



## Test Use and Abuse

Christopher H. Tienken

To cite this article: Christopher H. Tienken (2015) Test Use and Abuse, Kappa Delta Pi Record, 51:4, 155-158, DOI: [10.1080/00228958.2015.1089617](https://doi.org/10.1080/00228958.2015.1089617)

To link to this article: <http://dx.doi.org/10.1080/00228958.2015.1089617>



Published online: 14 Oct 2015.



Submit your article to this journal [↗](#)



Article views: 81



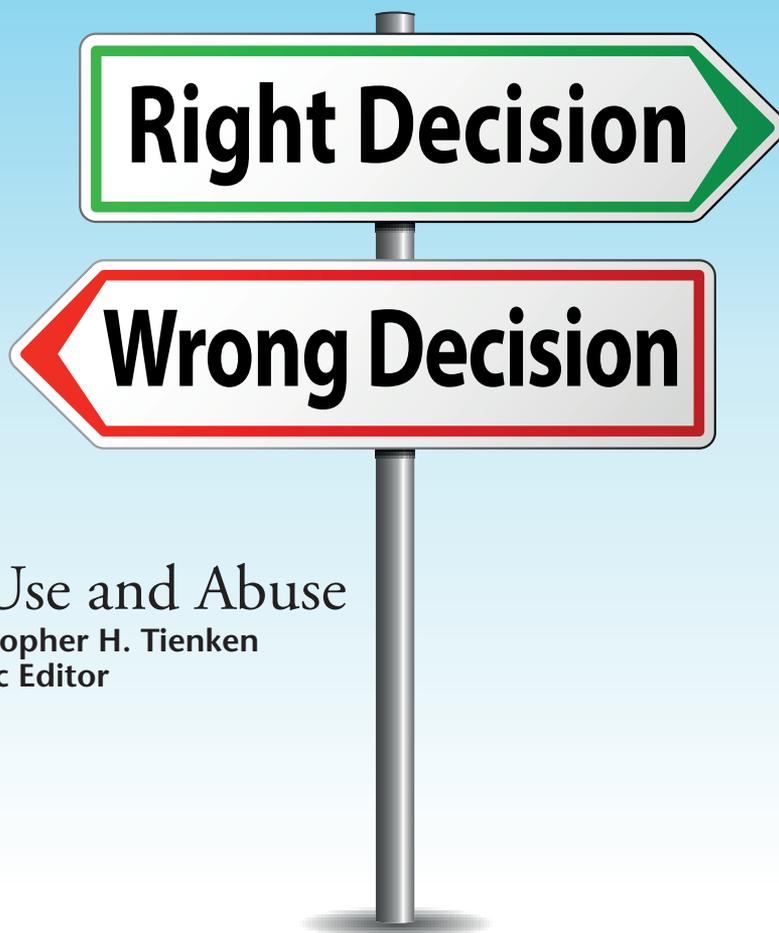
View related articles [↗](#)



View Crossmark data [↗](#)

## For the Record

*Kappa Delta Pi Record*, 51: 155–158, 2015  
 Copyright © Kappa Delta Pi  
 ISSN: 0022-8958 print/2163-1611 online  
 DOI 10.1080/00228958.2015.1089617



Test Use and Abuse  
 by Christopher H. Tienken  
 Academic Editor

### Abstract

*The ubiquitous use of standardized test results to make varied judgments about educators, students, and schools within the public school system raises concerns of validity. If the test results have not been validated for making multiple determinations, then the decisions made about educators, students, schools, and school districts based on the results could be flawed. The author discusses specific standards that guide the research-based uses of test results.*

**Key words:** *assessment, high-stakes testing, school reform, standardized testing*

More than 40 states have been awarded Race to the Top (RTTT; U.S. Department of Education, 2009) competitive grants or No Child Left Behind (NCLB, 2003) waivers from the

## For the Record

U.S. Department of Education since 2010. Each of those programs has required the use of yearly student results from state-mandated standardized tests of mathematics and English language arts to rate the effectiveness of teachers, principals, and assistant principals. Provisions embedded in each program also have required that state education policy makers use standardized test results yearly as part of their public school education accountability schemes to determine whether students in Grades 3–8 and high school are college and career ready. Clearly, standardized test results play an important role in the education reforms driven by both programs.

