



AUTHORS

Anna Zilberberg, M.A.
James Madison University

Robin D. Anderson, Psy.D.
James Madison University

Peter J. Swerdzewski, Ph.D.
Regents Research Fund

Sara J. Finney, Ph.D.
James Madison University

Kimberly R. Marsh, M.A.
James Madison University

CORRESPONDENCE

Email
zilberax@jmu.edu

Abstract

Despite the extensive testing for federal accountability mandates, college students' understanding of federal accountability testing (e.g., No Child Left Behind, Race to the Top, Spellings) has not been examined, resulting in a lack of knowledge regarding how such understanding (or lack thereof) impacts college students' behavior on accountability tests in higher education contexts. This study explores college students' understanding and misconceptions of federal accountability testing in K-12. To this end, we crafted nine multiple choice items with four distracters and piloted these items with two college student samples.

The results indicated that college students tend to be moderately confident in their responses regardless of the accuracy of the response. These findings imply that educating students on the purpose and process of accountability testing will require not only imparting correct information, but also debunking misconceptions.

Growing Up with No Child Left Behind: An Initial Assessment of the Understanding of College Students' Knowledge of Accountability Testing

“**M**any of the criticisms we hear about educational assessments appear to be based on misconceptions. Some of them are due to persons simply misunderstanding the meaning of test scores and their implications for instructional improvement and school accountability” (Goodman & Hambleton, 2005, p. 107).

Accountability testing in educational settings has been on the increase due to federal mandates, such as No Child Left Behind (NCLB, 2002), Race to the Top-related testing initiatives (Obama, 2009), and the Spellings report (2006). Despite the widespread use of accountability testing, little is known about students' understanding of accountability testing, and even less is known about how this understanding (or lack thereof) impacts students' test-taking behavior (e.g., effort, honesty). For example, do students who understand how K-12 accountability test results are used give more test-taking effort on the accountability assessments they complete in college than those students who do not know how K-12 accountability test scores are used? If students understand the role of the federal government in the K-12 accountability process, are they more or less likely to give their best effort on accountability tests they encounter in college? These are all empirical questions. However, before answering these questions, a more fundamental question must be answered—do students understand accountability testing mandates at all? It may be the case that students have very limited understanding of these mandates. On the other hand, given their extensive experience of being tested in K-12, they may have learned the purposes behind the testing process.

Of course, whether or not students possess an understanding of K-12 testing mandates is an empirical question.

The purpose of the current study is to provide an initial assessment of college students' understanding of testing associated with federal K-12 institutional accountability mandates (e.g., NCLB)—testing these students experienced for numerous years. That is, although the current generation of college students has experienced accountability testing from elementary school to college, little is known about how well students understand K-12 accountability testing, and how this understanding, or lack thereof, impacts students' test-taking behavior on the accountability tests they complete in college. An item-by-item examination of responses to carefully crafted items representing key aspects of accountability testing provided insight into college student misconceptions regarding such testing. Furthermore, by examining the confidence students have in the correctness of their responses, we begin to understand how difficult it might be to change these misconceptions. Before presenting our findings, we first emphasize the importance of examining college students' knowledge of institutional accountability testing in K-12 and review the literature in this domain.

Misconceptions about Institutional Accountability Testing

So, what are the core concepts of accountability testing imperative for students (and teachers and the public at large) to know? Sireci (2005) discusses six fundamental concepts about assessment. A basic understanding of these concepts is necessary for forming “intelligent opinions about the quality and appropriateness of tests” (Sireci, 2005, p. 112). These concepts are: (a) what is a standardized test; (b) the difference between norm-referenced and criterion-referenced tests; (c) reliability; (d) validity; (e) the setting of passing test scores; (f) obtaining more information about the test (e.g., where to find information about the test development process). Basic understanding of these concepts is a necessary precursor to a critical evaluation of the worth of accountability testing. However, Sireci noted that many criticize tests without adequate background knowledge of these critical concepts that underpin the testing process.

Criticisms voiced against accountability testing include narrowed curriculum, allocation of valuable instructional time toward testing and test preparation, high costs, increased cheating, over-reliance on a single test score, and biased test items (Goodman & Hambleton, 2005; Ravitch, 2010; Sireci, 2005). These criticisms spark debates in the arena of educational policy. Many educational professionals (e.g., teachers, administrators) question whether accountability programs actually serve to improve the quality of education (e.g., Abrams, Pedulla, & Madaus, 2003; Jones et al., 1999). However, there may be fundamental problems pertaining to the sources of many test-related criticisms if these criticisms are due to a lack of understanding of psychometric and policy-related concepts. Although some of the criticisms mentioned above embody legitimate concerns, many may be based on misconceptions about testing.

