., rote memorization), and to undermine intrinsic motivation (see Ryan & Grolnick, 1986; Ryan & La Guardia, 1999).

Finally, external events are experienced as amotivating when they convey incompetence or ineptitude. Tests that are too challenging and result in very negative feedback have a functional significance of being discouraging. In such cases, testing can undermine motivation and lead to a withdrawal of effort. Again, because HST reforms typically mandate the same standards

for all students regardless of learning styles, abilities or backgrounds, they can be amotivating to many.

The functional significance of HST has not been considered in most implementations. Recall that some goal theorists (e.g., Heidi, 2000) suggest linking performance outcomes with rewards to enhance achievement and motivation. Yet substantial empirical evidence suggests that such a linkage will lead to controlling forms of motivation that undermine both complex learning and persistence, as well as interest and enjoyment (see Deci et al., 1999; Ryan & La Guardia, 1999). Moreover, we suggest that because HST are standardized and "one size fits all" they not designed to be optimally challenging for all individuals, and thus readily can have an amotivating functional significance. In the SDT perspective, differentiating and predicting which forms of motivation will be incited thus requires an additional level of analysis; specifically, it requires understanding the functional significance that an exam has for an individual student.

Both experimental and field studies have provided substantial evidence supporting SDT's predictions of how feedback and external evaluations can have different functional significance, and thus differing impact upon ensuing motivation. Early experiments (e.g., Ryan, 1982; Ryan, Mims & Koestner, 1983) showed that rewards or feedback delivered in a controlling manner undermine intrinsic motivation whereas reward structures or feedback delivered in an informational style do not. For example, Grolnick and Ryan (1987) examined the controlling use of a test in an elementary school setting. Students were exposed to text book materials under three conditions: They were told to learn the material because they would be tested and graded (controlling condition); told that they would be tested, but only to identify what was learned (informational condition); or not told that they would be tested at all (comparison condition). It was found that the controlling use of the test resulted in less depth of processing and less conceptual integration. Students in the non-controlling, informational condition demonstrated, in contrast, higher levels of conceptual learning and reported more interest and enjoyment for material. Similar results have been found in college students by Benware and Deci (1984). SDT-based research by Kage (1991) found similar results in Japanese schools. Middle school students in controlling, evaluative conditions in which quizzes were administered with the expressed intent of grading their performance expressed less interest, less competence, and greater anxiety than students in an autonomy-supportive condition where the same quizzes were used as a means of monitoring their own learning. Further, students in the controlling condition also performed worse on summary exams, demonstrating how controlling tests can be counterproductive.

Ego-involvement

Advocates of HST such as Finn (1991) explicitly want to activate not only a desire to improve scores, but also a fear of failing. Unfortunately, it is often the threat of punishment rather than the pride and promise of success that is more salient in minds of those subjected to HST, especially in high poverty areas. Indeed, Miner (2000) notes that in actual implementations of HST, punishments are enacted twice as often as rewards. Teachers, students and administrators thus often experience such policies as primarily "shame-based" motivators.

Shame often arises from situations where one's self-worth is contingent on performance at a task. Within SDT such circumstances are referred to as ego-involving, which is considered a controlled form of motivation. Based on this formulation Ryan (1982) demonstrated that when subjected to ego-involving climates, students experience less interest, more pressure, and less desire to engage in an endeavor beyond what is needed to protect self-esteem. Numerous studies (e.g., Ryan Koestner, & Deci, 1991) have supported these hypotheses. The empirical evidence indicates that, like other controlled forms of regulation, ego-involvement undermines intrinsic motivation for learning (e.g., Golan & Graham, 1990) and leads to more superficial processing of information. Moreover, as Ryan and Connell (1989) showed, although parents report that ego-involved children apply effort in school, they also evidence higher school anxiety and more maladaptive coping when dealing with failure than students with more autonomous forms of motivation. These findings highlight the fact that although controlling regulatory styles such as ego-involvement can lead to "motivation", they also exact high collateral costs. Because HST puts the esteem of pupils on the line, it potentiates ego-involvement and its negative effects.

