Test Administration Procedures and Their Relationships with Effort and Performance on a College Outcomes Test

Jeffrey T. Steedle
Doris Zahner
Heather Kugelmass

Philadelphia, Pennsylvania

April 6, 2014

Copyright © 2014 Council for Aid to Education
Currently, Jeffrey T. Steedle is a Research Scientist with Pearson. Doris Zahner is the Director of Test Development and a Measurement Scientist at the Council for Aid to Education. Heather Kugelmass is a Ph.D. candidate in the Department of Sociology at Princeton University.

Correspondence concerning this article should be addressed to Jeffrey T. Steedle, Pearson, 400 Center Ridge Drive, Austin, TX 78753. Contact: jeffrey.steedle@pearson.com

Poor motivation is commonly cited as a concern when interpreting results from low-stakes standardized tests administered to postsecondary students. This study investigates the associations between test administration procedures and students’ self-reported effort and performance on the Collegiate Learning Assessment (CLA), an open-ended test of college students’ critical-thinking and writing skills. Coefficient estimates from a series of hierarchical linear models revealed that paying students to take tests and offering performance-based incentives were positively associated with effort and performance. Mandatory testing, however, was negatively associated with effort and performance. Faculty involvement in recruiting, giving extra course credit, and offering prize raffle entries were not associated with effort or performance. Effort appeared to mediate the relationship between some test administration variables (e.g., payment and mandatory testing) and performance.

Keywords: higher education, institutional assessment, motivation
TEST ADMINISTRATION PROCEDURES AND THEIR
RELATIONSHIPS WITH EFFORT AND PERFORMANCE
ON A COLLEGE OUTCOMES TEST

Motivation to perform well on a test is a potentially problematic source of construct-irrelevant variance (Messick, 1995), especially when there are no stakes attached to performance. That is, interpretations of test scores as indicators of examinees’ knowledge and skills may be compromised if those examinees do not put forth the effort necessary to demonstrate the full extent of their abilities. Low-stakes testing is commonplace in K-12 education in the United States, and concerns over suspect motivation are exacerbated when aggregate test results are used to measure teacher or school effectiveness (e.g., Guerriero, 2013). In postsecondary education, tests such as CAE’s Collegiate Learning Assessment (CLA), ACT’s Collegiate Assessment of Academic Proficiency, and the ETS Proficiency Profile are typically administered under low-stakes conditions to provide evidence of student learning for stakeholders, prospective students, and accreditors.

Poor motivation is commonly raised in critiques of postsecondary institutional assessment programs (Banta, 2008), and some colleges cite this concern as a reason for not using assessments such as the CLA. Statistical approaches to adjusting results for low motivation have been proposed (e.g., Sundre & Wise, 2003), but it would be better if no such adjustments were necessary. In addition to psychological variables, such as achievement goals and personality (Barry, Horst, Finney, Brown, & Kopp, 2010) and the format of the test (Sundre, 1999; Wise, 2006), test administration procedures potentially impact student motivation on tests. Indeed, inconsistencies in methods of recruiting and incentivizing examinees have been suggested as causes of variation in results across administrations (Hosch, 2010), and rigorous proctor training has been shown to increase self-reported effort (Lau, Swerdzewski, & Jones, 2009).

This study investigates test administration procedures and their possible associations with effort and performance on the CLA, a test of college students’ critical-thinking and writing skills. In 2008, a post-administration survey was delivered to administrators at all participating institutions. This survey asked whether CLA testing was mandatory, whether faculty were involved in recruiting students, and how students were incentivized. In this study, a series of hierarchical linear models were employed to determine whether test administration procedures were significantly associated with test performance and self-reported effort after controlling for prior ability. Results could inform best practices for recruiting and incentivizing students to maximize motivation, thereby improving the validity of inferences from results of low-stakes college outcomes tests.

LITERATURE REVIEW

Prior research has shown that attaching consequences to test performance can have large effects on motivation and test performance (Liu, Bridgeman, & Adler, 2012; Napoli & Raymond, 2004; Wise & DeMars, 2005; Wolf & Smith, 1995). Tests like the CLA, however, are almost exclusively administered in low-stakes contexts, so test administrators must rely on alternate means of fostering motivation. For example, it has been recommended to stress the importance and usefulness of a test because those feelings are correlated with effort, which partly mediates test performance (Cole, Bergin, & Whittaker, 2008). Alternatively, instilling a sense of competition with rival schools may have some effect on test performance (Bracey, 1996).

