

## SEEKING REDEMPTION FOR OUR PSYCHOMETRIC SINS\*

University of California, Los Angeles  
W. James Popham

I had a Catholic upbringing and, therefore, learned loads about sin. During countless catechism classes, taught by stern but kindly nuns, I learned to master such sin-related concepts as the distinction between *venial* sins (minor transgressions) and *mortal* sins (genuine whoppers). I also learned that a person's sins can be forgiven if sinners, at least Catholic ones, carry out a suitable *penance*.

Because the following essay revolves chiefly around *penance*, I need to clarify the meaning of that concept. According to my dictionary, and it is a secular one, penance is "a punishment undergone as penitence for sin." Penance, even when I was a child, always struck me as an eminently equitable kind of concept. Those who "do the crime" ought to "pay the time." And, as I learned in my catechism classes, the bigger the sin, the more significant the penance should be.

I submit, therefore, that because members of the educational measurement community have, during recent decades, committed some truly heavy-duty sins, they need to undertake some heavy-duty penance. Fair, after all, is fair.

### Assessment-Sired Sins

Many boys and girls in our nation's schools are currently receiving decisively lower quality educations as a direct consequence of the high-stakes tests now found in almost every state. The high-stakes tests to which I refer are those traditionally constructed achievement tests being used (1) to make important decisions about students or (2) to evaluate the teachers who taught those students. These traditionally constructed tests are either nationally standardized, "off-the-shelf" achievement tests or, increasingly, state-customized achievement tests. But even a state-specific, so-called "standards-based" test, despite its alluringly positive label, is apt to have been created according to a patently traditional measurement model.

A traditionally constructed test is one whose chief measurement mission is to provide a meaningful comparison between one test-taker's score and the scores of a previous group of test-takers, that is, the test's norm group. The comparative approach to measurement reflected in those traditional tests has an important role to play in education, because relative information about a student's status can be useful to the student's teachers as well as to the student's parents. However, a comparative approach to measurement is dead wrong for evaluating the quality of instruction. As a consequence, serious *instructional sins* are usually encountered in any classroom influenced by the presence of a traditionally constructed high-stakes achievement test.

---

\* A presentation at the 2002 CRESST Conference, Los Angeles, California, September 10-11, 2002.

These *instructional* offenses, then, are actually spawned by *assessment* offenses. Let's look at three of these instructional sins.

*Curricular reductionism.* The first assessment-sired instructional sin against students occurs in the form of dramatically reduced classroom attention to any curricular content not assessed by a locally operative high-stakes test. Skills and knowledge, that teachers once regarded as imperative for students to learn are either given short shrift or, more often, no shrift at all. And such curricular omissions occur *only* because teachers are so pressured to improve their students' scores on high-stakes tests that there's no time left over to address untested, albeit important skills and knowledge.

The result is that thousands of students are now being curricularly short-changed. They aren't learning what they ought to be learning. And this increasingly prevalent instructional sin is directly attributable to our psychometric sin of allowing schools to be evaluated with data derived from the wrong tests.

*Jettisoned joy.* A second assessment-caused instructional sin can be seen in many of today's classrooms where teachers have been pressured to boost their students' scores on a high-stakes test. To accomplish such score-boosting, some teachers have transformed their classrooms into drudgery-dominated drill factories wherein students must take part in seemingly interminable test-preparation practice sessions. In a proper educational world, students should find classroom instruction downright fascinating. Students can, and often should, derive genuine joy from what they are learning in class. Students can, and often should, develop a lasting interest in learning because of what goes on in their classrooms.

But none of these positive dividends of classroom instruction will be seen if students must undergo relentless test-preparation training. Because of today's pressures to raise scores on high-stakes tests, far too many teachers have abandoned the very activities that could foster students' positive affect toward learning. Learning-linked joy has been jettisoned in favor of drill-dominated test preparation. It is a terrible trade-off. It is an instructional sin, but it can be traced to our psychometric sin of allowing schools to be evaluated with the wrong tests.

*Modeled dishonesty.* A third assessment-rooted instructional sin seen in today's classrooms takes place when teachers, not students, cheat. Again, because of pressures to promote improved student scores on high-stakes tests, increasing numbers of teachers are engaging in test-preparation or test-administration practices that are simply dishonest. For instance, some teachers have copied the actual items from a currently operational form of a high-stakes test, then given their students practice—along with a correct-answer key—based on those items. (Talk about “guided practice!”) Later, when the high-stakes test is actually given, students discover that they have been made unwilling accomplices in a teacher-engineered fraud. We also see increasingly frequent instances of test-administration rule-breaking wherein teachers (1) give students more test-completion time than what's prescribed or (2) actually go around the room pointing out students' erroneous answers that should be “re-thought.”