Effects on teachers

SDT suggests that just as HSTs can undermine student learning, so too can they undermine best teaching practices and leadership principles. Placing controlling contingencies on teachers has been predicted within SDT to yield more controlling styles of teaching. For example, Deci, Spiegel, Ryan, Koestner, and Kauffman (1982) designed a teaching simulation study in which the teachers were asked to instruct students in a cognitive perceptual task. The teachers all had the same set of problems and were given the same preparation. Just before entering the teaching session, however, one group was explicitly told that it was their job to "make sure their student performed up to standards", whereas another group received no such pressure. The teaching sessions were recorded and rated for differences in teaching styles. Results showed that those who were explicitly pressured to produce high standards were more controlling. They engaged in more lecturing, criticizing, praising and directing – all techniques that have been shown to have a negative impact on students interest and willingness to undertake academic challenges. Flink, Boggiano, and Barrett (1990), examining a school-

based curriculum for elementary students across several schools, similarly showed that 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 performed more poorly on objective test outcomes. This is consistent with a wide body of literature linking evaluative pressure with poorer school performance (Kohn, 1993; Ryan & Stiller, 1991) and higher dropout (Hardre & Reeve, 2003).

In addition to the manner in which teachers instruct, SDT predicts that HSTs will have deleterious effects on the content of instruction as well. SDT argues that controlling rewards or contingencies typically lead people to take the shortest route to the end. Insofar as contingencies are focused on test outcomes, SDT suggests that HSTs will incite excessive test preparation activities, "teaching to the test", and a narrowing of the curriculum toward material that is expected on tests, among other practices. Moreover, the outcome focus of HST promotes any route to higher scores. HSTs may thus inadvertently reinforce negative behaviors such as encouraging low performers to leave school before testing, misreporting or distortion of test results, and controlling rather than supportive teaching climates. These predictions were made even before results from HST reforms were accumulating (see Ryan & La Guardia, 1999).

Finally, in most HST reforms all students are posed with the same "standardized" or "one size fits all" challenge. According to SDT this strategy will lead some students will be under-challenged, some over challenged, and few optimally challenged, lowering the intrinsic motivation associated with competence development.

In sum, SDT specifically pinpoints aspects of HST that can have potentially deleterious effects within educational settings. Although test–based information could have great informational value for reform, HST policies convert this potential information into controlling regulations, which tends to undermine autonomous forms of motivation and foster a more narrow, goal directed, and low quality approach to both learning and teaching focused on outcomes rather than process and "good practice" considerations.

2. *High-Stakes Testing: Evidence and Implications*

Although anecdotes and opinions abound, sound empirical research regarding the impact of HST policies is just emerging. In part this stems from the relatively recent implementation of HST policies within most countries. Yet, in the U.S. HST reforms have been popular for some time, and studies of their impact are now appearing. In what follows we review credible studies regarding HST policies, noting their relation to theory.

HST and teaching practices

The use of HST policies to effect changes in the curriculum is one of the first points on which advocates and critics differ. Without doubt HST fosters the use of a more standardized curriculum within schools. According to advocates such uniformity is key is assuring that all students are exposed to the same quality of learning experience. Critics argue that such uniformity is the result of a narrowing of curricula and an increased utilization of class time for test preparation. Some also believe a uniform curriculum is a step backwards from the promotion of differentiated approaches to the diverse needs of students.

McNeil and Valenzuela (2000) examined teacher reports on the effects of HST on their instructional practices. Teachers reported redistributing their classroom time such that the majority of instruction was focused on topics on targeted exams. Further, a significant amount of class time was reportedly spent on test taking strategies rather than substantive issues. This was especially true for schools serving the less affluent (who also tend to display lower performance). Similar results were obtained by Hoffman et al. (in press) who found that teachers in low-performing schools reported greater time spent on test preparation and test taking skills.

Moon, Callahan, and Tomlinson (2003) examined practices in a U.S. nationally stratified sample of teachers. Teachers subjected to HST were indeed altering their usage of instructional time, especially in schools with high concentrations of poverty. In addition to focusing more exclusively on topics expected to appear on the tests, teachers in low-income areas spent more time on test-taking skills. The authors concluded that HST policies may differentially deprive poorer students of exposure to challenging curricula and innovative instructional methods.

HST and high school completion/drop-out rates

An equally, if not more, important issue is the effect of HST policies on school completion and dropout. Recall that opponents of HST argue that such policies foster a standardized curriculum that is therefore non-optimal for many participants. They claim that one standard cannot fit all learners, nor can one approach be best suited for all students. SDT in particular suggests that the result of such non-optimal challenges will be decreased or impoverished motivation and also poorer retention. Moreover, because of the pressure posed by HST's on administrators to improve school rankings, there is incentive for schools to rid themselves of students who could potentially drag down scores. Such "push-outs" often occur by re-categorizing low-achieving students into special education programs, thereby rendering their scores exempt from accounts of high-stakes assessment scores (Schulte et al., 2001). In line with this, Haney (2000) found that exclusion rates explained score gains in Texas and Schulte et al. found similar results in North Carolina, both states where HST's were a strong policy focus. Other

practices include preventing students from passing on to a grade where high-stakes milepost assessments are given, a practice linked to an increased likelihood that students will drop-out (e.g., Clark, Haney, & Madaus, 2000).