Uniformly high motivation has been reported in some international testing contexts (Baumert & Demmrich, 2001; Eklöf, 2007), with examinees reporting social responsibility, competitive spirit, interest, and personal or intrinsic motivators as the reasons for their motivation (Eklöf, 2008). Such findings sharply contrast with U.S. students taking low-stakes tests like the National Assessment of Educational Progress, where a large percentage of students report a low sense of importance and low effort (ETS, 1993).

Although monetary incentives have been effective for improving test scores in some content areas (Bettinger, 2010), performance-contingent financial rewards appear to be more effective
than a reward for simple test completion (Braun, Kirsch, & Yamamoto, 2011). In fact, bigger financial incentives appear to be more effective than smaller financial incentives on math and reading tests, but only if given at the time of testing (Levitt, List, Neckerman, & Sadoff, 2011). Even offering additional entries in a prize lottery for high math test scores has been effective in increasing self-reported effort and performance (Cole, 2007). The effectiveness of performance-contingent rewards, however, is not universal. In a series of studies, neither small nor large incentives improved 12th graders’ math test performance (O’Neil, Abedi, Miyoshi, & Mastergeorge, 2005; O’Neil, Sugrue, & Baker, 1995/1996).

This study supplements prior research by examining the associations between monetary incentives (payment, performance-based incentives, and raffle entries), self-reported effort, and performance on a low-stakes college outcomes test. Additionally, this study included several variables not previously studied: the involvement of faculty in student recruitment, mandatory (versus voluntary) testing, and offering course extra credit. Separate analyses were conducted for freshmen and seniors due to different administration procedures for those two groups. Moreover, self-reported effort among older students may be different from younger students (Kiplinger & Linn, 1995/1996; Liu et al., 2012).

**METHOD**

**Sample**
The sample included 5,428 entering freshmen and 4,611 graduating seniors at 102 four-year colleges and universities (24 research universities, 47 master’s colleges and universities, 30 baccalaureate colleges, and one seminary). Of those schools, 41 had a total enrollment greater than 10,000 students, 58 were public institutions, and four were classified as Historically Black Colleges and Universities. In terms of geographic distribution, 15 were located in the mid-Atlantic or New England regions, 15 in the Great Lakes or plains, 38 in the southeast, and 34 in the west or southwest. They had admissions rates ranging from 18% to 100% (median 66%), six-year graduation rates ranging from 19% to 92% (median 55%), and a percentage of White students ranging from 5% to 95% (median 68%). Because the statistical analyses controlled for students’ prior academic abilities, only students with SAT or ACT scores on record were included in the data set.

**Measures**

**CLA Performance Task (PT).** Students had 90 minutes to analyze a set of documents representing a real-world problem and answer a series of essay questions asking them to analyze the provided information and then propose a solution. Students were randomly assigned one of many possible PTs. A group of trained scorers evaluated the responses using scales that described the quality of analysis, problem solving, and writing effectiveness and mechanics. Inter-rater correlations on PT total scores were typically around .85.

**Self-reported effort.** Following the PT, students completed a brief questionnaire, including an item that asked students how much effort they put into the test. The response options included “Made little effort,” “Made some effort,” “Mainly did my best,” and “Tried my best.”

**Prior ability.** As part of the regular CLA administration, SAT and ACT scores were collected from college registrars’ offices. ACT total scores were converted to the SAT score scale using a concordance table (ACT, 2008).

**Post-administration survey.** Completed surveys were received from 102 (52%) of the 195 institutions participating in the 2007-2008 CLA administration. Relevant to this analysis, the survey asked whether students were required to take the CLA, whether faculty participated in CLA outreach to students, and how participants were incentivized. The incentives included in this analysis were payment (money, gift certificate, or university cash), performance-based incentives (e.g., the top five scorers receive $100), raffle entry, and course extra credit. Other incentives were excluded from analyses because they were employed by fewer than 10% of institutions (food, priority course registration, or none).
Analysis
A series of five HLMs were fit to the data. The unconditional model was first fit to estimate the variance in PT scores that is between and within schools.

\[
\text{Level 1: } PT_{ij} = \beta_{0j} + r_{ij} \\
\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}
\]

The model shown below treats students (level 1) as nested within institutions (level 2). At the student level (level 1), PT<sub>ij</sub> is the PT score of student i at school j, and \(\beta_{0j}\) is the mean PT score at school j. At the school level, \(\beta_{0j}\) is modeled as the grand mean PT score plus the school-level residual, \(u_{0j}\).

