

*Australian Journal of Educational & Developmental Psychology*. Vol 12, 2012, pp. 71-82

## **An instrument to assess beliefs about standardized testing: Measuring the influence of epistemology on the endorsement of standardized testing**

Robert G. Magee<sup>1</sup> & Brett D. Jones  
*Virginia Tech*

### **ABSTRACT**

This article describes the development of an instrument to assess beliefs about standardized testing in schools, a topic of much heated debate. The Beliefs About Standardized Testing scale was developed to measure the extent to which individuals support high-stakes standardized testing. The 9-item scale comprises three subscales which measure beliefs about objectivity and knowledge, the utility of tests as a criterion for retaining students, and the use of tests in rewarding teachers. The scale's factorial structure and internal consistency were assessed in Study 1 ( $N = 142$ ). The scale's utility in predicting responses to items from the Phi Delta Kappa/Gallup Poll was demonstrated in Study 2 ( $N = 161$ ). Results provided evidence of the reliability and validity of the scale scores and its subscale scores. Respondents' acceptance of the notion that objective and unbiased knowledge is possible and desirable was associated with a greater likelihood of endorsing standardized testing. Effective dialogue about standardized testing should consider how differences in worldviews can shape individuals' beliefs about standardized testing.

### **INTRODUCTION**

The *No Child Left Behind Act of 2001* (NCLB, 2002) has been a controversial policy in the United States ever since it was passed (Jones, Jones, & Hargrove, 2003). The act was designed to increase accountability for student achievement by mandating that states implement statewide assessments (U.S. Department of Education, n.d.). Responses vary, but states have typically met this mandate by implementing statewide standardized tests that measure student learning in various subject areas at various grade levels. Much of the controversy around the increased use of standardized testing has been related to the reliance of test scores as the sole measure for making high-stakes decisions, such as whether to retain students in a grade, to allow students to graduate from high school, and to reward teachers and schools with money or ratings.

Researchers have documented that individuals' beliefs vary as to whether standardized tests should be used in making such high-stakes decisions. Many teachers and administrators have reported that standardized high-stakes testing is taking schools in the wrong direction, yet others believe that there are benefits to standardized testing (Jones & Egley, 2004a, b). Public opinion about NCLB is also mixed,

---

<sup>1</sup> Contact

Dr. Robert G. Magee  
108 Shanks Hall (0311), Virginia Tech, Blacksburg, VA, USA  
e-mail: rgmagee@vt.edu  
tel: (540) 231-7162  
fax: (540) 231-9817

according to the results of a Phi Delta Kappa/Gallup poll (Rose & Gallup, 2006). When asked to rate the overall impact of NCLB on the public schools in their community, 37% of the public reported that it was making no difference, 26% reported that it was helping, 21% claimed that it was hurting, and 16% did not know. When asked more specifically whether they would be in favor of requiring high school students to pass a test to graduate from high school, 63% of the public was in favour, 36% was opposed, and 1% did not know. Using standardized tests as a means to make high-stakes decisions about schools was less popular. Over two-thirds (69%) of the public reported that a single, statewide test would not provide a fair picture of whether or not a school needs improvement, with 28% reporting that it would provide a fair picture, and 3% reporting that they did not know.

Studies of teachers, administrators, and the public provide us with a general picture of how individuals view standardized testing in public schools. However, reasons for *why* beliefs about testing vary are largely unexplored, which led us to ask: Are there constructs such as worldviews or core values that can predict whether or not an individual supports high-stakes standardized testing? It might be the case that disagreement over the desirability, utility, and value of testing is in part a function of an individual's fundamental assumptions about the world around them, particularly with regard to the nature of truth and objectivity.