To remedy this, some HST policies in the U. S. rate schools not only on how well students perform on standardized measures, but also on dropout rates. Under NCLB, for example, schools with high dropout rates and/or poor test outcomes can face sanctions. The dual pressures have led districts to distort not only reporting of scores, but also their accounting practices regarding dropouts. Dropouts are notoriously hard to document, and typically there are great discrepancies between the reported dropout rate and the "disappearing rate" of school age pupils. In a study linking HST policies and school completion, Clark et al. (2000) found a strong link between attrition rates and the use of HST. They also reported that in Texas, where graduation from school requires satisfactory performance on high stakes exit exams, the average black or Hispanic student was three times more likely to drop-out, even when controlling for SES, academic track, language program participation, and school quality.

HST tests and transfer of learning

The effect of HST on teaching practices, inclusion rates and school completion are problematic for several reasons. For example, they often lead to an appearance of the reduction in the gap between minority and majority students when in fact no real change in the quality of learning has occurred. That is while these students' test-taking abilities may have increased their general knowledge base, or deeper-level learning will not, and score gains may also be a result of a changed pool of test takers. The question is whether increases on HST scores "generalize" or transfer to other contexts or to non-targeted measures (i.e., assessment measures without stakes attached to them). Klein, Hamilton, McCaffrey, and Stecher (2000) found, for example, that in Texas, where HST reforms are prominent, test scores on the HSTs have increased; so too has the amount of test preparation. Yet, score gains on HSTs did not result in parallel improvements on other indicators of learning, such as the NEAP. That is, the effects of test preparation did not appear to generalize, or "transfer".

Amrein and Berliner (2002) conducted one of the most methodologically solid studies on this issue to date. They collected test scores from 18 U.S. states with strong HST policies and compared scores on the states' exam to other national non-high stakes assessments; specifically the ACT, SAT and NAEP. They also used a combined national trend line for scores on the comparison measure to normalize any differences in year to year score gains that may occur in the high-stakes states relative to gains in the nation as a whole during the same period. Results indicated that when compared to the nation as a whole, HST policies did not lead to improved performance

on the SAT, ACT or NAEP. Further, gains and losses on the SAT and ACT were more likely to be related to who participated in the exams than to the implementation of HST policies. Similarly, gains and losses on the NAEP were more likely to be related to who was excluded from the exam than to the effects of HST. They concluded that data from HST may not be valid as indictors of genuine learning. Neil and Gaylor (2001) reported similar findings when using the NAEP as a comparison assessment. NAEP scores were not improved by HST policies and, in fact, states without HST policies were more likely to show improvements on NAEP scores than states with such policies. The authors also cited other negative consequences of HST (e.g., they may widen the achievement gap between high and low income students) in concluding that such reform policies were not conducive to learning improvements. Thus, while evidence regarding the effects of HST policies on teaching practices and score reporting may be mixed, there is little, if any, evidence that such policies reliably lead to genuine gains in learning.

This is a critical point. Whereas tests can supply information about schools in need of improvement or curricula that may work better than others, when high stakes are put behind the test results, the test results themselves may no longer have meaning. The stakes, that is, corrupt the criteria. Amrein and Berliner (2002) applied the Heisenberg Uncertainty Principle to describe this, the idea that attaching serious consequences for teachers and administrators to performance increases the probability that the meaning of test scores will be corrupted and the utility of test scores as indicators of reform becomes increasingly uncertain.

3. *Summary and Conclusions*

The long-term effects of HST have yet to be fully studied, and it will be several years before a more complete picture emerges. Yet the preliminary results available suggest that several concerns with HST, including those grounded in SDT, have merit. HST policies can beget a host of unintended consequences and collateral forms of damage. These include practices such as teaching to the test, excessive test preparation, and manipulation of eligibility and enrollments, and more controlling teacher styles. Such strategies may not foster meaningful integration of material, or a significant transfer of knowledge, and they may corrupt the interpretability of test results as measures of reform. Although these behaviors may increase performance on targeted measures, and thus be "reinforced" under HST policies, they paradoxically may foster the limited educational experience HST advocates often suggest they are seeking to prevent.

