Protecting the Veracity of Our Children's Test Scores: How Race to the Top Funding Can Spur Testing Integrity Reform

JEREMY M. KAUFMANN
Dartmouth College

This paper addresses the spate of high-profile teacher and administrator cheating scandals where educators stand accused of giving unauthorized help to students on state standardized exams. The paper explores why teachers feel the need to cheat, reviews the literature on how common experts believe the phenomenon to be, describes the methods teachers use to cheat, discusses cheating in the context of the psychology of motivation and rewards, and proposes policy options to reduce the incidence of cheating. The paper concludes with a proposal that the federal Department of Education include testing integrity policies as part of the holistic rubric that it uses to grade Race to the Top grant applications submitted by states.

Introduction

Over the past few months, a steady stream of eyebrow-raising reports of large-scale teacher cheating scandals has rocked our nation. *The New York Times*, *USA Today*, and *Atlanta Journal-Constitution* have all run feature stories on these scandals in the last year, and major cities including Atlanta, Los Angeles, New York, and Baltimore now face statewide probes into suspicious results (Bello and Toppo 2011). After living blindly through a decade that began with the signing of the No Child Left Behind Act in 2001 - the piece of legislation that ushered in the so-called new age of teacher accountability - the country is finally beginning to come to terms with the idea that accountability policies in the education system can have unanticipated consequences. From "erasure parties" in Atlanta, where teachers gathered at a principal's pool to erase and change their students' answers on standardized multiple-choice state exams (Bowers, Wilson, Hyde 2011), to classrooms in Chicago where underperforming educators were forced to sit on the floor during staff meetings if they refused a principal's request to alter their students' wrong answers (WSBTV, Atlanta), we have awakened to the fact that the teacher accountability movement has certainly spurred change, although not in unambiguously positive ways.

In fact, this issue has received so much attention in the press that Secretary of Education Arne Duncan recently instructed the Department of Education to issue a "Request for Information to Gather Technical Expertise Pertaining to Testing Integrity," (Federal Register) which is a step that often precedes an executive branch rulemaking process. Other signs also point to further federal involvement with this issue: the Department of Education recently collaborated with the National Center for

Education Statistics to host a conference bringing together experts on teacher cheating on February 28, 2012 and Arne Duncan is on record as saying that "valid, reliable data are absolutely essential to meaningful accountability and implementing education reform," (Marisol and Toppo 2011). Some policymakers have interpreted Duncan's statement as endorsing the idea that all states should be required to screen exams for telltale signs of cheating, a policy proposal that has begun to receive serious consideration over the last year (ProPublica 2011).

This paper will argue that the right reaction to these teacher cheating scandals is not the all-too-simple reactionary stance of declaring standardized testing a failure; rather it is to examine and re-evaluate the incentives that lead teachers to cheat (the benefit side of the equation) while also punishing the worst offenders as a lesson to anyone else who attempt to cheat. The goal is to spend money on software, to provide neutral proctors, and to undertake investigations to create a very credible threat that if a teacher cheats on a state exam he will get caught and prosecuted (the cost side of the equation). Because fundamental incentive problems exist at the school and district levels to eradicate cheating, as district administrators often want higher scores at all costs and can be complicit in cheating, the paper will argue that the federal government and Department of Education must play an important role in reversing this nefarious trend. They can do this by requiring certain test security measures to be in place before federal officials accept test scores as representative of the students in a given state when allocating grant money through programs such as Race to the Top. Just as Race to the Top has incentivized states to adapt certain tenure and merit pay policies, the federal government must take a lead in incentivizing schools to create testing environments where cheating is not acceptable.

Imagine a scenario in which there is a way to improve student performance on standardized tests by fifteen percentage points over the course of several years, but it comes with the downside that every year a small percentage of educators including teachers and principals would almost certainly try to game the system. What should a neutral policy maker decide to do? Clearly, there is no obvious answer to this type of scenario, but this paper will argue that the goal of a neutral policymaker should be to prioritize the educational attainment of the children over the pursuit of policies that put an end to teacher cheating. The truth is that if schools continue to reward or punish teachers based on scores, cheating will never be fully eradicated; however, by focusing on reducing the benefits of cheating accrued to teachers and increasing the costs of cheating, hopefully fewer teachers will cheat when confronted with the opportunity to improve their students' scores.

