8

Tests with Rewards

Our thesis up to this point has been that rewards work best when they are closely aligned with SMART goals (Specific, Measurable, Attainable, Realistic, and Time-bound). Testing for academic achievement is a critical part of the rewards process because it provides measurement, telling both the educator and the learner if goals are being achieved or what needs to be changed to make them achievable. Tests can be tailored to learners of every age and rewards can be attached to the test results.

The use of tests in K–12 education, while supported by parents and even students, is controversial in schools of education and therefore in teachers' lounges. In this chapter we review what critics of tests say, rebut the most common criticisms, and present best practices for attaching rewards to tests. We also describe the new (new for K–12 education, that is) technology of adaptive online testing and the importance of independent test audits.

Critics of Tests

Critics of rewards fiercely criticize standardized tests. For example, Alfie Kohn argued, "anyone who requires a formal test to know what is going on may need to reconsider the approach to instruction being used and whether he or she is talking too much and listening too little."[1] Monty Neil, executive director of FairTest, criticized the widespread use of tests, particularly the Massachusetts Comprehensive Assessment System, despite the state's repeated first place on the National Assessment of Educational Progress.[2] At a 1991 press conference organized by the American

Educational Research Association (AERA), a group composed mainly of education professors presented a long list of objections to the use of standardized tests, especially those connected to rewards for high scores.[3] Among their complaints: reliance on tests discourages teachers from teaching topics that don't appear on the tests; tests tap only "lower-order" thinking skills; and tests are unfair to minorities and women, costly in terms of money and time, and overused in the United States compared with other countries.

At best, the professors' criticisms apply only to poorly designed and administered tests. A frequent complaint about No Child Left Behind, a federal government education program, is that it relies too heavily on testing and has led to the adoption of "teacher-proof" curricula, or curricula so focused on specific learning outcomes that teachers were forbidden from making alterations in response to their own experience and understanding of their students' needs.[4] But the fact that tests sometimes are used poorly does not support wider criticism of their use as part of effective incentive programs.

Test critics ignore hundreds of published, well-designed studies complete with comparison groups that show the benefits of tests. Richard Phelps recently conducted a meta-analysis of the voluminous research and statistical analysis of the effects of standardized tests. His findings are summarized in the table on the following page.

Refuting the Critics

Six of the most frequent criticisms of standardized tests are presented below with rebuttals.[5]

Does testing take valuable time away from instruction?

Not when it is used correctly. Regular testing helps teachers determine if students have learned course material and are able to move on to more advanced skills and subject matter. Test results therefore *save* classroom time by helping teachers avoid needlessly reviewing or repeating lessons. Frequent testing increases the productivity of time spent on instruction by encouraging students to pay attention in class and do homework soon after it is assigned rather than "cramming" just before an exam. Well-designed tests encourage students to think through and practice material on their own time, either at school or at home.

Benefits of Standardized Tests

- Setting incentivized goals and measuring progress increases student motivation and performance in school achievement as well as sports and work settings.
- In school and college classes, frequent testing, including brief daily testing, results in greater learning.
- Giving students detailed analyses of their test results helps them identify their weaknesses and increases their learning; computer-assisted testing and information processing makes this quicker and more feasible.
- Learning is reinforced and enhanced by offering students overviews and details on what they have done well.
- Using tests to verify that students have mastered or nearly mastered specific content before introducing new material yields better results than traditional teaching that assumes a high degree of mastery.
- When studies focus on language learning, frequent testing has intensified and increased the speed with which students learn new languages.

Source: Richard Phelps "The Rich, Robust Research Literature on Testing's Achievement Benefits," in Richard P. Phelps, ed., *Defending Standardized Testing* (Mahwah, NJ: Lawrence Erlbaum Associates Publishers, 2005), pp. 55–90. See the source for original phrasing of these points.

Do testing programs divert scarce dollars from other more important uses such as classroom instruction?

No. The cost of testing is trivial compared to other school expenditures. Dollars spent on commercial firms for services including standardized testing, standards setting, and accountability reporting was less than 0.1 percent of total spending on K–12 education in 2000, amounting to only $5.81 per student on average.[6] It is doubtful that any other major enterprise in the U.S. spends a smaller percent of its total budget measuring its outputs. New technology, moreover, is reducing the cost of testing as well

as the amount of time teachers must spend calculating test scores and interpreting the results.[7]

Do standardized tests fail to cover important subject matter?