A worldview, what Kant called a *Weltanschauung*, comprises a set of untestable assumptions about the nature of reality and how the world "works" (Koltko-Rivera, 2004; Magee & Kalyanaraman, 2009). These assumptions, which are formed over time and usually remain tacit, shape a host of beliefs. Wittgenstein (1972) described these beliefs as a *Weltbild*, or world picture, and explained that they function as both a filter and a riverbed. As a filter, a *Weltbild* permits a person to see some things and ignore others. As a riverbed shapes the course of the flow of water, so also a *Weltbild* shapes one's thoughts or cognitive processes. Core elements of a worldview deal with human-nature relations, one's orientation toward time, attributions of causation, and the nature of truth (Kluckhohn & Strodtbeck, 1961) and influence how one makes sense of one's experience. The assumptions individuals make about the nature of truth and the possibility of objective knowledge should guide their beliefs about the desirability of standardized tests.

To answer the question of whether worldviews or core values are related to beliefs about standardized testing, it is necessary to have an instrument that produces valid scores about individuals' beliefs about standardized, high-stakes testing. The purpose of this study was to develop such an instrument.

## METHOD

### Participants and Procedure

The development and testing of the scale occurred in two phases. In the first phase, the scale items were administered to a sample of undergraduate students ( $N = 142$ ) to assess the scale's internal consistency and factorial structure. In the second phase, the scale and six items from the *2006 Phi Delta Kappa/Gallup Poll* (Rose & Gallup, 2006) were administered to a second sample of undergraduate students ( $N = 161$ ) to assess the scale's predictive utility. Across both samples, the mean age of the participants was 20.65 years, 74% of the participants were female, and 84% were Caucasian, with 8% Asian, with the remainder reporting another ethnicity. The scales were administered anonymously via paper and pencil along with measures from a separate study. In the second phase of the study, respondents completed the Beliefs About Standardized Testing scale on one sheet of paper before completing the items from the *2006 Phi Delta Kappa/Gallup Poll* (Rose & Gallup, 2006) on a separate sheet of paper.

### *The Beliefs About Standardized Testing scale (BAST)*

For a number of reasons, we were most interested in creating a measure of an individual's beliefs about standardized testing that related to the *accuracy* and *use* of standardized tests. First, although researchers have documented a variety of positive and negative effects of high-stakes standardized testing (Berliner, 2009; Evers & Walberg, 2002; Jones, 2007; Pedulla et al., 2003; Raymond & Hanushek, 2003),

issues related to the use and accuracy of the tests have been shown to be some of the most controversial (Jones & Egley, 2004a, b). Individuals often have strong feelings about whether or not standardized tests can accurately measure student learning and the purposes for which the test results should be used. Second, we believed that individuals' beliefs about the use and accuracy of tests might be impacted by their worldviews, which was one of the primary reasons we wanted to develop a beliefs about standardized testing instrument.

We suspected that differences over the desirability of standardized tests arise, in part, from differences in epistemology, or "how we know what we know." More specifically, these differences might be attributed to differences in individuals' worldviews regarding the nature of truth and objectivity. Individuals who believe that truth is universal and knowable and that objectivity is possible are likely to feel that standardized tests provide an accurate and objective measure of students' learning. In contrast, individuals who believe that truth is subjective, or socially constructed, and maintain that objectivity is a fiction, and that is impossible to know or ascertain absolute truth are more likely to distrust a one-time standardized test as a means to assess students' learning.

For these reasons, the BAST was not designed to measure *all* types of beliefs about standardized testing. Instead, we chose to create three subscales that measured: (a) beliefs about the nature of objectivity, knowledge, and universality (the Objectivity subscale); (b) beliefs about high-stakes decisions (the Retention subscale); and (c) beliefs about bonus money (the Reward subscale). See Appendix A for a list of all of the items.

The Objectivity subscale comprised five items that measured individual's beliefs about the possibility of unbiased knowledge that can be applied universally. The Retention subscale consisted of two items that measured individuals' beliefs about whether standardized tests should be used to make high-stakes decisions such as retaining students. The Reward subscale consisted of two items that assessed individuals' beliefs about whether or not schools and teachers should receive extra money when their students score highly on standardized tests. Participants responded to all 9 items on a 9-point Likert-type format that ranged from 1 (*strongly disagree*) to 9 (*strongly agree*).