If students are misinformed about the basics of testing, that would imply students believe they have some general knowledge about the fundamental assessment-related concepts outlined by Sireci (2005), but in fact that knowledge is, at least to some extent, inaccurate. Thus, altering negative attitudes about accountability testing entails not only imparting accurate knowledge, but also identifying and debunking misconceptions. Leading researchers in the field of psychometrics outline some of these misconceptions. Goodman and Hambleton (2005) draw from their experience in the field of psychometrics when discussing several assessment-related misconceptions that are due to “misunderstanding the meaning of test scores and their implications for instructional improvement and school accountability” (p.107). These authors identified four misconceptions held by the general public: (a) high-stakes assessments set everyone involved up for a failure; (b) a single test score is used to make high-stakes decisions; (c) test items are biased; and (d) performance standards are set too high. Although the anecdotal evidence pertaining to the general lack of knowledge and misconceptions about accountability testing is overwhelming and infor-

Despite the wide-spread use of accountability testing, little is known about students' understanding of accountability testing, and even less is known about how this understanding impacts students' test-taking behavior.

mative, empirical questions about how these misconceptions are related to attitudes toward tests, test-taking effort, and test performance cannot be addressed without a measure of student knowledge of accountability testing.

If We Care About Student Knowledge and Misconceptions, How Do We Assess It?

No measure of student knowledge of K-12 accountability testing currently exists. This is not surprising as it would be difficult to create given the breadth of the construct. This study is the first attempt to assess students' knowledge and should be viewed as such – a pilot study that provides initial insight into students' understanding and misunderstanding of accountability testing mandates. We created a set of items to address three aspects of students' knowledge. First, we were interested in investigating the extent to which students are aware of what exactly is mandated in terms of academic achievement in public schools (e.g., what “proficiency” entails in this context and what the test results are used for). Second, we were interested in learning whether students held any misconceptions in regard to how state-mandated testing was carried out in schools (e.g., what percentage of the academic year is taken up by testing). Third, we intended to learn whether students knew who the different stake-holders involved in state-mandated testing were and their respective roles (e.g., who sets the standards?).

...By examining the confidence students have in the correctness of their responses, we begin to understand how difficult it might be to change these misconceptions.

As higher education assessment practitioners, we believe that understanding college students' misconceptions about K-12 accountability testing is valuable in understanding college students' perceptions of higher education assessment testing. While we acknowledge that there are presently no specific nationally mandated tests for college students, increasing demands for accountability by the federal government essentially translates to mandates via the standards set by regional accreditors. This results in a K-16 continuum of assessment “mandates”. Failure to explore the possible impact of K-12 testing on college students' performance on higher education assessments could result in inappropriate inferences regarding students' progress and program effectiveness. For this reason, we are focusing on college students' understanding of K-12 accountability assessment.

Domains of Student Knowledge of Accountability Testing

Given the variation with which states implement state-mandated policy on testing, most investigations of accountability testing focus on a single state (e.g., Jones et al., 1999). Despite the many nuances in how states enact calls for accountability, several federal provisions apply equally to all states. In other words, there are common features pertaining to implementation of NCLB across all states. For this reason, it makes sense to examine students' knowledge of the basic premise underlying institutional accountability testing in K-12; that is, the purpose and the intended use of these federally mandated tests.

In order to develop a set of multiple choice items used in this study, a team of subject matter experts employed a careful and systematic approach. The team consisted of two faculty members with extensive expertise in psychometrics and higher education assessment policy and accountability issues, two advanced doctoral students in Assessment and Measurement, and a content expert in K-12 accountability issues. Combined, the team has thirty years of experience in assessment, accountability testing, and instrument design.

We constructed nine multiple-choice items to address the key aspects of NCLB universally applicable to all states. The content expert in K-12 accountability issues reviewed relevant literature and identified key aspects of NCLB applicable across states. The following key aspects were subsumed under the “What” category of K-12 accountability testing: Schools must experience growth (called Annual Yearly Progress, or AYP) toward proficiency each year; academic proficiency at each level is defined by the state; and the goal of NCLB is adequate education for all. The following key aspects were subsumed under the “Who” category: The federal government administers penalties to schools that fail to achieve proficiency; and the state government sets the learning standards with which the account-

ability tests must align. The following key aspects were subsumed under the “How” category: Detailed information of performance of each school and each of the four subgroups must be publicly available and readily accessible via the school’s report card, which must be provided to parents; NCLB states that school accountability is based only on student performance; factors such as resources, classroom sizes, parent involvement, etc. are not considered; and on average, students spend only about 1% of their total school year taking NCLB-required tests. Several concepts initially identified by the content expert were ruled out during the review process because they were deemed to be too specific or too advanced for students to know. For example, an interesting aspect of NCLB is that the federal act only disciplines schools and districts for poor student performance; whether or not an individual teacher is disciplined due to poor performance on NCLB tests is a state issue. Although a noteworthy side of NCLB, a multiple choice item was not created to address this aspect specifically. The process of delineating these key aspects listed above marked the beginning of an iterative item creation process. Next, the team members carefully crafted and reviewed the stems, distractors, and correct responses for each one of the multiple choice items. The resulting nine items are a product of this iterative and systematic process of item development. Nonetheless, we must stress that we do not assume that the sum of these items represent one test of a unidimensional construct of knowledge of accountability testing. Instead, these nine items allow initial insight into students’ misconceptions about specific testing issues.

