

**The Impact of School-Level Accountability  
on Local Test Preparation Practices**

Michelle Croft, Kris Waltman,  
Kyndra Middleton, and Erika Stevenson

The University of Iowa

Paper presented at the NCME Annual Meeting, Montreal, April 2005

## Abstract

The study was designed to develop a greater understanding of how test preparation practices/activities have changed in a state with an established testing program that has recently begun to use test scores for school-level accountability purposes. Teachers within 24 public high schools completed a questionnaire related to the use of test preparation activities, the ethicality of the activities, and motivational activities/incentives related to testing. Nonparametric statistics were used to compare responses among schools. The results of the study indicate that school achievement level is not related to the use of test preparation practices. However, the number of sources of pressure to increase test scores does contribute to the use of certain test preparation activities. Also, there were no state-wide trends in the use of motivational activities/incentives related to test scores, although over half of the schools in the sample did use some type of student incentives. Finally, there are suggestions related to future research and professional development regarding the appropriateness of certain test preparation activities.

This work was supported by the Iowa Department of Education (DE). The authors take full responsibility for the work and no endorsement from the Iowa DE should be assumed.

## Introduction

This study was designed to develop a greater understanding of how test preparation practices/activities have changed in a state with an established testing program that has recently begun to use test scores for school-level accountability purposes. Test preparation practices are an increasingly important issue because they often affect the validity of inferences based on the scores from the test and may undermine student learning. Thus, there is a need to determine what types and in what contexts (i.e., teacher-level and school-level characteristics) these practices are being employed so that corrective (or preventative) action can be taken, if needed, in order to maintain positive, productive instructional environments.

## Background

The main concern of measurement professionals and policy-makers regarding test preparation is the validity of the scores from the test because test preparation practices may influence the capability of the test to provide an accurate portrayal of a student's achievement. In some cases, test preparation may yield more valid scores (e.g., by increasing student familiarity with the test format, properly completing the answer sheets, and reducing student anxiety). However, more often there is a concern that test preparation may lead to "test pollution," described by Messick (1984) as an increase or decrease in test performance that is not connected to the construct represented on the test; thereby producing construct-irrelevant test score variance. Haladyna, Nolen, and Haas (1991) cite the following three main sources of test score pollution: a) the test administration conditions (e.g., student anxiety), b) external factors out of the school's control (e.g., English proficiency), and c) test preparation.

Past researchers have warned that certain preparation practices may produce artificial gains in test scores (Amrein & Berliner, 2002; Koretz & Barron, 1998; Koretz, McCaffrey, & Hamilton, 2001). Amrein and Berliner (2002) examined student achievement increases after the introduction of high-stakes tests. They found that the increases in test scores were most likely a result of a "training effect" and not substantial gains in achievement. Koretz and Barron (1998) found that the Kentucky's KIRIS (Kentucky Instructional Results Information) showed large gains in student performance, while the mean scores on other assessments such as the National Assessment of Educational Progress (NAEP) and ACT remained relatively unchanged among students who took both the ACT and the KIRIS, indicating that the gains in student performance on the KIRIS might not reflect legitimate increases in student learning. These initial findings

prompted Koretz et al. (2001) to assert that only “teaching more, working harder, and working more effectively” can produce unambiguous gains – all other methods that may result in increasing test scores, such as the reallocation of resources, alignment, coaching, and cheating, produce gains that may be suspect. With the exception of cheating, the other methods (i.e., reallocation, alignment, coaching) may produce actual gains in student learning, but the legitimacy of the gains is dependent upon how the methods are implemented. For instance, re-examining the alignment of the test to the curriculum may help teachers detect gaps in the curriculum, which may produce actual gains in student achievement; whereas realigning curriculum so that the only material taught is that which is on the test (e.g., “teaching to the test”), will most likely produce artificial gains in student achievement when the intended inference is to a broader domain of content and skill areas. Therefore, knowledge of the types of test preparation practices being employed is extremely useful when trying to accurately interpret score gains.

In an attempt to distinguish the teacher practices that would most likely contribute to test pollution, past researchers (Mehrens & Kaminski, 1989; Popham, 1991) have outlined those testing practices which are ethical/legitimate based on the *Standards for Educational and Psychological Testing* (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 1985). Mehrens and Kaminski (1989) developed a continuum of ethicality ranging from instruction on objectives that have been determined regardless of the test, which was always ethical, to practice on the same form of the test to be administered, which was always unethical. Practices become questionable in the middle of the continuum where content that is derivative of specific objectives on standardized tests is used as a test preparation activity. A similar continuum was developed by Haladyna, Nolen, and Haas (1991) in which only three activities were deemed ethical: teaching test-taking skills, checking answer sheets to ensure they were completed properly, and increasing student motivation to do well on the test. Like Mehrens and Kaminski (1989), Haladyna et al. (1991) believed that any time course objectives were modeled after the standardized test – excluding areas not covered by the test – then the practice was unethical.

Popham (1991) described five different types of test preparation practices and applied two criteria: measurement professional ethics and the educational defensibility of each practice. Popham’s approach arrived upon slightly different conclusions from both Mehrens and Kaminski

(1989) and Haladyna et al. (1991) because he did not consider the alignment of the curriculum to the test objectives. Popham concluded that using either the previous form or the current form of a test for preparation purposes were both *illegitimate* practices in terms of professional ethics and educational defensibility because neither shows a true growth in student learning. Instead, practicing with previous or current test forms represents a type of instruction that has the sole purpose of increasing test scores, not increasing mastery. In addition to the use of practice forms, Popham also criticized the use of exclusively same-format preparation in which all practice items are those in the same format as the items on the test. The practice of same-format preparation, according to Popham, is considered to be ethical by professional standards, but not in terms of educational defensibility. The reason for this is because same-format preparation does not allow for students to generalize their knowledge to other testing formats, which will be necessary for future situations. In contrast, using a varied-format preparation is both ethical and defensible because it provides instruction that is directly related to the test and provides other opportunities to allow students to adapt to new formats. Teaching general test-taking skills was the only other method of preparation that was considered to be both ethical and defensible.

Popham's (1991) description of ethical and educationally defensible test preparation practices was later refuted by Kilian (1992). First, Kilian disagreed that the use of previous form preparation was not educationally defensible. He argued that the use of previous form preparation allowed students to be given the opportunity to know "what is expected of them." Kilian also highlighted that Popham's article was appropriate primarily for criterion-referenced tests. For norm-referenced tests in order for the normative information to be valid, Kilian argued that the test preparation activities should be similar to those used by the norm group.

As can be seen from the different perspectives from measurement professionals concerning the ethicality of test preparation practices, there are a few practices where there is agreement that the practice is either ethical or unethical. All researchers included same-form preparation and exclusively same-format preparation as unethical. Teaching test-taking skills was the only activity that all measurement professional agreed upon as ethical.

Despite the efforts by measurement professionals to classify the appropriateness of test preparation activities, researchers have found that teachers and school administrators continued to be either unaware of which test preparation practices are appropriate or have beliefs that are different from the beliefs held by measurement professionals concerning which practices are

appropriate. For example, in a survey of teachers, Nolen, Haladyna, and Hass (1992) found that 25% of the teachers believed that teachers often taught students vocabulary items that would be used on the test. Popham (1991) found that 36% of the California teachers sampled believed that it was appropriate to use the same form of a test for preparation purposes, which was the one practice most clearly viewed by measurement professionals as being unethical. Furthermore, Cizek (1999) provides countless examples of “cheating” by both teachers and school administrators on standardized assessments.

Although understanding which practices teachers believe are ethical is important, the context of the schools regarding the use of test preparation practices must also be considered. Pedulla, Abrams, Madaus, Russell, Ramos, and Miao (2003) surveyed a nation-wide sample of teachers in order to determine what test preparation practices are being employed by teachers in different accountability systems (i.e., high, medium, or low stakes) for teachers and students. The researchers discovered that schools in which there were higher stakes, for either teachers or students, utilized test preparation practices to a greater extent than schools in which there were lower stakes. Specifically, there were more hours devoted to test preparation in high-stakes schools compared to low-stakes schools. Although Pedulla et al. (2003) does, for some items, disaggregate the responses by the grade levels served, there is no other information concerning the factors, such as the subject area taught or the achievement level of the school, which may be related to the use of test preparation practices.

Similar findings of the relationship between pressure to raise scores and the use of test preparation practices have also been documented by Nolen, Haladyna, and Haas (1992). The researchers found that almost 66% of the 1,373 elementary school teachers and over 40% of the 508 secondary school teachers surveyed reported feeling pressure to increase test scores from their school’s administration. More importantly, the teachers felt pressured to raise the scores through means other than instruction. Therefore, the focus of the school (i.e., on either test scores or student achievement) may contribute to the use of particular test preparation practices. The researchers did not examine the relationship between the extent of pressure teachers feel and the use of various test preparation practices or if certain teachers, particularly those who teach a subject that is covered by the standardized test, are more likely to feel pressured than other teachers. This knowledge would be useful in order to better examine which test preparation practices teachers will adopt when pressure to increase test scores surfaces.

As shown in Pedulla et al. (2003) and Nolen et al. (1992), there exists a relationship between pressure and the use of test preparation practices. Due to heightened levels of accountability associated with the standards-based reform movement, it is increasingly important that the scores from achievement tests accurately represent the achievement of students. The introduction of construct irrelevant-variance leads to mismeasurement; therefore all test preparation practices should be studied with respect to the introduction of construct-irrelevant variance. Haladyna and Downing (2004) assert that more research on unethical testing practices is needed due to the variation in test preparation practices among schools. Past research has cited variation among schools in their testing practices but failed to clearly outline what school-level or teacher-level characteristics were related to the particular testing practices (Diamond & Spillane, 2004). The current study compares the types of test preparation practices and their frequency of occurrence across various teacher-level and school-level characteristics.