This criticism misrepresents the intent and value of most tests. A good multiple-choice test necessarily asks questions pertaining only to a small fraction of the content and skills a student is expected to acquire. It is not a check-list of all the things a student should know. Such tests can quickly and reliably sample much more content than a few essay questions. Just as a telephone poll of a small percentage of registered voters can produce reliable estimates of vote totals on election day, so can a multiple-choice test generate a good estimate of a student's overall achievement.

Do tests overemphasize factual knowledge and low-level skills?

Dismissing the importance of factual knowledge is part of the "progressive" education philosophy that lies beneath most objections to testing, but it is a view not shared by parents and employers. Multiple-choice tests, moreover, are used successfully for licensing advanced professions such as law and medicine, providing proof that such tests can measure highly advanced knowledge, understanding, and skills. Tests assessing complex achievement can require respondents to select the best idea from a group of different and compelling positions and identify the best reason for action, the best interpretation of a set of ideas, or the best application of important principles.[8]

Does testing place excessive pressure on students?

Some students find taking tests so stressful that it disrupts their study habits and ability to focus on the questions being asked. Too much focus on a single high-stakes test can be counterproductive even for students who don't react this way. But with these concerns in mind, policymakers can expand or contract testing requirements based on feedback from parents and educators. For example, Texas rolled back from 15 to five the number of tests required for K–12 students in 2013 following protests from educators and parents.[9] Teachers can relieve students' stress by explaining the purpose of the tests is to make learning in the future easier. Giving frequent short quizzes and low-consequence tests can accustom students to taking tests. Remediation efforts and second-chance exams can minimize the degree to

which students feel they are unfairly punished for scoring poorly on a single exam. Experience in taking low- and high-stakes tests helps students for college entry tests as well as job entry tests, professional examinations for entry into law and medical schools, and professional licensing tests.

Does testing threaten or undermine the confidence of teachers?

Teacher opposition to tests often is based on past experiences with poorly designed tests that were difficult to administer or time-consuming to score and interpret, or tests that were not aligned with the curriculum. No doubt, such tests still exist and fuel teacher opposition to proposals to rely more heavily on test results. But over time, as tests are improved and the curriculum is altered to enable students to anticipate questions and answer them correctly, teacher opposition should diminish. Good tests help teachers measure their student's performance, an indispensable part of planning lessons and improving instruction techniques.

Best Practices

In previous chapters, the impact of a specific reward on academic achievement often was measured by changes in the number of students taking an optional test or passing a test required for advancement or their scores. From this literature as well as the work of authors who specifically address the best design of reward systems, we find agreement on best practices that fall into five general categories.

Testing for Grade Promotion

Requiring students to pass tests before advancing them to the next grade level can be a powerful incentive for paying attention in class, reviewing and rehearsing what was learned, and making a focused effort to show progress toward academic goals. Teachers benefit because such tests help ensure that all the students in a classroom at the beginning of a school year have the necessary skills to begin the year at grade level. Such tests are rarely used for each grade promotion decision, though a number of states require students to pass tests to graduate from elementary to middle school and then from middle school to high school. Some states, such as California, require annual tests of all students but prohibit basing grade-to-grade promotion decisions solely on test results.

What if students fail the required tests? A program used by the Chicago

Public Schools gave students the choice between being retained in a grade or taking a sharply focused six-week summer course and then taking a grade-level examination for promotion.[10] Independent researchers found that students who attended the six-week program made as much as a year-and-a-half of achievement progress. At a cost of six weeks of intensive work, the students saved themselves, teachers, and taxpayers a year of inefficient remediation.

Testing for Graduation

Requiring students to pass tests before awarding them a diploma can create positive incentives for improved behavior and performance in high school. Such tests send signals to students, teachers, parents, college admissions officers, and employers that a diploma certifies an accomplished standard of achievement. This belief in turn creates incentives for positive behavior by each group: students work harder to graduate knowing their diplomas are viewed as meaningful achievements by others; parents are given a goal they can encourage their children to reach; and college admissions officers and employers can reward students who earn their degrees by giving them preferential treatment in admissions and hiring decisions.