In order to gauge the degree of confidence that students possess with respect to their answers, a Likert-type item, prompting students to rate their level of confidence in their response to each knowledge item, was included after each one of the multiple-choice items. We were interested in how strongly students held their misconceptions regarding accountability testing, as strongly held misconceptions may be more difficult to correct than those held with less confidence. That is, it was of interest to gauge the strength of students’ beliefs in the accuracy of their knowledge. The items can be found in the Appendix.

...Altering negative attitudes about accountability testing entails not only imparting accurate knowledge, but also identifying and debunking misconceptions.

Methods

The items were administered as part of a large-scale university assessment effort at a mid-sized, mid-Atlantic four-year institution. Two samples of college students completed the items: (a) incoming college freshmen, and (b) college sophomores.

A total of 3606 incoming college freshmen were administered the items the summer before attending college. A total of 3196 attempted all nine items, thus this sample serves as the sample under study. Females comprised 62.47% of the sample, with less than 1% of students not indicating their gender. About 9.5% were 17 years of age or younger, 85.5% were 18, 4.5% were 19, less than 1% were over 20, with less than half-a-percent choosing not to indicate their age. Of the 3191 students who reported their ethnicity, 85% were Caucasian, 5.6% were Asian, 3.5% were African-American, 2.4% were Latino, and 5.75% were Native-American or multiracial. The majority of students (70%) were from Virginia, 29% were from outside of Virginia, and less than 1% were from outside of the United States, with about 0.59% of students choosing not to indicate their geographic area. Most students reported their high school GPA to be A- or above (59.79%), followed by B- and above (40.12%); less than 1% of students reported a GPA of C+ and below.

The sophomore student sample consisted of 424 students who were administered the items as part of a university-wide assessment day. A total of 382 attempted all nine items. Demographic information was available for 380 participants: 62.8% were female, 77% were Caucasian, average age was 19.15 (SD=0.88), average GPA was 3.05 (SD=0.56), and less than 1% were 18 or younger.

Results

Analyses were conducted at the item level, providing specific information regarding knowledge and confidence of distinct aspects measured by each one of the items. The distinctiveness of the nine items is empirically supported via weak relationships among these

diverse items: correlations ranged from -0.07 to 0.39 for freshmen and from -0.111 to 0.315 for sophomores.

Descriptive statistics for the items and corresponding confidence items for both freshman and sophomore samples can be found in Table 1. Overall, the items ranged in difficulty from 0.20 to 0.74 for the freshman sample and from 0.20 to 0.74 for the sophomore sample. Next, we examined students' responses to the items to highlight students' misconceptions. The item results are organized by domain of knowledge (i.e., the "what", "who", and "how" of accountability testing).

Table 1

Descriptive Statistics for Nine Multiple-Choice and Confidence Items for Freshman and Sophomore Samples

Item	Freshman Sample (N = 3196)				Sophomore Sample (N = 382)			
	Knowledge Items		Confidence Items		Knowledge Items		Confidence Items	
	Item Difficulty	SD	Mean	SD	Item Difficulty	SD	Mean	SD
1	0.24	0.43	5.07	1.27	0.36	0.48	4.90	1.32
2	0.25	0.43	4.01	1.51	0.26	0.44	3.98	1.42
3	0.56	0.50	4.95	1.33	0.50	0.50	4.91	1.34
4	0.28	0.45	4.70	1.32	0.28	0.45	4.66	1.36
5	0.37	0.48	4.94	1.27	0.45	0.50	4.70	1.25
6	0.48	0.50	3.60	1.58	0.55	0.50	3.70	1.52
7	0.20	0.40	3.80	1.38	0.20	0.40	3.73	1.37
8	0.62	0.48	4.01	1.64	0.67	0.47	3.70	1.66
9	0.74	0.44	4.75	1.40	0.74	0.44	4.60	1.42

Note. Confidence items reflect students' confidence in their response on each one of the items; higher scores indicate a greater degree of confidence (Likert-type scale ranging from 1 - *not confident*, 4 - *moderately confident*, 7 - *completely confident*).

“What”

Failure to explore the possible impact of K-12 testing on college students' performance on higher education assessments could result in inappropriate inferences regarding students' progress and program effectiveness.

For both the freshman and sophomore students, less than 45% of the students answered each of the three “what” items correctly. Furthermore, two of the three items were answered correctly at a guessing rate of 0.25 or very close to it, supporting our conclusion that students do not know the correct responses to these items.

What: Item 1. Item 1 evaluated whether students could correctly identify the goal of institutional accountability testing, with the correct response being that testing is used to determine if a given student is on-track for proficiency. Approximately 24% (below the guessing rate) of freshmen answered the item correctly, whereas 36% (SD = 0.48) of the sophomores answered this item correctly (see Table 2). Most students incorrectly endorsed the response option indicating that the most important goal of the state is to ensure that “every student answer enough questions correctly to indicate the student is proficient in the subject every year” (freshmen = 72%; sophomores = 61%).