### **Purpose of the Study**

The purpose of this study was to examine how testing practices at the high school level are impacted by attaching school-level accountability consequences to the scores from a long-standing, low-stakes, state-wide testing program. The specific questions addressed are as follows:

1. Do teachers in schools that serve generally low-, moderate-, or high-achieving students have similar views regarding the ethicality of the test preparation activities? Do teachers from these three types of schools use particular test preparation activities to the same extent? How often do teachers use test preparation activities that they believe are unethical?
2. How is the use of particular test preparation activities related to the following factors: a) participation in checking the alignment between their district's content standards and the content covered by the test being used for school-level accountability, b) the extent of pressure felt to increase student test scores, c) belief that their school focuses more on increasing student scores than on improving student learning, and d) the content area being taught?
3. Has the amount of time spent on test preparation this year changed compared to the amount of time spent the previous year? Are there specific subgroups of students being targeted for special assistance in test preparation? If so, which subgroups are most often targeted for assistance?

4. What types of motivational activities and/or incentives are being used by schools and teachers related to student scores on the test being used for school-level accountability? How do these practices compare across schools serving generally low-, moderate-, or high-achieving students?

### **Context of the Study**

Historically, the *Iowa Tests of Basic Skills (ITBS)* and *Iowa Tests of Educational Development (ITED)* have been used voluntarily by Iowa schools primarily to obtain information for supporting instructional decisions. In nearly every Iowa school, the “stakes” associated with the outcome of administering the *ITBS* and *ITED* were low—for students, teachers, and administrators. However, this is no longer true. The stakes associated with the use of these tests have been incrementally increased as a result of the national accountability movement, which resulted in the passing of state and federal legislation (1994 Elementary and Secondary Education Act, Chapter 12 of the Iowa Code, and the 2002 No Child Left Behind Act) that attached “consequences” to achievement test scores. Although the use of the *ITBS* and *ITED* are not specifically mandated via state legislation, it is the *expectation* that all districts will administer the *ITBS* and *ITED* in order to comply with the Iowa accountability plan for *No Child Left Behind* (NCLB).

It is important to note that although Iowa schools have a common measure by which to evaluate student achievement (i.e., the *ITBS/ITED*), there is not a common set of standards or curriculum that must be followed by each school. Instead of having state mandated standards and curriculum, each school district has been given the authority to determine how best to serve its students. In the context of this local control, establishing the extent of alignment between standards and accountability measures—a requirement of NCLB—had to be completed separately for each of the roughly 370 school districts during the 2002-03 academic year. To accomplish this requirement, training was provided to educators from each school district as how to formally check the alignment between its content standards and assessment system using a common set of criteria to evaluate the sufficiency of this alignment. (See <http://projects.education.uiowa.edu/itap> for details on this training.)

Although most districts used the alignment checking procedure modeled during this training, they differed greatly in terms of the extent of teacher involvement in the process. For example, some schools had all teachers participate in checking alignment whereas in other



districts, only school administrators were involved with the process. However, prior to this “formal” alignment checking training, teachers were long encouraged to review the content and skills measured by the *ITBS/ITED* so as to use assessment data in instructional decision making (i.e., the primary purpose of these achievement test batteries). Given these opportunities, teachers have been provided with ample exposure to the content and skills measured by the assessment used for school-level accountability purposes — more exposure than that afforded to teachers in most, if not all, other states.

Given the significant change in the way scores from the *ITBS* and *ITED* are being used, the Iowa Department of Education (DE) has made a commitment to monitoring the impact of the NCLB legislation on Iowa schools. The Iowa DE has contracted with the Center for Evaluation and Assessment at the University of Iowa to conduct a statewide study, called the “Iowa Accountability Research Study,” to examine the effects of the *No Child Left Behind* legislation on teaching and testing practices in Iowa. The primary focus of the study is the early detection of consequences (positive and negative) resulting from the use of the *ITBS/ITED* for accountability purposes. The research results also are essential for making accurate and realistic district-level and state-wide interpretations of NCLB assessment information.

The design for the full study involves sampling schools at three different grade spans (elementary schools, middle schools/junior highs, and high schools). In addition to asking all of the teaching and administrative staff within a school to complete a questionnaire, participating schools are to administer one or two additional achievement tests to their 4<sup>th</sup>-, 8<sup>th</sup>-, or 11<sup>th</sup>-grade students. Baseline data is being collected from teachers and administrators throughout this academic year (i.e., 2004-2005) using questionnaires, with the following years being committed to focus group interviews and additional target questionnaires to obtain a more complete understanding of a) the nature of teacher practices, b) how these practices have been impacted by NCLB, and c) how these practices have changed over time.

The questionnaires are to be completed shortly after the school administers the *ITBS/ITED* this academic year. Because schools are able to administer the tests during the fall, winter, or spring, the data collected to date has been for only a portion of participating schools. Of the three types of schools (i.e., elementary schools, middle schools/junior highs, high schools), the high schools were the most complete sample of schools available because nearly 75% of the high schools in the state administer the *ITED* during the fall, compared to only 50%

of the elementary schools that administer the *ITBS* during the fall. Thus, for the purposes of this paper, the results from a sample of high schools that tested during the fall are being used. Once information has been obtained from the complete sample of schools, the analyses will be repeated and comparisons will be made across the three types of schools.

## Method

### Sample

A representative sample of Iowa public high schools was selected for participation in this study by using a stratified random sampling scheme to take into consideration both the size of the school and the overall level of achievement for students within the school. To classify all Iowa public high schools according to size (excluding those schools serving exclusively special student populations), data from the Basic Educational Data Survey (BEDS) report from the 2003-04 academic year was used to determine the smallest 25% of schools, the middle 50% of schools, and the largest 25% of schools based on the number of students enrolled at grade 11. The overall achievement level of each school was defined using its 2003-04 performance on the *ITED*. Median percentile ranks based on national student norms (NPRs) corresponding to the *Core Total (CT)* score were computed for each grade level (i.e., 9-12) in the school that took the *ITED*. (The CT is a composite score based on the Reading, Language, and Mathematics tests.) A median CT score was then calculated for each school based on all grade levels within the school that took the *ITED* during the 2003-04 academic year. The schools were then rank ordered based on these median CT scores. The lowest 25% of schools were classified as “low,” the middle 50% of schools were classified as “moderate,” and the highest 25% of schools were classified as “high.” The sampling procedure involved randomly selecting schools within each of the nine cells (three levels of achievement by three levels of size) and then contacting the school to determine its interest in participating in the study. Random selection within each cell was repeated until the required number of schools was obtained. The sample of 24 schools presented in Table 1 reflects about 50% of the total number of high schools that are participating in this study during this academic year. As can be seen in Table 1, the resulting sample of schools is very similar to all Iowa public high schools in terms of socio-economic status (as measured by percent eligibility for free or reduced lunch) and overall achievement (as measured by the median CT).

**Table 1.**  
**Description of Schools**

	School Achievement Level <sup>a</sup>			Total
	Low	Moderate	High	
Number of Schools				
Sample	6	12	6	24
Population	91	181	89	361
Median % Fr/Red Lunch Elig.				
Sample	27.1	21.9	14.7	22.0
Population	30.2	22.1	16.2	22.5
Median CT (NPR)				
Sample	51.3	65.2	76.3	65.2
Population	53.3	64.0	73.0	64.0

<sup>a</sup> The ranges of median CT for each school achievement level are as follows:  
Low = 39 to 57.5, Moderate = 57.8 to 69.0, and High = 69.3 to 88.0

Within each school, all teachers were asked to complete the Teacher Questionnaire, regardless of teaching assignment. A description of how these teachers were distributed across subject areas and years of experience have been provided in Table 2 for the total group of teachers, as well as for teachers within each of the three subgroups based on school achievement level. As can be seen in the last row of the table, the distribution of teachers across these three subgroups approximated the percentage of schools represented by each type of achievement level (i.e., 25%, 50%, 25%). Inspection of the percentages of teachers who taught various subject areas across the three different types of schools indicates that the subject areas are similarly represented in each of the three subgroups. The three subgroups of teachers also had very similar distributions of years of experience.

**Table 2.**  
**Description of Teachers**

	School Achievement Level						Total N
	Low		Moderate		High		
	N	%	N	%	N	%	
Subject Area Taught <sup>a</sup>							
English/Lang. Arts	30	13.7	69	15.6	39	13.8	138
Mathematics	28	12.8	53	12.0	43	15.2	124
Science	24	11.0	52	11.7	41	14.5	117
Social Studies	21	9.6	49	11.1	31	11.0	101
Fine Arts/Foreign Lang.	38	13.4	77	17.4	52	18.4	167
Vocational	41	18.7	85	19.2	29	10.3	155
Other	17	7.8	41	9.3	13	4.6	71
Special Needs <sup>b</sup>	49	22.4	89	20.1	69	24.5	207
Years Experience <sup>c</sup>							
1 to 5	50	22.8	96	21.7	55	19.5	201
6 to 10	32	14.6	71	16.0	47	16.7	150
11 to 20	60	27.4	109	24.6	93	33.0	262
21 to 30	58	26.5	113	25.5	58	20.6	229
> 30	19	8.7	54	12.2	29	10.3	102
Total Number of Teachers	219	22.4	443	45.3	282	28.9	977

<sup>a</sup> Some teachers teach multiple subject areas, thus the percentages do not sum to 100%.

<sup>b</sup> Teachers identifying teaching special education, resource/remedial, at-risk, and/or talented and gifted students exclusively.

<sup>c</sup> Missing values for 33 teachers.

## Questionnaire Development

The development of the questionnaire was based on information from a thorough review of the related literature and collaboration with the Iowa DE. During questionnaire construction, care was taken to refrain from using jargon in hopes of increasing the teachers' understanding of the questions. To check on readability, the questionnaire was piloted with a small group of teachers and reviews were solicited from several measurement specialists. The full questionnaire contained questions organized by the following five sections: 1) teacher background information, 2) instructional practices, 3) testing practices, 4) professional development and resources, and 5) perceptions regarding the impact of NCLB. For this paper, responses to selected questions from the sections covering background information, testing practices, and perceptions have been used. The sections of primary focus contained questions related to the following variables: a) perceived ethics of test preparation activities, b) use (frequency and timing) of test preparation activities, c) student populations targeted for test preparation activities, and d) motivational activities/incentives. A copy of the corresponding sections of the questionnaire can be found in Appendix A.