The Role of Quantitative Indicators in the Social Sciences

From a social science perspective, it is certainly not shocking news that high stakes incentives induce cheating; what is surprising though is that policy makers were blind to these inevitable occurrences. Social scientists have understood from the 1970s that "when a quantitative indicator [such as student test scores] is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be

to distort and corrupt the social processes it is intended to monitor" (Campbell 1976). This principle is broadly referred to as Campbell's Law and often it is used to point out the negative consequences of high stakes testing as "achievement tests may well be valuable indicators of general school achievement under conditions of normal teaching aimed at general competence. But when test scores become the goal of the teaching process, they both lose their value as indicators of educational status and distort the educational process in undesirable ways" (Campbell 1976). This principle is analogous to Goodhart's Law, which is used by economists to explain that when an economic indicator is made a target for the purpose of conducting economic policy, then it loses the information content that would qualify it to play that role (Goodhart 1975) and to the Heisenberg Uncertainty Principle in physics. For the sake of intellectual honesty, we must admit that as long as high stakes testing exists in its current form, some cheating is inevitable; however, policy makers should look at ways to nudge the marginal teacher (the teacher most torn between cheating or not cheating) away from betraying their students by cheating on their behalf.

How Teacher Cheating Scandals Harm Our Educational System

Certainly such an effort would take considerable resources, resources that cash-strapped states and school districts might be hesitant to mobilize in an economic recession. Therefore, it is essential to understand the social costs associated with teacher cheating scandals when governments attempt to think through a cost-benefit calculus of preventing cheating. First and most obviously, teacher cheating is a harmful phenomenon as it ultimately hurts the students themselves: if students appear to be doing well, they will then develop a false sense of security. Parents will not try to help them in the subjects with which they struggle; the school will not feel obliged to offer extra resources or help, and the child will continue to lag behind grade level expectations. We are well aware of the crucial fact that once a child falls behind, it is exceptionally challenging to reverse this deleterious trend (Balfanz, Herzog and MacIver 2007). Next, cheating makes it harder to identify struggling schools, students, and ineffective teachers, which is the entire motivation for the emphasis that has been placed on testing over the last decade. Finally, reports of systematic cheating by administrators undermine popular support for the education reform movement. If citizens cannot trust the scores that are used by policymakers to measure certain educational interventions, then they begin to question why the government should intervene at all and to lose hope that the education reform movement can produce better student outcomes.

The Prevalence of Teacher Cheating

When reflecting on the incidence of these teacher-cheating scandals, one is naturally bound to wonder how prevalent this form of cheating is and what forms it takes. The best answer to the question of how prevalent erasure-driven cheating is comes from Brian Jacob and Steven Levitt's paper, "Rotten Apples: An Investigation

of the Prevalence and Predictors of Teacher Cheating" which uses an algorithm based on the number of erasure marks on student answer sheets to statistically assign a probability to the chance that an exam was altered by an individual who did not take the exam, using thousands of Chicago public school exams. Jacob and Levitt find cheating in four-to-five percent of the classes that they examine, which translates to over 1,000 separate instances of cheating in their sample. However, this is likely to be a lower bound because their experimental design focuses only on cheating where teachers alter student test forms and ignores the possibility that teachers could give hints to students, allow students to have extra time, or even providing correct answers on scrap paper before students bubble in their choices, which are all recorded occurrences.

Overall, the results show that teacher cheating is found to be quite elastic to minor changes in incentives. When school accountability policies create strong incentives to cheat without instituting safeguards against cheating, an increase in administrator and teacher cheating results. In Chicago, a change in superintendant leadership in 1996 led to standardized exams becoming higher stakes events in two different ways: first, low scoring schools could now be placed on probation and face the threat of reconstitution and secondly, low performing students would not be automatically promoted to the next grade. The findings were stark: a ten percentage point increase in the proportion of students in a class whose test scores "count" will increase likelihood of teacher cheating by roughly 20 percent. Bizarrely, Jacob and Levitt show that prior to introduction of an accountability policy, teachers were more likely to cheat for high achieving students, but were now more likely to cheat on the exams of low achieving students. In a similar vein, the new Chicago policies caused cheating to rise sharply in classrooms with many low achieving students while teachers of high performing students did not alter their behavior.

Additionally, the authors highlight that teacher cheating is not a statistically independent event, as it is correlated with how many other teachers are cheating in a school and that teachers that cheat once are more likely to cheat again. They find that a ten percentage point increase in cheating classrooms in a school raises the likelihood that a teacher cheats by roughly 16 percentage points, suggesting centralized cheating by a school counselor, test administrator, or principal. Moreover, they find that classrooms that tested poorly last year are more likely to cheat this year. If the classroom is one standard deviation below the mean, then teachers are 23 percent more likely to cheat. This could indicate a positive feedback effect where a teacher is faced with a struggling class so he cheats, leaving the teacher in the next grade with a dilemma: either cheat to maintain the "high" scores or don't cheat and be blamed for the inevitable decline in scores, a scenario which presents teachers with a serious moral quandary. They also find that teachers in classrooms with lower achievement, a high poverty rate, and a higher proportion of black students are more likely to cheat; in addition, younger teachers are also more likely than more senior teachers to cheat. Finally, they demonstrate that teachers are less likely to cheat for students in the lowest quartile (as they are likely to fail anyway) and are less likely to cheat for males and older students, hypothesizing that teachers have some increased fear of being

caught. All in all, the results clearly illustrate that teachers respond to the "benefits" of cheating: they target students who are right on the line of passing or failing, cheat specifically at a time when it is necessary to do so (when sanctions are on the line), and cheat more when their class is less capable.