What: Item 4. Item 4 examined students' knowledge of the purpose of the NCLB Act (which is to ensure adequate access to education for all students). About 28% of freshmen and sophomores answered this item correctly (just a few percentage points above the guessing rate). Notably, the majority of freshmen (55%) and sophomores (58%) endorsed the incorrect response option indicating that the act is specifically designed to ensure that all students in the United States are meeting the same national standards of learning in academic areas including math, reading, and science.

What: Item 5. Item 5 evaluated students' knowledge of what is meant by proficiency as operationalized by the state-mandated tests (i.e., mastery of grade-level work as defined by state). Approximately 37% of freshmen correctly responded to the item and about 45% of sophomores answered this item correctly. However, more students endorsed

the incorrect response option that defined proficiency as having enough knowledge and skill to be successful in the next grade level (freshmen = 53% and sophomores = 51%).

“Who”

For both the freshman and sophomore students, the percentage of students answering each of the three “who” items correctly varied widely (e.g., 26% of students answering correctly on one item compared to 74% answering correctly on another). One of the three items was answered at a guessing rate or close to it, suggesting that students most likely do not know the correct response to that item.

Who: Item 2. Item 2 examined students’ knowledge of the repercussions associated with students not performing well on the tests, with the correct response option being that schools are penalized in various ways. About 25% of freshmen answered the item correctly (just at the guessing rate); likewise, approximately 26% (just above the guessing rate) of the sophomores answered this item correctly. The most frequently endorsed option was the incorrect response that students get held back a grade until the student learns enough to pass the test (47% of freshmen and 41% of sophomores endorsed this option). This finding reflects students’ confusion regarding the federal mandates versus the implementation of these mandates in certain states and districts.

Who: Item 3. Item 3 examined students’ knowledge regarding who sets the standards for the state-mandated tests, with the correct response option being that specific standards are set by the state. About 56% of freshmen selected the correct answer and about 50% of sophomores answered this item correctly. The most frequently selected incorrect answer among freshmen (36%) and sophomores (45.8%) was that the U.S. Department of Education is the standard-setting body.

Who: Item 9. Item 9 evaluated students’ knowledge regarding which governing body selects the specific content for federal accountability tests, with the correct response option being that content is set by the state. About 74% of freshmen and sophomores answered this item correctly. The second-most-frequently endorsed answer was the incorrect response option suggesting that the federal government is responsible for the specific content on federal accountability tests (19% of freshmen and 21% of sophomores endorsed this option).

“How”

For both the freshman and sophomore students, the majority of students answered only one of the three “how” items correctly. One of the three items was answered below the guessing rate, further suggesting that students most likely do not know the correct response to that item.

How: Item 6. Item 6 examined students’ knowledge of the reporting requirements on a school’s “report card” (i.e., report average scores by grade and by ethnic group). About 48% of freshmen answered the item correctly, whereas approximately 55% of sophomores answered this item correctly. The second-most-frequently endorsed answer was the incorrect response that the individual scores, with names concealed, are listed on the “report card” (29% freshmen and sophomores endorsed this option).

How: Item 7. Item 7 evaluated students’ knowledge of factors used to evaluate the effectiveness of schools, with the correct response option being that test scores are the only factor used for the purposes of federal accountability. Only about 20% of freshmen and sophomore students answered this item correctly (below the guessing rate). The three distracters, which focused on financial resources, SES of students, and school size and location, were almost equally endorsed. This was true for both the freshman and sophomore samples.

How: Item 8. Item 8 examined students’ knowledge regarding the average amount of time students spend annually taking state-mandated tests, with the correct response option being that about 1% of the academic year is used for the administration of federal accountability tests. About 62% of freshmen answered the item correctly, whereas about

67% of sophomores answered this item correctly. The most-often-endorsed incorrect response indicated that 7% of the school year is devoted to testing (26% of freshmen and 23% of sophomores endorsed this option).

Table 2
Percent of Students Endorsing Each of the Nine Items by Area for Freshman and Sophomore Samples

Area	Test Items	Percent			
		Freshmen: (N = 3196)	Sophomores (N = 382)		
WHAT	Item 1	Response (a)	3.63	2.62	
		Response (b)	0.59	0.79	
		Response (c)	72.06	60.99	
		Response (d)*	23.72	35.60	
	Item 4	Response (a)	2.69	1.05	
		Response (b)	55.48	57.59	
		Response (c)	13.86	13.61	
		Response (d)*	27.97	27.75	
	Item 5	Response (a)	53.41	50.79	
		Response (b)	4.32	1.31	
		Response (c)	5.41	3.4	
		Response (d)*	36.86	44.5	
	WHO	Item 2	Response (a)	4.1	3.4
			Response (b)	46.53	40.84
			Response (c)	24.19	29.84
Response (d)*			25.19	25.92	
Item 3		Response (a)	2.94	2.09	
		Response (b)	4.88	1.83	
		Response (c)*	55.73	50.26	
		Response (d)	36.45	45.81	
Item 9		Response (a)	0.81	0.79	
		Response (b)	5.73	3.4	
		Response (c)*	74.28	74.35	
		Response (d)	19.18	21.47	
HOW		Item 6	Response (a)	13.61	10.99
			Response (b)	28.91	28.8
			Response (c)	9.89	4.97
	Response (d)*		47.59	55.24	
	Item 7	Response (a)	23.81	26.18	
		Response (b)	24.69	22.51	
		Response (c)	31.07	31.15	
		Response (d)*	20.43	20.16	
	Item 8	Response (a)*	62.27	66.75	
		Response (b)	26.1	23.3	
		Response (c)	8.48	7.85	
		Response (d)	3.16	2.09	