## Procedures

Schools participating in the study were committed to administering one or two additional tests to the 11<sup>th</sup>-grade students, in addition to asking all of the teaching to complete the questionnaire. The data collection procedures were designed so that students were to take the additional test within two weeks of taking the operational version of the *ITED* (i.e., the test being used for accountability purposes at grade 11). In addition, teachers were to complete the questionnaire shortly after administering the *ITED* so that they could more easily recall the types of activities used with their students in preparation for taking these tests. The typical amount of time needed to complete the questionnaire was 30 minutes.

Once a teacher completed the questionnaire, he or she was to seal it in an envelope and return it to the building administrator (who returned the complete set) or to mail it directly to the researchers. Teachers and administrators were aware that if at least 90% of the teachers in the school submitted completed questionnaires, their school would receive additional compensation for participating in the research study (i.e., beyond what they were receiving for the additional testing). The percentage of teachers within a school that returned completed questionnaires

ranged from 66% to 100%, with the median across the full sample of 24 schools being about 96%. As can be seen by the rates presented in Table 3, the three subsamples of schools had similar response rates.

**Table 3.**  
**Questionnaire Completion Rates:**  
**Percentage of Teachers within a School who Completed the Questionnaire**

	School Achievement Level			Total
	Low	Moderate	High	
Median	98.8%	95.7%	89.1%	95.7%
Range	65.9 to 100%	83.6 to 100%	68.7 to 100%	65.9 to 100%

Responses to each questionnaire were entered into a database and each response was then verified by two people for accuracy. Responses to the open-ended questions were then extracted into a spreadsheet so that codes could be assigned to each response. The coding was completed by four members of the research team, with each response being independently coded by two researchers. Comparisons of the codes were then made, and, in cases where there was not perfect agreement, a consensus process was used to determine a final code. The portion of the codebook used for the responses corresponding to the relevant sections of the questionnaire can be found in Appendix B.

### Data Analysis

For the purpose of this paper it is assumed that the responses from teachers within a particular school are not independent. For example, a school's climate is likely to directly impact the extent of pressure a teacher might feel to increase the scores of his or her students, and many schools have promoted the use of building-wide initiatives in response to NCLB. Therefore, for most of the analyses the *school* was used as the unit of analysis instead of the *teacher*. Results from the questionnaires were analyzed separately for each school and medians (Mdn) are reported for the combination of schools. In most cases the scales being used are ordinal in nature, there is evidence indicating that the distribution functions are not normal, and/or there were extreme outliers due to the small number of teachers in some of the schools. Thus, when significance tests were called for, the following nonparametric techniques were used: the Kruskal-Wallis and Sign Test. Analyses were first made to determine if the teacher responses were similar across all three types of schools (i.e., low, moderate, or high). If the

responses were sufficiently similar, follow-up analyses were made based upon the full sample of 24 schools.

Analyses were based on a reduced sample of 864 teachers who responded to the complete set of questions pertaining to the legitimacy of test preparation activities and their use of these activities, instead of the total number of teachers who participated in the study ( $n = 977$ ). Before excluding cases with missing data, the characteristics of the teachers who omitted responses were analyzed to ensure that the sample remained representative. The analysis indicated that teachers who omitted items tended to be from high-achieving schools at a slightly higher rate (12%) than low- or moderate-achieving schools (8% and 7%, respectively). There appeared to be no differences in the number of years teaching for those who omitted responses compared to those who provided complete responses to all of the test preparation items. Finally, there were only minor differences in terms of the teacher's subject area between the group of teachers with complete responses and the group with incomplete responses. Specifically, there were slightly larger numbers of fine arts and foreign language teachers who omitted responses compared to teachers responsible for other subject areas. Despite these differences, it is reasoned that the remaining sample based on a complete set of responses (i.e., 88% of the total group) remained representative.

## **Results**

### **Legitimacy and Use of Test Preparation Activities**

#### Ethicality of Test Preparation Activities

In order to determine if teachers in schools that serve generally low-, moderate-, or high-achieving students have similar views regarding the ethicality of particular test preparation activities, teachers were asked to rate nine test preparation activities in terms of their personal belief regarding the ethicality of the practice using a 5-point scale, where 1 = "very ethical" and 5 = "not at all ethical." Within each school, the median rating across all teachers was calculated for each activity. To determine if the sets of median ratings from low-, moderate-, and high-achieving schools were similar, the Kruskal-Wallis test was used. The results for these nine significance tests have been summarized in Table 4. As seen in the last column of the table, only two of the test preparation activities had significant differences among the three types of schools. The first activity in which there was a significant difference is the perceived legitimacy of the "use of practice tests within one month of testing" ( $p < .05$ ,  $\chi^2 = 7.271$ ). When the median

response for each type of school was compared for this activity, it could be seen that teachers in high-achieving schools were more likely to rate the activity as being less ethical ( $Mdn = 3.0$ ) compared to teachers in either moderate-, or low-achieving schools (both with  $Mdn = 2.0$ ). A significant difference of the ethicality rating was also found in the perceived legitimacy of “the use of the previous year’s *ITED* data to inform instruction” ( $p < .05$ ,  $\chi^2 = 6.531$ ). As with the use of practice tests, the use of the previous year’s *ITED* data was also rated as being *less* legitimate by teachers in high-achieving schools ( $Mdn = 1.8$ ) than teachers in low- or moderate-achieving schools (both with  $Mdn = 1.0$ ).

Table 4.

Comparison of Teacher Beliefs Regarding Ethicality of Test Preparation Practices by School Type			
Test Preparation Activity	df	$\chi^2$	<i>P-value</i>
Practice with <u>exactly the same form</u> of the <i>ITED</i> administered this year	2	4.023	.134
Practice with the <i>ITED</i> form used <u>last year</u>	2	2.090	.352
Routinely provide instruction <u>only</u> on the content areas tested on the <i>ITED</i>	2	2.881	.237
Routinely use classroom tests in the same format as the <i>ITED</i>	2	2.877	.237
Use <u>practice tests</u> within one month of testing	2	7.271	.026 *
Provide a refresher on content/skills areas within one month of testing	2	3.484	.175
Teach test-taking skills	2	3.000	.223
Use previous year’s <i>ITED</i> data to inform instruction	2	6.531	.038 *
Provide instruction without checking <i>ITED</i> test content	2	2.138	.343

Based on the analysis presented above, it appears that teachers in the three different types of schools generally tend to have similar beliefs regarding the ethicality of particular test preparation activities. Thus, it is reasonable to pool the results from all 24 schools when describing the typical beliefs for teachers regarding these practices. To do this, the median of the school median ratings for each test preparation activity was calculated. These median teacher ratings are presented in Table 5 along with the minimum and maximum school median ratings. As presented in Table 5, the median ethicality rating for each test preparation practice indicates that, in general, teachers are consistent with measurement professionals concerning the legitimacy/ethicality of the test preparation practices, particularly for activities towards the ethical side of the continuum. For example, “teaching test-taking skills” was viewed as “very ethical” and “practicing with exactly the same form” received the median rating closest to “not at all ethical” ( $Mdn = 4.0$ ). The rating for the latter activity (i.e., “practicing with exactly the same form”) was lower than expected because it was anticipated that teachers would be almost unanimous in their belief that the activity was “not at all ethical” (i.e., 5.0). As seen in the

second column of Table 5, the median for this activity is 4.0, which is similar to the rating obtained for “routinely providing instruction only on the content areas test on the *ITED*.” Upon further inspection it was found that two of the schools had a median rating of 3.0 (half-way between “very ethical” and “not at all ethical”) for “practicing with exactly the same form,” raising the suspicion that the teachers might have interpreted the statement in a manner other than had been intended.

**Table 5.**  
**Median Teacher Ratings Regarding the Ethicality of Test Preparation Activities**

Test Preparation Activity	Median	Min.	Max.
Practice with <u>exactly the same form</u> of the <i>ITED</i> administered this year	4.0	3.0	5.0
Practice with the <i>ITED</i> form used <u>last year</u>	3.0	2.0	4.0
Routinely provide instruction <u>only</u> on the content areas tested on the <i>ITED</i>	4.0	3.0	5.0
Routinely use classroom tests in the same format as the <i>ITED</i>	2.0	1.0	3.0
Use <u>practice tests</u> within one month of testing	2.0	1.0	3.0
Provide a refresher on content/skills areas within one month of testing	2.0	1.0	3.0
Teach test-taking skills	1.0	1.0	1.5
Use previous year’s <i>ITED</i> data to inform instruction	1.0	1.0	3.0
Provide instruction without checking <i>ITED</i> test content	2.0	1.0	3.0

<sup>a</sup> Ethicality scale ranges from 1 = “very ethical” to 5 = “not at all ethical”

### Use of Test Preparation Activities

In addition to examining the perceived ethicality of the test preparation activities, identifying which activities were being used most frequently and if the use of these practices varied across school type was examined. To do this, the percentage of teachers within a school who used a given test preparation activity was calculated and the Kruskal-Wallis test was used to determine if the sets of percentages for the three types of schools differed. The results of the Kruskal-Wallis test, as seen in Table 6, indicate that there were no significant differences in the use of the various test preparation practices across the different achievement levels of the schools. Thus, results from the 24 schools have been pooled together to describe the typical usage of these activities.



