

UC Davis

UC Davis Previously Published Works

Title

Bad Apples or Bad Orchards? An Organizational Analysis of Educator Cheating on Standardized Accountability Tests

Permalink

<https://escholarship.org/uc/item/4444h68j>

Journal

Sociology of Education, 93(4)

ISSN

0038-0407

Authors

Hibel, Jacob

Penn, Daphne M

Publication Date

2020-10-01

DOI

10.1177/0038040720927234

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

**Bad Apples or Bad Orchards? An Organizational Analysis of Educator Cheating on
Standardized Accountability Tests**

Jacob Hibel
Department of Sociology
University of California, Davis
jhibel@ucdavis.edu

Daphne M. Penn
Graduate School of Education
Harvard University

Forthcoming in *Sociology of Education*, 2020

Keywords: Accountability, Cheating, Organizational Theory

Acknowledgments

Work for this project was made possible by a William T. Grant Foundation Scholars Award to the first author. We are grateful to Tom Beamish, Dan Koretz, Bill McCarthy, John Meyer, Dick Murnane and participants in the CREO seminar series at Notre Dame for their valuable feedback. We wish to thank the anonymous reviewers and the editor, Linda Renzulli, for their valuable suggestions throughout the review process.

ABSTRACT

Using an explanatory sequential mixed methods design, we analyze quantitative administrative and survey data and qualitative archival data to examine the organizational character of standardized test cheating among educators in Georgia elementary schools. Applying a theoretical typology that identifies distinct forms of rule-breaking in bureaucratic organizations, we find that teacher-focused, individual-level explanations for cheating are inadequate, particularly in the context of large-scale cheating outbreaks. Our findings suggest that cheating scandals tend to arise when rule-breaking decisions shift toward higher levels of the educational bureaucracy, and school and district leaders enact multiple strategies to motivate coordinated cheating efforts among lower-level educators. In these scenarios, a “bad apples” explanation focused on rogue teachers fails to account for the systematic organizational underpinnings of standardized test cheating. We describe the institutional and organizational predictors of organized adult cheating on standardized tests, and conclude with a discussion of our findings’ implications for education policy and research.

The No Child Left Behind Act of 2001 (NCLB) targeted 2014 as the year by which 100 percent of students were expected to test at or above proficiency thresholds on state standardized tests. During the lead-up to this deadline, standardized test cheating by U.S. K-12 educators grew increasingly prevalent. According to the U.S. Government Accountability Office (2013), 40 states suspected and 33 states confirmed accountability test cheating during the two-year period from 2010-2012 alone. The most widely publicized case of educator cheating occurred in Atlanta Public Schools (APS), where more than 50 schools were implicated in a school district-wide cheating scandal (Georgia Office of the Governor, 2011a). Considered by many to be the most egregious academic cheating scandal in U.S. history, the Atlanta case formally implicated 178 educators at 44 schools, and resulted in the conviction of 11 teachers and administrators on racketeering charges, with all but one receiving prison sentences.

In the years following the APS scandal, additional cases of cheating have emerged in schools and districts that, like APS, predominantly serve populations of urban, low-income, students of color. For instance, a 2013 investigation of District of Columbia Public Schools found serious test security violations in 11 schools, which resulted in the invalidation of mathematics and reading test scores of the identified testing groups (Office of the State Superintendent of Education, 2013). A more recent example occurred in Gary, Indiana, where one school's principal and six teachers had their teaching licenses suspended following a state investigation into an unusually high number of wrong-to-right erasures in 2018. The school, which had enjoyed public accolades for its students' high test scores, produced test forms with wrong-to-right erasures six to 13 times higher than the statewide average (Colias-Pete and Carlson 2019)

The primary focus of any accountability system is to alter human behavior to produce more desirable outcomes (Smith and Fey 2000). In the case of accountability-based education reforms, the central goal is to improve instruction and thus increase student learning. Cheating undermines this process, decoupling students' test performance from their actual learning and thereby rendering meaningful instructional improvements irrelevant to teachers' and schools' pursuit of accountability rewards or avoidance of sanctions.

Cheating has particularly negative consequences for students in need of additional learning supports. Standardized tests have been used to evaluate students for academic track placement, promotion to a higher grade, and high school graduation in many states (Darling-Hammond 2004). Test scores have also served as a partial determinant of eligibility for special education, tutoring, English acquisition, and other remedial services (Stuebing et al. 2002). Students with artificially inflated standardized test scores may not only be prematurely promoted to higher grade levels or academic tracks with more challenging curricula, but many are forced to face these challenges without the support services for which they would otherwise qualify. Studies that have examined the consequences of cheating for students suggest that test score manipulation is associated with decreased subsequent achievement, reduced access to remedial services, and distorted measurement of student progress leading to increased achievement disparities (Apperson, Bueno and Sass 2016; Battistin 2016).

Reducing educator cheating requires that we understand the mechanisms and social processes driving the problem and the conditions under which it is most likely to emerge. The few empirical studies that have explored the predictors and prevalence of cheating have conceptualized cheating as an isolated classroom practice perpetrated by individual teachers as

an “amoral calculation” in response to performance incentives (e.g., Amrein-Beardsley, Berliner, and Rideau 2010; Jacob and Levitt 2003).

In this study, we examine educator cheating through an organizational lens, moving beyond conceptualizations of cheating as an individual deviant act to examine the interplay among regulatory structures, power dynamics within hierarchical education institutions, and individual decision-making as determinants of educators’ participation in this form of deviance. We focus on a set of questions regarding the organizational nature of standardized test cheating. In the complex, highly bureaucratized institutional setting of a public school system, at what hierarchical level(s) does cheating originate? Are rogue teachers the primary perpetrators, or might it instead be more appropriate to search for “cheating schools” or “cheating districts?” Under what conditions does large-scale, organized cheating emerge? How is cheating facilitated at different levels of the bureaucratic organization?

Resolving these research questions requires us to both quantitatively assess cheating’s patterns and prevalence and qualitatively examine educators’ experiences related to the decision to engage in cheating. We therefore use an explanatory sequential mixed-methods research design (Creswell and Plano Clark 2011) to quantitatively document patterns of standardized test cheating across different levels of the organizational hierarchy and gain qualitative insight into the conditions that contribute to educator cheating as an organizational, rather than individual, phenomenon. Our findings indicate individual “bad apple” teachers account for only a small share of cheating’s total prevalence. By contrast, cheating grows far more frequent in cases in which school or, especially, district leaders engage in cheating, and propagate this form of rule-breaking throughout their organizations through both passive and active means.

ORGANIZATIONAL PERSPECTIVES ON RULE-BREAKING

Sociologists have approached the study of rule-breaking from a variety of sub-disciplinary perspectives. A key insight from this diverse body of work is reflected in the “situated action” perspective on rule-breaking by and within organizations. Developed largely during the 1980s, this point of view bridges micro-, meso-, and macro-level causes of deviance (Vaughan 1998; 2007). Among the macro-level processes examined as causes of organizational deviance, competitive organizational environments and a strong, broadly endorsed capitalist achievement ideology emerged as the primary structural forces providing the incentive for rule violations (e.g., Coleman 1987). At the meso-level, researchers identified organizations’ structures, goals, norms, and their regulatory environments as central causes of deviance (e.g., Kramer 1982). Micro-level theories of deviance also evolved during this period, as classical rational choice explanations were modified to recognize actors’ embeddedness in organizational fields that differentially shape and constrain the range of available choices as well as their context-specific utility (e.g., DiMaggio and Powell 1991).

Inhabited institutionalism (Hallett and Ventresca 2006) provides an especially useful theoretical framework for understanding individual and collective deviance in schools and other hierarchical organizations. This perspective draws attention to the relationships and processes that shape organizational responses to external pressures and changes therein. Applied to organizational deviance, an inhabited institutional perspective resists reducing rule-breaking decisions to discrete, utility maximizing choices, but instead considers them as organizational processes. Taking standardized test cheating as an example, the “standards and accountability” era of U.S. education reform produced a macro-level shift in collective understandings of schooling’s purpose, away from holistic individual development and toward human capital

production. With NCLB, this focus was narrowed to a preoccupation with standardized test scores. In neo-institutional parlance, the ‘collective myth’ of schools as producers of testable knowledge became predominant, and accountability reforms caused the technical (i.e., within-classroom) process of education to become “recoupled” (Hallett 2010) with these macro-institutional myths.

In response to macro-institutional shifts, schools directed increased effort toward raising test scores. Once legitimate approaches became insufficient to generate continued test score gains, certain schools and districts proceeded to boost scores through illicit means for the sake of satisfying regulatory demands. Once the wrongdoing boundary has been breached within an organization, “collective corruption” (Ashford and Anand 2003) has the potential to spread through institutionalization, rationalization, and socialization processes, until what began as “idiosyncratic corrupt practices become institutionalized over time” (Ashford and Anand 2003: 4). Through such a process, cheating can become entrenched as an organizational-level, as opposed to individual-level, form of deviance.

Contentiousness and Hierarchy in Deviant Organizations

In an effort to synthesize sociological theories of organizational deviance, Martin and colleagues (2013) formulated an integrative theoretical framework for characterizing rule-breaking in bureaucratic organizational contexts. Their typology focuses on two closely associated dimensions: *contentiousness*, the extent to which rule-breaking is tacitly allowed within an organization, and *hierarchy*, the bureaucratic level(s) at which rule-breaking occurs.

The contentiousness dimension defines two basic types of rule-breaking: *permitted* and *contested*. *Permitted* rule-breaking is that which occurs when rule enforcement exists in name

only, or when rules are sufficiently ambiguous as to make their violation routine. *Contested* rule-breaking occurs in defiance of authority, in cases in which enforcement exists but is ineffective in preventing misconduct. Educator cheating on standardized tests would be considered *permitted* if organizational leaders (e.g., principals, superintendents) were cognizant and accepting of cheating by teachers and other lower-level workers, or if external enforcement of rules was sufficiently weak to empower organizational leaders to ignore them. In these scenarios, higher-tier members of the organization might be expected to establish “ceremonial” or “mock” rules for the sake of appeasing external authorities while failing to apply actual rules and consequences within their district or school. *Permitted* rule-breaking thus emerges as a normalized form of deviance within an organization, tolerated by bureaucratic authorities for the sake of achieving their instrumental institutional or individual goals.