Note. *indicates correct response

Students' Confidence in the Accuracy of their Responses

Recall that both freshman and sophomore students were asked to rate the confidence level they had in the accuracy of their responses. Even though students from both samples answered most of the items incorrectly, their confidence ratings reflected that they were moderately confident in the accuracy of their responses (on a 7-point scale with a value of 4 representing *moderately confident*, average confidence ratings for items ranged from 3.60 to 5.07 for freshmen and from 3.70 to 4.91 for sophomores). Upon examining

student confidence separately for those who responded to the item correctly versus those who responded incorrectly, we noted several trends across the two samples (see Table 3). First, effect size estimates (Cohen's d) indicated that there were negligible differences between student confidence ratings from those who responded correctly versus from those who responded incorrectly to items 1, 2, 5, and 7. In other words, although we would hope that those responding correctly would have more confidence in their response than those responding incorrectly, that did not occur for four of the items. Importantly, these four items did not represent the same domain (e.g., "What") but instead cut across each of the aspects of accountability testing (with items 1 and 5 corresponding to the "What", item 2 corresponding to the "Who", and item 7 corresponding to the "How" domains of accountability testing, respectively).

Second, for items 6 (factors used to evaluate the effectiveness of schools), 8 (the amount of time spent on the administration of federal accountability tests during the year), and 9 (what the tests are designed to measure), freshman and sophomore students who responded correctly were significantly more confident in the accuracy of their response than those who responded incorrectly. It should be noted that the largest effect size estimates were consistently observed for items 6 and 8 across both samples. These two items correspond to the "How" domain of accountability testing, ultimately indicating that students may be more accurate in their appraisal of their knowledge in relation to this specific domain.

Not only did they not know the basic premises underlying federal institutional accountability testing in K-12, but they also believed that their misconceptions were correct.

Table 3
Differences in Average Confidence Ratings between Students Who Answered Each Item Correctly and Incorrectly

Freshmen ($N = 3196$)										
Item	Correct			Incorrect			Difference	t	p	Cohen's d
	Mean	SD	N	Mean	SD	N				
1	4.88	1.30	758	5.12	1.25	2438	0.25*	4.67	<0.01	0.19
2	4.09	1.58	805	3.98	1.49	2391	-0.11	1.69	0.09	-0.07
3	5.08	1.26	1781	4.78	1.40	1415	-0.30*	6.33	<0.01	-0.23
4	4.61	1.32	894	4.74	1.32	2302	0.13	2.50	0.01	0.10
5	5.03	1.26	1707	4.83	1.28	1489	-0.20*	4.45	<0.01	-0.16
6	3.92	1.61	1521	3.31	1.50	1675	-0.61*	11.07	<0.01	-0.39
7	3.86	1.44	653	3.79	1.36	2543	-0.07	1.14	0.26	-0.05
8	4.38	1.63	1990	3.41	1.47	1206	-0.97*	17.31	<0.01	-0.62
9	4.89	1.34	2374	4.35	1.50	822	-0.53*	9.01	<0.01	-0.39

Sophomores ($N = 382$)										
Item	Correct			Incorrect			Difference	t	p	Cohen's d
	Mean	SD	N	Mean	SD	N				
1	4.82	1.14	136	4.95	1.41	246	0.12	0.88	0.38	0.09
2	3.93	1.47	99	3.99	1.40	283	0.06	0.38	0.70	0.05
3	4.92	1.25	192	4.89	1.42	190	-0.03	0.24	0.81	-0.02
4	4.41	1.23	106	4.76	1.40	276	0.36	2.30	0.02	0.26
5	4.78	1.18	170	4.63	1.31	212	-0.14	1.12	0.26	-0.12
6	3.95	1.62	211	3.4	1.33	171	-0.54*	3.53	0<.01	-0.36
7	3.68	1.53	77	3.74	1.34	305	0.07	0.37	0.71	0.05
8	4.02	1.66	255	3.05	1.48	127	-0.97*	5.60	<0.01	-0.61
9	4.69	1.38	284	4.34	1.49	98	-0.35	2.14	0.03	-0.25

Note. Confidence items reflect students' confidence in their response on each one of the multiple-choice items; higher scores indicate a greater degree of confidence (Likert-type scale ranging from 1 - *not confident*, 4 - *moderately confident*, 7 - *completely confident*). Positive difference indicates that those who answered the item incorrectly were more confident in their response than those who answered the item correctly. An asterisk (*) signifies statistical significance at the $p < 0.01$ level. Cohen's d , a practical significance measure of the magnitude of an observed difference between two means on a standardized metric, (Cohen, 1988) is calculated based on pooled standard deviation.