**Table 6.**  
**Comparison of Teacher Use of Test Preparation Activities by School Type**

Test Preparation Activity	df	$\chi^2$	<i>P</i> -value
Practice with <u>exactly the same form</u> of the <i>ITED</i> administered this year	2	.110	.946
Practice with the <i>ITED</i> form used <u>last year</u>	2	1.167	.558
Routinely provide instruction <u>only</u> on the content areas tested on the <i>ITED</i>	2	1.803	.406
Routinely use classroom tests in the same format as the <i>ITED</i>	2	.695	.706
Use <u>practice tests</u> within one month of testing	2	3.141	.208
Provide a refresher on content/skills areas within one month of testing	2	.822	.663
Teach test-taking skills	2	.705	.703
Use previous year's <i>ITED</i> data to inform instruction	2	2.938	.230
Provide instruction without checking <i>ITED</i> test content	2	.214	.899

The median percentages of teachers within a school who use a particular test preparation activity across all 24 schools are presented in Table 7, along with the minimum and maximum school-level percentages. Column two of Table 7 shows that the most commonly used test preparation activities were “providing instruction without checking *ITED* test content” (66.7%), “teaching test-taking skills” (58.1%), and “using the previous year’s *ITED* data to inform instruction” (53.3%). The least used test preparation activities included “practicing with exactly the same form of the *ITED* that was to be administered this year” (8.0%), “providing instruction only on the content areas tested on the *ITED*” (12.9%), and “practicing with last year’s form of the *ITED*” (16.3%). Although the median percentage of teachers within a school indicating that they use one of these three activities is quite low, it is still of concern when the range of the percentages is also analyzed. For example, in one school approximately 25% of the teachers indicated that they “practiced with exactly the same form of the *ITED*.” Likewise in at least one school, approximately 38% of the teachers indicated that they “practiced with last year’s form of the test.”

**Table 7.**  
**Percentage of Teachers within a School Using Various Test Preparation Activities**

Test Preparation Activity	Median	Min.	Max.
Practice with <u>exactly the same form</u> of the <i>ITED</i> administered this year	8.0	0.0	25.5
Practice with the <i>ITED</i> form used <u>last year</u>	16.3	0.0	38.5
Routinely provide instruction <u>only</u> on the content areas tested on the <i>ITED</i>	12.9	0.0	32.0
Routinely use classroom tests in the same format as the <i>ITED</i>	29.5	8.0	60.5
Use <u>practice tests</u> within one month of testing	20.8	0.0	56.6
Provide a refresher on content/skills areas within one month of testing	21.5	0.0	64.0
Teach test-taking skills	58.1	18.8	80.9
Use previous year's <i>ITED</i> data to inform instruction	53.3	4.0	72.0
Provide instruction without checking <i>ITED</i> test content	66.7	43.8	78.7

It should be noted that for Iowa schools practicing with last year's test form is equivalent to practicing with next year's test form because these parallel forms are administered in alternating years. In addition to compromising the security of next year's test, this practice is inappropriate because within a given test form the tests for adjacent grade levels contain a set of common items. Therefore, if a 9<sup>th</sup>-grade student practices using last year's 9<sup>th</sup>-grade level of the *ITED*, when this student is in 10<sup>th</sup>-grade he or she will take a test comprised of approximately 50% of the items they had practiced the previous year. It should also be noted that for most of the activities there is a large difference between the minimum and maximum usage by school. For example, considering the "use of previous year's *ITED* data to inform instruction," in one school only 4% of the teachers reported using the activity, compared to 72% of the teachers in another school reportedly using the previous year's *ITED* data to inform instruction.

### Use and Ethicality

Given the increased stakes associated with the *ITED* scores, there is reason to believe that some teachers might feel compelled to use particular test preparation activities in the desire to increase student scores, even though they believe these practices to be inappropriate and/or unethical. In order to determine the extent to which this phenomenon might be occurring, the distribution of ethicality ratings for those teachers indicating that they used the particular practice was examined. This comparison was made using the teacher as the unit of analysis, not the school. The results have been presented in Table 8.

**Table 8.**  
**Ethicality Rating by the Use of the Test Preparation Activity**

Test Preparation Activity	Teachers Using the Activity		Percentage <sup>a</sup> of Teachers by "Ethicality" Rating				
	N	%	Very Ethical			Not at all Ethical	
			1	2	3	4	5
Practice with <u>exactly the same form</u> of the <i>ITED</i> administered this year	95	11.0	25.3	18.9	26.3	12.6	16.8
Practice with the <i>ITED</i> form used <u>last year</u>	135	15.6	34.8	26.7	19.3	10.4	8.9
Routinely provide instruction <u>only</u> on the content areas tested on the <i>ITED</i>	125	14.5	16.8	16.0	31.2	19.2	16.8
Routinely use classroom tests in the same format as the <i>ITED</i>	294	34.0	47.3	29.3	16.7	5.8	1.0
Use <u>practice tests</u> within one month of testing	240	27.8	56.7	26.7	11.3	3.3	2.1
Provide a refresher on content/skills areas within 1 month of testing	220	25.5	52.7	24.6	16.8	4.5	1.4
Teach test-taking skills	494	57.2	76.7	14.8	6.7	1.2	0.6
Use previous year's <i>ITED</i> data to inform instruction	410	47.5	67.8	24.1	4.9	2.4	0.7
Provide instruction without checking <i>ITED</i> test content	570	66.0	53.9	20.7	18.6	4.0	2.8

<sup>a</sup> Based on number of teachers reportedly using the activity

<sup>b</sup> Based on the entire group of 864 teachers

For the most part, the data in Table 8 indicate that teachers are using test preparation activities that they believe are ethical (receiving either a 1 or 2). For example, of the 494 teachers (57% of all teachers) that teach test-taking skills, 96.5% (i.e., 76.7% and 14.8%) believe that the practice is ethical, compared to about 2% (1.2% and 0.6%) that indicated the activity was on the less ethical end of the scale. In contrast, the most common occurrence of teachers using practices that they noted as being less ethical was related to “routinely providing instruction only on the content areas tested on the *ITED*.” From the combination of the last two columns (i.e., ethicality ratings of 4 or 5) it can be determined that 36% of the teachers who used the activity reported that this activity was unethical, in contrast to 33% (16.8% and 16.0%) of the 125 teachers using this activity viewed it as ethical. Likewise, 29% (12.6% and 16.8%) of the 95 teachers practicing with exactly the same form of the test reported that they believed this practice to be unethical, contrasted with 44% (25.3% and 18.9%) believing it was ethical. Although there are some instances of teachers using test preparation practices they believe are unethical, it appears that generally teachers use test preparation practices that they deem to be ethical.

### **Factors Related to the Use of Test Preparation Activities**

Because the previous results indicate that test preparation activities are being utilized by a large number of teachers, it is beneficial to determine if there are any particular factors that are related to the use of these activities. The analyses in the following section are aimed at investigating the differences between teachers who use a particular activity compared to teachers who do not use the activity. Specific analyses conducted include the relationship between the use of test preparation activities and participation in alignment checking, the extent of pressure felt to increase student test scores, the teacher’s belief that their school is more interested in focusing on student scores than on improving overall student learning, and the content area for which the teacher is responsible.

Investigating the relationship between the use of particular test preparation activities and the four factors (i.e., alignment checking, pressure, school focus, and content area) was complicated by the fact that the group of teachers who use a particular activity cannot be viewed as being independent of the group of teachers not using the activity. Teachers are nested within a school, thus the set of teachers within a particular school who use an activity are not independent of the set of teachers within that school who do not use a particular activity. Therefore, the Sign Test was used as a significance test based on the school being the unit of analysis to determine if

the differences, for example, in the percentage of teachers who believed their school was more interested in increased test scores was systematically higher or lower for the subgroup of teachers who used the test preparation activity compared to the subgroup of teachers who did not use the activity.

### Alignment

Teachers were asked if they had ever taken part in checking the alignment (formally or informally) between their district’s content standards and the content and skills covered by the *ITED*. Within each school, two percentages of teachers who had participated in alignment checking were calculated – one based on the subgroup of teachers who used the particular test preparation activity and one based on the subgroup of teacher who did not use the activity. The Sign Test was applied to determine if there was a systematic difference between the two sets of percentages across the 24 schools. The results of the Sign Tests, presented in Table 9, indicate that there was a significant difference for only one of the test preparation activities – teachers who used the previous year’s *ITED* data to inform their instruction. Teachers who have participated in this practice were more likely to have participated in alignment checking than teachers who did not use the previous year’s *ITED* data to inform their instruction ( $p < .01$ ). Further analysis of the data reveal that of the teachers within the school who used the previous year’s *ITED* data to inform instruction, typically 71.4% had participated in alignment checking, compared to 50% of teachers who participated in alignment checking but did not use *ITED* data to inform instruction.

**Table 9.**  
**Comparison of Participation in Alignment Checking by Test Preparation Use**

Test Preparation Activity	N <sup>a</sup>	Number of Differences		P-value (2-tailed)
		Positive <sup>b</sup>	Negative <sup>c</sup>	
Practice with <u>exactly the same form</u> of the <i>ITED</i> administered this year	20	12	6	.238
Practice with the <i>ITED</i> form used <u>last year</u>	21	9	10	1.000
Routinely provide instruction <u>only</u> on the content areas tested on the <i>ITED</i>	23	10	11	1.000
Routinely use classroom tests in the same format as the <i>ITED</i>	24	15	7	.134
Use <u>practice tests</u> within one month of testing	23	15	6	.078
Provide a refresher on content/skills areas within one month of testing	23	12	8	.503
Teach test-taking skills	24	15	6	.078
Use previous year’s <i>ITED</i> data to inform instruction	24	18	4	.004 *
Provide instruction without checking <i>ITED</i> test content	23	14	6	.115

<sup>a</sup> For some practices  $N < 24$  because no teacher within the school reported using the activity.

<sup>b</sup> Teachers using the practice are more likely to have participated in alignment checking than those who did not.

<sup>c</sup> Teachers not using the practice are more likely to have participated in alignment checking than teachers who did.