In explaining these trends, SDT suggests that although high stakes can "motivate" school systems, teachers and some students to increase test scores, it is a controlling form of motivation. Controlling incentives thus can drive

schools toward test-focused curricula that may be neither optimally challenging nor well tailored to the diverse interests and needs of students, decreasing engagement and persistence. HST policies may also drive teachers toward more controlling classroom methods and less intrinsically engaging, instructional practices. Finally HST can lead to systemic problems, such as misreporting of scores, push-outs and other problems that corrupt the informational value of tests as indictors of quality, progress or school reform.

From an SDT perspective the problem with HST reforms is not in the T, but in the HS. Tests can have tremendous value, and standardized tests – especially if used along with multiple other indicators of school quality – could be an instrument of reform, helping to identify schools in need of resources, curricula that work better than others, and students in need of more intensive intervention or alternative approaches. However, when high stakes are attached to tests, their informational value can become corrupted, and tests change from instrument which record the effects of education initiatives to intrusive devices which themselves shape practice (Airasian, 1988; Canadian Teacher Federation, 2004). HST policies do "re-form" education, placing excessive emphasis on outcomes, and a paradoxical inattention to process and methods.

HST fits within the many other examples of how attempting to foster learning and development with externally controlling means can lead to degraded forms of motivation based on a short term, instrumental focus. As the focus on standardized tests becomes more universal, SDT suggests a need for critically studying how such tests can be better applied and implemented, emphasizing their informational rather than controlling or amotivating possibilities.

References

Airasian, P.W. (1988). Symbolic Validation: The case of state-mandated, high-stakes testing. Educational Evaluation and Policy Analysis, 10, 303-313.

Ames, C. (1992). Classrooms: Goals, Structures, and Student Motivation. Journal of Educational Psychology, 84, 261-71.

Amrein, A. L. & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning Education Policy Analysis Archives, 10 (18). Retrieved January, 2003 from http://epaa.asu.edu/epaa/v10n18/.

Benware, C. & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. American Educational Research Journal, 21, 755-765.

Canadian Teachers Federation. (2004). Educational accountability with a human face. Ottawa: CTF.

Clark, M., Haney, W., Madaus, G. (2000). High Stakes Testing and High School Completion. National Board on Education and Testing Policy 1 (3). Retrieved June, 2004 from http://www.bc.edu/research/nbetpp/publications/v1n3.

Deci, E. L., Koestner, R., & Ryan, R.M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin, 125, 627-668.
Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. New York: Plenum.
Deci, E. L., Spiegel, N. H., Ryan, R. M., Koestner, R., & Kauffman, M. (1982). The effects of performance standards on teaching styles: The behavior of controlling teachers. Journal of Educational Psychology, 74, 852-859.
Dewey, J. (1938). Experience and education. New York: Collier.
Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to personality and motivation. Psychological Review, 95, 256-273.
Elliot, A. J., & Moller, A. C. (2003). Performance-approach goals: Good or bad forms of regulation? International Journal of Educational Research 3, 339.
Finn, C. (1991). We must take charge: Our schools and our future. New York: Free Press.
Flink, C., Boggiano, A. K., & Barrett, M. (1990). Controlling teaching strategies: Undermining children's self-determination and performance. Journal of Personality and Social Psychology, 59, 916-924.
Golan, S., & Graham, S. (1990). The impact of ego and task-involvement on levels of processing. Paper presented at the annual meeting of the American Educational Research Association, Boston.
Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. Journal of Personality and Social Psychology, 52, 890-898.
Haney, W. (2000). The myth of the Texas miracle in education. Education Policy Analysis Archives, 8 (41). Retrieved from http://epaa.asu.edu/epaa/v8n41
Harackiewicz, J. M., Barron, K. E., Carter, S. M., Lehto, A. T., & Elliot, A. J. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. Journal of Personality and Social Psychology, 73, 1284-1295.
Hardre, P., & Reeve, J. (2003). A motivational model of rural students' intentions to persist in, versus drop out of, high school. Journal of Educational Psychology, 95, 347-356.
Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. Review of Educational Research, 70, 151-179.
Hidi, S. (2002). An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation. In C. Sansone, & J. M. Harackiewicz (Eds.), Intrinsic and extrinsic motivation (pp. 311-342). San Diego, CA: Academic Press.
Hoffman, J.V., Assaf, L., Pennington, J., & Paris, S.G. (in press). High stakes testing in reading: Today in Texas, tomorrow? The reading Teacher.
Kage, M. (1991, September). The effects of evaluation on intrinsic motivation. Paper presented at the Japan Association of Educational Psycholgy, Joetsu, Japan.
Keller, B. (2000). Incentives for test-takers run the gamut. Education Week, Wed, May 3.