Third, freshman and sophomore samples differed in their confidence rating patterns for only two items: items 3 (who deems what students are supposed to learn and what test content is aligned to) and 4 (the overall purpose of NCLB). Freshmen responding correctly to item 3 indicated significantly higher levels of confidence in the accuracy of their response in comparison to freshmen that responded incorrectly. Conversely, there were negligible differences between sophomore student confidence ratings from those who responded

correctly versus incorrectly to item 3. For item 4, there were negligible differences between freshman student confidence ratings from those who responded correctly versus incorrectly. Interestingly, this was not the case for sophomore students, with students responding incorrectly to item 4 indicating significantly higher levels of confidence in the accuracy of their response in comparison to those who responded correctly. Overall, students' confidence ratings indicate strongly held misconceptions regarding accountability testing.

Discussion

The purpose of the current study was to provide an initial assessment of college students' understanding of K-12 accountability mandates. The nine items piloted in this study were carefully crafted to assess college students' knowledge of, and misconceptions about, K-12 institutional accountability testing associated with NCLB mandates. Specifically, the following three aspects of accountability testing were addressed: what such tests entail, how the results are used, and who mandates testing. In addition, a Likert-type scale confidence item accompanied each of the multiple choice items to allow for the measurement of the degree of confidence that students had in the accuracy of their responses.

Results pertaining to both knowledge (i.e., correctness of response) and confidence followed a similar pattern for both freshman and sophomore samples. More specifically, students hold misconceptions in all three areas addressed by the items: the "what", "who", and "how" of accountability testing. Pertaining to "what", the majority of freshmen and sophomores erroneously believe that the purpose of NCLB is to impose national standards of learning (as opposed to providing equal access to adequate education as defined by the state), to define proficiency as being successful at the next level (as opposed to staying on track to proficiency). Pertaining to "who", a majority of students erroneously believe that the federal government holds students back a grade if test results do not meet the standards (as opposed to schools receiving various penalties); approximately half of all students believe that the U.S. Department of Education sets the standards (as opposed to state), and about 20% of students believe that the federal government (as opposed to state) selects the content to be covered on the tests. Pertaining to "how", only about half of the students know that the school "report card" includes average scores broken down by ethnic group (about 29% of all students think that individual student scores are reported) and both freshmen and sophomores hold misconceptions as to what factors are used for evaluating school effectiveness. Fortunately, the majority of students in both samples know how much time is devoted to federal accountability test administration (i.e., about 1% of the academic year). On average, both freshmen and sophomores hold common misconceptions regarding institutional accountability testing.

In addition, it appears that students tend to confuse the actual state mandates with the schools' practice or implementation. For example, students' responses to items 2 and 5 illustrate students' experience of needing to pass the accountability test in order to advance to the next grade. Although the federal mandate does not require individual students to pass the test in order to advance to the next grade level, many states and districts do impose this requirement. Thus, students assume that the passing requirement is due to the mandate, whereas in reality it is due to the state- or district-specific implementation of the mandate.

Evaluation of students' confidence levels in their responses reveals that students are confident in these beliefs even when the students are wrong. On average, both freshmen and sophomores were moderately confident in their responses (with variability being slightly higher in the freshman sample), even though both groups of students answered the majority of the items incorrectly. In other words, students' evaluation of their own knowledge was inaccurate; it was biased upward. That is, judging by the students' self-reported moderate confidence in their erroneous responses, not only did they not know the basic premises underlying federal institutional accountability testing in K-12, but they also believed that their misconceptions were correct.

The current study is not free of limitations, but there are several ways in which future research can remedy these limitations and build on our findings. The purpose of this

Educating students about the purposes of assessment might result in more accurate test scores.

preliminary investigation was to examine students' understanding of several key aspects of K-12 accountability, as well as identify the misconceptions most common among students, as represented by the distracters. Distracter analyses not only allowed the researchers to identify gaps in student knowledge of accountability testing, but also students' common misconceptions. Future instrument development studies should continue to give careful attention to the distracters. Furthermore, an examination of test-retest reliability should be conducted in order to garner evidence for the stability of scores across administrations. In addition, future item construction initiatives would benefit from item reviews conducted by independent content experts. Even though the current study is not a full scale development study, we believe that this initial investigation sets the stage for such future research endeavors.

Implications and Conclusions

In the context of accountability testing, it might be the case that college students who are ill-informed about what K-12 accountability tests entail, how the results are used, and who is mandating the tests are more likely to develop negative attitudes toward all large-scale accountability testing, are less likely to alter their attitudes toward such tests, and are therefore less likely to exert effort on the accountability tests they complete both in K-12 and college, jeopardizing the validity of inferences made based on these test scores. That is, educating students about the purpose of assessment might result in more accurate test scores. For example, Huffman, Adamopoulos, Murdock, McDermid, and Cole (2011) found that college students who were exposed to an informative motivational presentation about the purpose of a program-level assessment scored higher on average on this assessment than students who received a monetary incentive, or those students who received no treatment (no presentation, no money). Given this finding, it is not surprising that others have called for informing students about the purpose of accountability testing (Leveille, 2006; Zilberberg, Brown, Harmes, & Anderson, 2009).