#### *The Phi Delta Kappa/Gallup Poll (PDK/GP)*

To provide a measure to which we could compare the BAST scores, we selected six items from the *Phi Delta Kappa/Gallup Poll of the Public's Attitudes Towards the Public Schools*. The PDK/GP is a research instrument that is conducted annually to allow the public, educators, and policy makers to track public opinion about public schools (Phi Delta Kappa International, 2009). The six items we selected from this instrument (see Appendix B) were implemented verbatim from the *2006 Phi Delta Kappa/Gallup Poll* (Rose & Gallup, 2006), with one minor exception. For the three items that included the word "community," we added the word "home" before "community" to ensure that the college students in our sample would refer to the community they considered their permanent home rather than their relatively new university community, which might be very different and about which they likely had less knowledge.

## RESULTS

The BAST scale was scored such that a high score represented favorable beliefs about standardized testing (see Table 1 for means and standard deviations), and the scale's properties were assessed using a sample of college students ( $N = 142$ ). Because the scale comprises three subscales, we tested the scale's factor structure through a confirmatory factor analysis on the item covariances. A rotated 3-factor solution, with an oblique Quartimin rotation, yielded a rotated factor pattern matrix (Table 2) that was clearly interpretable and accounted for 85% of the scale's variance.

The nine-item scale also appeared to be internally consistent (Cronbach's  $\alpha = .83$ ). In addition, the subscales appeared to be internally consistent. The five-item Objectivity subscale yielded a Cronbach's  $\alpha$  of .76. The two items of the Retention subscale correlated strongly,  $r = .65$ ,  $p < .001$ , as did the two items of the Reward subscale,  $r = .71$ ,  $p < .001$ .

**Table 1:** Means and Standard Deviations for the BAST Items

	<i>M</i>	<i>SD</i>
Item 01	2.56	1.96
Item 02	2.52	1.80
Item 03	4.42	2.25
Item 04	2.58	1.54
Item 05	3.35	1.92
Item 06	4.07	2.33
Item 07	3.45	2.04
Item 08	3.28	2.18
Item 09	3.61	2.39

**Table 2:** Pattern Matrix for the BAST Items

	Objectivity	Retention	Reward
Item 01	1.491		
Item 02	1.279		
Item 03	1.012		
Item 04	1.148		
Item 05	0.852		
Item 06		1.648	
Item 07		1.434	
Item 08			1.734
Item 09			1.864

*Note.* Loadings < .30 suppressed.  $N = 142$ .

The utility of the scale and its subscales as a predictor of responses to the six items from the PDK/GP was assessed with a second sample of undergraduate students ( $N = 161$ ). Except where, logistic regression was used to test if the BAST was a significant predictor of respondents' answers to the PDK/GP items. Although the sample was not representative of the voting public, the sample was particularly useful for testing the scale's predictive validity. If the scale exhibits the ability to discriminate

among members of a relatively homogeneous population, it should demonstrate even greater discriminant validity among members of a more heterogeneous population.

It bears noting that most of the respondents (79%) had graduated from a Virginia high school, and the majority (65%) indicated that they had to pass one or more examinations to graduate from high school. It is reasonable to suspect that their beliefs about standardized testing and their opinion of the NCLB might be coloured by their recent experience with standardized testing. Nevertheless, their experience with standardized testing did not affect their mean scores ( $M_{\text{yes}} = 3.42$ ,  $SE = 0.12$ ;  $M_{\text{no}} = 3.26$ ,  $SE = 0.16$ ) on the BAST scale,  $F(1, 159) = 0.663$ ,  $p = .42$ .

#### *PDK/GP Item 1.*

The BAST predicted the degree of respondents' satisfaction with the amount of perceived emphasis on testing  $\chi^2(3, 160) = 25.02$ ,  $p < .001$ ,  $R^2(U) = .09$ . Tests of parameter estimates with "Too much" as a baseline suggested that the "About the right amount" response was significantly different (Table 3). The BAST was not a significant predictor of the odds of a respondent choosing one of the other responses,  $ps > .08$ . When the respondents who selected "Don't Know" ( $n = 7$ ) were excluded, the BAST remained a significant predictor of respondents' dissatisfaction with the emphasis on testing, with a modest gain in the model's explanatory power,  $\chi^2(2, 153) = 24.43$ ,  $p < .001$ ,  $R^2(U) = .11$ . Compared to respondents who indicated "Too much" as a baseline, respondents who scored higher on the BAST were significantly more likely to say that emphasis on testing was "about the right amount."