John Bishop's research on 17 states that require such tests showed they increase mathematics and science achievement and lead to good habits such as completing homework on time and talking with parents about schoolwork. Other researchers are less certain, expressing concern that such tests reduce the motivation of students already struggling with coursework to finish high school.[11] Creating alternative paths to a diploma, as Florida has done (see Chapter 7), is a possible solution to this potential problem.

Rather than requiring all students to complete their final four years of schooling, policymakers should allow advanced students to take high-school exit exams early. Passing comprehensive content-based exams would ensure these students mastered the high-school academic curriculum.[12] Given reasonably rigorous standards, exit exams could reduce the need for remedial instruction in colleges and prepare non-college-bound students for technical training, apprenticeships, or employment.

Test Incentives for Young Children

Young children can be encouraged to prepare for and take tests by receiving points, stickers, praise, and other small rewards for good scores. Tests teach

children that goals and progress toward their achievement are measurable, that measurement can be objective, and that good performance often is rewarded. Rewards can be readily tailored to the interests and needs of children, something parents and observant teachers are usually able to do.

One stay-at-home mom capitalized on children's enthusiasm for token rewards by developing a point system to reward her children and their friends for passing short comprehension tests on books they read at their leisure. Her project grew to become the Accelerated Reader Program, which is available in about half of the schools in the U.S.[13]

Accelerated Reader consists of a search system that helps teachers select books suited to a student's interests and reading ability, five-item tests on more than 30,000 books generally available in school and community libraries, and a computer test with instant scoring. As students become stronger readers, they are encouraged to read more difficult books because points are awarded in proportion to the difficulty of the book and the number of correct answers. The program, which costs only $100 per year per student, keeps a detailed record of each student's progress, minimizing the amount of bookkeeping required by teachers and parents. While not intended to take the place of in-class instruction, Accelerated Reader has shown excellent results in a number of rigorous (control-group) studies conducted by independent scholars.[14]

Test Incentives for Adolescents

As children move into adolescence, rewards must become larger to account for the growing "wealth" of the students, the rising value of their time, and their ability to defer gratification and be motivated by rewards that are further in the future. Token awards and public recognition attached to tests can still incentivize academic progress, but financial rewards and increased responsibility such as permission to use a car or to choose other activities may be more motivating.

In Chapter 3 we reviewed research by John Bishop, Victor Lavy, and Roland Fryer Jr. on how adolescents respond positively to being paid to take or score high on Michigan's Educational Achievement Program (MEAP) tests, Advanced Placement exams, and the New York Regents Examination. In Chapter 7 we reported how KIPP Academies pay middle-school students with "KIPP dollars" for effort, good behavior, and completed homework. We also saw how strategies for "self-teaching" reward students by lowering the students' cost of learning and allowing

them to compete with other students in learning programs such as Lego Mindstorms.

Many teenagers are able to see a connection between the tests they must take and possible careers, such as positions in medicine, science, or law enforcement. Some charter schools have shown remarkable success by getting adolescents to focus on the importance of attending college after graduation and the need to meet college admission standards. Students who are not college-bound can be incentivized to stay in school and graduate by giving them a path to a high-school diploma that gives them the experience or industry certifications they need to get the jobs they want after graduating. Florida's three paths – standard, scholar, and merit – show how this can be done without lowering graduation requirements.

Tests and Advanced Accomplishment

Grade-level advancement and high-school graduation exams typically establish low thresholds that students must pass rather than high targets to which they should aspire. Advanced academic accomplishment such as honors courses, showing high foreign language proficiency, and passing Advanced Placement (AP) exams requires much more from students. Programs that encourage students to make extra effort include making access to such courses convenient and free, teaching them in especially comfortable or high-status classrooms, and supplying tutoring and other support activities. Schools can support student clubs based on the coursework and interests of high-achieving students, such as debate and foreign languages. Tournaments, class trips, parties, and other activities organized for high achievers can take the place of some of the school activities dominated by lower-achieving students.

Most students and parents are aware of AP programs and the possibility of attaining college credits, but they do not fully comprehend how large the financial rewards can be when a student places out of one or more college courses. With ever-rising college tuition levels, the savings can reach $40,000 or more. Teachers and school administrators can make this incentive much more powerful by documenting the dollar value of college credits and describing how much past students and their parents saved by taking the exams.