In addition to drawing conclusions on how teachers respond to the potential benefits of cheating, the Jacobs and Leavitt paper also highlights that teachers often respond in predictable and rational ways when it comes to dealing with the costs of cheating. For example, cheating prevalence is systematically lower in cases where the costs of cheating are higher such as in mixed grade classrooms, where two different exams are administered simultaneously. In fact, teachers proctoring examination rooms with students in multiple grades are 65 percent less likely to cheat than teachers in classrooms where all students are in the same grade. This is not a surprising finding, though, because with two different sets of questions and answers it would certainly take longer for teachers to erase answer sheets if they were overseeing multiple tests and also it would be more difficult for teachers to provide unauthorized help to the students, as teachers would not have as much time to become familiar with test questions. In addition, the study finds that teachers who administer the exam to their own students are 50 percent more likely to cheat than teachers who are assigned a random room to proctor, a finding that could potentially provide the platform for mandating that teachers should not proctor their own students (a position discussed later in the paper).

How Teachers Cheat: Understanding the Techniques

One of the shortfalls of this paper is that it does not consider the other methods that teachers use to cheat besides simply altering answers on their students' "bubble sheets." Several statewide investigations in places like Georgia and New York found that teachers are cheating through two additional major channels. First, teachers may have students record answers on a sheet of loose-leaf paper and then correct the students' answers before students fill in the bubble sheets so they avoid detection if an erasure analysis were to be conducted (Loughran and Comiskey 1999). Secondly, some teachers abuse their power as proctors by giving students additional time, "reviewing" important concepts on the board during the exam, and reading through students' answer sheets and telling them to "re-check" certain questions (Loughran and Comiskey 1999). One board of education report on a cheating school in the Bronx in New York City reported that "methods depended on proctors who were willing to cheat. It was not unusual to pull students from their regular class on test day if their proctor refused to cheat and place them with an educator who agreed to" (Loughran and Comiskey 1999).

Furthermore, this investigation of behaviors revealed that teachers are tremendously responsive to the costs of cheating: they took the necessary steps, often ingenious ones, to avoid getting caught. For example, in one district a principal put a dot on the answer key to questions that were supposed to be hard questions so that teachers wouldn't give help on that question, as to alleviate the suspicion that comes from

below average students getting "hard" questions correct (Beckett 2011). This type of behavior shows that educators are tremendously savvy, as Jacobs and Levitt's identification of cheating teachers in their previously discussed paper came mainly through an algorithm that detected whether students were essentially getting too many hard questions right relative to the number of questions on an exam. Other examples of strategic behavior include a testing coordinator at one school who gave teachers advanced warning as to when board of education monitors would show up, information that the testing coordinator was legally obligated to keep confidential (Beckett 2011).

Why Teachers are Cheating

The extent to which this type of behavior has infiltrated our nation's school systems certainly begs the question of why teachers are cheating and the sources of the pressure placed on them to cheat. Shephard and Dougherty surveyed teachers in two large districts to answer this question and found that 17 percent of teachers say parents are a source of pressure to cheat, 24 percent of teachers said other teachers, 56 percent said principals, 66 percent said the media, and a whopping 79 percent said district level administrators and the board of education were a source of pressure (Shephard and Dougherty 2001). To answer this question more holistically though, I examine the recent scandals in the Atlanta and Dougherty County Public School systems in Georgia, scandals that resulted in over 170 teachers and administrators being accused of cheating.

In a statewide report on the scandal (commissioned by Governor Sonny Perdue), the Inspector General lays out three main reasons for the cheating. First, was the intense fear of failure that teachers had of meeting AYP or "Annual Yearly Progress" as defined under the No Child Left Behind Act. Targets for student improvement set by the district were unrealistic and the administration put unreasonable pressure on teachers and principals to achieve targets, targets that increased every year at an increasing pace, making it harder and harder for schools to keep up (Bowers, Wilson, and Hyde 2011). Second, the report faulted the Atlanta Public Schools superintendent, Dr. Beverly Hall, for creating a culture of fear, retaliation, and intimidation in the district. This culture of fear was marked by a three-year principal replacement rate of over 90 percent, staff meetings where low performing teachers were forced to sit under desks and crawl on the floor like dogs, and a district-wide motto of "No Exceptions. No Excuses" (Bowers, Wilson, and Hyde 2011). Finally, the report commented on a total failure of leadership at the principal level and administrative level. The administration made use of a harmful incentives policy; for example, if a school met a 70 percent proficiency target, every school employee from bus drivers, to the school nurse, to guidance counselors, and teachers would receive a cash bonus so everybody had an incentive to cheat. In fact "numerous teachers raised concerns about cheating and other misconduct to the principal or test coordinator only to end up disciplined or fired" (Bowers, Wilson, and Hyde 2011).