The Objectivity subscale was a significant predictor of respondents' opposition to the perceived emphasis on testing,  $\chi^2(3, 160) = 17.97$ ,  $p < .001$ ,  $R^2(U) = .06$ . Tests of parameter estimates with "Too much" as a baseline suggested that the "About the right amount" response was significantly different (Table 3). The subscale was not a significant predictor of the odds of a respondent choosing the "Not enough" response,  $p > .20$ , but the subscale seemed to predict the choice of the "Don't know" response,  $p = .06$ . When the respondents who selected "Don't know" ( $n = 7$ ) were excluded, the Objectivity subscale remained a significant predictor of respondents' opposition to the perceived emphasis on testing, and the model's explanatory power improved only slightly,  $\chi^2(2, 153) = 16.35$ ,  $p < .001$ ,  $R^2(U) = .07$ .

The Retention subscale, likewise, was a significant predictor of respondents' opposition to the perceived emphasis on testing,  $\chi^2(3, 160) = 22.02$ ,  $p < .001$ ,  $R^2(U) = .08$ . Compared to "Too much," the likelihood that respondents would select the "About the right amount" option was significantly greater the more they agreed that retention decisions should be made using standardized tests (Table 3). When the respondents who selected "Don't know" ( $n = 7$ ) were excluded, the Retention subscale remained a significant predictor,  $\chi^2(2, 153) = 21.06$ ,  $p < .001$ ,  $R^2(U) = .09$ .

The whole model of the Reward subscale was not a significant predictor of the responses ( $p = .18$ ), and the pattern did not change when the respondents who selected "Don't Know" ( $n = 7$ ) were excluded. However, the Reward subscale was a significant predictor of whether respondents would choose "About the right amount" over "Too much" (Table 3).

#### *PDK/GP Item 2.*

The BAST predicted respondents' opposition to requiring a graduation exam,  $\chi^2(2, 161) = 21.25$ ,  $p < .001$ ,  $R^2(U) = .07$ . Compared with the "Oppose" response, the odds of respondents supporting the requirement differed significantly, and the odds of respondents choosing "Don't know" were significant as well (Table 4). Excluding the respondents who selected the "Don't know" response ( $n = 30$ ) to compare support and opposition directly, the scale predicted opposition versus support for requiring a graduation exam,  $\chi^2(1, 131) = 14.44$ ,  $p < .001$ ,  $R^2(U) = .10$ . The higher a person's score on the BAST, the more likely he or she was to support requiring a graduation exam.

The Objectivity subscale was a significant predictor of opposition, too,  $\chi^2(2, 161) = 9.88$ ,  $p = .007$ ,  $R^2(U) = .03$ . Compared to the "Oppose" response, support for requiring a graduation exam was more likely the more a person said that objective and unbiased assessments were possible (Table 4). Likewise, the subscale was a significant predictor of the "Don't know" responses. Excluding the respondents who

**Table 3:** PDK/GP Item 1: Opinion of Amount of Emphasis on Achievement Testing

	<i>Not enough</i>		<i>About the right amount</i>		<i>Don't know</i>	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
BAST	.40	.30	.91***	.21	.64	.37
Objectivity	.32	.27	.66***	.18	.59	.32
Retention	.21	.18	.52***	.13	.36	.22
Reward	.19	.17	.21*	.10	.04	.20
Excluding “ <i>Don't know</i> ” responses						
BAST	.40	.30	.95***	.22		
Objectivity	.32	.27	.67***	.18		
Retention	.21	.18	.53***	.13		
Reward	.19	.17	.21*	.10		

Note. Compared to “Too much.”