We can understand why teachers might succumb to such inappropriate conduct. It's because traditionally constructed tests typically provide altogether inadequate descriptions (at least for purposes of instructional decision-making), of the skills and knowledge they attempt to assess. As a result, many test-pressured teachers really don't know how to go about promoting students' higher scores on such comparatively oriented traditional achievement tests. The teachers simply don't know where to aim their instruction.

But what an awful thing it is to have teachers modeling blatant dishonesty for their students. If we still believe that, while in their classrooms, teachers ought to function *in loco parentis*, then we don't want these proxy parents to provide their students with models of deceitfulness. Such conduct is seriously damaging to students. It is, for certain, another instructional sin. But, as noted before, this sin is also caused by our psychometric sin of allowing schools to be evaluated with the wrong tests.

Taken together, then, the instructional sins of curricular reductionism, joy eradication, and modeled dishonesty are currently having a profound negative effect on the children in our schools. And, because all of these problems are assessment-engendered, all members of the measurement community need to think carefully about their personal contributions to this increasingly serious educational calamity. These classroom instructional shortcomings arise, I believe, because of the flat-out sinfulness of psychometricians. And, if I'm correct, then the real culprits in today's widespread degradation of education are not classroom teachers but, instead, educational measurement specialists.

## **The Sounds of Silence**

When an individual commits an overt sin such as telling a lie, robbing a store, or murdering someone, it is all too clear that something reprehensible has been done. *Sins of commission* are usually easy to spot.

But *sins of omission* can also have serious consequences. And those are the kinds of sins that the educational measurement crowd has been committing during recent decades. We have been silent while certain sorts of assessment tools, the very assessment tools that we know the most about, have been misused in ways that harm children. We have sat on our duffs while traditional, comparatively oriented achievement tests have been increasingly employed to evaluate schools and, as a consequence, to lower educational quality. And I do not excuse self, for I was right in their duff-sitting along with the rest of the gang. We have sinned by omission against the nation's children.

The comparatively oriented educational achievement tests that sprang up in the U.S. soon after World War I, then flourished in subsequent years, accomplished an altogether praiseworthy measurement mission. Such tests allow an individual examinee's performance to be contrasted with the performances of a prior group of examinees. These "norm-referenced" inferences regarding a student's relative

performance, as I indicated earlier, provide both educators and parents with useful comparative interpretations.

Yet, the very test-construction procedures that render traditional tests most suitable for such comparative assessment actually inhibit the usefulness of these tests for purposes of evaluating educators' effectiveness.\* And yet, because off-the-shelf nationally standardized tests always came to us from first-rate measurement organizations and were accompanied by all sorts of supporting technical data, many educational policymakers regarded such tests as suitable assessment instruments for evaluating schools. And, by our silence, we let them think it.

Most of the measurement-caused problems now seen in American schools arose because members of the measurement community were mute during the years that the evaluative misuses of traditional tests become increasingly common. And that was a major measurement sin. Even though only a sin of omission, it was a truly mortal one.

If you prefer a somewhat more secular approach to psychometric matters, simply consult the most recent version of the *Standards for Educational and Psychological Testing*.\*\* You will see that Validity Standard 1.4 asserts "if a test is used in a way that has not been validated, it is incumbent on the user to justify the new use, collecting new evidence if necessary." Well, there may be scads of empirical evidence indicating that valid *relative* inferences can be drawn from traditionally constructed tests, but try lining up solid empirical evidence such tests can provide valid *evaluative* inferences about a school staff's instructional effectiveness. Lots of luck!

### **For Your Penance**

Way, way back when I was a youngster, after I had made my routine Saturday-night confession to our parish priest, he would first give me absolution for my sins (which became increasingly more serious soon after I hit adolescence) then say something such as: "For your penance, say 10 Hail Marys." Thus, to really make amends for my sins, I needed to say, at least mentally, the Hail Mary prayer 10 times.

Well, for our past sins of omission, and for the resultant educational harm our silence has caused for so many students, I believe that members of the measurement community must now pay a proper penance. And for that penance, I propose that all of us *need to promote increased assessment literacy* on the part of educational policymakers, educational practitioners, the public, and especially parents of school-age children. A more assessment-literate citizenry is less likely to permit measurement misuses to take place. A more assessment-literate citizenry will be able to push for the kinds of educational tests that not only supply accurate accountability evidence by

---

\* I address this issue in greater detail in Popham, W. James, *The Truth About Testing: An Educator's Call to Action*, Association for Supervision and Curriculum Development, Alexandria, Virginia, 2001.