### Pressure to Increase Test Scores

Because of the increased stakes associated with scores from the *ITED*, it was desirable to detect if there was a relationship between the use of various test preparation activities and the amount of pressure teachers feel to increase test scores. The teachers were asked to indicate the extent of pressure they feel to increase test scores from each of seven sources (i.e., self, colleagues, administration, school board, parent, general public/media, and government) using a 3-point scale (0 = “Not at all”, 1 = “A little”, and 2 = “A lot”). Due to the scale being ordinal in nature, a general indicator of the extent of pressure was based on the number of sources from which a teacher felt *at least* “a little” pressure. Then, within each school, the median number of sources from which the teachers feel pressure was calculated based on the two subgroups of teachers (i.e., those who use the activity and those who do not). The Sign Test was used to compare the median number of sources of pressure teachers across the 24 schools. The results for these nine significance tests are presented in Table 10.

**Table 10.**  
**Comparison of Amount of Pressure to Increase Scores by Test Preparation Use**

Test Preparation Activity	N <sup>a</sup>	Number of Differences		P-value (2-tailed)
		Positive <sup>b</sup>	Negative <sup>c</sup>	
Practice with <u>exactly the same form</u> of the <i>ITED</i> administered this year	20	9	6	.607
Practice with the <i>ITED</i> form used <u>last year</u>	21	5	10	.302
Routinely provide instruction <u>only</u> on the content areas tested on the <i>ITED</i>	23	12	6	.238
Routinely use classroom tests in the same format as the <i>ITED</i>	24	16	3	.004 *
Use <u>practice tests</u> within one month of testing	23	15	3	.008 *
Provide a refresher on content/skills areas within one month of testing	23	16	3	.004 *
Teach test-taking skills	24	16	4	.012 *
Use previous year’s <i>ITED</i> data to inform instruction	24	15	3	.008 *
Provide instruction without checking <i>ITED</i> test content	23	12	10	.832

<sup>a</sup> For some practices N < 24 because no teacher within the school reported using the activity.

<sup>b</sup> Teachers using the practice are more likely to feel pressure to increase test scores from a greater number of sources than those who did not use the practice.

<sup>c</sup> Teachers not using the practice are more likely to feel pressure to increase test scores from a greater number of sources than those who did use the practice.

As can be seen in Table 10, there were significant differences for five of the activities. The activities were as follows: a) “routinely using classroom tests that are in the same format of the *ITED*” ( $p < .01$ ), b) “use practice tests within one month of testing” ( $p < .01$ ), c) “provide refreshers on the content/skills areas on the *ITED* within one month of testing” ( $p < .01$ ), d) “teaching test-taking skills” ( $p < .05$ ), and e) “use the previous year’s *ITED* data to inform

instruction” ( $p < .01$ ). For each of these activities, teachers who used the activity were more likely to feel pressure from a greater number of sources than teachers who did not use the activity.

For each of the five practices mentioned above, the median number of sources of pressure within each school was calculated for those who reported using the activity and those who did not use the activity. In each case, teachers who use the activity (e.g., teach test-taking skills) reported feeling pressure to increase test scores from typically six different sources, whereas teachers who do not use the activity reported feeling pressure from typically five different sources.

### School Focus

Another factor potentially related to the use of test preparation activities is the school’s climate, as measured by teacher perceptions that the school is more interested in increasing test scores than improving overall student learning. Teachers were asked to indicate whether their school was more interested in increasing test scores or more interested in improving overall student learning. The percentage of teachers within each school indicating that their school was more interested in increasing test scores was calculated based on the two subgroups of teachers – those who used the activity and those who do not use the activity. The Sign Test was again used to determine if the sets of percentages across the 24 schools differed systematically. The results of the Sign Test, as seen below in Table 11, indicate that using the previous year’s *ITED* data to inform instruction was the only activity significantly different for the two subgroups of teachers ( $p < .05$ ). The Sign Test showed that teachers who used the previous year’s *ITED* data to inform their instruction were *less* likely to say that their school focuses more on increasing student test scores than on improving overall student learning than teachers who did not use the previous year’s *ITED* data to inform their instruction ( $Mdn = 33\%, 48\%$ , respectively).

**Table 11.**  
**Comparison of Belief that School Focus is on Increasing Test Scores by Test Preparation Use.**

Test Preparation Activity	N <sup>a</sup>	Number of Differences		P-value (2-tailed)
		Positive <sup>b</sup>	Negative <sup>c</sup>	
Practice with <u>exactly the same form</u> of the <i>ITED</i> administered this year	20	10	9	1.000
Practice with the <i>ITED</i> form used <u>last year</u>	21	11	9	.824
Routinely provide instruction <u>only</u> on the content areas tested on the <i>ITED</i>	23	11	11	1.000
Routinely use classroom tests in the same format as the <i>ITED</i>	24	9	13	.523
Use <u>practice tests</u> within one month of testing	23	9	13	.523
Provide a refresher on content/skills areas within one month of testing	23	8	14	.286
Teach test-taking skills	24	8	15	.210
Use previous year's <i>ITED</i> data to inform instruction	21	6	17	.035 *
Provide instruction without checking <i>ITED</i> test content	23	12	10	.832

<sup>a</sup> For some practices N < 24 because no teacher within the school reported using the activity.

<sup>b</sup> Teachers using the practice are more likely to believe that their school focuses more on test scores than teachers who did not use the practice.

<sup>c</sup> Teachers not using the practice are more likely to believe that their school focuses more on test scores than teachers who did use the practice.

### Content Area

In addition to analyzing differences in the use of test preparation practices for all teachers within a school, it was also desired to detect if there were differences in the use of these activities among teachers who teach subjects or special student populations that would be directly involved in *ITED* testing. These areas include English/Language Arts (ELA), Mathematics, Social Studies, and Science. In addition, teachers of Special Needs Students (including Special Education, At-Risk, and English Language Learners) were also included because they are typically responsible for providing or reinforcing instruction in the core curricular areas. This collection of teachers was further divided to detect differences in the use of test preparation practices by teachers who are subject to federal accountability requirements (i.e., ELA, Mathematics, and Special Needs) and teachers responsible for tested content that is not subject to federal accountability requirements (i.e., Social Studies and Science). These two groups are referred to here as the “accountability” and “non-accountability” groups. The reason for making the distinction between these two areas (i.e., accountability vs. non-accountability) is to determine if there are any differences in the use of test preparation activities for teachers who are responsible for teaching a tested content area/student population, but where the stakes associated with the test scores are different.

The two subgroups – accountability and non-accountability – were formed using information regarding the subject area and student level for which the teacher reported being responsible. In the case of multiple responses (e.g., Math and Science), if at least one of the subject areas was Math, ELA, or Special Needs, the teacher was assigned to the accountability group. Then for each of the two subgroups, the percentage of teachers using a particular test preparation practice was calculated. As with the previous analyses, because teachers are nested within school buildings, the Sign Test was used to detect differences in the use of the test preparation activities for the accountability versus the non-accountability groups of teachers across the sample of 24 schools. The results of the Sign Test, as seen in Table 12, show that there was only one practice, “providing a refresher on tested content areas,” for which there was a significant difference in the percentage of teachers using the activity between the accountability and the non-accountability subject areas. Teachers responsible for an accountability subject area were more likely than non-accountability content area teachers to provide a refresher on tested content areas ( $p < .05$ ). Additional comparisons of the proportions found that 41.4% of the teachers in the accountability content area provided refreshers while only 21.1% of the non-accountability content area teachers used the activity.

**Table 12.**  
**Comparison of Test Preparation Use by Content Area (i.e., accountability vs. non-accountability).**

Test Preparation Activity	N	Number of Differences		P-value (2-tailed)
		Positive <sup>a</sup>	Negative <sup>b</sup>	
Practice with <u>exactly the same form</u> of the <i>ITED</i> administered this year	24	13	7	.263
Practice with the <i>ITED</i> form used <u>last year</u>	24	14	7	.189
Routinely provide instruction <u>only</u> on the content areas tested on the <i>ITED</i>	24	13	9	.523
Routinely use classroom tests in the same format as the <i>ITED</i>	24	13	10	.678
Use <u>practice tests</u> within one month of testing	24	13	8	.383
Provide a refresher on content/skills areas within one month of testing	24	17	5	.017 *
Teach test-taking skills	24	17	7	.064
Use previous year's <i>ITED</i> data to inform instruction	24	15	8	.210
Provide instruction without checking <i>ITED</i> test content	24	13	10	.678

<sup>a</sup> Accountability group is more likely to use the test preparation activity than the non-accountability group.

<sup>b</sup> Non-accountability group is more likely to use the test preparation activity than the accountability group.



## **Conducting of Test Preparation Activities**

The prior analyses have described the types of test preparation practices being employed by teachers and identified that there are particular factors, such as the amount of pressure or stakes associated with the content area, which are associated with the use of particular test preparation activities. The following analyses are aimed at understanding if there have been any changes in the time spent on test preparation and which student populations the activities affect.

To determine if the amount of time spent on test preparation activities has changed compared to last year, teachers were asked to indicate the extent to which they believed that the time spent on test preparation had changed since the previous year. The median response across all schools, using the school as the unit of analysis, indicated that generally there had been “no change” in the amount of time devoted to test preparation compared to last year. However, when analyzing the frequency of teacher responses presented regardless of school only about 32% of the teachers indicated that there has been no change in the amount of time spent on test preparation activities. A total of 29.1% of teachers reported some type of increase, either significant (8.6%) or slight (20.5%), and only 3% of the teachers reported a decrease (slight or significant). With the large number of “don’t know” responses (36.7%), it is difficult to determine the exact trend. However, teachers who responded in such a manner were largely from fine-arts, foreign language, and vocational areas, which are content areas not measured by the *ITED*.

Teachers were also asked if they targeted any subgroups of students, including English Language Learners and Special Education students, for special assistance in test preparation. Their responses clearly indicate that no specific subgroups of students are being heavily targeted. The largest subgroup reported as being targeted for special assistance in test preparation was special education students (18.8%), with borderline non-proficient and English Language Learners students being reportedly targeted to a lesser extent (9.6% and 8.8%, respectively).