Yet, there are some other eyebrow-raising facts about the Atlanta scandal that didn't make the official statewide report, details that suggest a more complex and nuanced situation than teachers simply responding to administrative pressure to meet standards. For example, 62 percent of elementary school teachers who admitted to cheating were teaching first or second grade, yet No Child Left Behind standards only apply to the third grade and above: not first and second grade (Bowers, Wilson and Hyde 2011). In other words, if fear of not meeting AYP was all that was driving the cheating, none of these teachers should have had a reason to cheat. Yet further examination of these cases reveals that these teachers cheated for more nuanced reasons: they wanted to move up to the older grades, as teaching the lower grades was seen as a demotion; they "wanted to be first and nobody wanted to be last," and it was easier to cheat. Specifically, exams at the youngest grades were actually read to the kids and the answers were in the teacher manual, lending support to the idea that teachers respond to the costs of cheating: if we make it easy for cheating to happen, it will inevitably follow. Teachers in these youngest grades also admitted that those who refused to cheat were assigned to problem classes or given no chance to move up to grades where exams actually counted while compliers received plum after-school positions and preparatory periods. This is a story that has not yet been told in the newspaper and the academic literature: reasons for teacher cheating extend far beyond the pressure to meet AYP.

Adapting Policies to Prevent Teacher Cheating

In light of these scandals, it is incumbent upon educators and policymakers to think about what policy changes can be adapted to reduce the incidence of teacher cheating. First, to make it less rewarding to cheat, policies should be adapted that restructure how teachers are currently rewards, so that we decrease the potential benefits of cheating. Second, we need to make it harder to cheat by increasing the cost of cheating with both ex-ante policies, or policies that prevent cheating by dealing with the cheating before it happens as well as critically think about ex-post solutions, or what to do after cheating has taken place. The situation has never been more urgent on the policy front; in fact, Secretary of Education Arne Duncan just recently sent a letter to all state education superintendents letting them know how serious the problem of cheating infractions has become (Pro Publica 2011). Nevertheless, while addressing the problem, one must be careful to frame the issue as a desire to protect children, as opposed to a goal of catching cheating teachers, as this is likely to be a way to get more educator support. Similarly, it will be essential to emphasize that we are not blaming teachers and that the vast majority of teachers are honestly administering these exams. Reformers will need as much support in tackling this issue as they can get; after all there is no natural constituency of those who will support these reform incentives except for perhaps the legal community, making reform even more difficult (Wilson et al 2012).

Rethinking the Psychology of Motivation: What Role Should Rewards Play

When thinking about how to reduce the benefits to cheating, it is important to understand that teachers are cheating because standardized tests have now become measurement tools of teachers and not just students. They are now used in teacher evaluations, determine teacher pay, and determine school status under No Child Left Behind (Cizek 1999). This shift has represented a fundamental societal rethinking of incentives and motivation, yet the psychology behind motivation is much less clear than economists would have us believe in their models of rational beings responding predictably to carrots and sticks. The current understanding of motivation and drive began with experiments in the 1950s where psychologist Dr. Harry Harlow showed that monkeys attempting to solve a maze made more errors and solved the problem less frequently with rewards (Pink 2011). Similarly, experiments by Edward Deci in the 1960s pointed to the destructive nature of rewards: in his many trials he noted that subjects attempting to solve puzzles who received monetary rewards for successful completion spent significantly less time playing with the puzzle when rewards were not offered as compared to a group of subjects who never received rewards in the first place, who never lost their internal, intrinsic motivation to solve the puzzle anyway. Deci went on to claim that "when money is used as an external reward, the subjects lose intrinsic interest for the activity" (Pink 2011).

Daniel H. Pink summarizes the current understanding of rewards in his book Drive: The Surprising Truth about What Motivates Us and classifies tasks into two different groups, arguing that the effect of a reward depends on the type of task. First is the algorithmic task, where one follows a set of established instructions down a single pathway to one conclusion and the second is the heuristic task, where one has to experiment with possibilities to devise a novel solution. Teaching students is best thought of as a heuristic task since there is no one-size-fits-all approach and as Pink describes, when it comes to heuristic tasks, contingent rewards ("if you do this, you'll get that" type rewards) often fail terribly. Pink argues that rewards create inevitable side effects just like medicines and sarcastically writes that

> Rather than being offered as an over the counter salve for boosting performance, goal setting should be prescribed selectively, presented with a warning label, and closely monitored. Goals may cause systematic problems for organizations due to narrowed focus, unethical behavior, increased risk taking, decreased cooperation, and decreased intrinsic motivation. Use care when applying goals in your organization.