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

selected the “Don't know” response ( $n = 30$ ) to compare support and opposition directly, the Objectivity subscale remained a significant predictor, with a slight gain in the model's explanatory power,  $\chi^2(1, 131) = 7.60, p = .006, R^2(U) = .05$ .

The Retention subscale was a significant predictor of opposition to a graduation exam requirement,  $\chi^2(2, 161) = 36.36, p < .001, R^2(U) = .12$ . Support for the requirement, compared to the “Oppose” baseline, was significant, as was the odds that a respondent would select the “Don't know” response (Table 4). A direct comparison of support and opposition, excluding the “Don't know” responses ( $n = 30$ ), was significant and improved the model's explanatory power,  $\chi^2(1, 131) = 30.06, p < .001, R^2(U) = .20$ .

A test of the Reward subscale as a predictor of participants' responses was less conclusive,  $\chi^2(2, 161) = 4.91, p = .086, R^2(U) = .02$ . However, the odds of a participant expressing uncertainty concerning to the requirement, compared to “Oppose,” was significant (Table 4). Nevertheless, a direct comparison of support to opposition, excluding the “Don't know” responses ( $n = 30$ ), was non-significant,  $p > .40$ .

**Table 4:** PDK/GP Item 2: Opinion about Requiring an Exam to Graduate From High School

	<i>Favor</i>		<i>Don't know</i>	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
BAST	.70***	.19	.66**	.20
Objectivity	.45**	.16	.37*	.17
Retention	.66***	.13	.46***	.13
Reward	.09	.10	.24*	.11
Excluding “ <i>Don't know</i> ” responses				
BAST	.68***	.19		
Objectivity	.44**	.17		
Retention	.65***	.14		
Reward	.09	.10		

Note. Compared to “Oppose.”

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

*PDK/GP Item 3.*

Neither the BAST nor the subscales were significant predictors of respondents' level of knowledge of the No Child Left Behind Act, all  $ps > .40$ . Excluding respondents who indicated "Don't know" ( $n = 4$ ) did not change the test results, all  $ps > .40$ .

*PDK/GP Item 4.* Although the BAST did not predict respondents' knowledge of the No Child Left Behind Act, the scale was a significant predictor of respondents' opinion of the act,  $\chi^2(5, 160) = 13.47, p = .019, R^2(U) = .03$ . The BAST predicted differences between strong opposition to the Act and levels of support for the Act (Table 5). When respondents who indicated "Don't know" ( $n = 5$ ) and "Don't know enough to say" ( $n = 41$ ) were excluded and the remaining responses were treated as an ordinal scale with four levels, the BAST was a significant predictor of respondents' opinion of the Act,  $\chi^2_{\text{model}}(1, 114) = 6.94, p = .008, R^2(U) = .03$ , and the scale yielded a significant coefficient,  $B = .41, \chi^2 = 7.21, p = .007$ .

A model of the Objectivity subscale was not a significant predictor,  $p = .13$ . However, as noted in Table 5, the subscale predicted the likelihood that, compared to the "Very unfavorable" option, respondents would choose either the "Somewhat favorable" or "Somewhat unfavorable" responses. An ordinal logistic regression model that excluded the "Don't know" ( $n = 5$ ) and "Don't know enough to say" ( $n = 41$ ) responses was not significant,  $p = .13$ .

The Retention subscale was a significant predictor of respondents' opinion of the act,  $\chi^2(5, 160) = 11.80, p = .038, R^2(U) = .02$ . Like the Objectivity subscale, the Retention subscale predicted the likelihood that a respondent would choose either the "Somewhat favorable" or "Somewhat unfavorable" responses (Table 5). Unlike the Objectivity subscale, though, a model of the Retention subscale was a significant predictor of the ordinally scaled dependent variable,  $\chi^2_{\text{model}}(1, 114) = 8.07, p = .005, R^2(U) = .03$ , with a significant coefficient for the subscale,  $B = .28, \chi^2 = 8.45, p = .004$ .