\*\* American Educational Research Association. *Standards for Educational and Psychological Testing*. Washington, DC: Author, 1999.

which educators can be evaluated, but also can support improved instructional practices by those educators.\*

Let me close out this sin-soaked analysis by listing several practical forms of penance through which members of America's educational measurement community can promote assessment literacy for those who should possess it. I realize that different people's conceptions of what constitutes assessment literacy will vary, but I'm thinking about such basic considerations as (1) educational measurement's imprecision, (2) ways to judge the suitability of an educational test, (3) proper uses of certain kinds of tests, and so on. In short, I'm thinking of the sorts of *practical* measurement-related information that citizens, and especially parents of school-age kids, need to possess.

### **Pick Your Penance**

To promote assessment literacy, therefore, I suggest the following possibilities (not listed in order of import):

1. **Magazine Articles.** *Writing articles about educational measurement fundamentals for general-audience magazines, especially those magazines apt to be read by parents.*
2. **Op/Ed Essays.** *Authoring op/ed essays about educational assessment for newspapers, not just an author's local newspaper.*
3. **Oral Presentations.** *Making oral assessment-focused presentations to varied audiences, for example, teachers and administrators (who can, in turn, communicate to parents), members of school boards, legislators, parents, and so on.*
4. **Isolating Relevant Resources.** *Identifying resources (books, articles, videos, audiotapes, etc.) appropriate for the promotion of assessment literacy, then relaying a description of such resources to colleagues who, if they possess proper "penitential proclivities," might personally use such resources in promoting assessment literacy.*
5. **Assessment-Literacy Councils.** *Establishing state and local assessment-literacy councils whose chief mission would be to promote greater understanding about educational measurement among the general populace.*
6. **Organizational Allies.** *Stimulating the interest of education-related organizations, for instance, school or district PTA groups, in carrying out assessment-literacy enhancement activities.*

---

\* Illustrative of the kinds of educational assessments to which I refer are the sorts of tests advocated by the Commission on Instructionally Supportive Assessment. *Building Tests That Support Instruction and Accountability: A Guide for Policymakers*. Washington, DC: Author, 2001.

7. **Materials Development.** *Joining forces with like-minded assessment personnel to secure foundation or governmental funding to create materials, e.g., videos or internet interventions, that can then be used by others in the promotion of assessment literacy.*
8. **Assessment Conferences.** *Organizing conferences for citizens, especially parents and policymakers, unabashedly aimed at the enhancement of attendees' assessment literacy.*

I am certain that there are other, more inventive ways to promote assessment literacy. However, the eight suggestions cited above should convey a general idea of the sorts of penance I have in mind. In all such promotional efforts, of course, one key theme must be to employ educational tests properly. If more citizens begin to understand how educational tests should and shouldn't be used, it seems certain that, over time, we will see a reduction in the sorts of assessment-triggered instructional problems described earlier.

Similarly, if more educational administrators and classroom teachers understand how today's high-stakes tests are being misused, then they can become an *informed* constituency whose views, properly rooted in measurement realities, can become more influential. Yet, because the accountability-related views of educators will always be regarded as potentially self-serving, the promotion of greater assessment literacy in the population at large is likely to have a more positive impact on educational practice.

## Choices, Choices

If certain members of the educational measurement community haven't really recognized the nature of the issue I've been describing, then their ignorance of this situation essentially lets them off the hook. If someone doesn't recognize an act as sinful, if I remember correctly from my catechism days, this markedly diminishes that person's culpability.

However, once you've finished reading this analysis, culpability-reducing ignorance can no longer be claimed. What this means, in straightforward language, is that *it's too late for you*. Thus, if you don't tackle some of the penance options I've cited here, you are surely headed straight down the "sinning by omission" trail toward perdition. And, because your psychometric sins of omission will most likely lead to all sorts of classroom wrongdoing, the quality of your afterlife is, obviously, in serious jeopardy.

Just as Sisyphus, who was assigned in Hades to eternally roll a stone toward the top of a slope, whereupon the stone always rolled back down again, you might end up in the same lower world and be obliged to engage in some form of never-ending punishment. For instance, you could be required to *eternally* explain to bystanders "the nature of construct-related validity." Consider that sort of eternity, and you'll surely choose some sort of penance-paying option!