By contrast, *contested* standardized test cheating arises when teachers or administrators circumvent their bureaucratic superiors’ strictly enforced testing procedures to boost their students’ scores. Examples could include a rogue teacher violating the school’s testing rules to inflate their students’ scores, a principal breaking school district testing rules to make their school appear to be higher performing, or a district superintendent violating state or federal testing rules. Implicit in both *contested* and *permitted* rule-breaking are relationships, either conflictual or collaborative, between actors at different levels of the bureaucracy. *Permitted* rule-breaking depends on the complicity of organizational leaders who would otherwise be responsible for enforcing rules, whereas *contested* rule-breaking is driven by choices and behaviors of lower-level workers that contravene leadership’s wishes. This hierarchical element of bureaucratic rule-breaking comprises the second dimension of Martin and colleagues’ typology.

Rule-breaking may occur at the *individual*, *group*, or *organizational* levels of a bureaucratic organization. *Individual-level* rule-breaking is, as its name suggests, a type of deviance undertaken by individual actors, typically for their own enrichment. An example of *individual* rule-breaking is the quintessential “bad apples” scenario (Jacob and Levitt 2003) in which an individual teacher makes a strategic decision to violate testing protocols to artificially inflate their students’ scores, thereby appearing to be a more effective educator than accurate test scores would suggest. Alternatively, a single teacher might engage in *individual* rule-breaking for more altruistic reasons, perhaps manipulating students’ scores to facilitate their future educational opportunities (e.g., Diamond and Persson 2016) or to present the school at large in a more positive light.

If the decision to cheat on the standardized test were coordinated by a group of teachers rather than independent actors, this would represent *group-level* rule-breaking. *Group-level* rule-breaking occurs when workers collude to circumvent organizational rules for the sake of achieving the collective’s broader goals or shared individual goals that are more readily attained through co-offending. In APS, for example, district officials assigned each school grade-level targets, minimum average proficiency levels that students in each grade were expected to achieve. Thus, teachers who taught at the same grade-level shared a collective goal, one that some groups might have more reliably met by artificially inflating students’ scores than by adhering to testing rules. If groups of teachers reached a collective strategic decision to cheat on the test to meet grade-level targets, it would be appropriate to consider this an instance of *group-level* rule-breaking rather than purely coincidental *individual-level* action taken by teachers in the same grade-level and school.

Organizational rule-breaking occurs at a still higher bureaucratic level. It arises when top management violates externally imposed rules, making rule-breaking a matter of unofficial organizational policy. Public schools, like other complex bureaucracies, comprise nested layers of authority and action, and therefore the *organizational* form of cheating could occur at multiple levels. For example, under NCLB, each school was required to meet Adequate Yearly Progress (AYP) benchmarks tied to standardized test results. Principals faced powerful incentives to ensure a particular proficiency rate among their students lest the school be labeled a “failing school” and the principal removed from their position. If principals succumbed to AYP pressure and directed their teachers to engage in cheating as a result, this scenario would constitute *organizational* rule-breaking at the school level.

District superintendents are also routinely evaluated on the average test performance of their district’s students. As a consequence, just as principals might encourage cheating for the sake of meeting AYP, superintendents could encourage principals (and, by extension, teachers) to generate artificially high test scores for the sake of avoiding accountability sanctions while also furthering their own careers. This scenario would constitute *organizational* rule-breaking at the district level. In such a case, the superintendent would likely work to maintain an outward façade of lawfulness while inculcating an internal ethos of rule-breaking, a phenomenon Meyer and Rowan (1977) referred to as “institutional decoupling.”

<Table 1 About Here>

These two rule-breaking dimensions yield a matrix of six distinct varieties of cheating, represented in Table 1. *Permitted* rule-breaking arises when the individual(s) who decide to cheat

have relative confidence that they can do so without repercussions. This could take the form of a classroom teacher making the independent decision to cheat, not because they were instructed to do so, but because they could rely on rule-enforcers turning a blind eye to their behavior.

Similarly, *permitted* rule-breaking might arise among a group of teachers who share responsibility under a test-based accountability system (e.g., teachers in the same grade-level or subject area). *Organizational, permitted* rule-breaking could occur at the school level if principals understood that regulatory agents would not closely monitor their testing procedures, or that penalties for cheating would not be enforced. *Organizational, permitted* rule-breaking at the district level might have been most likely to arise prior to the institutional recoupling that accompanied the standards and accountability reform movement (Hallett 2010), when district leaders often had little expectation that standardized testing would carry high stakes or be tightly controlled.

Whereas *permitted* rule-breaking is defined by the absence of meaningful rule enforcement, *contested* rule-breaking is defined by conflict surrounding enforcement. At the *individual* and *group* levels, this conflict exists within schools, as deviant teachers – independently or in conspiratorial groups – break rules that their organizational leaders are committed to enforcing. At the *organizational* level, *contested* rule-breaking occurs when principals (in violation of district-level rule enforcement) or superintendents (in violation of state or federal rule enforcement) knowingly and actively engage in efforts to break testing rules to artificially raise students' scores.

Drawing on Martin and colleagues' theoretical typology of rule-breaking in bureaucratic organizations, we examine the phenomenon of educator cheating on standardized accountability tests in Georgia during the spring of 2009. Our analysis characterizes standardized test cheating

along the hierarchy and contentiousness dimensions, and explains the organizational sources of variation in cheating forms across districts and schools. The presence of a large-scale cheating scandal in the state during this period provides the opportunity to draw comparisons between districts in which cheating was rampant and those characterized by far less frequent violations of testing protocol.

METHODS

This study uses an explanatory sequential mixed-methods research design (Creswell and Plano Clark 2011) to examine the prevalence and pattern of standardized test cheating among educators in Georgia elementary schools as well as the social processes that shape cheating's organizational nature. We began our investigation by conducting quantitative analyses of wrong-to-right erasure data to assess whether patterns of cheating across Georgia were consistent with the *organizational, group, or individual* type of rule-breaking within districts and schools, and to examine the relationship between rule-breaking form and the extent of cheating that took place within these settings. Subsequent qualitative analyses draw on a variety of archival data sources to examine the *contentiousness* dimension, identifying organizational characteristics that promoted or inhibited cheating at different bureaucratic level(s) in two Georgia school districts with especially high wrong-to-right erasures: APS and Dougherty County School System (DCSS).

Quantitative Data and Analyses

In the spring of 2010, the Georgia Governor's Office of Student Achievement (GOSA) and the testing company CTB McGraw-Hill conducted a statewide analysis of student answer forms for

the 2009 Criterion-Referenced Competency Test (CRCT), the test used to assess Georgia public schools' AYP. This analysis focused on identifying suspicious patterns of wrong-to-right erasures: test questions on which an incorrect response was erased and replaced with the correct response. Several cheating strategies produce an inflated number of wrong-to-right erasures, the two most common being direct teacher-to-student instruction to change a response during test administration and test form alteration by educators following the testing period¹. We obtained the full record of the 2009 CRCT erasure analysis via an open records request filed with GOSA in 2012. From these data, we calculated state population means and standard deviations for the number of wrong-to-right erasures per test-taker on the reading, language arts, and math CRCT tests, separately for grades one through five. We then compared each classroom's mean number of wrong-to-right erasures on each subject-area test to these population means and standard deviations, using Student's t-distribution to adjust for varying classroom sample sizes. Based on these mean comparisons, we identified as cheating sites those classrooms whose two-tailed probabilities were less than 1%; that is, those classrooms whose mean wrong-to-right erasure scores exceeded t_{critical} at $\alpha = .01$. In all, we analyzed 91,424 classroom mean erasure scores from CRCT tests taken in 1,203 schools in 171 districts. Across the three subject areas, we flagged 6,052 of these classrooms, or 6.62 percent, for exceptionally high average numbers of wrong-to-right erasures per student².

The first objective of our quantitative analysis was to test three distinct hypotheses framed by Martin and colleagues' theoretical typology of rule-breaking: an individual-level hypothesis, a group-level hypothesis, and an organizational hypothesis.

¹ Not all forms of cheating produce wrong-to-right erasures, however, and our analysis therefore presents a conservative estimate of the extent of standardized test cheating in Georgia public schools.

² This proportion is comparable to the cheating prevalence reported by Jacob and Levitt (2003) in their analysis of Chicago public schools.

Individual-level rule-breaking hypothesis: If test score cheating were predominantly an *individual-level* form of rule-breaking carried out by rogue teachers, we would expect to find a random, approximately³ normal distribution of cheating across schools. Given that 6.62 percent of Georgia classrooms showed evidence of cheating, we would expect to find a roughly 6.62 percent level of cheating in all Georgia schools in the event that cheating were strictly an *individual-level* act, assuming that unethical teachers are randomly distributed across Georgia schools. Cheating should show little non-random variation between schools or districts, or between grade-levels within schools, as an individual teacher's odds of rule-breaking should be statistically independent of higher-level organizational contexts. Instead, the great majority of variation in the probability of cheating should lie *between teachers in the same school*.