The Reward subscale was not a significant predictor of respondents' choices,  $p > .20$ . However, when the "Don't know" and "Don't know enough to say" responses were excluded and the remaining responses were scaled ordinally, a model of the Reward subscale was a significant predictor of respondents' opinion of the act,  $\chi^2_{\text{model}}(1, 114) = 4.18, p = .041, R^2(U) = .02$ , and the subscale's coefficient was significant,  $B = .18, \chi^2 = 4.08, p = .043$ .

*PDK/GP Item 5.*

Neither the BAST nor any of the subscales predicted respondents' perceptions of the valence of the impact of the No Child Left Behind Act on public schools, all  $ps > .20$ . Excluding the respondents who indicated "Don't Know" ( $n = 49$ ) did not improve the model, all  $ps > .20$ .

**Table 5:** PDK/GP Item 4: Opinion of NCLB Act

	<i>Very favorable</i>		<i>Somewhat favorable</i>		<i>Somewhat unfavorable</i>		<i>Don't know</i>		<i>Don't know enough to say</i>	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
BAST	.70	.45	.75**	.25	.65*	.25	.85	.45	.33	.25
Objectivity	.35	.42	.52*	.24	.54*	.24	.51	.41	.21	.24
Retention	.52	.28	.41**	.16	.32*	.16	.52	.28	.16	.16
Reward	.32	.27	.35*	.15	.24	.16	.39	.26	.20	.16

Note. Compared to "Very unfavorable."

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

*PDK/GP Item 6.*

The BAST predicted respondents' opposition to using a statewide exam to assess school effectiveness,  $\chi^2(2, 160) = 24.78, p < .001, R^2(U) = .09$ . High scores on the BAST were associated with increased odds that respondents would favour (compared to "No") using a statewide exam (Table 6). High scores were also associated with a greater likelihood that respondents would select the "Don't know" option. Excluding the respondents who selected the "Don't know" response ( $n = 30$ ) to focus on support vis-à-vis opposition, the scale still predicted support for requiring a graduation exam, and the model's explanatory power increased markedly,  $\chi^2(1, 130) = 21.68, p < .001, R^2(U) = .18$ . A higher score on the BAST significantly increased the odds that a respondent would express support (versus opposition) for using a single test as a measure of whether a school needed improvement.

Likewise, the Objectivity subscale was a significant predictor of the responses,  $\chi^2(2, 160) = 22.77, p < .001, R^2(U) = .08$ . The more a respondent endorsed the idea that knowledge can be measured in an objective and unbiased fashion, the more one supported (compared to "No") requiring a graduation exam. Higher scores were also associated with increased odds of indicating "Don't know." When the "Don't know" responses ( $n = 30$ ) were excluded, the subscale improved as a predictor of a person's support or opposition,  $\chi^2(1, 130) = 21.68, p < .001, R^2(U) = .14$ .

The Retention subscale was a significant predictor of the responses,  $\chi^2(2, 160) = 15.67, p < .001, R^2(U) = .06$ . Compared to opposition, the subscale predicted the odds that a respondent would favor requiring a graduation exam. The subscale was not a significant predictor of the odds that respondents might select the "Don't know" response,  $p = .13$ . When the "Don't know" responses ( $n = 30$ ) were excluded, the subscale improved as a predictor,  $\chi^2(1, 130) = 15.09, p < .001, R^2(U) = .13$ , but the coefficient for the subscale remained virtually unchanged.

The Reward subscale was not a significant predictor of support or opposition to requiring a graduation exam,  $p = .12$ , and removing the "Don't know" responses did not improve the model,  $p = .12$ .

**Table 6:** PDK/GP Item 6: Opinion about Whether a Single Test Provides a Fair Picture of Whether a School Needs Improvement

	Yes		Don't know	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
BAST	.95 <sup>***</sup>	.24	.64 <sup>**</sup>	.20
Objectivity	.77 <sup>***</sup>	.21	.63 <sup>***</sup>	.18
Retention	.51 <sup>***</sup>	.14	.17	.11
Reward	.18	.12	.18	.11
Excluding "Don't know" responses				
BAST	1.17 <sup>***</sup>	.30		
Objectivity	.88 <sup>***</sup>	.24		
Retention	.52 <sup>***</sup>	.15		
Reward	.17	.12		

Note. Compared to "No."