For example, the Great Teachers and Leaders category of the RTTT competitive grant program application required grant recipients to use results from tests aligned with college- and career-ready standards as part of teacher and principal evaluation schemes. The Great Teachers and Leaders category was weighted the heaviest within the application, worth 138 of the application's 500 possible points (U.S. Department of Education, 2009, p. 3). The Great Teachers and Leaders category contained five subcategories, each worth a certain portion of the points. Improving Teacher and Principal Effectiveness Based on Performance (worth 58 points) was the heaviest weighted subcategory in the entire application (U.S. Department of Education, 2009, p. 3).

Additionally, school administrators in some school districts use results from state-mandated standardized tests to make determinations about student placement into various academic tracks and programs, such as remedial courses, gifted and talented programs, or honors level and Advanced Placement courses. Also, bureaucrats in more than 20 states have mandated the use of results from state standardized tests in

high school as exit exam criteria for graduation. The test results are even used as part of school district quality ratings for federally and state-mandated performance reports.

The results from state-mandated standardized tests are used to make multiple determinations and interpretations about teachers, school administrators, students, and school quality. In most cases, state education bureaucrats use the results from one mandated standardized test in mathematics and one test in English language arts for multiple purposes to meet the various RTTT grant and NCLB waiver reporting requirements for teacher and principal effectiveness as well as college and career readiness for students. In a state like New Jersey, the results from the state-mandated high school mathematics test in Grade 11 could be used to make determinations about (a) the effectiveness of the high school principal, (b) the effectiveness of the high school math teachers, (c) the quality of the school district's mathematics program, (d) whether a Grade 11 student is college ready, (e) whether that student is career ready, (f) a student's strengths and weaknesses in math, (g) Grade 12 course placements for that student, and (h) whether the student can graduate high school. That is eight determinations made totally or in part from one test score.

As one can imagine, if the test results have not been validated for making multiple determinations, then the decisions made about educators, students, schools, and school districts that are based on the results could be flawed. A current example includes the use of state test results to rank schools and school districts and to reward and punish them. As I (Tienken, in press) elaborate upon in *Education Policy Perils: Tackling the Tough Issues*, results from state standardized tests can be predicted with a great deal of accuracy at the school and district levels, using

only community demographic data. Some school and district educators are needlessly critiqued, replaced, or put on corrective action while others receive praise, all based on test results that have not been validated for making those types of determinations.

In this essay I argue, based on professional standards for testing, that school administrators and state bureaucrats should ensure that the test results they use for multiple determinations have been validated for those determinations. They also should use multiple indicators when making decisions about students and other educators.

## Professional Standards for Testing

The seventh edition of the *Standards for Educational and Psychological Testing*, developed by a joint committee represented by members of the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), contains 12 categories of standards and provides specific guidance on topics that include appropriate test design, development, validity, and use of standardized tests and results (AERA, APA, & NCME, 2014). The standards also describe the rights and responsibilities of test takers.

## General Guidance

Standard 1.0 provides general guidance regarding validity of results for uses related to various types of standardized testing contexts such as employment, education program placement, college entrance, and diagnostics. The Standard states, "Clear articulation of each intended test score interpretation for a specified use should be set forth, and appropriate validity evidence in support of each intended interpretation should be provided" (AERA et al., 2014, p. 23). Standard 1.1 expands on this guidance:

"No test permits interpretations that are valid for all purposes or in all situations. Each recommended interpretation for a given use requires validation" (AERA et al., 2014, p. 23). Standard 1.1 further recommends, "A rationale should be presented for each intended interpretation of test scores for a given use, together with a summary of the evidence and theory bearing on the intended interpretation" (AERA et al., 2014, p. 23).