## **Motivational Activities and Incentives**

In addition to investigating activities occurring within the classroom setting, it was also desirable to detect the occurrence of any school-wide activities that are related to testing. In particular, the types of activities of interest include motivational activities used prior to testing, teacher incentives related to students’ scores on the *ITED*, as well as student incentives related to *ITED* performance.

### Motivational Activities

Teachers were asked to describe any special activities related to testing that take place prior to the administration of the *ITED*. Once coded, these responses were summarized for each school. The median percentages of teachers using a given activity across all schools and within each of the three school types (i.e., low-, moderate-, and high-achieving) have been presented in Table 13, along with the minimum and maximum school-level percentages. Based on these data, where the median percentage of teachers within a school reporting using a particular type of motivational activity is at or close to 0%, it appears that there are not any motivational activities that are being widely used throughout the schools in this sample. Only four schools had at least 20% of its teachers indicating that a special activity was used. Of these four schools, three are lower-achieving schools and one is a higher-achieving school. One of the lower-achieving schools had 26% of its teachers report that they engage students in exercises before testing, as well as providing a healthy snack during testing (reported by 30.5% of the teachers). A second lower-achieving school also reported offering fruit and other snacks during testing (reported by 32% of the teachers). The third lower-achieving school reported the use of “test talks” by 44% of the teachers where the test was discussed among small groups of students. These pep talks were described as discussing the importance of the test as well as making personal goals for each student’s performance on the test. In the high-achieving school that reported the use of motivational activities, 47% of the teachers reported having students exercise by stretching before testing or between tests and providing a snack including fruit and water to the students. In all, there does not appear to be widespread use of any particular type of motivational activity across the sample of 24 high schools.

**Table 13.**  
**Percentage of Teachers within a School Reporting Use of Motivational Activities**

Type of Motivational Activity	Type of School						All Schools	
	Low		Moderate		High		Mdn	Range
	Mdn	Range	Mdn	Range	Mdn	Range		
Breakfast	0	0 to 1.7	0	0 to 7.7	0	0 to 0	0	0 to 7.7
Snacks	5.0	0 to 32.0	0	0 to 12.1	0.6	0 to 47.1	0.6	0 to 47.1
Exercise	0	0 to 25.6	0	0 to 0	0	0 to 47.1	0	0 to 47.1
Conference w/Student	0	0 to 12.2	0	0 to 8.3	0	0 to 1.5	0	0 to 12.2
Test Talks	6.2	0 to 44.4	2.9	0 to 15.4	0	0 to 18.9	3.2	0 to 44.4
Posters	0	0 to 9.8	0	0 to 1.8	0	0 to 0	0	0 to 9.8
Letter to Parents	0	0 to 1.7	0	0 to 1.4	0	0 to 2.0	0	0 to 2.0

### Teacher Incentives

Teachers were also asked to describe any types of incentives they were offered (publicly or privately) that were related to their students' *ITED* scores. Their responses indicate that no school in this sample has implemented a system for rewarding teachers based on their students' test scores. At most, only one or two teachers indicated that some type of incentive was offered. These responses were primarily related to their personal desire to "not be on the list" of schools in need of assistance or the personal pride they take in their work.

### Student Incentives

Teachers were also asked to describe any incentives students were offered related to their individual or group performance on the *ITED*. Once coded, these responses were summarized for each school and then across schools within a given achievement level. The median percentage of teachers reporting the use of a particular type of student incentive are reported in Table 14 for each type of school and for the full sample. Unlike teacher incentives, there were a variety of student incentives that were listed as being used by schools in the sample, although there does not appear to be any specific type of student incentive that is pervasive throughout the entire sample. For example, although privileges such as open campus for lunch, free time, or a special non-academic activity were the more commonly cited student incentives by teachers in each of the three school subgroups, the median percentage of teachers citing these types of activities across the entire sample of 24 schools was just 7.4%.

**Table 14.**  
**Percentage of Teachers within a School Reporting Use of Student Incentives**

Type of Student Incentive	Type of School						All Schools	
	Low		Moderate		High		Mdn	Range
	Mdn	Range	Mdn	Range	Mdn	Range		
Monetary/Gifts	2.0	0 to 48.1	0	0 to 3.6	5.9	0 to 82.4	0	0 to 82.4
Privileges	29.8	0 to 85.2	7.4	0 to 155.4	12.9	1.5 to 82.4	7.4	0 to 155.4
Recognition	0	0 to 29.6	3.7	0 to 9.1	0.6	0 to 50.5	0.6	0 to 50.5
Treats	0	0 to 6.3	0	0 to 22.2	0	0 to 8.4	0	0 to 22.2
Warnings/Threats	0	0 to 14.8	0	0 to 14.3	0	0 to 3.7	0	0 to 14.8
Other	2.3	0 to 6.3	0	0 to 5.6	0.6	0 to 8.4	0	0 to 8.4

**Note** Percentages can be larger than 100% due to teachers identifying more than one incentive within a given category.

Within the individual schools, there does not appear to be a noticeable difference between the three types of schools in the types of incentives being offered to students related to their *ITED* performance. For example additional analyses indicated that 50% of the low-achieving schools, 50% of the moderate-achieving, and 67% of the high-achieving schools provided some type of incentive to their students. There also appeared to be little difference between the achievement level of the school and the types of activities that the schools were providing. Out of the 13 schools that provided some type of student incentives, all but two provided some type of privilege for the students as an incentive. The most common of the privileges included some type of “fun” activity such as a pizza party for score improvement. Three of the schools also incorporated a school specific privilege such as having lunch with the principal, parking spaces for upperclassmen, and having an unstructured study hall. Other privileges offered included being awarded extra credit that could be applied to a course grade, the opportunity for open-campus lunch, allowing students to leave school early, or a trip, which was typically to the bowling alley.

Additional incentives offered in the specific schools included monetary/gifts such as a gift card to a local bookstore or drawing for prizes for students who scored at or above the 90<sup>th</sup> percentile. Other incentives included treats such as cinnamon rolls and school wide recognition via awards or t-shirts. Very few schools claimed to employ warnings/threats, but those that were mentioned included requiring students to complete additional coursework (either content-related or test prep-related) based on their *ITED* performance.

## **Discussion**

### **Factors Related to the Use of Test Preparation Activities**

The present study has provided a detailed description of the test preparation activities that are being implemented in response to the current school-level accountability legislation. Whereas past research has described the use of test preparation activities by school achievement level (Taylor, Shepard, Finner, & Rosenthal, 2003) or stakes of accountability for teachers and students (Pedulla et al., 2003), the current study has been able to contend with a variety of teacher-level and school-level factors which may influence the use of particular test preparation practices in the current accountability environment. As such, we are better able to determine which of these factors significantly contributes to the use of test preparation practices. The

findings of the study are somewhat encouraging concerning the trends in the use of test preparation activities; however, some of the teacher responses have raised concerns in regards to teachers' perceptions of the ethicality of certain test preparation practices, the types of activities that are being used, and the types of motivational activities/incentives that are being used by schools.

Contrary to previous research in which the achievement level of the school influenced the use of test preparation activities (Taylor et al., 2003), the current study found no relationship between the achievement level of the school and the use of test preparation activities. The finding is of particular interest because follow-up analyses indicated that there was no relationship between the types of school (i.e., low-, moderate-, and high-achieving) and the number of sources of pressure from which teachers feel pressure to increase scores or the extent of pressure. It appears as if teachers in all schools – not just those serving students with lower achievement – are feeling pressure to increase test scores from a number of sources of pressure (e.g., school administration).

Although the achievement level of the school did not provide much information regarding the factors related to the use of test preparation activities, there were particular teacher-level factors that did contribute to the understanding of the use of test preparation activities. First, contrary to concerns that the participation in alignment checking may lead to “teaching the test” through providing instruction on only tested content thus contributing to artificial gains in test scores (Koretz et al., 2001), in this study the only test preparation activity related to participation in alignment checking was the use of the previous year's *ITED* data to inform instruction. This finding is understandable considering those who are likely to want to use data to inform their instruction would be interested in having a better understanding of what the test is measuring. Therefore, alignment checking, as operationalized by teachers in this study, does not appear to be related to teachers providing instruction related to *only* the content and skill areas being tested.

Similarly, a teacher's subject area also had a very small effect on the use of test preparation activities by teachers responsible for an area that is covered by the *ITED* (i.e., ELA, Math, Science, Social Studies). However, additional analyses are needed to determine if the ELA, Mathematics, or Special Education teachers within the accountability group or if the Science and Social Studies teachers in the non-accountability group differ with respect to their

test preparation practices. The disaggregation is particularly of interest because in the state of Iowa, schools are required to report the performance of their 11<sup>th</sup>-grade students in the area of Science, which may have contributed to the lack of differences in the use of test preparation activities between the accountability and non-accountability groups. In addition to detecting differences within the subject areas associated with the content areas being tested, it may also be desirable to include other subject areas such as foreign language, vocational, etc. It is possible that teachers in these areas have implemented test preparation activities as part of a school-wide initiative in support of increasing reading and mathematics performance despite not being directly responsible for teaching content areas covered by the *ITED*.

One teacher-level factor in which there was a relationship with the use of test preparation activities was the relationship between the use and perceived ethicality of test preparation. The results suggested that teachers use practices they deem as ethical. However, as the ethicality ratings indicate, teachers do not always possess views regarding the ethicality of certain test preparation practices that are congruent with those of measurement experts. For example, some teachers believed that practicing with “exactly the same form of the *ITED* to be administered this year” and “practice with the form of the *ITED* used last year” were ethical activities. Therefore, although teachers tend to use practices they perceive as ethical it is important to ensure that teachers do not have misconceptions of the ethicality of these practices.