Group-level rule-breaking hypothesis: If cheating were a *group-level* form of rule-breaking organized by teachers in the same grade-level, we should find that teachers who belong to common workgroups share statistically dependent probabilities of cheating. Whereas schools and districts would demonstrate approximately normally distributed aggregate cheating probabilities centered around the state average of 6.62 percent, teachers within a particular school would not demonstrate randomly distributed probabilities of cheating. Recall that *group-level* rule-breaking stems from organized, cooperative action for the sake of achieving common goals. Teachers in the same grade-level experience common AYP (and in districts such as APS, additional district-wide) targets for student achievement. If cheating were organized at the workgroup level, we should find that a given teacher's probability of cheating depends on the grade-level to which they belong, either a "cheating" group of teachers or a "non-cheating" group. The bulk of variation in teachers' latent cheating probabilities should therefore exist

³ Although teachers' latent cheating probabilities may be normally distributed, the observed proportion of cheating is bounded at zero and 100 percent of classrooms.

within schools but *between grade-levels*. After accounting for teachers' grade-level affiliation, we should observe little residual variance in between-teacher cheating probabilities.

Organizational rule-breaking hypothesis: Unlike the aforementioned scenarios, we would not expect to observe a normal distribution of cheating across schools or districts if cheating were an *organizational* type of rule-breaking. In this case, cheating would be approved by a bureaucratic authority, either at the district or school level. If these high-level managers were effective in permitting or encouraging lower-level workers to engage in rule-breaking, cheating probabilities should be relatively homogenous within the lower-tiers (i.e., across schools in the same district in the event that the superintendent was the primary rule-breaker or across grade-levels and teachers within the same school in cases in which the principal was the primary rule-breaker). As a result, certain districts or schools would emerge as “cheating” organizations and others would be “non-cheating.” If this were the case, the state-wide proportion (.0662) of classrooms in which cheating was apparent would be generated by a large number of districts/schools in which cheating was all but absent, and a comparatively small group of districts/schools in which cheating was the norm, yielding a heavily right-skewed cheating distribution. In terms of variance partitioning, *organizational* rule-breaking would be evidenced by the bulk of cheating variation lying *between districts* or *between schools*, depending on the level of the organizational hierarchy at which the rule-breaking decision originated. In either case, variation within lower levels of the bureaucracy would be comparatively minimal and stochastic.

Qualitative Data and Analyses

Whereas quantitative analyses allow us to document the pattern and hierarchical type of cheating that emerged across Georgia in 2009, understanding the organizational and social conditions that led standardized test cheating to be *permitted* or *contested* requires us to qualitatively examine educators' experiences in organizations marked by high levels of cheating. APS and DCSS are two such districts.

Located in Georgia's major urban center, APS is one of the largest public school districts in the state, comprising more than 49,000 students, nearly 4,000 teachers, and 114 schools in 2008-2009 (National Center for Education Statistics 2009). DCSS, which centers on the city of Albany some 200 miles southwest of Atlanta, is a far smaller and more rural school district, with 16,222 students, 1,114 teachers, and 30 schools during 2008-2009 school year (National Center for Education Statistics 2009). Although GBI investigators uncovered evidence of systematic cheating in both districts (Georgia Office of the Governor, 2011a; 2011b), the system-wide pattern of cheating in DCSS garnered substantially less media attention than the APS scandal.

Our archival data sources consisted of organizational documents, investigative reports, court records, and media reports regarding the cheating scandal and subsequent investigation. Through an open records request, we obtained organizational documents regarding testing and accountability policies in both districts. Key to our understanding of conditions that encouraged cheating was a set of APS accountability documents that outlined a secondary high-stakes accountability system, beyond NCLB's AYP requirements, that set progressive achievement targets for each school and grade-level. APS's target system was considered a more rigorous accountability system than AYP because it not only set annual targets for the number of students expected to meet the basic standards set forth by the state, but also for the number of students expected to exceed the standards. APS grade-level targets were set annually based on the CRCT

passing and exceeding rate of students from the previous cohort. Investigators, witnesses, and implicated educators cited this secondary accountability system as an organizational characteristic that encouraged cheating in APS.

In addition to organizational documents, we consulted primary and secondary sources associated with the GOSA/GBI investigation that followed the statewide erasure analysis's identification of 49 APS and 11 DCSS elementary schools⁴ as likely cheating sites (Georgia Office of the Governor, 2011a; 2011b). The GBI investigation consisted of interviews with teachers, administrators, and, in several schools, students, from every APS and DCSS school identified as a potential cheating site. As these interviews were conducted by a state law enforcement entity, interviewees were under oath and subject to prosecution for perjury for providing false information. At the conclusion of the investigation, the Office of the Governor Special Investigators released two reports based on more than 2,100 interviews with district and school personnel and analysis of over 800,000 organizational documents (Georgia Office of the Governor, 2011a; 2011b). The APS and DCSS reports are 413 and 293 pages long, respectively, and include copies of official correspondences from teachers to district-level leaders regarding school culture, official ethics and compliance reports filed by educators about cheating, written statements by individuals who witnessed cheating, interview summaries from witnesses and implicated individuals, and testing manuals for the CRCT. We also examined publicly available court documents related to the indictments, pleas, and/or subsequent trials of educators implicated in the cheating scandal.

In an effort to contextualize the cheating scandal and examine how the investigation

⁴ Our qualitative analysis focuses on 33 APS and 10 DCSS elementary schools with detailed evidence and testimony regarding cheating in schools. We excluded 16 APS schools and 1 DCSS school from the analysis because they were either not investigated or because the GBI determined that the evidence presented provided insufficient proof that cheating had occurred.

evolved over time, we also created a media coverage dataset that included all articles in *The Atlanta Journal-Constitution* (AJC) from December 2008 to December 2015 with a title that directly referenced the cheating scandal, investigation, or trial. The AJC is an Atlanta-based newspaper that partially prompted the statewide cheating investigation with its 2008 investigative report about suspicious test score gains in APS schools.

Qualitative analyses. We analyzed archival data using inductive and deductive coding techniques, and assigned codes using *Nvivo for Mac* software. We initially performed incident-by-incident coding (Glaser 1978) of each suspected or reported case of cheating identified in the GBI report, which allowed us to become familiar with each school as an individual case. The second step was focused coding, an iterative process informed by our research questions and patterns that emerged during incident-by-incident coding. Through focused coding, we developed general categories of codes related to who engaged in rule-breaking, educators' motivations for cheating, modes of cheating, characteristics of the school environment, and the consequences of cheating for students, teachers, and the school (see Online Appendix for details).

During our final phase of coding, we drew on Martin and colleagues' typology of rule-breaking to conduct theoretical coding and develop an analytic story to identify the social processes through which large-scale, organized, standardized test cheating emerged. Specifically, we coded the data based on whether cheating was *permitted* or *contested* within schools and the bureaucratic level(s) at which cheating occurred. Rule-breaking was coded as *permitted* if enforcement efforts by bureaucratic authorities around testing protocols were minimal and/or reflected "myths and ceremonies" established to appease external actors (Meyer and Rowan 1977). We coded rule-breaking as *contested* if cheating occurred despite school or district

leaders' sincere efforts to enforce testing protocols.

Concerning the hierarchical dimension of rule-breaking, we coded cheating as *organizational* if bureaucratic superiors actively encouraged, allowed, demanded, or otherwise coerced the educators they oversaw to violate testing rules. *Group-level* rule-breaking reflected collective efforts by classroom teachers to influence peers to violate rules for the purpose of achieving grade-level or subject area testing goals. *Individual-level* rule breaking referred to an educator's decision to act alone in inflating test scores without active coordination by administrators or peers. Using this framework, we were able to examine the pathway of social influence (e.g., top-down versus bottom-up) through which cheating originated and diffused in APS and DCSS schools.

QUANTITATIVE RESULTS

<Figure 1. About Here>

Empirical Cheating Distributions

Figure 1 presents the empirical standardized test cheating distribution aggregated across Georgia elementary schools. If cheating were a random teacher- or workgroup-level binary variable (i.e., if cheating were an *individual* or *group* form of rule-breaking) we would expect schools' proportions of classes flagged for cheating to be approximately normally distributed around the population proportion of .0662. This is the hypothetical normal density function illustrated by the solid line in Figure 1. However, the empirical distribution is highly skewed, as the large majority of schools have low proportions of cheating classrooms and a small number of schools

have quite high proportions of cheating classrooms. Forty-eight percent of schools contained zero cheating classrooms, an additional 26 percent had classroom cheating proportions between .001 and the state mean of .0662, and a small minority of Georgia elementary schools, six percent, experienced proportionate cheating levels above .25. This distribution, therefore, is more closely aligned with the hypothetical pattern of cheating that would arise under *organizational* rule-breaking than the *individual* or *group* forms. The state-level proportion of .0662 does not arise from approximately 6.62 percent of teachers cheating in most elementary schools; it is instead the arithmetic mean of many low proportions and a few very high values, evidence that organizational context is consequential for a given teacher's odds of cheating.

<Figure 2. About Here>

Figure 2 presents the empirical cheating distribution at the district level. Like the school-level distribution, the district-level cheating distribution presented in Figure 2 is a poor fit to the hypothetical pattern that would emerge under *individual* or *group-level* rule-breaking. Districts' proportionate levels of cheating are not randomly distributed about the state-level proportion, but instead follow a positively skewed distribution similar to that of schools. A large segment of the school district population evinced low levels of cheating, with 13 percent of districts demonstrating no cheating classrooms and 82 percent reporting levels lower than the state mean. By contrast, a small collection of districts have proportionate cheating levels far above the state average, with 10 percent of districts reporting cheating proportions greater than .10, and three percent of districts reporting cheating proportions greater than .20.

Taken together, Figures 1 and 2 lend support to the *organizational* cheating hypothesis, a type of bureaucratic rule-breaking in which a relatively small number of actors at high levels of the bureaucracy – superintendents, principals, or both – permit, encourage, or require lower-level employees to circumvent externally imposed rules surrounding standardized tests administration. Rather than appearing as a consistent, if regrettable, characteristic of public elementary schools in general, cheating demonstrates wide variability across organizational units. A substantial portion of Georgia schools and districts demonstrated no cheating whatsoever and a comparatively small minority witnessed especially high levels of cheating in 2009.