Next, a conditional model including only grand-mean centered SAT scores at level 1 was fit to see how much of the between-school variance was accounted for by SAT scores.

\[
\text{Level 1: } PT_{ij} = \beta_{0j} + \beta_{1j}(SAT_{ij} - \bar{SAT}) + r_{ij} \\
\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \\
\beta_{1j} = \gamma_{10}
\]

\(SAT_{ij}\) is the SAT score of student i at school j. Note that grand-mean centering has the effect of making \(\beta_{0j}\) equal to school j’s PT mean adjusted for SAT scores.

Model 3 added test administration dummy variables at the school level (level 2).

\[
\text{Level 1: } EFF_{ij} = \beta_{0j} + \beta_{1j}(SAT_{ij} - \bar{SAT}) + r_{ij} \\
\text{Level 2: } \\
\beta_{0j} = \gamma_{00} + \gamma_{01}MAN_j + \gamma_{02}FAC_j + \gamma_{03}PAY_j + \gamma_{04}PER_j + \gamma_{05}RAF_j + \gamma_{06}CRE_j + u_{0j} \\
\beta_{1j} = \gamma_{10}
\]

Here, adjusted mean PT scores (\(\beta_{0j}\)) are modeled as a function of six dummy variables indicating whether testing was mandatory (MAN), faculty were involved in recruitment (FAC), students were paid (PAY), performance-based incentives were employed (PER), raffle entries were offered (RAF), and course extra credit was offered (CRE) at school j.

Model 4 is identical to model 3 except that it adds self-reported effort \(EFF_{ij}\) at the student level.

\[
\text{Level 1: } PT_{ij} = \beta_{0j} + \beta_{1j}(SAT_{ij} - \bar{SAT}) + \beta_{2j}(EFF_{ij} - \bar{EFF}) + r_{ij} \\
\text{Level 2: } \\
\beta_{0j} = \gamma_{00} + \gamma_{01}MAN_j + \gamma_{02}FAC_j + \gamma_{03}PAY_j + \gamma_{04}PER_j + \gamma_{05}RAF_j + \gamma_{06}CRE_j + u_{0j} \\
\beta_{1j} = \gamma_{10} \\
\beta_{2j} = \gamma_{20}
\]
Grand-mean centering on SAT scores and effort has the effect of making $\beta_{0j}$ equal the mean PT score at school j adjusted for SAT scores and effort.

As demonstrated in previous research, effort may mediate the relationship between administrative procedures and CLA performance (Cole, Bergin, and Whittaker, 2010). When that is the case, one should expect coefficients for dummy variables to be significant in Model 3 but not when self-reported effort is included (Model 4). Model 5 used self-reported effort as the outcome to directly investigate the associations between administrative procedures and self-reported effort.

**Level 1:** $PT_{ij} = \beta_{0j} + \beta_{2j}(SAT_{ij} - \bar{SAT}) + r_{ij}$

**Level 2:**

$$\beta_{0j} = \gamma_{00} + \gamma_{01}MAN_{j} + \gamma_{02}FAC_{j} + \gamma_{03}PAY_{j} + \gamma_{04}PER_{j} + \gamma_{05}RAF_{j} + \gamma_{06}CRE_{j} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

**RESULTS**

Table 1 shows the percentages of schools that employed the administration procedures examined in this study. Note that schools employed these procedures more frequently with seniors than with freshmen. This finding is consistent with common anecdotal reports from schools suggesting that seniors are more challenging to recruit than freshmen.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Freshmen</th>
<th>Seniors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandatory testing (MAN)</td>
<td>18%</td>
<td>33%</td>
</tr>
<tr>
<td>Faculty involvement (FAC)</td>
<td>60%</td>
<td>68%</td>
</tr>
<tr>
<td>Payment (PAY)</td>
<td>50%</td>
<td>70%</td>
</tr>
<tr>
<td>Performance-based incentives (PER)</td>
<td>15%</td>
<td>19%</td>
</tr>
<tr>
<td>Raffle (RAF)</td>
<td>24%</td>
<td>30%</td>
</tr>
<tr>
<td>Course extra credit (CRE)</td>
<td>21%</td>
<td>21%</td>
</tr>
</tbody>
</table>