The teacher-level information that was perhaps the most useful in understanding the use of test preparation activities was the relationship between school climate, as measured by the amount of pressure to increase test scores and the teachers’ perceptions of school focus. Consistent with past research, the amount of pressure a teacher feels to increase test scores does appear to contribute to the use of test preparation practices (Nolen et al., 1992; Pedulla et al., 2003). Thus far, the results were mildly encouraging because most of the significant differences related to pressure and use were for activities that were not on the least ethical end of the continuum. It was encouraging to observe that teachers who feel more sources of pressure have not yet resorted to using unethical activities such as practicing with either the current or last year’s form of the test. However, the practices will need to be closely monitored as schools begin to struggle to meet their Adequate Yearly Progress (AYP) goals. In contrast, the perception of school focus was quite encouraging because the only significant finding was for using the previous year’s data to inform instruction. Teachers who used the *ITED* scores as they

were originally intended (i.e., to inform instruction) were more likely to believe their school was more interested in overall student learning than simply increasing test scores.

### Trends in Test Preparation Practices

Overall, it appears as if teachers' perceptions, both of the ethicality of test preparation practices and the pressure to increase test scores, are most directly related to the use of test preparation activities compared to the other factors studied. Due to the limited impact of the other factors (e.g., achievement level of the school), it is necessary to analyze the overall trends in these practices in order to better understand test preparation activities. The trends are important so that one is able to have a holistic view of the state of test preparation practices as well as motivational activities/incentives. Doing so also allows for a better description of what activities are being used in the schools, what types of students are exposed to test preparation, and how the amount of time spent on test preparation has changed in the last year.

When considering the amount and duration of test preparation practices, there appears to be a slight increase in the time spent on test preparation activities compared to the 2003-04 school year (i.e., the first year schools knew how scores from the *ITED* would be used for accountability purposes.) Furthermore, teachers appear to be targeting all students with test preparation activities with no specific types of students being singled out for specialized activities. Concerning the types of activities being used, there was a trend in which teachers were using multiple test preparation activities instead of only one or two. Among the most commonly reported activity was the teaching of "test-taking skills" (Mdn = 58%) which is comparable to the frequency reported by both Pedulla et al. (2003), where between 54%-71% of the secondary teachers reported teaching test-taking skills, and Taylor et al. (2003), in which as many as 78% of the teachers in "excellent schools" taught test-taking skills. Although the teaching of test-taking skills has been a widely accepted practice (Haladyna et al., 1991; Kilian, 1992; Mehrens & Kaminski, 1989; Popham, 1991), one must question the practical value of other practices mentioned by teachers in the sample, particularly the "use of practice tests." Typically 20% of the teachers within a school in this study reported using practice tests, but up to 50% of teachers within a school have reported using these aids. The use of practice tests are often reported by teachers in other research (Herman & Golan, 1993; Pedulla et al., 2003; Taylor et al., 2003), but their use is of particular significance in the current context because most students in Iowa have been taking the *ITBS/ITED* since at least the third grade. One might

reasonably question the educational value of administering practice tests to help high school students gain familiarity with the format of a test that they have taken for seven to nine years.

Other trends identified by this study focused on the types of motivational activities and incentives being used in schools in an attempt to increase test scores. The teacher responses suggest that although there are a variety of motivational activities and student incentives being used in the schools, there are no systematic trends regarding the specific types of activities/incentives being used. However, the types of activities that are being used appear to vary in terms of their defensibility. Some practices appear to be appropriate (e.g., snacks before testing) while others might be considered more questionable (e.g., providing money to the class that has the largest increase in scores or awarding extra credit that can be applied to a course grade). Although it is important that students be motivated to do well on the tests in order for the scores from the test to more accurately reflect what the student knows and is able to do, increases in student motivation may interfere with the inferences that can be made based on the test scores regarding student achievement. For example, inferences regarding the student's status compared to a norm group may be affected if motivation is increased relative to the level exhibited by students in the norm group. Such a discrepancy is likely to yield an artificially higher status, reflecting greater motivation and not necessarily a higher standing in the norm group. Interpreting student growth over time, either longitudinal or cross-sectional, may also be affected if levels of motivation change from year to year. If special activities or incentives being employed are successful at increasing a student's motivation to do well on the tests, schools need to sustain these activities over time or incentives might need to be increased in order to sustain student motivation to perform well on the tests. For example, having a special assembly to motivate students to perform to the best of their ability on the *ITED* may be effective for the first year or two that it is implemented, but the effectiveness of the activity to motivate students may diminish over time. As a consequence, it is possible the scores may actually decline. In this scenario, it will be extremely difficult, if not impossible, for a school to know if the decrease in scores is attributable to a decrease in student motivation, student achievement, both motivation and achievement, or some other factors. The possibility of this phenomenon will be more closely analyzed by the larger accountability project of which the current research is a component.



## Limitations

The results of the study are helpful in understanding test preparation activities, but there are a few limitations associated with this study that should be mentioned. First, the data are self-report data. Although anonymity was promised to all participants, some of the information on the survey may be of a highly sensitive nature (e.g., practicing with exactly the same form of the *ITED* that is to be administered this year). As such, it is possible that the occurrence of test preparation activities in schools, particularly those that are unethical, may be underestimated.

The context of the study might be considered both an advantage and a potential weakness of the study. By having only Iowa schools included in the sample, we have been able to detect how changing the way scores from a long standing testing program are used (i.e., from low-stakes to high-stakes for schools) have impacted test preparation practices. In other states, however, there is frequent change in the testing program being utilized. In these settings, teachers may be inclined to engage in more (and different) test preparation activities than teachers in this study because of the lack of student and/or teacher familiarity with the test's content and format. Thus, results from this study may not generalize to other states due to the consistency of the testing program. In addition, although there is considerable variability across schools in terms of the achievement levels of the students being served, the overall achievement level of Iowa students is quite high when compared to students nationally. Also, there is little racial/ethnic diversity in Iowa, which may affect the types of students who are targeted for test preparation. Finally, the context of accountability may be different in Iowa than in other states. In Iowa, accountability is school-level, not student- or teacher-level, making the results ungeneralizable to a different accountability context. Overall, future research may need to consider other geographic and demographic areas to determine the extent to which the trends presented in the current study can be generalized to a broader context.

A final limitation of the study is the limited sample size – 24 schools. Although the sample is representative of Iowa schools in regards to achievement level and socioeconomic status, a better description of the trends in test preparation practices may be seen if more schools were included in the study. For example, schools that test in the spring may have more of an opportunity to institute special activities or incentives compared to schools that tested in early fall. Also, the limited sample size affects the power of the statistical tests. It is for these reasons

this paper will be revised once data has been collected for the complete sample of 48 high schools.

### **Implications for Future Research and Professional Development**

Despite these limitations, we believed that the results of the study provide a useful examination of the impact of school-level accountability on local testing practices. Future research should expand the scope of the study to include schools with varied demographics as well as different grade levels to determine what test preparation activities are being used in different contexts. Also, an analysis of the long-term impact of the use of motivational activities and special incentives is needed. Specifically a comparison of achievement trend data for schools with teachers reporting extensive use of test preparation activities and/or motivational activities/student incentives compared to schools without this focus. As such, the impact of test preparation practices on student scores could be better analyzed. Finally, investigation into effective ways to help teachers gain knowledge about which test preparation practices are appropriate and which practices should be avoided needs to take place because the use of some of practices is likely to corrupt the meaning of the scores.

Results from this study also indicate that there is a great need for professional development in the area of test preparation. Approximately 20% of the teachers sampled believed that practicing with exactly the same form of the *ITED* that is to be administered this year was ethical, which was alarming. It is possible that the teachers misinterpreted the statement. However, if they did interpret the statement as intended, efforts need to be made to help teachers understand why practicing with the same form of the test to be administered should never be considered as being defensible. Similarly, there were large numbers of teachers within a school that reported practicing with current forms of the *ITED*: 25.5% of teachers within a school reported using the same form and 38.5% reported using the previous year's form. The prevalence of the activities in particular schools suggests that professional development is needed to help teachers and administrators understand why these practices should not be used. Likewise, for certain ethical forms of test preparation, professional development may be needed to encourage such behavior. For instance, in some schools only 4% of the teachers reported using the scores from the *ITED* to inform their instruction. Educators in schools such as these should understand the possible informative uses of the *ITED* as a tool instead of viewing the *ITED* as an inconvenience.

In addition to correcting misconceptions regarding test preparation activities, it is also important to obtain information through interviews with the teachers to determine how they interpreted the meanings of these test preparation activities as they were presented in the questionnaire. If teachers interpreted activities to be something other than what was intended, the types of inferences that can be made based on their responses are greatly limited. In addition to how the teachers interpreted the activities presented; it is also necessary to determine if teachers responded according to the “ethicality” dimension. Popham (1991) described two dimensions related to test preparation, ethicality and educational defensibility. It is possible that teachers are responding to the educational defensibility dimension rather than the ethicality of the test preparation practice or perhaps responding using both dimensions (e.g., educational defensibility and ethicality). Through interviews it would be possible to detect the underlying dimensions to their responses regarding the ethicality of the test preparation activities, and this information would be of great assistance when helping educators understand the consequences of using particular types of activities.

Future research also needs to address the relationship between curriculum alignment and test preparation. While Mehrens and Kaminski (1989) and Haladyna et al. (1991) both agreed that teaching without checking the alignment between the test content and curriculum was an ethical practice, with the presence of standards-based reform, the practice may cease to be perceived as ethical by teachers responsible for teaching tested content. Furthermore, there may be confusion among teachers regarding alignment between curriculum and the test, where teachers are concerned about the distinction between proper alignment and cheating (Heldt, 2005). Additional professional development may be needed to clarify proper practices related to alignment, such as “at what point has alignment gone too far.”