Variance Partitioning

Although the cheating pattern revealed in Figures 1 and 2 is suggestive of a strong organizational influence on cheating, there is little reason to expect that a single type of rule-breaking predominates across all districts and schools. Cheating may instead be a top-level organizational phenomenon in some districts or schools, a workgroup-level phenomenon in others, and a random, teacher-level activity in still other locations. To investigate the possibility that districts vary in the type of rule-breaking occurring within their schools, we examine heterogeneity in the patterns of wrong-to-right erasure variation across different sets of public school districts. We focus on two sets of districts in particular: those with classroom-average wrong-to-right erasures greater than the statewide mean, and those with erasure averages less than the state mean.

We begin by estimating four-level regression models with grade-level, school, and district random intercepts. The dependent variable in these models is measured at the classroom level, and is a measure of the average number of wrong-to-right erasures per test form,

standardized across all Georgia classrooms in the same grade level and subject area ($\bar{x} = 0, s = 1$).

Educators in higher performing schools are better equipped to respond to testing pressure than those in schools with lower student achievement, as the task of maintaining high performance is less burdensome than generating it anew (Carnoy and Loeb 2002). Thus, compared to low-achievement contexts, we expect accountability pressure to be less severe in high-achievement settings. To account for school- and grade-level variation in the pressure to produce high test scores, we control for each school's percentage of students who failed to meet AYP standards on each grade-level and subject-specific CRCT in 2008, the year preceding the 2009 erasure analysis. Our variance partitioning estimates are therefore conditional on schools' and grade-levels' risk of failing to meet accountability standards.

We present the intraclass correlations (ICCs) estimated via these conditional random-intercept regression models in Table 2. The ICC provides two key pieces of information: an estimate of the correlation among responses at a given hierarchical level of nesting, as well as an estimate of the proportion of variance in cheating explained by the corresponding random effect. Examining the ICCs therefore allows us to evaluate the variance partitioning implications of the three hypotheses described above. To recapitulate: *individual-level* rule-breaking should be characterized by statistical independence within districts, schools, and grade-levels, with the large majority of cheating variance lying between classrooms; *group-level* rule-breaking should produce weak correlations at the between-school and between-district levels, but strong correlations within grade-levels; and *organizational* rule-breaking should produce strong correlations within districts and/or schools, with comparatively little variation residing at the between-grade or between-classroom levels.

<Table 2 About Here>

The first column in Table 2 presents ICCs estimated from analysis of the complete population of Georgia elementary schools as a base for subsequent comparison. These results indicate that 43 to 52 percent of the classroom-level variation in cheating is attributable to the clustering of classrooms within grade-levels, schools, and districts, with the remaining 48 to 57 percent of variation existing between classrooms (i.e., between teachers, calculated as $1 - \text{Total ICC}$). Shared school district-level factors account for approximately 9 percent of variation in classrooms' wrong-to-right erasure frequencies, school-level factors account for 21 to 27 percent of variation, and grade-level factors account for 13 to 17 percent of variation. Thus, each contextual setting is consequential for the number of suspicious erasures that occurred in a given classroom, though nesting within schools explains a greater share of the total variation in cheating than grade-level or district nesting.

As described above, however, it is possible that this population average variance structure masks meaningful variability in the forms of cheating that emerged across districts. The second column of results in Table 2 presents ICCs from models restricted to school districts with classroom cheating proportions greater than the subject-specific state average. The amount of variance explained by between-district factors in this sample is comparable to the overall population. Similarly, the grade-level variance component represents between 14 and 19 percent of total variation in wrong-to-right erasure prevalence in above-average cheating districts, a figure comparable to the grade-level ICC in the overall population. However, between-school factors account for between 25 and 31 percent of total wrong-to-right erasure variance in high-cheating districts, a comparatively greater school-level influence than that evidenced in the

overall population or in below-average cheating districts. Teacher-level factors account for somewhat less of the total variation in wrong-to-right erasures in high-incidence districts than in the overall population of schools, with the between-classroom component accounting for 41 to 51 percent of overall variance.

The third column of results in Table 2 presents ICC estimates for a sample restricted to districts with below-average cheating frequencies. The cheating variance breakdown in this subsample is markedly different from that of the entire state or the subset of districts with above-average cheating levels. In low-cheating districts, the correlation between scores of classrooms in the same district is relatively weak, falling between .02 and .03. Likewise, the within-school correlation is far weaker than that observed in the population or the high-cheating sample, amounting to only an additional 4 to 7 percent of the total erasure variation. The between-grade-level correlation is also comparatively smaller in magnitude, accounting for between 9 and 11 percent of the total variation in wrong-to-right erasures in low-incidence districts. Cheating in these comparatively low-incidence districts appears to be primarily a classroom-level phenomenon, with the portion of residual variance attributable to between-teacher variation ranging from 80 to 82 percent across subject areas. School-level factors do little to explain variation in erasure behavior in low-incidence cheating contexts. By contrast, the school-level influence is approximately six to nine times greater in districts with above-average levels of cheating.

These key differences suggest that cheating in low-incidence districts represents a distinct form of rule-breaking from cheating in high-incidence districts. If cheating were an *organizational* form of rule-breaking in all cases, low-incidence districts would evince similar within-district variance patterns to districts with higher cheating frequencies, but would simply

contain fewer “cheating schools.” These data, however, suggest that low-incidence cheating arises when the decision to cheat is independently made by teachers, and little correlation in erasure behaviors exists among teachers in common settings. In such cases, our analysis suggests that relatively few teachers elect to engage in cheating of their own accord. Districts with above-average cheating frequencies, on the other hand, are characterized by substantially less classroom-level variation in wrong-to-right erasures being attributable to teacher-level factors. Taken together, these results support the *individual-level* rule-breaking hypothesis in districts with comparatively low levels of cheating. However, cheating outbreaks such as those observed in APS and DCSS demonstrate a pattern consistent with the *organizational* form of rule breaking. Our results suggest that high-incidence cheating is not likely a matter of more individual “bad apples” clustering in particular schools and districts, but is instead a result of organizational deviance in which a few “bad orchards” yield correspondingly large rotten crops.

QUALITATIVE RESULTS

Although our quantitative analysis of statewide CRCT erasure data revealed patterns broadly consistent with the *organizational* type of rule-breaking in districts with high rates of cheating and the *individual* type in low-cheating districts, these test score data are poorly suited to uncovering the social processes that facilitated cheating or the organizational conditions that established cheating as an organizational, rather than individual, phenomenon. To these ends, we turn to our qualitative analyses and discuss the bureaucratic level at which the cheating decisions originated, cheating’s contentiousness within schools, and educators’ motivations for cheating.

Bad Apples or Bad Orchards?

Our analysis revealed stark differences between APS and DCSS with regard to rule-breaking coordination and the contentiousness of rule enforcement at different levels of the organizational bureaucracy. Within APS, the mandate to cheat was passed down to schools from district leadership, including the superintendent, Beverly Hall, who had been named the 2009 National Superintendent of the Year by the American Association of School Administrators. Hall and four other district administrators were indicted on federal racketeering charges related to the cheating scandal, and all except Hall, who died before her trial, were convicted or pleaded guilty. In one plea agreement, a district official stated that she was ordered by the superintendent to destroy evidence related to investigations of cheating in the district. In contrast to APS, our results yielded no evidence that DCSS district personnel actively permitted or encouraged cheating in their district. Principals in DCSS anticipated that their bureaucratic superiors would enforce testing rules. Therefore, although both districts were subject to enforcement efforts by external actors within their broader institutional environment, differences in rule enforcement by district officials established school-level, *organizational* cheating as *contested* in DCSS and *permitted* in APS and contributed to cheating as a wide-spread, multi-level phenomenon in APS.

We found evidence of *organizational* cheating at 20 of the 33 APS elementary schools included in our analysis, compared to only 3 out of 10 DCSS schools. However, unlike DCSS school leaders, APS administrators facilitated cheating with the understanding that their district-level supervisors had no intention of enforcing testing rules. In both APS and DCSS schools characterized by *organizational* rule-breaking, cheating teachers did not act in opposition to school leadership, but were instead actively encouraged or directed by at least one administrator to commit testing violations. For example, at one school, the principal instructed proctors to prompt students to change incorrect responses while monitoring CRCT administration.

Additionally, in multiple APS schools, administrators instructed teachers to pick up their tests early or keep them beyond the testing window in order to “clean up the tests,” which teachers described as code for correcting students’ responses.

In *organizational* rule-breaking schools, cheating efforts were highly organized, involved sophisticated techniques for making the cheating process more expedient, and often involved a core group of teachers and administrators who helped to change answers. In several APS *organizational, permitted* rule-breaking schools, administrators provided the “clean-up committee,” a group of teachers whose ostensible task was to erase stray marks from students’ test forms, with answer form transparencies to allow the committee to make efficient corrections to large numbers of student tests. At one school, a group of educators went so far as to hold a “changing party” over the weekend at one teacher’s home. Several teachers who did not report directly participating in cheating noted that their students’ answer documents appeared in a different order or disappeared entirely after returning them to school administrators. Teachers also frequently voiced concerns to colleagues and GBI investigators regarding discrepancies between their students’ performance in class and their high scores on the CRCT. These suspicions informed educators’ often-expressed sentiment that school or district officials used a trusted circle of individuals to help them change answers school-wide, an efficient method of raising students’ scores that did not require acquiescence from all classroom teachers.