Based on general guidance from Standard 1.0 and Standard 1.1, state education bureaucrats and school administrators should provide transparent evidence of validity for each way test results would be interpreted. The guidance also implies that using results from one test for multiple interpretation purposes might not be valid. It is incumbent upon those who use the results from one test in multiple ways to present appropriate evidence that demonstrates the results from the test can be used in those ways.

For example, using a standardized test administered in Grade 3 to determine college and career readiness would potentially require a validation period of 8 years for the college readiness determination and perhaps longer for career readiness validation. College readiness and career readiness are two different determinations and require two separate validations of the test results to make those determinations. Similarly, one might argue for more evidence of validity in the case where an elementary school principal receives an ineffective rating based on school standardized test scores while the majority of her teachers are rated effective via the same test results. Again, a determination about school administrator effectiveness is different than teacher effectiveness, yet the same results are used for dual purposes. In many states, the state test



**Christopher H. Tienken** is an Assistant Professor at Seton Hall University in the College of Education and Human Services, Department of Education Leadership, Management, and Policy. He is a former public school teacher and administrator. You can contact him at [christopher.tienken@shu.edu](mailto:christopher.tienken@shu.edu) and [www.christienken.com](http://www.christienken.com).

The opinions expressed in this editorial are those of the author and do not reflect any official position of Kappa Delta Pi, International Honor Society in Education.

## For the Record

results have not been validated to make those dual determinations.

The authors of the standards present specific cautions about using results from standardized tests for multiple purposes in educational settings like P–12 public schools. Standard 12.2 states, “In educational settings, when a test is designed or used to serve multiple purposes, evidence of validity, reliability/precision, and fairness should be provided for each intended use” (AERA et al., 2014, p. 195). The authors go on to explain, “Choices in test design and development that enhance validity for one purpose may diminish validity for other purposes” (AERA et al., 2014, p. 195).

In other words, a test designed to measure the effectiveness of a school principal may not be valid for measuring the effectiveness of a classroom teacher. The authors state clearly that one test cannot be a valid measure of multiple complex behaviors or tasks such as the effectiveness of a teacher, the effectiveness of a principal, and the college and career readiness of a student: “No one test will serve all purposes equally well” (AERA et al., 2014, p. 195).

### Three-Legged Stool

The authors caution in Standard 9.13 that standardized test results “should not be interpreted in isolation” (AERA et al., 2014, p. 145). Users of standardized test results should attempt to confirm the results for groups and individuals by obtaining multiple forms of data about those groups or individuals. The three-legged stool provides an appropriate visual representation that state bureaucrats and school administrators might want to keep in the backs of their minds as they make policies and practices that use standardized test results to make important determinations about educators and children.

Data from various sources should be triangulated so that a decision is not made based only upon the results from a state-mandated standardized test.

### Kick the Habit

Teachers and school administrators should resist the urge to rely too heavily on the results from state-mandated standardized tests or any one test. In many cases, the results from the standardized tests being used across the country have not been fully validated for the determinations being made with them. Because most teachers and administrators have been trained within an increasingly standardized education environment, they sometimes get hooked on state-mandated standardized test results.

An interesting exercise for educators in a school could be to develop a menu of other indicators one could use to make important decisions about students and teachers without using any results from state-mandated standardized tests. They could create a simple matrix with the type of determination to be made listed on the left side of the matrix and all the existing sources of data at hand running along the top of the matrix. Then they would be able to easily identify determinations that lack at least three different types of data. That could help alert educators to the types of assessments they might have to develop in-house. More importantly, the exercise will help educators kick the habit of using results from one test for multiple purposes for which the test was not designed. ■

### References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing* (7th ed.). Washington, DC: AERA.
- No Child Left Behind Act of 2001, 20 U.S.C.A. § 6301 et seq. (West 2003).
- Tienken, C. H. (in press). Standardized test results can be predicted, so stop using them to drive education policymaking. In C. H. Tienken & C. A. Mullen (Eds.), *Education policy perils: Tackling the tough issues*. New York, NY: Routledge.
- U.S. Department of Education. (2009). *Race to the top program: Executive summary*. Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>