Students and teachers also can be motivated by modest financial awards each time a student passes an AP exam, as the O'Donnell Foundation's program in Dallas demonstrated. Recall from our discussion of the program

in Chapter 3 that paying students and teachers $100 for each passing score and providing a $2,500 stipend to teachers who undergo training to teach the courses increased the number of students who passed AP exams more than 12-fold.[15] Large-scale expansion of the O'Donnell example would seem likely to reduce the time students require to gain diplomas as well as college degrees, and in the process significantly cut the costs of their education.

Adaptive Online Testing

As mentioned earlier, new technology is reducing the cost of testing and saving teachers time by automating test administration, scoring, and interpreting the results. This is part of a bigger transformation technology is bringing to K–12 education that we describe in greater depth in Chapter 10. The following observations focus on how online adaptive testing is giving testing with rewards greater importance than ever before.

Adaptive online tests are typically administered via the Internet from a central site outside the classroom and school. Students take the tests alone or in a group depending on the number of Internet-connected computers available to the class. Large numbers of test questions are stored in databases and a computer program uses a student's performance on each item to select the next item. Each correct answer leads to a more difficult question while wrong answers lead to easier questions. This process quickly reveals the student's ability level, reducing the number of questions a student has to answer and cutting testing time by as much as two-thirds.

Adaptive testing has been used for some years for college, graduate, and professional school admission, providing valuable experience for the companies that now offer the service to K–12 schools. Where paper-and-pencil tests may take weeks to be scored and returned, adaptive tests provide instant scores and quick, detailed reports for teachers to pinpoint gaps in students' understanding. Schools can receive reports comparing each student's performance with class, school, district, and state results. Because adaptive tests take less time, they may be given more frequently. Students cannot cheat as easily on adaptive tests as they can on traditional paper-and-pencil tests since nearby students are usually answering different questions.[16]

The Khan Academy, a free library of online videos and problem sets, uses an adaptive learning system for its widely praised tutorials on mathematics and other subjects.[17] Students watch short videos on core

mathematics concepts, starting at the beginning with addition, subtraction, and the idea of numbers. They then take tests to determine how well they have understood the material. The software tracks the errors the student makes and cues up the next set of videos and concepts to target the student's weaknesses or to move on to more advanced lessons. As students progress, they earn badges and points for concepts they have mastered.

Independent Test Audits

A critical problem facing those who would align rewards with test results is the temptation faced by policymakers and educators to make the tests easier to create the appearance of progress or to hide declining performance. The reliance of No Child Left Behind on state-designed tests made this problem especially apparent as states lowered their standards and schools manipulated the administration of tests in order to meet the law's achievement requirements.

Many states make their tests easier over time to produce the appearance of student progress. Others allow the tests to remain unchanged and their content items become better known in schools, thus enabling "teaching to the test" and making it appear that student achievement is rising. Still other states introduce a new curriculum and set of tests every three or four years, conveniently making it impossible to compare academic progress over periods of time that would show the actual trends in student achievement. And if all this weren't bad enough, some educators actually conspire and cheat to avoid reporting a lack of progress.[18]

To minimize such problems, accounting and auditing procedures analogous to those in business could be established for schools, districts, and whole states (or at least problematic ones). Independent auditing firms could review and report on plans and the execution of testing programs. When fraud is detected, perpetrators could face the same types of consequences as those in business – that is, firing and prosecution for fraud. In schools suspected of fraud, the independent auditors would administer the tests as well.

Conclusion

Critics of tests present a litany of objections, many of them based on past experiences with poorly designed tests, but ways to improve tests are well specified. Some students respond poorly to high-stakes exams, but there are ways to mitigate their anxiety. Simply letting them avoid tests during their

K–12 careers isn't doing them any favors since tests with consequences will be a challenge they will face later in higher education and adult life. Some states and schools have manipulated their tests in order to avoid federal sanctions or to escape accountability to voters, but this is only an argument for relying on national tests or turning to independent agencies to administer and audit testing.