It is important to note that, despite district leaders’ support for cheating, there were 13 APS schools in which teachers cheated independently or in groups (i.e., *individual* or *group-level* rule-breaking schools). Although administrators in seven of those schools declined to actively organize cheating efforts, they provided tacit permission to engage in rule breaking by turning a blind eye to testing violations, a type of *permitted* rule-breaking. For instance, one principal

reportedly told a potential whistleblower, “If you don’t tell me anything, I don’t have to report anything.” Cheating in the remaining six APS cheating sites and the majority of DCSS schools was both *individual* and *contested*. In these cases, school leaders worked to enforce testing regulations, and teachers who engaged in cheating did so surreptitiously. In both districts, *individual*, *contested* cheating among educators often took the form of prompting, voice inflections, or non-verbal cues by individual teachers as a way to encourage students to change incorrect answers during test administration as opposed to educators changing answers after administration. Compared to DCSS and APS contexts of *organizational* or *permitted* rule breaking, these “bad apple” schools had a lower percentage of classrooms flagged by the GBI for wrong-to-right erasures (see Table 3).

<Table 3 About Here>

Table 3 presents a descriptive comparison of APS and DCSS elementary schools according to the form of rule-breaking we observed. Martin and colleagues’ rule-breaking typology predicts that cheating frequencies will be higher in *organizational* rule-breaking schools, where cheating is a matter of unwritten policy, than in *individual-level* schools, where rule-breaking is limited to relatively few rogue teachers. This hypothesis is borne out in our data; across both districts, a pooled average of 63.8 percent of classrooms in *organizational* rule-breaking schools demonstrated evidence of cheating, compared to 37 percent of classrooms in the average *individual-level* rule-breaking school.

Despite substantial differences in cheating behaviors between *individual* and *organizational* rule-breaking schools, these institutions did not demonstrate large disparities in

the structural or compositional measures presented in Table 3. *Individual-level* and *organizational* rule-breaking schools enrolled comparable numbers of students, had similar staffing levels, and served similarly low-income, mostly African American student populations. Accountability pressure, as measured by the share of students who failed to meet proficiency thresholds as well as schools' AYP status in the year preceding the erasure analysis, did not differ substantially between these two groups of schools. Of note, however, is the large increase in the percentage of students who failed to meet proficiency targets in 2010, the year following the public revelation of Atlanta's cheating scandal, when educators could anticipate heightened scrutiny of their testing procedures. Indeed, within-cohort comparisons (e.g., comparing the 2008 proficiency levels of a school's third graders to their 2010 proficiency rates as fifth graders) reveal an average increase in the share of students failing to meet proficiency targets of more than nine percentage points across both institutional types. This equates to more than a 50 percent increase in the share of students failing to pass one or more portions of the CRCT immediately following the exposure of large-scale cheating in APS, providing further indication of the legacy of cheating that had become entrenched in many of these schools.

Taken together, our findings from both quantitative and qualitative analyses suggest that large-scale cheating outbreaks arise not from large numbers of teachers making concurrent, independent decisions to violate testing rules, but rather from a sea change in schools' normative structure initiated by a comparatively small number of district and/or school administrators for the purpose of facilitating rule-breaking. This process becomes clearer upon examining the various motivations to cheat affecting actors at all levels of the institutional hierarchy.

<Table 4 About Here>

Sources of Cheating Motivation

The second phase of our qualitative analyses explored why cheating became a multi-level phenomenon in which individuals across different levels of the organization cooperated to break rules. Table 4 presents a summary of the social and organizational conditions that precipitated *organizational* rule-breaking in APS and DCSS. Three dominant themes emerged: the presence of strong, top-down pressure to meet unrealistic accountability targets; a climate of fear tied to job security and professional standing; and opportunities for professional gain or recognition tied to standardized test scores. These themes share a dependence on organizational leaders' decisions and behaviors. When school leaders stoked teachers' concerns around accountability pressure, job security, or professional recognition, large-scale *organizational* rule-breaking tended to ensue. In schools that were not characterized by these themes, cheating generally conformed to *individual* rule-breaking, which impacted fewer students, on average.

Pressure. The power of top-down, external pressure was a major theme that emerged from our analysis of archival data. Principals and teachers alike suggested that they felt powerful, unambiguous pressure to boost CRCT scores. At one APS school, for example, the regional school director pressured the principal to sign a guarantee that 100-percent of the students at her school would at least meet expectations on the CRCT, an unrealistic goal given the school's historical performance. After refusing to sign the document, the principal began receiving threatening emails from the regional director. Similarly, one principal emphasized to teachers that their classroom should have a 100-percent CRCT pass rate, demonstrating how pressure from the district was passed down to both principals and teachers.

The pressure to produce high test scores was often coupled with explicit pressure from district and school officials to violate CRCT testing policies. For instance, one administrator repeatedly asked a fourth grade teacher to participate in erasing until he finally agreed to assist. A teacher at the same school expressed that, after being directed to change answers by an administrator, she sat in the room with others and pretended to erase answer sheets because she felt that saying no was not an option. In a letter to the Georgia Professional Standards Commission, one teacher stated that, after rejecting an administrator's offer to check her students' answers, the administrator instructed her to "just look over them anyway" and left the teacher alone with the CRCT test booklets. Another teacher, who confessed to cheating, indicated that he followed the principal's directives to use voice inflection and non-verbal cues to alert students if they selected the wrong answer. The above are just a few examples of the direct pressure administrators placed on teachers to break the rules around CRCT testing.

Importantly, external pressure did not emerge as a major theme in schools characterized by *individual, contested* rule-breaking. Specifically, pressure was only cited as a motivation for cheating in two out of six DCSS schools and three out of six APS schools in which rule-breaking was both *individual* and *contested*.

Fear. A pervasive climate of fear driven by the perceived negative consequences of low standardized test scores and retaliation for reporting unethical testing behavior emerged as a second major theme. The climate of fear surrounding testing sent a clear message to teachers from school administrators: failure to meet accountability standards would have serious negative implications for their careers. In these settings, administrators exploited their situational power advantage to compel classroom teachers and other school personnel to engage in rule-breaking by threatening them with demotion to a lower grade level, being assigned a professional

development plan (PDP, a formal record of educators' underperformance and the first step in the termination process), and job loss. At one such school, the principal threatened to negatively alter teachers' past evaluations if their students' test scores did not meet expectations at the end of the year. Among principals at APS schools characterized by *organizational* or *permitted* rule-breaking, six were on PDPs due to school performance and 15 threatened teachers with a PDP for low test scores. According to teachers, principals frequently stated, "if I am going to be placed on a PDP for low test scores, then so will you." This sanction carried the unmistakable message that the recipient's job was in jeopardy, lending administrators a powerful coercive tool.

In APS, the most severe penalties came in response to attempted whistleblowing. In instances where a teacher or staff member reported cheating, officials took no action, dismissed the claim after a limited investigation, and/or retaliated against the whistleblower, with lower-status employees being the most frequent and severe targets of retaliation. For example, a paraprofessional who objected to her principal's directive to prompt students during CRCT testing reported her concerns to district officials. The following day, a district official called a meeting with both the paraprofessional and the principal and warned the whistleblower to "keep her mouth shut" and to stop causing problems. The paraprofessional's employment was terminated at the end of the school year. A second paraprofessional reported to her principal that she witnessed a teacher changing test responses, but stated that the principal "acted like she didn't hear it." Later that day, the principal called the paraprofessional's home and told her to report to the district office the following morning, at which point her position was terminated. Additionally, teachers at two separate schools suggested to investigators that their contracts were not renewed in retaliation for speaking out about cheating.

Considering the role of power adds substantial complexity to our understanding of educator cheating. In certain settings, teachers operate simultaneously as perpetrators and victims, forced into a binary choice between cheating to retain their jobs or defying coercive pressure from their organizational leaders, thereby putting their careers at risk. Coupled with the organized nature of rule-breaking and pressure to further increase students' inflated scores, fear created an environment in which cheating was the "least worst option" (Martin et al. 2013) available to teachers.

Professional Gain. Finally, many educators in APS cited personal and professional recognition as a motivation to cheat. The annual APS convocation epitomized the district's climate of heightened competition surrounding standardized tests. Held in a cavernous indoor professional football stadium, the annual convocation was a mandatory event for all APS employees. Educators in schools that met 70 percent or more of the previous year's proficiency targets were invited to sit in chairs on the main floor of the arena during a celebration of their achievements, while educators from schools that failed to meet targets received no recognition and were seated in progressively distant reaches of the grandstand as their schools' proficiency rates declined. "Making the floor" was a commonly referenced goal of teachers and administrators in the GBI report, as they viewed this achievement as a mark of distinction, whereas failure to make the floor was considered a correspondingly negative indicator of professional competence.

Although rewards occasionally came in the form of financial bonuses, the more salient prizes in the eyes of many teachers and administrators interviewed by the GBI included professional esteem, public recognition, and prestigious job promotions. Educators at multiple schools indicated that professional advancement in APS was linked to growth in students' test

scores. However, for many educators in APS, the challenging realities of urban schooling all but precluded the possibility of meeting high and annually increasing achievement targets by legitimate means.

A climate of competition linked strictly to standardized test scores was a common theme in APS schools in which *organizational* or *individual, permitted* rule-breaking was the norm. In contrast, competition and professional recognition were cited as sources of cheating motivation in only one DCSS school. The professional culture among educators in DCSS, which did not dedicate much energy to publicizing student test scores, appears to have avoided developing the competitive bent that took root in APS.

In all respects, our archival data analysis suggests that large-scale cheating in APS primarily took the form of *organizational* rule-breaking that was *contested* at the district level and *permitted* within schools. APS district and school officials created an organizational testing culture marked by pressure, fear, and competition, and instituted cheating as a normative practice in schools throughout the district. In contrast, cheating only took the form of *organizational* and/or *permitted* rule-breaking in the minority of DCSS schools in which pressure and fear were characteristics of the school culture.

DISCUSSION

Our organizational analysis of standardized test cheating provides several insights regarding the nature of cheating outbreaks. Foremost, our findings suggest that cheating should not be considered a strictly teacher-level form of deviance. A given teacher's probability of cheating is correlated with that of the other teachers who share the same district, school, and grade-level contexts. Yet this contextual dependence varies across settings. In districts in which more than

the statewide mean of 6.62 percent of classrooms showed evidence of tampering, 49 to 59 percent of the classroom-level variance in wrong-to-right erasures resided at bureaucratic levels above the individual teacher, compared to 17 to 19 percent of the cheating variance in lower-incidence school districts.