Table 2 provides descriptive statistics on the outcome measures. As would be expected due to learning in college, the seniors scored higher on the CLA, but some of this difference can be accounted for by differences in ability, which were apparent from the mean difference in SAT scores. On average, freshmen and seniors reported similar levels of effort, with more than 70% of students reporting that they mainly did their best or tried their best. The correlation between self-reported effort and CLA performance was .24 ($p < .001$) in both samples, and this result is fairly typical (Steedle, 2014).
Results from the analysis of freshmen are shown in Table 3. Models 1 and 2 indicated that 25% of the variance in CLA scores was between schools, but, after controlling for SAT scores, 7% was between schools. In Model 3, only the coefficient for payment had a significant positive coefficient (p < .05), but the negative coefficient for mandatory testing was nearly significant (p = .07). When self-reported effort was added in Model 4, only mandatory testing had a nearly significant coefficient (p = .09). Significant coefficients for payment in Models 3 and 5 but not Model 4 suggested effort as a mediator between paying students and test performance. Model 5 also had a significant positive coefficient for performance-based incentives (p < .01).

In the senior analysis, 20% of the CLA score variance was between schools, and 5% was between schools after controlling for SAT scores (Table 4). In Model 3, there was a significant negative coefficient for mandatory testing (p < .05) and a significant positive coefficient for performance-based incentives (p < .01). Only performance-based incentives had a significant coefficient (p < .001) in Model 4. Model 5 had significant coefficients for mandatory testing (p < .01), payment (p < .01), and performance-based incentives (p < .05). The comparison of Models 3, 4, and 5 reveals the possible mediating effect of effort between mandatory testing and performance.
## Table 3
Coefficients from models fit to entering freshman data

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
<th>Model 4</th>
<th></th>
<th>Model 5 (Outcome variable: EFF)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAT</td>
<td>$\gamma_{10}$</td>
<td>0.57***</td>
<td>0.01</td>
<td>0.57***</td>
<td>0.01</td>
<td>0.57***</td>
<td>0.01</td>
<td>0.00009</td>
<td>0.00007</td>
<td></td>
</tr>
<tr>
<td>EFF</td>
<td>$\gamma_{20}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>42.82***</td>
<td>2.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>School level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$\gamma_{c0}$</td>
<td>456.5***</td>
<td>16.07</td>
<td>1069.8***</td>
<td>10.79</td>
<td>948.3***</td>
<td>12.29</td>
<td>2.83***</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>MAN</td>
<td>$\gamma_{c1}$</td>
<td>-23.01</td>
<td>12.40</td>
<td>-19.44</td>
<td>11.36</td>
<td>-0.08</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAC</td>
<td>$\gamma_{c2}$</td>
<td>-5.27</td>
<td>10.32</td>
<td>-3.31</td>
<td>9.43</td>
<td>-0.04</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAY</td>
<td>$\gamma_{c3}$</td>
<td>20.89*</td>
<td>10.44</td>
<td>11.89</td>
<td>9.56</td>
<td>0.22***</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PER</td>
<td>$\gamma_{c4}$</td>
<td>18.18</td>
<td>13.00</td>
<td>10.10</td>
<td>11.87</td>
<td>0.19**</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAF</td>
<td>$\gamma_{c5}$</td>
<td>12.88</td>
<td>11.27</td>
<td>12.60</td>
<td>10.33</td>
<td>0.00</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRE</td>
<td>$\gamma_{c6}$</td>
<td>-17.06</td>
<td>11.86</td>
<td>-14.24</td>
<td>10.87</td>
<td>-0.07</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student level</td>
<td>$\sigma^2$</td>
<td>29328</td>
<td>23800</td>
<td>23803</td>
<td>22714</td>
<td>0.62403</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School level</td>
<td>$\tau_{c0}$</td>
<td>9859</td>
<td>1854</td>
<td>1462</td>
<td>1168</td>
<td>0.03912</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001, ** p < .01, * p < .05
Table 4
Coefficients from models fit to graduating senior data