Tests are essential if rewards are to be used to accelerate learning since tests measure progress toward academic goals. The extensive research reviewed by Richard Phelps shows how setting goals and measuring progress increases student motivation and performance, frequent testing results in greater learning, and giving students detailed analyses of their test results helps them identify their weaknesses and increases their learning. Many cities and states are using tests with rewards successfully in five general areas: for grade advancement, for graduation, with young children, with adolescents, and for advanced academic achievement. Adaptive online testing and independent test audits promise to cut the time and cost of testing and to boost dramatically the power of rewards to accelerate learning.

Notes

1. Alfie Kohn, *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes* (Boston, MA: Houghton Mifflin, Co., 1993), p. 203.

2. Monty Neil, testimony, June 23, 2009, http://blip.tv/film-our-way-films/monty-neil-mcas-hearing June-23-2009-2547640.

3. Richard P. Phelps, "Why Testing Experts Hate Testing," in Williamson M. Evers and Herbert J. Walberg, eds., *Testing Student Learning, Evaluating Teaching Effectiveness* (Stanford, CA: Hoover Institution Press, 2004), p. 29.

4. Thomas Ahn and Jacob L. Vigdor, "Were All Those Standardized Tests for Nothing? The Lessons of No Child Left Behind," American Enterprise Institute, May 2013.

5. This section is adapted from Herbert Walberg, *Tests, Testing, and Genuine School Reform* (Stanford, CA: Hoover Institution Press, 2011), Chapter 2.

6. Caroline M. Hoxby, "The Cost of Accountability," in Williamson M. Evers and Herbert J. Walberg, eds., *School Accountability* (Stanford, CA: Hoover Institution Press, 2002), pp. 47–74.

7. See Herbert Walberg, *supra* note 5, Chapter 9.

8. Norman E. Gronlund and C. Keith Waugh, *Assessment of Student Achievement, 9th edition* (Boston, MA: Allyn and Bacon, 2008).

9. Erik Robelen, "Testing Rolled Back, Graduation Path Revised Under Texas Bill," *Education Week*, May 28, 2013.

10. Melissa Roderick, Jenny Nagaoka, and Elaine M. Allensworth, "Is the Glass Half-full or Mostly Empty? Ending Social Promotion in Chicago," in Edward H. Haertel and Joan Herman, eds., *Uses and Misuses of Data for Educational Accountability and Improvement; The 104th Yearbook for the National Society for the Study of Education, Part 2* (Malden, MA: National Society for the Study of Education, 2005), pp. 223–59.

11. Center on Education Policy, "State High School Exit Exams: A Policy in Transition," Graduate School of Education and Human Development, The George Washington University, September 2012.

12. See the discussion of curriculum-based external exams in Chapter 3.

13. Details on the company and the program can be found at http://www.renlearn.com/ar/.

14. Keith J. Topping and Terry D. Paul, "Computer-Assisted Assessment of Practice at Reading: A Large-Scale Survey Using Accelerated Reader Data," *Reading and Writing Quarterly* 15 (1999): 213–31; Stacy R. Vollans, Keith J. Topping, and Ryka M. Evans, "Computerized Self-Assessment of Reading Comprehension with the Accelerated Reader: Action Research," *Reading and Writing Quarterly* 15 (1999): 197–211; John A. Nunnery, Steven M. Ross, and Aaron McDonald, "A Randomized Experimental Evaluation of the Impact of Accelerated Reader/Reading Renaissance Implementation on

Reading Achievement in Grades 3 to 6," *Journal of Education for Students Placed at Risk* 11, no. 1 (2006): 1–18.

15. C. Kirabo Jackson, "Cash for Test Scores: The Impact of the Texas Advanced Placement Incentive Program," *Education Next* 8, no. 4 (Fall 2008). See also Herbert J. Walberg, "Incentivized School Standards Work," *Education Week*, November 4, 1998, p. 48.

16. Cheating is still possible, though. See Cameron McWhirter, "High-Tech Cheaters Pose Test," *The Wall Street Journal*, June 11, 2013, p. A3.

17. See www.khanacademy.org.

18. For example, see Ned Resnikov, "Atlanta Cheating Scandal Puts National Education Policy on Trial," MSNBC, http://tv.msnbc.com/2013/04/01/atlanta-cheating-scandal-puts-national-education-policy-on-trial/, viewed on June 10, 2013.