

Problems with Using Low-Stakes Tests to Assess Student Learning

Rosemary E. Sutton
Cleveland State University

1

Testing and Consequences

- High Stakes Tests
 - Consequences for test performances
- Accountability Movement in K-16
 - Low Stakes For Students
 - High Stakes for Institutions

2

Ex. Collegiate Learning Assessment (CLA)

- The CLA is unique in its combination of skills measures, value-added approach, use of performance tasks, web-based administration, matrix sampling strategies, and adjustments made to control for initial ability.” (CLA brochure)
- Constructed response test - 120 minutes
- Two areas: Performance Task and Analytical Writing
- www.cae.org/cla

3

CLA cont.

- *Performance task (90 minutes)*
 - students are required to use an “integrated set of critical thinking, analytical reasoning, problem solving and communication skills about a hypothetical but realistic situation.”
 - e.g., the student is asked to prepare a summary memo about a plane crash based on a variety of documents including Newspaper articles about the accident, a Federal Accident Report on in-flight breakups in single engine planes, emails, and charts on the airlines performance characteristics.

4

CLA cont.

- *Analytical Writing* - 2 two components
 - Make-an-argument prompt (45 minutes)
 - Critique-an-argument prompt (30 minutes)
- Excellent publicity
 - “Spellings Commission” cites CLA as example of assessment
 - Kevin Carey (2006) *College Rankings Reformed*, Education Sector Report
 - Voluntary System of Accountability (2007)
<http://www.aascu.org/accountability/>

5

Test Validity

- Old Meaning
 - Does the test measure what it is intended to?
- Current Meaning
 - “Adequacy and appropriateness of the *interpretations* and *uses* of assessment results” for a given group of individuals (Linn & Miller, 2005, p. 68).
 - **Test validity is not only about the nature of the test.**

6

Slide 6

A1 Administrator, 3/4/2007

A2 Administrator, 3/4/2007

Can Low-Stakes Tests have Good Validity?

- On low-stakes tests students' motivation varies
 - ❑ Motivation on the CLA is "a major challenge" (Ewell, 2005)
 - ❑ Difficult to get students to take the CLA
 - ❑ When student do take it, some students stay 3 hours – some leave after 20 minutes
- **How can any interpretations of test scores be adequate or appropriate if many students don't show up or try?**

7

Ideal and Real Test Takers*

- Reasons for investing time and effort in a low-stakes test:
 - ❑ Invested in intellectual skills
 - ❑ Used to trying hard
 - ❑ Importance of test
 - Reasons against:
 - ❑ Many students balance school, work and families responsibilities
 - ❑ Time management and setting priorities (e.g., in Freshman seminars)
 - ❑ Not important
- * (Ryan, Ryan, Arbuthnot & Samuels, 2007)

8

Strategies to Reduce Problems with Low Stakes Tests

Make tests high stakes

- Assessments should be based on what has been taught
- Appropriate for curriculum-based tests (devised by faculty or those used more widely)
- Not appropriate for general skills test such as CLA
 - ❑ All the skills cannot be taught at freshman level
 - ❑ Problems on equity as CLA correlated highly with SAT/ACT
 - ❑ "Unfair penalization" (Popham, 2005)

9

Strategies - 2

Provide incentives for students to do well

- Incentives for taking test
 - ❑ Taking test vs. taking test seriously.
- Incentives for doing well
 - ❑ Equity problems
- Graduated incentives
 - ❑ Those who took the test got \$25:00, drawings for prizes for those who stayed 2 and 3 hours.

10

Strategies - 3

Provide students feedback on their scores

- Limited data suggests this is not effective (Wise & DeMars, 2005)
- If the test is not important to students then feedback may be meaningless
- Major content tests may be important to students and so feedback may be helpful

11

Strategies - 4

Make tests simpler and shorter

- Evidence suggests reduced student effort on
 - ❑ Constructed response tests vs. multiple choice
 - ❑ Longer tests
- Longer tests create practical problems
 - ❑ e.g. GRE subject matter tests are over 2 hrs
- BUT complex long tests may be needed to assess higher level cognitive skills.

12

Strategies – 5a

Motivation Filtering

- Examine test scores of groups who report varying amount of motivation.
 - Student Opinion Survey: 10 item scale, 2 factors
 - Importance: e.g., “Doing well on this test was important to me”
 - Effort: e.g., “I engaged in good effort throughout these tests.”
- <http://jmu.edu/assessment/>

13

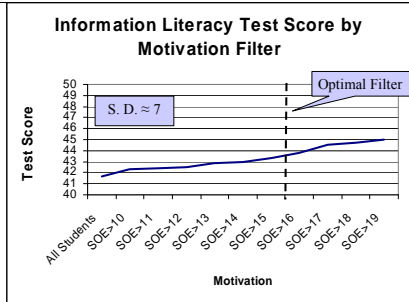
Strategies - 5b

Assessment Test	Correlation with Motivation (SOS)		
	Test Score	SAT-V	SAT-Q
Information Literacy	.45*	.08	-.04
Fine Arts	.33*	.13	.09
Natural World	.31*	.03	.07

* P < .001. Adapted from Wise, Wise and Bhola (2006)

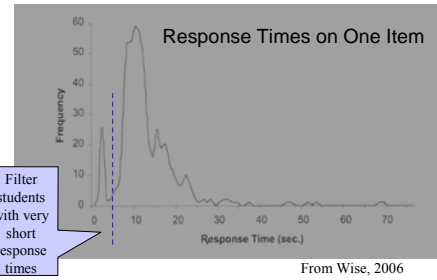
14

Strategies- 5c



15

Strategies – 5d



16

Strategies -5e

- Rapid Response related to
 - Item length
 - Item position
- Small # of rapid responses – 3% and 6%
- Filtering
 - lowers reliability (e.g. .88 to .75)
 - enhances validity (correlations with SAT)

17

Strategies 5f

- Motivation Filtering
 - Existing research is within institution
 - No between-college research
 - Probably greater differences
- BUT Between-institution comparisons are made**

18

Conclusion

- Low-Stakes Tests
 - Variable student motivation
 - Reduced validity
- Strategies to increase motivation have problems
- Motivation filtering is an intriguing area of research.

19

References

- Ewell, P. T. (2006, May-June). From the States: Smoke signals from the hills of Texas. *Assessment Update*, 18, (3), 12-13, 15.
- Linn, R. L., & Miller, M. D. (2005). *Measurement and Assessment in Teaching* 9th ed. Upper Saddle River, NJ: Pearson.
- Popham, W. J. (2005). *Classroom Assessment: What teachers need to know*. Boston, MA: Pearson.
- Ryan, K. E., Ryan, A. M., Arbuthnot, K., & Samuels, M. (2007). Students' motivation for standardized math exams. *Educational Researcher*, 36(1), 5-13
- Wise, S. L. (2006). An investigation of the differential effort received by items on a low-stakes computer-based test. *Applied Measurement in Education*, 19(2), 95-114.
- Wise, V. L., Wise, S. L., & Bholra, D. S. (2006). *Educational Assessment*, 11(1), 65-83.
- Wise, S. L. & DeMars, C. W. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1-17.

20