One may assume that simply educating students about the mandates will resolve the issue of undesirable attitudes and allow students to form appropriate attitudes based on accurate information, which would subsequently improve test-taking behavior. However, in addition to the concern that students may lack knowledge about accountability testing, there is the equally worrisome concern that students may falsely believe they understand the core concepts of accountability assessment, when in reality their understanding is flawed and based on misconceptions. In other words, the issue at hand is more complicated if students are not merely uninformed about these concepts, but instead misinformed.

It follows that if one's goal was to educate students on the basics of accountability testing so that students can hold well-informed, intelligent opinions and develop appropriate attitudes, the challenge will not be just imparting knowledge; educational intervention will also entail debunking pre-existing misconceptions and shattering students' ungrounded confidence. As American historian Boorstin noted, "The greatest obstacle to discovery is not ignorance - it is the illusion of knowledge". Keeping this challenge in mind, future research endeavors can focus on developing a measure and using it for designing and evaluating such educational interventions. No time is more important than the present. Just as this article is being submitted for publication, President Obama has begun to implement the NCLB waiver program (McNeil & Klein, 2011). As states consider tailored plans for accountability that comply with the NCLB waiver requirements, it is critical that states seeking the waivers and the federal government granting the waivers understand what students know about accountability testing and—just as importantly—what students misunderstand about accountability testing. The successful adoption and implementation of revised accountability structures are predicated on knowing what students know about accountability testing, and understanding what actions students take based on this knowledge. As educational policy changes so must the multiple choice items assessing students' understanding of this educational policy.

Importantly, the current findings may also be relevant for developing a measure of knowledge of accountability testing in higher education. College students may differ in

It is critical that states seeking the waivers and the federal government granting the waivers understand what students know about accountability testing and – just as importantly – what students misunderstand about accountability testing.

what they know about accountability testing in K-12 versus accountability testing in college. Moreover, the relationships between such knowledge (or lack thereof) and test-taking behavior (e.g., effort, honesty) may differ depending on the context. That is, knowledge regarding accountability testing in higher education may have a stronger relationship with test-taking behavior on higher education accountability testing than knowledge of K-12 accountability testing. It would also be interesting to assess if knowledge of K-12 accountability testing is related to knowledge of higher education accountability testing. To answer these empirical questions, a higher education version of the items is needed.

In closing, we believe that the results of this pilot study provide an initial assessment of college students' understanding of accountability testing in K-12. As a preliminary investigation of the construct not previously researched, this study sets the stage for future full-scale test development studies, which would entail independent content review of the items as well as gathering reliability and validity evidence for the measure. After a reliable and valid method for measuring students' understanding of K-12 accountability is developed, numerous empirical questions can be addressed, such as the relationship between students' knowledge of accountability testing, students' attitudes toward such tests, and students' test-taking effort.

References

- Abrams, L. M., Pedulla, J. L., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice, 42*(1), 18-29.
- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*, 888-918.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Goodman, D., & Hambleton, R. K. (2005). Some misconceptions about large-scale educational assessments. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 91-110). Mahwah, NJ: Lawrence Erlbaum Associates.
- Huffman, L., Adamopoulos, A., Murdock, G., Cole, A., & McDermid, R. (2011). Strategies to motivate students for program assessment. *Educational Assessment, 16*(2), 90-103.
- Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impacts of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81*(3), 199-203.
- Leveille, D. E. (2006). *Accountability in higher education: A public agenda for trust and cultural change*. Berkeley: Center for Studies in Higher Education.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 23*(9), 4-16.
- McNeil, M., & Klein, A. (2011, September 28). Obama outlines NCLB flexibility; plan waives cornerstone provisions of law. *Education Week*, pp. 1-3.
- No Child Left Behind Act of 2001, 20 U.S.C.S. Section 6301 et seq. (2002). Retrieved March 2nd, 2011, from <http://www.ed.gov/nclb/overview/importance/difference/index.html>
- Obama, B. (2009, July 23). Interview by M.D. Shear & N. Anderson [Audio recording]. President Obama Discusses New 'Race to the Top' Program. Washington Post Education Archives, Washington, DC. Retrieved from <http://www.washingtonpost.com/>
- Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of Social Psychology*, 4th ed., vol. II (pp.323-390). New York, NY: McGraw-Hill.
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.
- Sireci, S.G. (2005). The most frequently unasked questions about testing. In R. Phelps (Ed.), *Defending standardized testing* (pp. 111-121). Mahwah, NJ: Lawrence Erlbaum.
- Spellings, M. (2006). *A test of leadership: Charting the future of U.S. higher education*. A report of the commission appointed by the Secretary of Education. Washington DC: U.S. Department of Education.
- Zilberberg, A., Brown, A. R., Harmes, J. C., & Anderson, R. D. (2009). How can we increase student motivation during low-stakes testing? Understanding the student perspective. In D. McInerney, G. Brown, & G. A. Liem (Eds.), *Research on Sociocultural Influences on Motivation and Learning* (Vol. IX). Greenwich, CT: Information Age Press.