Not surprisingly, the psychological effects of rewards hypothesized by Pink hold true in the case of teacher cheating scandals. Pink notes that "rewards are addictive in that once offered, a contingent reward makes an agent expect it whenever a similar task is faced, which in turn compels the principal to use rewards over and over again." This idea is interesting in the context of merit pay and bonuses as it is not too difficult to imagine a scenario where a school district in a budget shortfall has to cut back on

bonuses. In this type of scenario, it is more than reasonable to expect the teachers to respond in ways that are likely to be harmful to their students and society, just as Deci observed in his laboratory experiments of people losing their internal motivation to solve puzzles.

The contemporary psychological understanding of incentives leads us to two guiding principles when it comes to offering rewards, and not surprisingly the current use of teacher performance incentives fit neither principle. The first idea is that any extrinsic reward should be unexpected and offered only after the task is complete (Pink 2011). Holding out a prize at the start of a project and offering it as a contingency will inevitably focus people's attention on obtaining the reward rather than on attacking the problem. In the specific case of performance incentives for teachers, a concrete example of these side affects would be teachers wanting money, focusing on raising test scores, possibly cheating, and neglecting the overarching educational goals. On the contrary, introducing the subject of rewards after the job is done is less risky because these rewards are less likely to be experienced as the reason for doing the task and are thus less likely to be detrimental to intrinsic motivation. The second psychological guiding principle when it comes to rewards is that they should provide useful information (Pink 2011). In the workplace, people often look for signs of how they are doing and rewards can answer the question as long as they are not a tacit effort to manipulate behavior. Rather, rewards must give meaningful information about work performance.

Overall, teacher performance incentives neglect the central tenets of Self-Determination Theory (SDT), the theory that underpins psychologists' current conceptions of motivation. SDT begins with the idea that we need to understand universal human needs when thinking about motivation and posits that the three innate psychological needs of competence, autonomy, and relatedness guide our actions (Pink 2011). Pink acknowledges the intrinsic motivations that underpin our actions and writes that "If there's anything fundamental about our nature it's the capacity for interest: some things facilitate it, some things undermine it." In light of SDT, it is essential to focus on creating environments for innate psychological needs to flourish. The high-pressure environment that Dr. Beverly Hall created through her "No Excuses" campaign is a perfect example of how those plans run counter to current conceptions of intrinsic motivation. After all, when law firms in the 1980s placed greater emphasis on increasing the number of billing hours, the amount of billing fraud rose dramatically (Pink 2011). Chief Justice William Rehnquist went on record saying that if one is expected to bill more than 2,000 hours per year, there are bound to be temptations to exaggerate the hours actually worked. Furthermore, Rehnquist went on to describe that these sorts of high stakes, measurable goals can drain intrinsic motivation, sap individual initiative in addition to encouraging unethical behavior, which is identical to what we have seen in the education reform movement.

Policy Solutions That Increase the Costs of Cheating

Moving beyond the guiding principles that policymakers should consider when reforming teacher incentive structures, it is also important to consider several ex-ante and ex-post policy solutions that will increase the costs of cheating. The goal is that by making it more difficult for teachers to cheat, teachers will therefore be less likely to cheat. These solutions begin with the assumption that cheating is common because it is so easy: in the words of one Los Angeles teacher who confessed to cheating, "the testing procedures haven't been secure over the past 10-plus years and teachers are all to aware of this" (Los Angeles Times). In terms of these solutions, it is important to recognize that prevention is the first-best solution because it is so much more cost effective. In the words of cheating expert Lou Fabrizio, effective anti-cheating policies begin with creating a culture of honesty. This can be accomplished at a bureaucratic level by hammering out a testing code of ethics as seen in the state of North Carolina with a manual that lays out the standards of professional conduct to all and creates a credible threat of license suspension and revocation if one is found guilty of violating the code. At a school level, a culture of honesty means that there is no punishment or retaliation for reporting suspicious behaviors. The side effect of such a policy is that this kind of culture will allow other teachers, parents, and even students to come forward and report any suspicious behavior.