Although we interpret the clustered pattern of standardized test cheating as evidence of coordinated organizational deviance, there are alternative processes that could produce the same distribution of suspicious erasures. The most straightforward hypothetical scenario is one in which the subset of teachers who are strongly predisposed to deviant behavior actively seek district or school settings in which they can effectively exercise their preference to cheat. Our analysis of the GBI interview summaries does not lend support to this hypothesis, however, as APS and DCSS teachers overwhelmingly describe their schools as sites of conversion to cheating as opposed to safe havens for deviant teachers.

A different teacher sorting process might also lead to the emergence of “cheating” schools or districts, however. Even if there are not groups of inherently deviant teachers who seek out common school settings, the teacher labor market tends to match teachers with the least experience and qualifications to schools that enroll the highest proportions of low-achieving students (Lankford, Loeb, and Wyckoff 2002). Similar processes operate within schools, placing the least experienced teachers in classrooms with the lowest-achieving students (Kalgorides, Loeb, and Bételle 2013). Thus, the population of teachers who face the highest test score improvement hurdles are precisely those with the fewest pedagogical resources upon which they might draw to produce the achievement gains necessary to meet accountability targets. It may therefore be the case that teachers who have limited access to legitimate means of increasing student achievement (as opposed to a greater desire to cheat) are sorted into schools and

classrooms that pose the greatest achievement challenges, which could contribute to the clustered cheating pattern we observe. Future research that adjusts for teacher-level covariates, including experience and skill, will be necessary to assess this hypothetical mechanism.

In light of continued political pressure to leverage standardized test scores as indicators of teacher and administrator efficacy, our review of archival data from APS and DCSS schools points toward strategies that will be especially effective in reducing the prevalence of cheating in the future. In particular, educators reported being most likely to engage in cheating when they viewed achievement targets as both unrealistic and highly consequential for their professional security. Results from our qualitative data analysis suggest that reducing this pressure, perhaps by adjusting achievement targets to reflect districts' and schools' student body compositions to make targets more realistic, should reduce the motivation to cheat at all organizational levels. Florida, Virginia, and Washington D.C. attempted to make adjustments along similar lines, instituting race-specific achievement targets to account for average achievement differences between students from different racial backgrounds (Postal 2012). Although this race-focused strategy proved tremendously controversial, our analysis suggests that an alternative departure from uniform achievement targets could allay a prime source of pressure on teachers to inflate their students' test scores.

Our finding that large-scale cheating is associated with the *organizational* form of rule-breaking leads to additional implications for education policy. Qualitative data suggest that, as standardized accountability tests gained influence over the career trajectories of high-level members of the educational bureaucracy, ensuring high scores became an increasingly pressing matter for these organizational leaders. However, due to the nested layers of authority and action that characterize American public school systems, a superintendent's or principal's decision to

promote cheating created a cascade, resulting in a far greater prevalence of cheating than would have ensued if cheating decisions had been determined by individual teachers. Organizational leaders who held responsibility for ensuring test security enacted multiple strategies to provide lower-level school personnel with opportunities and incentives to cheat. Thus, efforts to mitigate standardized accountability tests should be most effective to the extent that they are directed toward higher levels of the educational bureaucracy, focusing at least as much attention on superintendents and principals as classroom teachers.

Ensuring high-stakes accountability tests' validity is essential to the provision of equitable and excellent educational opportunities, particularly for students who may be entitled to remedial services due to lagging achievement. For example, a fifth-grade teacher interviewed by the GBI expressed shock and dismay at the fact that eight out of 16 boys in her class could barely read, yet each one passed the CRCT and was promoted to the sixth grade. Teachers in 28 of the APS and DCSS schools we examined either identified students whose academic ability and performance did not match their CRCT scores, expressed concern that they were unable to recommend grade retention or refer a student for special education services due to their suspiciously high test scores, or expressed surprise that students with formally identified disabilities excelled on the CRCT, potentially causing them to prematurely exit the special education program.

It should be clear that reducing opportunities to cheat is important, yet the powerful motivation to cheat speaks to a larger, more pressing issue within contemporary education reform debates: the emphasis on standardized test scores as indicators of educational effectiveness. We do not mean to suggest that standardized tests are inherently problematic, as valid and reliable student achievement data surely have a place as an evaluative and planning

tool for educators and administrators. Yet the sheer number of documented cases of cheating in American public schools suggests the need for more appropriate, holistic approaches to teacher, administrator, and school evaluation. A “cult of data” (Georgia Office of the Governor, 2011a) has arisen in many public school systems in which standardized test scores are reified as a gold-standard metric of educational success. The consequent emergence of cheating scandals across the country suggests that a more balanced system of measurement and accountability will be necessary to move public schools in the direction of greater excellence and equity.

REFERENCES

- Amrein-Beardsley, Audrey, David C. Berliner, and Sharon Rideau. 2010. "Cheating In the First, Second, and Third Degree: Educators' Responses to High-Stakes Testing." *Educational Policy Analysis Archives*, 18(14), 1–36.
- Apperson, Jarod, Carycruz Bueno and Tim R. Sass. 2016. "Do the Cheated Ever Prosper? The Long-Run Effects of Test-Score Manipulation by Teachers on Student Outcomes." CALDER Working Paper No. 155. National Center for Analysis of Longitudinal Data in Education Research.
- Battistin, Erich. 2016. "How Manipulating Test Scores Affects School Accountability and Student Achievement." *IZA World of Labor*, 295.
- Booher-Jennings, Jennifer. 2005. "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal* 42:231-68.
- Cizek, Gregory J., and James A. Wollack, 2017. *Handbook of Quantitative Methods for Detecting Cheating on Tests*. New York, NY: Routledge.
- Coleman, James S. 1987. "Toward an Integrated Theory of White-Collar Crime." *American Journal of Sociology* 93:406-439.
- Colias-Pete, Meredith and Carole Carlson. 2019. "Ex-principal, 6 Gary Teachers Suspended After Mccullough Academy Standardized Test Cheating Scandal," *Chicago Tribune*, January 16. Retrieved May 2, 2019 (<https://www.chicagotribune.com/suburbs/post-tribune/ct-ptb-gary-mccullough-test-suspensions-st-20190116-story.html>)
- Creswell, John W., and Vicki L. Plano Clark. 2011. *Designing and conducting mixed methods research* (2nd ed.) Thousand Oaks, Calif.: SAGE Publications.

- Darling-Hammond, Linda. 2004. "Standards, Accountability, and School Reform." *Teachers College Record* 106(6):1047-1085.
- Dee, Thomas S., Will Dobbie, Brian Jacob, and Jonah Rockoff. 2016. "The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examination." *NBER Working Paper* 22165.
- Diamond, Rebecca, and Petra Persson. 2016. "The Long-Term Consequences of Teacher Discretion in Grading Of High-Stakes Tests." *NBER Working Paper* 22207.
- DiMaggio, Paul J. and Walter Powell. 1991. "Introduction." Pp. 1–40 in *The New Institutionalism in Organizational Analysis*, edited by W. W. Powell and P. J. DiMaggio. Chicago, IL: University of Chicago Press
- Duncan, Arne. 2011. "Despite Cheating Scandals, Testing and Teaching Are Not At Odds." *The Washington Post*. Retrieved December 1, 2011 (http://www.washingtonpost.com/national/on-leadership/despite-cheating-scandals-testing-and-teaching-are-not-at-odds/2011/07/19/gIQADUb3NI_story.html).
- Georgia Office of the Governor. 2011a. "Atlanta Public Schools Investigation Report: Volumes 1-3." Retrieved October 5, 2012 (http://www.gadoe.org/_documents/gaosa/APS%20Investigation%20Volume%201.pdf)
- , 2011b. "Dougherty County Investigative Report: Volumes 1 and 2." Retrieved October 5, 2012 (http://www.gadoe.org/_documents/gaosa/APS%20Investigation%20Volume%201.pdf)
- Glaser, Barney. 1978. *Theoretical sensitivity: advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press

- Government Accountability Office. (2013). "K-12 education: States' Test Security Policies and Procedures Varied (GAO-13-495R)." Washington, DC: Author. Retrieved from <http://www.gao.gov/assets/660/654721.pdf>
- Hallett, Tim. 2010. "The Myth Incarnate: Recoupling Processes, Turmoil, and Inhabited Institutions in an Urban Elementary School." *American Sociological Review* 75(1):52–74
- Hallett, Tim, and Marc J. Ventresca. 2006. "Inhabited institutions: Social interactions and organizational forms in Gouldner's Patterns of Industrial Bureaucracy." *Theory and Society* 35(2): 213-236.
- Jacob, Brian A. and Steven Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *The Quarterly Journal of Economics* 118(3):843-878.
- Jennings, Jennifer L. and Heeju Sohn. 2014. "Measure for Measure: How Proficiency-based Accountability Systems Affect Inequality in Academic Achievement." *Sociology of Education* 87:125-141.
- Kagan, Robert A. and John T. Scholz, 1984. "The Criminology of the Corporation and Regulatory Enforcement Strategies," in *Enforcing Regulation* , edited by K. Hawkins and J. Thomas, Boston: Kluwer Academic Publishers
- Kalogrides, Demetra, Susanna Loeb, and Tara Bételle. 2013. "Systematic Sorting: Teacher Characteristics and Class Assignments." *Sociology of Education* 86: 103-123.
- Kramer, Ronald C. 1982. "Corporate Crime: An Organizational Perspective." Pp. 75-94 in *White-Collar and Economic Crime*, edited by Peter Wickman and Timothy Dailey. Lexington, Mass.: Lexington Books.