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th></th>
<th>Model 2</th>
<th></th>
<th></th>
<th>Model 3</th>
<th></th>
<th></th>
<th>Model 4</th>
<th></th>
<th></th>
<th>Model 5 (Outcome variable: EFF)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAT</td>
<td></td>
<td>0.56***</td>
<td>0.02</td>
<td></td>
<td>0.55***</td>
<td>0.02</td>
<td></td>
<td>0.54***</td>
<td>0.02</td>
<td></td>
<td>0.00035***</td>
<td>0.00008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43.06***</td>
<td>3.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>School level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>561.9***</td>
<td>18.28</td>
<td></td>
<td>1183.2***</td>
<td>12.21</td>
<td></td>
<td>1057.6***</td>
<td>14.35</td>
<td></td>
<td>2.91***</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-20.06*</td>
<td>9.87</td>
<td></td>
<td>-11.90</td>
<td>9.18</td>
<td></td>
<td>-0.18**</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAC</td>
<td></td>
<td>-11.05</td>
<td>9.82</td>
<td></td>
<td>-9.21</td>
<td>9.09</td>
<td></td>
<td>-0.04</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAY</td>
<td></td>
<td>10.11</td>
<td>10.45</td>
<td></td>
<td>3.64</td>
<td>9.71</td>
<td></td>
<td>0.16**</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PER</td>
<td></td>
<td>41.03***</td>
<td>10.63</td>
<td></td>
<td>33.69***</td>
<td>9.84</td>
<td></td>
<td>0.16*</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAF</td>
<td></td>
<td>6.84</td>
<td>9.69</td>
<td></td>
<td>8.12</td>
<td>9.00</td>
<td></td>
<td>-0.03</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRE</td>
<td></td>
<td>-3.47</td>
<td>11.49</td>
<td></td>
<td>-3.95</td>
<td>10.69</td>
<td></td>
<td>0.01</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student level</td>
<td></td>
<td>32371</td>
<td>26880</td>
<td></td>
<td>26907</td>
<td>25842</td>
<td></td>
<td>0.60117</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School level</td>
<td></td>
<td>8236</td>
<td>1600</td>
<td></td>
<td>1036</td>
<td>824</td>
<td></td>
<td>0.0453</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001, ** p < .01, * p < .05
DISCUSSION

This study investigated associations between test administration procedures, self-reported effort, and performance for a low-stakes test of college outcomes. The analysis of senior data revealed a significant negative association between requiring students to take the CLA and performance after controlling for other variables (nearly significant for freshmen). That is, results are consistent with the notion that mandating testing, rather than soliciting volunteers, can negatively impact student effort and performance.

Some have suggested that institutional assessment programs would benefit from faculty buy-in and involvement (Steedle, 2010). In this study, faculty involvement in recruitment was not significantly associated with effort or performance, after controlling for other variables. However, the extent of faculty involvement was unknown.

Of the incentives, payment for task completion (money, gift certificate, or university cash) and performance-based incentives (e.g., the top five scorers receive $100) showed positive associations with effort and performance, after controlling for other variables. Raffle entries and course extra credit did not. Performance-based incentives for freshmen and payment for seniors were correlated with self-reported effort but not test performance. In some cases, there was evidence that effort acted as the mediator between administration procedures and performance. For example, results are consistent with the notion that mandatory testing of senior students reduced effort, thereby reducing performance. In a similar fashion, higher performance among freshmen who were paid can be accounted for by the association between payment and effort.

The practices of paying students to participate and offering performance-based incentives appear to support testing effort and performance, thereby strengthening the validity of test-score interpretations on low-stakes tests of college outcomes. In contrast, the practice of mandating testing (e.g., in randomly selected freshman seminars or senior capstone courses) appears to depress testing effort and performance, which could negatively impact validity. This conclusion could be subjected to experimentation in future studies. Until then, it is recommended that students be solicited to volunteer for testing and offered financial incentives, while making every effort to ensure that the tested sample is representative of the larger student body.
REFERENCES