A second concrete solution that states have implemented to prevent cheating has been using outside proctors to administer state exams. Steven Leavitt in his paper actually recommends hiring an outside agency with no relation to school districts to proctor exams to avoid the incentive problems inherent in using school officials to proctor exams. In fact, New York state just this year adopted a version of this policy by banning any teacher from grading his or her students' state Regents exams after evidence pointed to possible teacher cheating (Otterman 2011). However, the cost and effort required to implement this type of policy is probably higher than most schools are willing to go and several districts have pursued variants of this policy. One school district in North Carolina actually has teachers in one school administer exams at another school, as testing often occurs across states at the same time for all schools and in theory this reduces the likelihood of cheating since the proctoring teachers won't suffer if children taking the exam under their supervision perform badly (Wilson et al 2012). As well, the City of Baltimore has recently adopted a policy of using co-proctors in wake of reports of several cheating incidents (Wilson et al 2012). Under this system, multiple teachers are essentially assigned to the same room so it becomes impossible to cheat unless you conspire with your other proctor. This operates under the principle that teachers in general are not willing to be so open with their cheating that they would put their career in the hands of one of their colleagues. Clearly, it is not a foolproof method, but Baltimore has ramped up their policy from having only 13 schools with co-proctors two years ago to co-proctors in nearly all classrooms today (Wilson et al 2012).

Along similar lines, some advocate putting students taking different tests in the same room with a proctor, as it becomes harder to help students on two or more exams

simultaneously, a claim which received support in the Levitt empirical study. Another variation on this technique would be to ask testing companies to create multiple versions of the exam so that teachers can't just simply change student answers because the answers would be different on different versions on the exams and in theory a cheating teacher only has a very limited amount of time in which he or she is in possession of the exams (Jacobs and Levitt 2003). Finally, one of the more simple solutions would be to forbid teachers and administrators from erasing stray erase marks and darkening answer bubbles. Currently these procedures are done to make the tests easier for the scoring machine to read, but in actuality this is unnecessarily leading to a major incentive misalignment. Georgia's handbook of testing integrity, which is distributed to all proctors every year, says it is allowable "for the erasure of all stray pencil marks and smudges from the answer documents. On the other hand, it is a breach of test security if anyone alters or interferes with examinees' responses in any way." There is no reason why teachers should be given the ability to erase stray erase marks; if stray marks are a serious concern, an unrelated third party employed by the testing company would be in a better position to do so.

However, it is important to recognize that it is impossible to prevent all cheating from occurring and it is therefore necessary to put an administrative system into place that can deal with charges of teacher cheating. According to cheating expert Steve Ferrara, one of the most effective ways to handle the issue ex-post is to set up anonymous tip hotlines where anyone can leave an anonymous tip or complaint with the guarantee that his or her concerns will be investigated (Wilson et al 2011). While this is not ideal, as it may lead to many false positives, at least people can feel free to report their suspicions and avoid retaliation. However, this is a problem if districts are ignoring the obvious like one district in Atlanta, which denied cheating despite going from 784th to 4th place in math scores in one year (WSBTV 2011). People will choose not to call into tip lines if they don't think somebody will do a thorough investigation

Additional strategies that policymakers can use include revising investigative procedures such that the people running investigations would not be under the auspices of a school district to avoid any improper conflicts of interest. Cheating expert Lou Fabrizio recommends that the investigative chain of command should run from local education officials who would compile a local review to accountability services staff employed by the state who would be responsible for reviewing all complaints. Any complaints with merit should then be forwarded to the state board of education attorney and then onto the Attorney General's office (Wilson et al 2011).

Creating an effective chain of command in these investigations is of the utmost importance as there are too many cases of states sitting on complaints and doing nothing to investigate. In New Jersey, the state Department of Education conducted erasure analysis of state exams since 2008 but did not investigate any of the schools flagged in the analysis until the Asbury Park Press successfully sued the Department to obtain a copy of the analysis and made it public (Pro Publica 2011). Not surprisingly, once the results were made public 34 schools were placed under investigation. Similarly, Pennsylvania's Department of Education received an erasure analysis report in 2009 that flagged dozens of schools for potential cheating violations

but left the results untouched for two years and did not even notify school districts about the anomalies. After reporters for an education blog obtained the report and made it public, the state ordered initial investigations of 89 schools (Pro Publica 2011).

In addition to implementing a stronger chain of command, it is also important that the people running these investigations have an investigator's background as opposed to just a teacher's background because the two jobs require vastly different skill sets. For example, investigators learned from the Atlanta cheating scandal that getting a confession from an accused educator on the first interview was not the norm. Of the 86 educators who confessed, all of them did so after multiple rounds of interviews when surrounded by lawyers and state agents (Wilson et al 2012). This is not the kind of investigation somebody without the requisite background could effectively run.