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .



## DISCUSSION

The confirmatory factor analysis and the item correlation results provide evidence of the reliability of the BAST scale scores and its subscale scores. The comparison of the BAST scale and subscale scores to scores on the PDK/GP item scores through regression provides evidence for the validity of the BAST scale scores and its subscale scores. These results are promising and suggest that the BAST could be used to assess the extent to which individuals support standardized testing.

Respondents' opposition to high-stakes standardized testing stemmed from worldview-related beliefs about truth and objectivity. Those who believed that knowledge is complex, and that students are unique opposed efforts to assess students' knowledge via a single test. Further, doubts that standardized tests can provide an objective and unbiased assessment also had an impact on respondents' opposition to standardized tests. Not surprisingly, this opposition also extended to policy efforts to use standardized tests as a criterion for making decisions about retention and rewards. Importantly, neither respondents' knowledge of the NCLB nor their personal experience with testing predicted their beliefs about standardized testing. These findings provide evidence that individuals' worldviews, and not their knowledge about high-stakes testing nor their experiences of testing affected their beliefs about standardized testing.

Because support for and opposition to standardized testing are affected strongly by individual's epistemological beliefs, discussion on testing should take into account how differences in worldviews can shape the assumptions that individuals make about testing. Changing an individual's worldviews does not occur quickly or easily (Murphy & Mason, 2006). Given this, it will be difficult to change someone's beliefs about standardized testing. It might involve making individuals aware of their epistemological stances and how their assumptions about the world shape their views on public policy.

Future directions for research in this area include studying whether a change in epistemological beliefs is necessary for individuals to change their beliefs about standardized testing, or whether a person's stance on testing can be uncoupled from epistemological beliefs. Understanding some of the processes individuals go through when they change their beliefs about standardized testing would be useful and would help to answer questions such as what worldview-related issues must be addressed to change individuals' perceptions about standardized testing and what types of dialogue are needed for individuals to change their beliefs about standardized testing? Answering these questions would help to explicate the relationship between individuals' worldviews and their beliefs about standardized testing. There are many reasons for supporting or opposing standardized testing. We argue that addressing the role of epistemology in these reasons will advance the discussion of standardized testing.

## REFERENCES

- Berliner, D. C. (2009). MCLB (Much Curriculum Left Behind): A U.S. calamity in the making. *The Educational Forum*, 73, 284-296.
- Evers, W. M., & Walberg, H. J. (2002). *School accountability*. Stanford, CA: Hoover Institution Press.
- Jones, B. D., & Egley, R. J. (2004a). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives*, 12(39). Retrieved from <http://epaa.asu.edu/epaa/v12n39/>
- Jones, B. D., & Egley, R. J. (2004b). Is testing the right direction? Administrators share their thoughts. *ERS Spectrum*, 22(3), 16-25.
- Jones, B. D. (2007). The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology*, 23(2), 65-86.
- Jones, M. G., Jones, B. D., & Hargrove, T. Y. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman & Littlefield.
- Kluckhohn, F. R., & Strodtbeck, F. L. (1961). *Variations in Value Orientations*. Evanston, IL: Row, Peterson.