This study has provided a detailed description of the types of test preparation activities being utilized by teachers and schools when the scores from an established testing program begin to be used for high-stakes, school accountability purposes. However, additional research is needed to determine the impact of these practices on student achievement and the learning environment, as well as if the same types of practices are used in other school-level contexts. The primary focus, however, should be on obtaining evidence associated with the effect these practices have on the validity of the inferences of the scores, particularly with respect to the interpretation of score gains. As noted by Koretz et al. (2001), the inclusion of activities other

than “teaching more, working harder, and working more effectively” make interpreting score gains more ambiguous, because it is unknown if the gains are due to increased student learning, increased student motivation, changes in the alignment of the curriculum, or test preparation practices. Results from this study will be used in conjunction with evidence regarding curricular changes made by teachers in these schools (Stevenson, Waltman, Middleton, & Croft, 2005) to provide a framework for interpreting changes in student scores across time, as well as identifying some of the positive and negative consequences associated with the implementation of high-stakes testing for school-level accountability purposes. It is hoped that other testing programs (at the state or local levels) might use a similar design to collect this important validity evidence.

## References

- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Amrein, A. L., & Berliner, D. C. (2002). *The impact of high-stakes tests on student academic performance: an analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP test results in states with high school graduation exams. (EPSL-0211-126-EPRU)*. Tempe, AZ: Education Policy Research Unit (EPRU) Arizona State University.
- Cizek, G. J. (1999). *Cheating on tests: how to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, 106(6), 1145-1176.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.
- Heldt, D. (2005, February 27). Are schools 'teaching to the test'? *The Gazette*, pp. 1, 16.
- Herman, J. L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*, 12(4), 20-25,41-42.
- Kilian, L. J. (1992). A school district perspective on appropriate test-preparation practices: a reaction to Popham's proposals. *Educational Measurement: Issues and Practice*, 11(4), 13-15, 26.
- Koretz, D. M., & Barron, S. I. (1998). *The validity of gains in scores on the Kentucky Instructional Results Information System (KIRIS)*: RAND.
- Koretz, D. M., McCaffrey, D. F., & Hamilton, L. S. (2001). *Toward a framework for validating gains under high-stakes conditions. CSE Technical Report 551*. Los Angeles, CA: University of California, Los Angeles.
- Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice*, 8(1), 14-32.
- Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, 21, 215-237.
- Nolen, S. B., Haladyna, T. M., & Haas, N. S. (1992). Uses and abuses of achievement test scores. *Educational Measurement: Issues and Practice*, 11(4), 9-15.
- Pedulla, J. F., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers* (No. BBB36747). Chestnut Hill, MA: National Board on Educational Testing and Public Policy.

- Popham, W. J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice*, 10(4), 12-15.
- Stevenson, E., Waltman, K., Middleton, K., & Croft, M. (2005, April 11-15). *Using test scores for school-level accountability: Analyzing the impact on curriculum and instructional practices*. Paper presented at the American Educational Research Association, Montreal.
- Taylor, G., Shepard, L., Finner, F., & Rosenthal, J. (2003). *A survey of teacher's perspectives on high-stakes testing in Colorado: what gets taught, what gets lost*.

**Appendix A.**  
**Testing practices portion of teacher questionnaire**

3.1 The following are types of activities that are sometimes used with students. These practices vary in terms of their “legitimacy.” Please rate these practices in terms of how ethical <u>you believe</u> each practice is, using 1= <i>very ethical</i> and 5 = <i>not at all ethical</i> .	Very Ethical		Not at all Ethical		
	1	2	3	4	5
3.1.1 Provide practice on questions from <u>exactly the same form</u> of the ITBS/ITED that was administered this year.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.1.2 Provide instruction on the skill areas associated with your district’s content standards and benchmarks (or grade-level indicators) <u>without</u> checking to see which specific skill areas are covered by the ITBS/ITED.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.1.3 Teach test-taking skills, such as completing bubble sheets, pacing/timing, strategies for answering multiple-choice questions, etc.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.1.4 Implement instructional interventions based on a review of ITBS/ITED test results from the previous year in an effort to improve students’ areas of relative weakness.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.1.5 Provide practice on questions from the form of the ITBS/ITED that was administered during the previous year.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.1.6 Within 1 month of testing, use <u>practice</u> exercises/tests that are in the same format and use language similar to test questions found on the ITBS/ITED.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.1.7 Within 1 month of testing, provide a “refresher” on content and/or skill areas that specifically match those on the ITBS/ITED.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.1.8 Routinely provide instruction on <u>only</u> the content and skill areas that specifically match those areas measured by the ITBS/ITED.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.1.9 Routinely use classroom tests that are in the same format and use language similar to test questions found on the ITBS/ITED.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3.2 For each of the following activities, specify the amount of time you have spent in your classroom engaged in each of the activities since testing occurred last year. Then, for those activities on which you spend at least some amount of time, identify the number of school years you have used this activity in this school.	Frequency					# of Years Used		
	No Time	≤ 1 day	2-5 days	2-3 weeks	≥ 4 weeks	1 year	2 years	≥ 3 years
3.2.1 Provide practice on questions from <u>exactly the same form</u> of the ITBS/ITED that was administered this year.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.2.2 Provide instruction on the skill areas associated with your district’s content standards and benchmarks (or grade-level indicators) <u>without</u> checking to see which specific skill areas are covered by the ITBS/ITED.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.2.3 Teach test-taking skills, such as completing bubble sheets, pacing/timing, strategies for answering multiple-choice questions, etc.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.2.4 Implement instructional interventions based on a review of ITBS/ITED test results from the previous year in an effort to improve students’ areas of relative weakness.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.2.5 Provide practice on questions from the form of the ITBS/ITED that was administered during the previous year.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.2.6 Within 1 month of testing, use <u>practice</u> exercises/tests that are in the same format and use language similar to test questions found on the ITBS/ITED.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.2.7 Within 1 month of testing, provide a “refresher” on content and/or skill areas that specifically match those on the ITBS/ITED.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.2.8 Routinely provide instruction on <u>only</u> the content and skill areas that specifically match those areas measured by the ITBS/ITED.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.2.9 Routinely use classroom tests that are in the same format and use language similar to test questions found on the ITBS/ITED.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**3.3 Who conducts most of the activities (as described in 3.1 and 3.2) with your students in preparation for testing?**

You and/or other classroom teachers

Guidance counselor

Other: \_\_\_\_\_

Not applicable, none of these activities are used with my students

Don’t know



**3.4 Which subgroup(s) of students do you engage in unique activities (as described in 3.1 and 3.2) in preparation for testing? Mark all that apply.**

- English Language Learners
- Special education students
- Students identified as being on or below the border of “proficient”
- Not applicable, I do not target any of the specific subgroups of students identified above

**3.5 How does the amount of time spent this school year on activities (as described in 3.1 and 3.2) in preparation for testing compare to the amount of time spent on these types of activities last school year?**

- Increased significantly
- Increased slightly
- About the same
- Decreased slightly
- Decreased significantly
- Don’t know

**3.6 In some schools, special activities related to testing (other than those described in 3.1 and 3.2) are conducted immediately prior to and/or during the administration of the ITBS/ITED. In the space below, please describe all such special activities conducted in your school this academic year. Then, for each activity, identify the number of school years that this activity has been used in your school.**

Description of Special Activities	# of Years
	—
	—
	—

**3.7** In some school, administrators have offered teachers incentives for increasing scores on standardized tests. In the space below, please describe any incentives you or your colleagues were offered (publicly or privately) related to your students' scores on the ITBS/ITED this academic year. Then, for each type of incentive, identify the number of school years that this incentive has been offered to teachers in your school.

Description of Incentives for <u>Teachers</u>	# of Years
	_____
	_____
	_____

**3.8** In some schools, administrators or teachers have offered students incentives for increasing their scores on standardized tests. In the space below, please describe any incentives that your specific students were offered (publicly or privately) related to their (individual or group) performance on the ITBS/ITED. Please specify the grade level of the students receiving the incentives. Then, for each type of student incentive, identify the number of school years that this incentive has been used in your school.

Description of Incentives for <u>Students</u>	Grade Level(s)	# of Years
	_____	_____
	_____	_____
	_____	_____

## Appendix B. Testing practices codebook

Description of special activities	
pep	pep rallies and/or motivational talks
indiv	individual meetings with students
break	breakfast
snack	snacks/refreshments before/during test administration
exer	exercises before/during test administration
visual	visuals/posters for motivational purposes
quiet	quiet time for students' mental preparation
parent	parent newsletter/letter regarding ITBS/ITED
instr	ongoing instructional activity ( <u>not</u> immediately prior to or during administration)
t-prep	activity related to section 3.2 (i.e., test preparation)
context	testing context (e.g., testing in small groups)
sched	scheduling change
o	Other
dk	Don't know
na	Not applicable
no	none

Description of teacher incentives	
<b>F</b>	<b>Financial rewards</b>
b-incr	bonuses for increasing scores
b-high	bonuses for high scores
pd	money for professional development activities
supply	money for classroom supplies or curricular materials
o	Other
<b>P</b>	<b>Privileges</b>
exempt	exemption from meetings
r-load	reduced class load
o	Other
<b>D</b>	<b>Desire/Perception</b>
list	desire to not be on "the list"
job	maintain job
o	Other
<b>T</b>	<b><u>Treats</u></b>
<b>NA</b>	<b>Not applicable</b>
<b>O</b>	<b>Other</b>
<b>DK</b>	<b>Don't Know</b>
<b>NO</b>	<b>None</b>

## Description of student incentives

### F Financial Rewards

- class monetary awards to class funds
- cert gift certificates/coupons
- sch scholarship
- gift special gifts
  - o Other

### P Privileges

- a assemblies
- trip field trips
- act special instructional activities
- fun special non-academic activities (e.g., pizza party, movie)
- open open campus (for lunch, etc.)
- time free time
- exempt exemptions from certain classes
- credit extra credit, exemption from course exam
  - o Other

### R Recognition

- award awards (e.g., certificate) for increased/high achievement
  - o Other

### W Warnings/Threats

- record scores go on student's permanent record
- retain grade-level retention and/or graduation requirement
- summer require/recommend summer school &/or special coursework during the year
- grade scores part of grade
  - o Other

### T Treats

### O Other

- discuss discussion of importance (e.g., to school, student)
  - o Other

NA Not applicable

DK Don't Know

NO None