As well, it is essential that states conduct these erasure analysis tests because without these studies, it is impossible to detect cheating unless some accuser actually comes forward. Erasure analysis tests done in Atlanta and Dougherty County, Georgia were the foundations of the investigative cases made against teachers; yet in an environment of budget cuts and downsizing, test security measures are some of the first items in the state budget to lose money. For example, California substantially reduced its budget for testing integrity in the wake of the financial crises and stopped testing for anomalies despite a relatively minimal cost of $105,000 (Pro Publica 2011). Only twenty states and the District of Columbia performed erasure analysis on paper and pencil tests during the 2010-2011 school year (Pro Publica 2011). However, in the wake of mounting scandals across the nation, five more states have recently joined the movement and in the past year Kentucky, Maryland, Missouri, New Mexico, and North Carolina have begun conducting these erasure tests. In an interview this year, Secretary of Education Arne Duncan said that a school's contract with the testing companies that score the exams should require statistical study that can detect cheating (Bello and Toppo 2011).

The problem, though, is that it remains incredibly difficult to catch cheating teachers. For example, in Fairfield, CT, an erasure analysis conducted on exams originating from Stratfield Elementary School lead to a retest with lower scores. One observer familiar with the evidence concluded that "the probability of tampering was 95% certain but officials would never find the smoking eraser "(Lindsay 1996). The reason why educators often get away with cheating is that statewide Offices of Testing Security often refer problems to a principal or superintendent at the district or school where the complaint originates (Wilson et al 2012). However, neither the principal nor superintendent has an incentive to fix the problem. As well, the Office of Testing Security often accepts the findings of investigations handled at the school level without any independent inquisition and parents are not required to be notified if the discovery of suspicious erasures leads to a retest, even if their own children are being interviewed to determine if cheating actually took place! Overlapping governmental jurisdictions also make it hard to conduct a comprehensive investigation.

Making Reform Happen: A Solution Involving Race to the Top

Today, the issue of teacher cheating has never been more salient. Although it is a topic that has received relatively limited scholarly attention and consideration in the educational policy community, it is becoming more and more apparent that these instances of cheating are too numerous to ignore and occur too often for people to make the claim that they are randomized events. Rather, the systemic nature of these events has prompted the Department of Education to issue a formal Request for Information seeking input from experts on this issue. The U.S. Department of Education now says it is looking for ways to share best practices, which would allow states and school districts to reduce the incidence of the problem. Specifically, the request for information mentions the federal government is interested in ways to stop the prevalence of cheating, best practices in teacher cheating investigations, and ways to limit cheating given the new focus on online tools in assessment. The federal Department of Education is justifying their authority to become involved in this issue under section 1111(b)(3)(C)(iii) of the Elementary and Secondary Education Act of 1965, as amended, (ESEA) and 34 CFR 200.1-200.24, which requires states to establish and maintain assessment systems that are valid, reliable, and consistent with nationally recognized professional and technical standards (Federal Register). Furthermore, the Department argues they have power under Title I of the ESEA, to review and approve each State's assessment system. Accordingly, the Department examines evidence compiled and submitted by each State about its process for monitoring and improving the technical quality of its system. During the review of State assessment systems, the Department specifically examines procedures and policies for test security and data quality, including the training and monitoring of staff (Federal Register).

Given that the Department of Education has the statutory mandate to ensure test security and is distributing grant money based on test results, it has a role to play in keeping the states honest. As a way to encourage states to address the problem of teacher cheating, the Department should include the existence of strong testing integrity policies as part of the rubric it uses to grade state submissions for Race to the Top grant funding. The Department has already had great success in getting states to adapt novel tenure policies, curricular decisions, and charter school policies through its Race to the Top competition and it could deal with teacher cheating in a similar manner. It should assign a point value to states that include a testing integrity reform plan in their Race to the Top applications, a plan that would hopefully contain many of the policies and strategies that were described throughout the paper. This outside motivation would force states to take on a problem that they would otherwise choose to avoid; after all if their scores are inflated they look good and have little incentive to make a meaningful effort at testing integrity reform. Since testing reform has no natural political constituency, this is a way to get states caring about reform: when hundreds of millions of dollars of grant money hang in the balance, states can be incentivized to take on these hard issues.

Examining the effect Race to the Top has had on state education policy across the country illustrates the dramatic power the federal government has when incentivizing states indirectly through grant money. When Secretary Duncan announced the first winners in March 2010 - $100 million to Delaware and $500 million to Tennessee - it became pretty clear what he wanted: the finalists had lifted limits on charter schools, found some way to tie teacher ratings to students' test scores and signed on to the Common Core Standards, a national curriculum movement that sets benchmarks in english and math through the 12th grade (Paulson 2010). North Carolina state lawmakers cited the White House when they tried to lift the state's cap on the number of charter schools in May 2010. Seeing what was required to win, former Colorado Gov. Bill Ritter signed a teaching overhaul bill weeks before the round two Race to the Top deadline, despite opposition from the state's largest teachers union (Paulson 2010). Along the same lines, lawmakers in Louisiana and Minnesota are considering similar measures, and performance statutes have passed in Connecticut, Maryland, Michigan, Tennessee and Washington (Paulson 2010). In fact, a total of 46 states have submitted applications and in the words of the Department of Education,