- Koltko-Rivera, M. E. (2004). The psychology of worldviews. *Review of General Psychology*, 8, 3-58.
- Magee, R. G., & Kalyanaraman, S. (2009). Effects of worldview and mortality salience in persuasion processes. *Media Psychology*, 12, 171-194.
- Murphy, P. K., & Mason, L. (2006). Changing knowledge and beliefs. In P. A. Alexander, & P. H. Winne (Eds.), *Handbook of educational psychology*. (pp. 305-324). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pedulla, J. J., 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*. National Board on Educational Testing and Public Policy. Retrieved from <http://www.bc.edu/research/nbetpp/statements/nbr2.pdf>
- Penfield, R. D. (2010). Test-based grade retention: Does it stand up to professional standards for fair and appropriate test use? *Educational Researcher*, 39(2), 110-119.
- Phi Delta Kappa International. (2009). *PDK/Gallup Poll*. Retrieved from <http://www.pdkintl.org/kappan/poll.htm>
- Raymond, M. E., & Hanushek, E. A. (2003). High-stakes research. *Education Next*, 3(3), Retrieved from <http://educationnext.org/>
- Rose, L. C., & Gallup, A. M. (2006). The 38<sup>th</sup> annual Phi Delta Kappa/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 88(1), 41-53.
- Wittgenstein, L. (1972). *On Certainty*. G. E. M. Anscombe & G. H. von Wright (Eds.) D. Paul & G. E. M. Anscombe (Trans.). New York: Harper and Row.

#### **Biographical Notes:**

Robert G. Magee, Ph.D., is an Assistant Professor of Persuasion and Media Effects in the Communication Department at Virginia Tech. His research includes examining how people's worldviews and values influence persuasion processes and outcomes.

Brett D. Jones, Ph.D., is an Associate Professor of Educational Psychology in the Learning Sciences and Technologies Department at Virginia Tech. His research includes investigating how students' beliefs impact their motivation and examining methods instructors can use to design instructional environments that support students' motivation and learning.

## APPENDIX A

### Beliefs About Standardized Testing items

*Objectivity Subscale: Beliefs about the nature of objectivity, knowledge, and universality*

1. Standardized tests are the best way to objectively measure how much a student knows.
2. Students are too different for a single standardized test to really be useful in measuring their abilities. (Reversed)
3. A good standardized test can provide a fair (unbiased) indication of the quality of education a student receives in school.
4. True knowledge is too complex to be measured by a standardized test. (Reversed)
5. It is impossible for a standardized test to really be unbiased. (Reversed)

*Retention Subscale: Beliefs about high-stakes decisions*

6. It is a good idea for states to require all high school students to pass a standardized test or set of standardized tests in order to graduate from high school.
7. Students, in any grade above third grade, who do not pass their grade-level standardized test should have to repeat that grade level.

*Reward Subscale: Beliefs about bonus money*

8. Schools whose students have the highest scores on standardized tests should receive bonus money.
9. Teachers whose students score the highest on standardized tests should receive bonus money.

**APPENDIX B**  
Phi Delta Kappa/Gallup Poll items

1. In your opinion, is there too much emphasis on achievement testing in the public schools in your home community, not enough emphasis on testing, or about the right amount?
  1. Too much
  2. Not enough
  3. About the right amount
  4. Don't know
  
2. Some states are requiring high school students to pass a graduate qualifying exam in order to graduate from high school. Would you favor or oppose having such a requirement for all the high schools in your home community?
  1. Favor having required exam
  2. Oppose having required exam
  3. Don't know
  
3. Now, here are a few questions about the No Child Left Behind Act. How much, if anything, would you say you know about the No Child Left Behind Act – the federal education bill that was passed by Congress in 2001 – a great deal, a fair amount, very little, or nothing at all?
  1. A great deal
  2. A fair amount
  3. Very little
  4. Nothing at all
  5. Don't know
  
4. From what you know or have heard or read about the No Child Left Behind Act, do you have a very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable opinion of the act – or don't you know enough about it to say?
  1. Very favorable
  2. Somewhat favorable
  3. Somewhat unfavorable
  4. Very unfavorable
  5. Don't know enough to say
  6. Don't know
  
5. Just your impression, how would you rate the overall impact of the No Child Left Behind program on the public schools in your home community? Would you say it was helping, hurting, or making no difference in the performance of the local public schools?
  1. Helping
  2. Hurting
  3. Making no difference
  4. Don't know
  
6. According to the No Child Left Behind Act, determining whether a public school is or is not in need of improvement will be based on the performance of its students on a single statewide test. In your opinion, will a single test provide a fair picture of whether or not a school needs improvement?
  1. Yes, would
  2. No, would not
  3. Don't know