Appendix

Nine Multiple-Choice Items

Directions: Below are a series of questions designed to examine your understanding of state-mandated tests – tests that students must take in public elementary, middle, and high school (For example, in Virginia these tests are called Standards of Learning (SOL); we are not referring to tests such as SAT, PSAT, or ACT).

After selecting your response to each of the multiple choice items, please rate the level of confidence in your response.

1. For the state-mandated tests that students must take in public elementary, middle, and high school, the most important goal for the state is:

- (a) For every student to answer every question correctly every year.
- (b) For those students who are going to college to answer every question correctly every year.
- (c) For every student to answer enough questions correctly to indicate the student is proficient in the subject every year.
- (d) For every student to answer enough questions to ensure the student is on track to being proficient in the subject by a certain year (e.g., two or three years in the future).

1C. Please rate how confident you are that your response to the question above is correct.

2. For state-mandated tests that students must take in public elementary, middle, and high school, if students do not perform as expected, the Federal government (as opposed to the state or the school) mandates:

- (a) The student’s teacher move to a grade in which the teacher is better at teaching.
- (b) The student gets held back a grade until the student learns enough to pass the test.
- (c) The school must purchase new educational materials such as textbooks that are more appropriate for the learning styles of the students at the school.
- (d) The school receive a penalty, such as being required to provide tutoring to all students, firing administrators at the school, or closing the school altogether.

2C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident
1	2	3	4	5	6	7

3. The state-mandated tests that students must take in public elementary, middle, and high school are created to align to what students are supposed to learn according to:

- (a) The student’s teacher.
- (b) The student’s school.
- (c) The state in which the student’s school is located.
- (d) The U.S. Department of Education.

3C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident
1	2	3	4	5	6	7

4. Which of the following most accurately describes the goal of the No Child Left Behind Act, which is the Federal law that mandates state tests that students must take in public elementary, middle, and high school?

- (a) The act is specifically designed to help ensure the United States has a more competitive science and technology workforce compared to emerging nations such as China and India.
- (b) The act is specifically designed to ensure that all students in the United States are meeting the same national standards of learning in academic areas including math, reading, and science.
- (c) The act is specifically designed to ensure no student is left without the critical resources that are needed to learn, including current textbooks, laboratory equipment for science classes, and at least some access to the Internet within the school building.
- (d) The act is specifically designed to ensure that all students have access to an adequate education as defined by each individual state.

4C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident
1	2	3	4	5	6	7

5. For state-mandated tests that students must take in public elementary, middle, and high school, there are certain levels of proficiency. If a student scores at the “proficient” level or higher, the student is said to:

- (a) Have enough knowledge and skill to be successful in the next grade level.
- (b) Be on track to not take remedial courses in college.
- (c) Be sufficiently proficient to succeed in college.
- (d) Have mastered grade level work as defined by the state.

5C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident	
1	2	3	4	5	6	7	

6. For state-mandated tests that students must take in public elementary, middle, and high school, the Federal government requires the following be publicly available via a school’s “report card”, which must be provided to parents and is often featured on the Internet:

- (a) The average scores for all teachers in a school, separated out by subject area and whether or not the teacher is new to the teaching profession.
- (b) The individual scores for all students in the school, although the names of individual students are kept private.
- (c) The individual scores for those students in a school whose scores were not considered “proficient”, although the names of individual students are kept private.
- (d) The average score across all students in each grade in a school, as well as the average score across all students in each of four subgroups race/ethnic subgroups in a grade (African American, Asian / Pacific Islander, Caucasian, Hispanic / Latino).

6C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident	
1	2	3	4	5	6	7	

7. For state-mandated tests that students must take in public elementary, middle, and high school, the Federal government mandates that students’ scores are used in conjunction with the following when evaluating the effectiveness of a school:

- (a) The financial resources available to the school, especially state and local budget allocations.
- (b) The socio-economic status of students at the school, especially the level of parents’ education and parental involvement in the school.
- (c) School characteristics, especially class size and the location of the school in relation to urban or rural areas.
- (d) Test scores are the only factors used to evaluate the effectiveness of schools.

7C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident	
1	2	3	4	5	6	7	

8. For state-mandated tests that students must take in public elementary, middle, and high school, on average the amount of time that students spend taking the actual state test (i.e., excluding practice tests) is:

- (a) 1.0% of the school year
- (b) 7% of the school year
- (c) 12% of the school year
- (d) 18% of the school year

8C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident	
1	2	3	4	5	6	7	

9. For state-mandated tests that students must take in public elementary, middle, and high school, the tests are designed to measure:

- (a) What the teacher expects the student to learn.
- (b) What the school /school district expects the student to learn.
- (c) What the state expects the student to learn.
- (d) What the Federal government expects the student to learn.

9C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident	
1	2	3	4	5	6	7	