> with less than 1 percent of the annual K-12 education spending in our country, [Race to the Top] has given states the incentive to lead reform in a comprehensive and collaborative way. Race to the Top has helped advance reform more in the past 18 months than any other program in the history of the Department of Education (Duncan 2010)

Given the success that Race to the Top has had in encouraging states to tackle other challenging educational policy issues, there is no doubt that including testing integrity guidelines in the Race to the Top application would prompt schools to give this entrenched problem a second look. The truth is that this rather unpleasant problem of teacher cheating is comes down to incentives and responsibility: without some type of reform, teachers will continue to have incentives to cheat and district administrators will continue to look the other way.

WORKS CITED

"An Overview of Issues Concerning Cheating on Large-Scale Tests." Paper presented at ...National Council on Measurement in Education. National Association of Test Directors. Accessed March 12, 2012. http://www.natd.org/Cizek%20Symposium%20Paper.PDF.

Robert Balfanz, Liza Herzog, and Douglas MacIver, "Preventing Student Disengagement and Keeping Students on the Graduation Track in High-Poverty Middle-Grades Schools: Early Identification and Effective Interventions," Educational Psychologist 42, no. 4 (2007): 223–36.

Beckett, Lois. "America's Most Outrageous Teacher Cheating Scandals." Pro Publica. Accessed March 12, 2012. Last modified September 19, 2011. http://www.propublica.org/article/americas-most-outrageous-teacher-cheating-scandals/single.

"Despite Sweeping Scandals, Big States Don't Check for Cheating by Teachers." *Pro Publica*, September 13, 2011. http://www.propublica.org/article/despite-sweeping-scandals-big-states-dont-check-for-cheating-by-teachers.

Bello, Marisol, and Greg Toppo. "Few States Examine Test Erasures." *USA Today*, September 13, 2011.

Bowers, Michael J., Robert E. Wilson, and Richard L. Hyde. *Special Investigation into Test Tampering in Atlanta's School System*. Atlanta: Office of the Georgia Governor-Special Investigation, 2011.

Cizek, Gregory. *Cheating on Tests: How to Do It, Detect It, and Prevent It*. N.p.: Psychology Press, 1999.

"CRCT Report: Answers Changed at 'Erasure Parties.'" WSBTV. Accessed March 13, 2012. Last modified July 6, 2011. http://www.wsbtv.com/news/news/crct-report-answers-changed-at-erasure-parties/nDLhT/.

Homeroom (blog). http://www.ed.gov/blog/.

Jacob, Brian A., and Stephen D. Levitt. "'Rotten Apples: An Investigation Of The Prevalence And Predictors Of Teacher Cheating." *Quarterly Journal of Economics* 118, no. 3 (August 2003): 843-878.

Jacob, Brian A., and Steven D. Levitt. "Catching Cheating Teachers: The Results of an Unusual Experiment in Implementing Theory." In *Brookings-Wharton Papers on Urban Affairs*, edited by William G. Gale and Janet Rothenberg Pack. Washington, D.C: Brookings Institution Press, 2003.

Loughran, Regina, and Thomas Comiskey. *Cheating the Children: Educator Misconduct on Standardized Tests*. Accessed March 13, 2012. http://www.nycsci.org/reports/Testcheat.pdf.

Otterman, Sharon. "City Reports Increase in Allegations of Cheating by Educators." *New York Times*, August 22, 2011.

"Regents Set to Alter Rules for Grading State Exams." *New York Times*, October 17, 2011.

Paulson, Amanda. "How Race to the Top is Recasting Education Reform in America." *The Christian Science Monitor*, June 1, 2010.

Pink, Daniel H. *Drive: The Surprising Truth About What Motivates U*.: Penguin Group, 2011.a

"Request for Information to Gather Technical Expertise Pertaining to Testing Integrity." Federal Register. Accessed March 13, 2012. Last modified January 17, 2012. https://www.federalregister.gov/articles/2012/01/17/2012-753/request-for-information-to-gather-technical-expertise-pertaining-to-testing-integrity.

Shepard, L. A., & Dougherty, K. C. (1991). *Effects of high-stakes testing on instruction*. Chicago, Ill.: Spencer Foundation.

Wilson, Bob, Tisha S. Edwards, Steve Ferrara, and Lou Fabrizio. "Response and Investigation of Alleged and/or Actual Irregularities in Academic Testing." Panel Discussion, National Center for Education Statistics, Washington DC, February 28, 2012.