Klein, S. P., Hamilton, L. S. McCaffrey, D. & Stecher, B. (2000). What do test scores in Texas tell us? Education Policy Analysis Archives, 8 (49). Retrieved April, 2003 from http://epaa.asu.edu/epaa/v8n49/

Kohn, A. (1993). Punished by Rewards. Boston: Houghton Mifflin.

McNeil, L., & Valenzuela, A. (2000). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. Cambridge, MA: Harvard University Civil Rights Project.

Martin, A. J., Marsh, H. W. & Debus, R. L. (2001). Self-handicapping and defensive pessimism: Exploring a model of predictors and outcomes from a self-protection perspective. Journal of Educational Psychology, 93, 87-102.

Midgley, C., Kaplan, A., & Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, at what cost? Journal of Educational Psychology, 93, 77-86.

Miner, B. (2000). Testing: Full speed ahead. In K. Swope & B. Miner (Eds.) Failing our Kids: Why the testing craze won't fix our schools (pp. 114-117). Milwaukee, WI: Rethinking Schools, Ltd.

Moon, T. R., Callahan, C. M., & Tomlinson, C. (2003). Effects of state testing programs on elementary schools with high concentrations of student poverty-Good news or bad news? Current Issues in Education, 6(8). Retrieved May, 2003 from http://cie.ed.asu.edu/volume6/number8/

Neil, M., & Gayler, K. (2001). Do high-stakes graduation tests improve learning outcomes? In G. Orfield & M. L. Kornhaber (Eds.), Raising standards or raising barriers? Inequality and high-stakes testing in public education. New York: The Century Foundation Press.

Oakes, J. (1991). The many-sided dilemmas of testing. In Voices from the field: 30 expert opinions on America 2000, the Bush administration strategy to "reinvent" America's schools (pp. 17-18). New York: William T. Grant Foundation.

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. Journal of Educational Psychology, 92, 544-555.

Rogers, C. (1969). Freedom to learn. Columbus, OH: Merrill.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. Journal of Personality and Social Psychology, 43, 450-461.

Ryan, R. M. , & Brown, K. W. (in press). Legislating competence: The motivational impact of high stakes testing as an educational reform. In C. Dweck & A. E. Ellitot (Eds), Handbook of Competence. New York: Guilford Press.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reason for acting in two domains. Journal of Personality and Social Psychology, 57, 749-761.

Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development and well-being. American Psychologist, 55, 68-78.

Ryan, R. M., Koestner, R. & Deci, E. L. (1991). Ego-involved persistence: When free-choice is not intrinsically motivated. Motivation and Emotion, 15, 185-205.

Ryan, R. M., & La Guardia, J. G. (1999). Achievement motivation within a pressured society: Intrinsic and extrinsic motivations to learn and the politics

of school reform. In T. Urdan (Ed.), Advances in motivation and achievement: Vol 11. (pp. 45-85). Stanford, CT: JAI Press.
Ryan, R. M., & Lynch, M. (2003). Motivation and classroom management. In R. Curren (Ed.), A companion to the philosophy of education (pp. 260-271). Malden, MA: Blackwell Publishing.
Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. Journal of Personality and Social Psychology, 45, 736-750.
Ryan, R. M., & Stiller, J. (1991). The social contexts of internalization: Parent and teacher influences on autonomy, motivation and learning. In P. R. Pintrich & M. Maehr (Eds.), Advances in motivation and achievement: Vol. 7. (pp. 115-149). Greenwich, CT: JAI Press.
Schulte, A. C., Villwock, D. N, Whichard, S.M., Stallings, C. F. (2001). High stakes testing and expected progress standards for students with learning disabilities: A five-year study of one district. School Psychology Review, 30 (4), 487-506.
Skinner, B. F. (1953). Science and human behavior. New York: Macmillan.
Thorkildsen, T. A., & Nicholls, J. G. (1991). Students' critiques as motivation. Educational Psychologist, 26, 347-368.
Thorndike, E. L. (1913/1962). Psychology and the science of education. New York: Columbia University.
Urdan, T. C., Kneisel, L., & Mason, V. (1999). Interpreting messages about motivation in the classroom: Examining the effects of achievement goal structures. In T.C. Urdan (Vol. Ed.), Advances in motivation and achievement: Vol. 11 (pp. 123-158). Greenwich, CT: JAI Press, Inc.
Utman, C. H. (1997). Performance effects of motivational state: A meta-analysis. Personality and Social Psychology Review, 1, 170-182.

Anschrift des Autors:

Richard M. Ryan, Professor of Psychology, Psychiatry and Education, Meliora Hall-CSP, University of Rochester, Rochester, NY, 14627
http://www. selfdeterminationtheory.org