- Lankford, Hamilton, Susanna Loeb, and James Wyckoff. 2002. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis." *Educational Evaluation and Policy Analysis* 24:37-62.
- Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83(2):340-63.
- McMillan, James H., Steve Myran, and Daryl Workman. 1999. "The Impact of Mandated Statewide Testing on Teachers' Classroom Assessment and Instructional Practices". Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada. Retrieved October 14, 2011 (<http://www.eric.ed.gov/PDFS/ED431041.pdf>).
- Martin, Andrew, Steven Lopez, Vincent Roscigno, and Randy Hodson. 2013. "Against the Rules: Synthesizing Types and Processes of Bureaucratic Rule-breaking." *Academy of Management Review* 38:550-574.
- Pedulla, Joseph J., Lisa M. Abrams, George F. Madaus, Michael K. Russell, Miguel A. Ramos, and Jing Miao. 2003. "Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers." *National Board on Educational Testing and Public Policy*. Retrieved October 14, 2011 (<http://www.bc.edu/research/nbetpp/statements/nbr2.pdf>).
- Postal, Leslie. 2012. "State Sets Reading, Math Goals Based on Race, Ethnicity." Orlando Sentinel. Retrieved January 8, 2012 (http://articles.orlandosentinel.com/2012-10-10/features/os-race-student-achievement-florida-20121009_1_grade-level-black-students-white-students)

- Resmovits, Joy. 2011. "Schools Caught Cheating in Atlanta, Around the Country." *The Huffington Post*. Retrieved December 1, 2012 (http://www.huffingtonpost.com/2011/08/08/atlanta-schools-cheating-scandal-ripples-across-country_n_919509.html)
- Sainz, Adrian. 2012. "Cheating Scandal: Feds Say Teachers Hired Stand-in to Take Their Certification tests." *NBC News*. Retrieved December 1, 2012 (http://usnews.nbcnews.com/_news/2012/11/25/15430647-cheating-scandal-feds-say-teachers-hired-stand-in-to-take-their-certification-tests?lite).
- Smith, Mary Lee and Patricia Fey. 2000. "Validity and Accountability in High-Stakes Testing." *Journal of Teacher Education* 51:334
- Stuebing, Karla K., Jack M. Fletcher, Josette M. LeDoux, G. Reid Lyon, Sally E. Shaywitz, and Bennett A. Shaywitz. 2002. "Validity of IQ-Discrepancy Classifications of Reading Disabilities: A Meta-Analysis." *American Educational Research Journal* 39(2): 469-518.
- Sutherland, Edwin. 1947. *Principles of Criminology*. Philadelphia: J.B. Lippincott.
- , 1949. *White Collar Crime*. New York: Dryden.
- Vaughan, Diane. 1992. "The Macro-Micro Connection in 'White-Collar Crime' Theory." Pp. 124-145 in *White Collar Crime Reconsidered*, edited by Kip Schlegel and David Weisburd. Boston: Northeastern University Press.
- , 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. Chicago: University of Chicago Press.
- , 1998. "Rational Choice, Situated Action, and the Social Control of Organizations." *Law & Society Review* 32(1): 23-61.

- , 2007. 'Beyond Macro- and Micro-Levels of Analysis, Organizations, and the Cultural Fix.' Pp. 3–24 in *International Handbook of White-Collar and Corporate Crime*, edited by Henry N. Pontell and Gilbert Geis. New York, NY: Springer
- Vogell, Heather. 2011. "Investigation into APS Cheating Find Unethical Behavior Across Every Level." *Atlanta Journal Constitution*. Retrieved November 14, 2011 (<http://www.ajc.com/news/investigation-into-aps-cheating-1001375.html>).

Figure 1. Proportion of Classrooms Flagged for Cheating by School

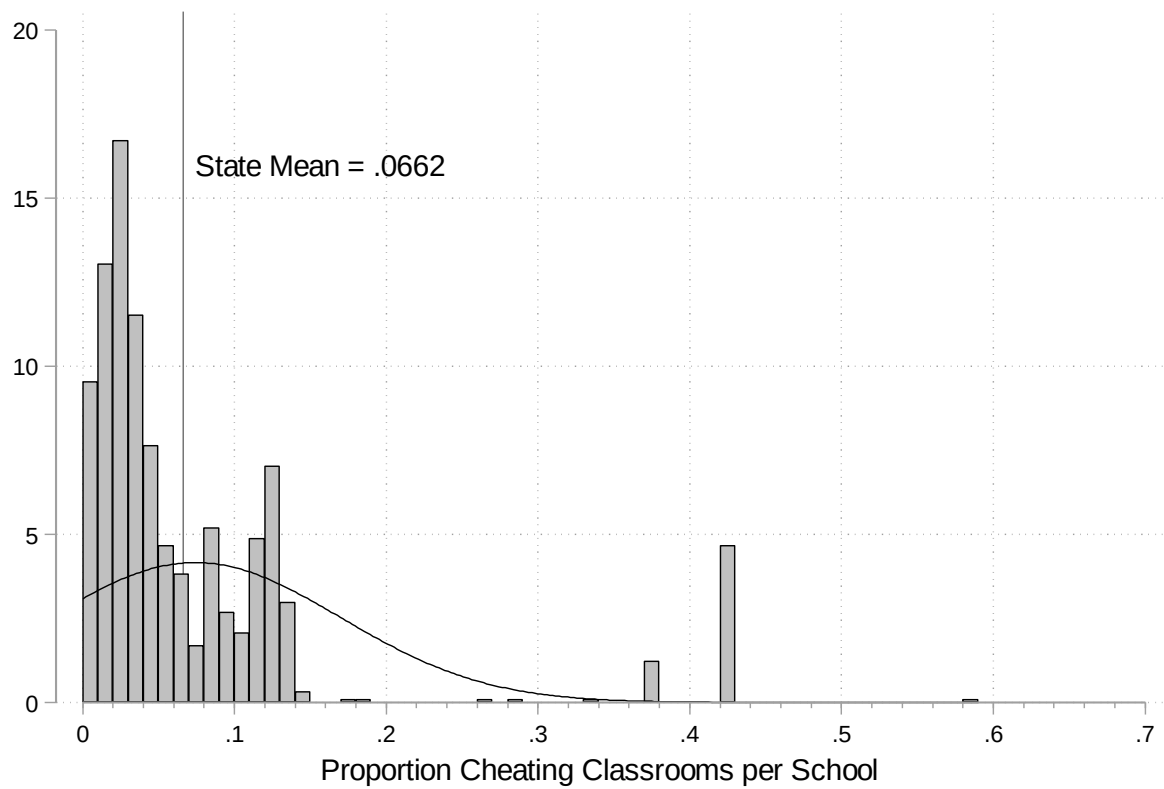


Figure 2. Proportion of Classrooms Flagged for Cheating by School District

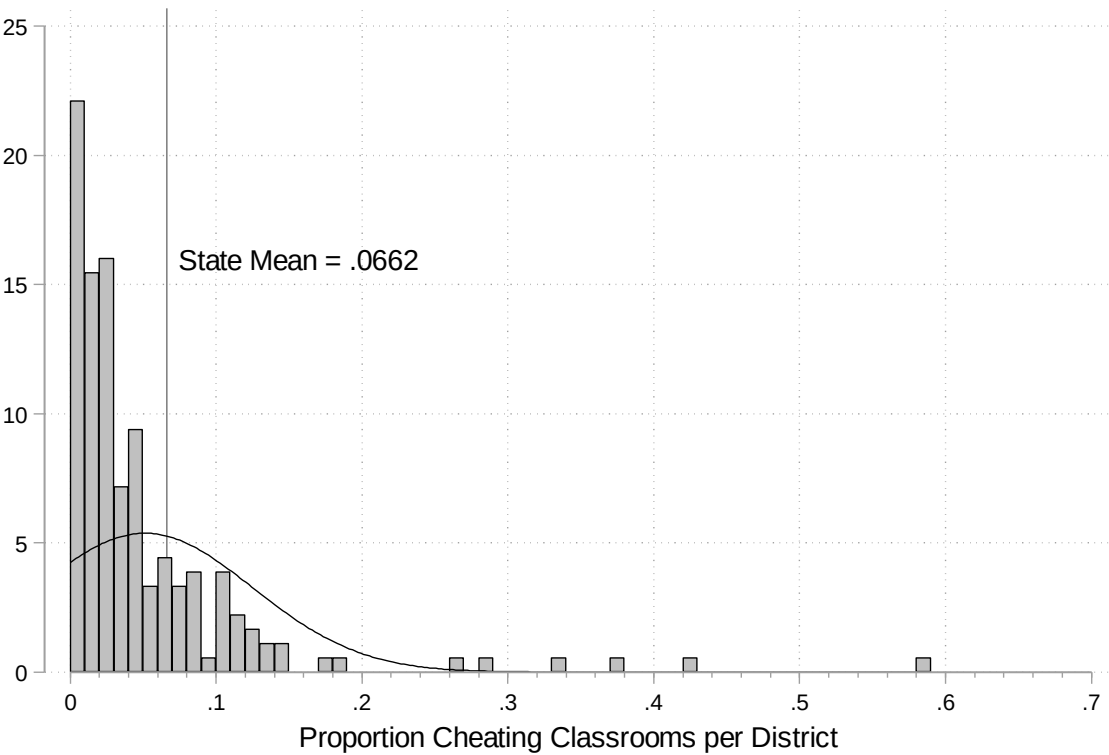


Table 1. Theoretical Typology of Standardized Test Cheating

Hierarchical Level	Contentiousness	
	Permitted	Contested
Individual	Individual teachers independently engage in cheating with the understanding that it is tacitly allowed by school leaders	“Bad apple” teachers undertake covert cheating efforts against school leaders’ wishes
Group	School leaders turn a blind eye to groups of teachers’ coordinated cheating efforts	Groups of teachers conspire to break school testing rules for the sake of meeting collective goals
Organization	Lack of regulatory oversight leads to institutional decoupling, creating opportunity for school or district officials to ignore externally imposed rules and promote cheating without fear of reprisal	School or district administrators intentionally violate external testing regulations, creating “bad orchard” districts/schools by incentivizing or coercing lower-level educators to engage in cheating

Adapted from Martin et al. (2013)

Table 2. Intraclass Correlations from Multilevel Regression Analysis of Classroom-Level Wrong-to-Right Erasures

Reading Test			
	All Districts	Districts with > P Cheating	Districts with < P Cheating
	ICC	ICC	ICC
School District Level	.086	.087	.025
School Level	.269	.315	.070
Grade-level Level	.170	.189	.090
Total	.525	.591	.185
N , Districts	171	39	132
N , Schools	1,203	402	801
N , Grade-Levels	5,612	1,968	3,644
N , Classrooms	30,443	9,230	21,213

Math Test			
	All Districts	Districts with > P Cheating	Districts with < P Cheating
	ICC	ICC	ICC
School District Level	.086	.089	.034
School Level	.223	.280	.048
Grade-level Level	.161	.162	.114
Total	.470	.531	.196
N , Districts	171	34	137
N , Schools	1,203	387	816
N , Grade-Levels	5,612	1,906	3,706
N , Classrooms	30,445	8,913	21,532

Language Arts Test			
	All Districts	Districts with > P Cheating	Districts with < P Cheating
	ICC	ICC	ICC
School District Level	.090	.088	.032
School Level	.207	.254	.044
Grade-level Level	.134	.143	.101
Total	.431	.485	.177
N , Districts	171	40	131
N , Schools	1,203	419	784
N , Grade-Levels	5,612	2,059	3,553
N , Classrooms	30,443	6,989	20,754

P = Statewide proportion of classrooms with mean wrong-to-right erasure counts greater than 2.32 standard deviations above their statewide grade-level and subject matter means. $P_{\text{reading}} = .063$; $P_{\text{math}} = .069$; $P_{\text{language arts}} = .067$

Table 3. Descriptive Comparison of *Individual-Level* and *Organizational* Rule-Breaking Elementary Schools, Atlanta and Dougherty County School Districts

Number of Schools by Rule-Breaking Type and District				
	Individual-Level Rule-Breaking Schools		Organizational Rule-Breaking Schools	
	APS	DCSS	APS	DCSS
Contested Rule-Breaking	6	6	0	3
Permitted Rule-Breaking	7	1	20	0
Variable Means and Standard Deviations				
	Individual-Level Rule-Breaking Schools		Organizational Rule-Breaking Schools	
	APS	DCSS	APS	DCSS
Cheating Frequency				
Proportion Cheating Classrooms ¹	.354 (.236)	.384 (.123)	.673 (.194)	.530 (.455)
WTR Erasures Per Test Form ²	2.93 (2.90)	3.40 (1.81)	7.37 (3.96)	5.70 (5.89)
Accountability Status				
Percentage Of Students Failing to Meet Proficiency Targets, 2008	15.53 (7.74)	18.41 (6.05)	16.66 (6.62)	14.01 (6.05)
Percentage Of Students Failing to Meet Proficiency Targets, 2010	22.99 (9.49)	23.75 (9.04)	26.70 (7.19)	28.25 (3.19)
Within-Cohort Increase in % Failing to Meet Proficiency, 2008-2010	7.65 (7.47)	7.94 (5.19)	9.66 (7.46)	16.31 (2.39)
School Designated "Needs Improvement" under NCLB, 2008	.077 (.277)	.143 (.378)	.050 (.224)	.000 (.000)
Demographic Composition				
Student Enrollment	492.69 (155.93)	523.86 (112.09)	404.15 (132.88)	490.67 (23.54)
Pupil/Teacher Ratio	13.30 (1.39)	15.27 (1.36)	13.38 (1.69)	16.50 (1.81)
Percentage Free/Reduced Price Lunch Eligible Students	84.62 (25.42)	89.63 (8.41)	92.22 (8.49)	94.57 (.815)
Percentage Black Students	88.15 (24.14)	89.86 (9.81)	94.55 (9.91)	94.33 (8.96)
Percentage White Students	6.08 (20.46)	6.86 (7.13)	.450 (1.19)	4.67 (7.23)
Percentage Latinx Students	3.69 (5.51)	2.14 (2.48)	4.00 (9.67)	.333 (0.58)
Percentage Asian Students	.538 (1.39)	.286 (.756)	.250 (.550)	.000 (.000)
Percentage English Learner Students	2.85 (4.47)	.857 (1.21)	2.20 (4.91)	.000 (.000)
Percentage Students w/ Disabilities	7.69 (1.80)	10.57 (3.05)	8.30 (4.24)	11.33 (4.04)
Observations	13	7	20	3

¹ Cheating classrooms are those with mean wrong-to-right erasure frequencies greater than 2.32 standard deviations above the state mean

² z- score; wrong-to-right erasure counts are standardized within grade levels and subject areas across all tests in Georgia

Table 4. Cheating Motivation Themes in Atlanta and Dougherty County Schools

Theme	Description	Example Evidence	Prevalence	Example Quotes	Relation to contentiousness/hierarchy
Pressure	Top-down pressure to meet unrealistic accountability targets, maintain artificially inflated test scores, deviate from established testing protocol, or alter CRCT test forms	The regional academic director forced one APS school principal to sign a guarantee that 100% of students would meet expectations on the CRCT; multiple principals told teachers to find another job if they could not meet AYP; teachers were directly and indirectly instructed to "clean" or "check" student answer sheets; during testing, some administrators asked teachers to identify students whose performance was not meeting or exceeding expectations; organizational superiors constantly reminded educators about "the numbers"	23 out of 33 schools in APS; 6 out of 10 schools in Dougherty	"At East Lake, we do whatever we have to do even if it means breaking the rules...See the scores? East Lake makes its targets." - APS teacher quoting principal "If Johnny does not know how to read, he had better know how on test day." APS teacher quoting principal	Pressure was not a primary theme in <i>individual-level</i> rule-breaking schools, but was virtually ubiquitous in <i>organizational</i> rule-breaking schools. Pressure coincided with <i>permitted</i> rule-breaking.
Fear	School climate characterized by teachers' perceived negative consequences of low standardized test scores and retaliation for reporting unethical testing behavior	Teachers and principals were threatened with demotion to a lower grade level, being assigned a PDP ¹ , and job loss in response to low test scores; teachers and other school personnel were compelled to resign rather than participate in organized cheating; educators/staff members were fired following attempts to report cheating to school or district officials; principals threatened to alter past teacher evaluations based on CRCT test scores	21 out of 33 schools in APS; 3 out of 10 schools in Dougherty	"Teachers compel themselves to do whatever they need to do to make sure that they do not lose their jobs... Everybody was in fear. It is not that the teachers are bad people and want to do it, it is that they are scared." - Implicated testing coordinator "If you want to keep your job, you better clean those tests"- Implicated teacher	Fear coincided with <i>organizational and permitted</i> rule-breaking forms, and frequently co-occurred with pressure. Fear's effectiveness derived from powerful actors coercing lower-status workers to engage in cheating.
Professional Gain	Principals and teachers were rewarded for meeting achievement targets, regardless of the means taken.	Teachers' and principals' rewards occasionally came in the form of financial bonuses, but professional esteem, public recognition, and job promotions were far more frequently cited by interviewees; The APS convocation ceremony glorified educators in high-scoring schools while shaming those whose students produced lower scores. Multiple educators cited the importance of "making the floor" during convocation	9 out of 33 schools in APS; 1 out of 10 schools in Dougherty	"Teachers who had high scoring classes were given such special privileges as bonus supply money and the opportunities to attend conferences" - APS teacher "If you cheat, you can move kids, you are going to get promoted." - APS Principal	Professional Gain emerged as a major theme only in APS, where test-based competition among schools drove many <i>organizational</i> , <i>permitted</i> rule-breaking efforts. Only one Dougherty County teacher referenced professional gain as a motivation, and that teacher worked in a school coded as <i>individual, contested</i> rule-breaking site.

¹ Professional Development Plan, a formal record of educators' underperformance and the first step in the termination process

Online Appendix: Select Codes and Descriptions

<i>General Category</i>	<i>Category Description</i>	<i>Codes</i>
School Characteristics	Descriptive characteristics of schools	District (APS or DCSS); Grades (Elementary, Middle, High); Title I (Y/N); Percentage of Flagged Classes (0-25, 26-50, 51-75, Above 75)
Mode of Cheating	Educators' descriptions of how cheating was facilitated	Pre-test (direct instruction using official CRCT materials); during testing (prompting, voice inflection; nonverbal cues to students); Post-test (erasing)
Motivation for Cheating	Educators' descriptions of the factors that contributed to their decision to break the rules	Indirect pressure ("by any means necessary"); Direct pressure/request to cheat; fear (of professional loss); professional gain; monetary incentives
Consequences of Cheating	Student outcomes associated with inflating test scores	Social promotion; missed remediation; grade-level/skill mismatch
Contentiousness of Rule-breaking	The intensity of rule enforcement by organizational actors with the power to contest it	Low intensity (permitted); high intensity (contested)
Hierarchy of Rule-breaking	Organizational level at which rule-breaking is organized and enforced	Individual (teacher or paraprofessional); workgroup (grade or subject area teams); organizational (principal, assistant principal, ILS, testing coordinator)
Overall Pattern of Cheating	The school's rule-breaking pattern according the Martin et al. (2013) rule-breaking typology	Indulgency Pattern (Individual + Low Intensity); Bad Apples (Individual + High Intensity); Making Out (Workgroup + Low Intensity); Internal Power Struggles (Workgroup + High Intensity); Decoupling (Organizational + Low Intensity); Organizational Deviance (Organizational + High Intensity).