



October 29, 2025

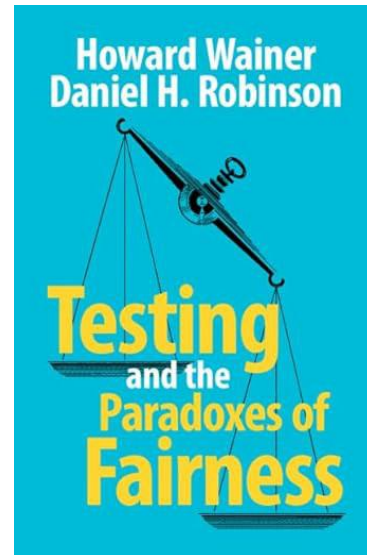
ISSN 1094-5296

**Wainer, H., & Robinson, D. H. (2025). *Testing and the Paradoxes of Fairness*. Cambridge University Press.**

231 pp.

ISBN: 1009576828

**Reviewed by Kurt F. Geisinger**  
**University of Nebraska–Lincoln**  
**United States**



I agreed to review this book having known Dr. Howard Wainer for many years and having met Dr. Daniel H. Robinson this past fall when he gave a lecture at my university. Wainer is without a doubt one of the most impactful members of the psychometric profession—knowledgeable, creative, erudite, and highly influential. I am much less familiar with Professor Robinson, but by many accounts he is a strong defender of testing and the use of tests. As a quantitative psychologist and educational researcher, I have learned from my qualitative colleagues that it is important to lay one's values on the table so that others can judge one's comments against

their orientations and perceptual lenses. I have directed the Buros Center for Testing for 20 years, served on the SAT Committee, chaired the College Board's Research Committee, and chaired both the GRE Board and its Technical Advisory Committee. I also built and consulted on many of the New York City Police and Fire departments civil service tests, including dozens of promotional examinations, for more than a decade and defended many of them in court. On the other hand, I was an expert witness against the United States Medical Licensing Test on behalf of a disabled medical student, the Medical School Admissions Test on behalf of students with learning disabilities, the SAT on behalf of students with learning disabilities, the California High School Exit Examination on behalf of ethnic minority students and students with disabilities, and the Law School Admission Test on behalf of two blind students. I would like to believe that I am balanced, a position you may dispute after reading this review. I believe strongly in the value of testing, and I also believe strongly in our need for diversity in the institutions of our country.

Wainer and Robinson's volume is a major and important book. It provides strong advocacy for several kinds of test and their use. As will become clear, I had a somewhat challenging time identifying, clarifying, or titling their orientation, and it seems to vacillate in places. The authors appear to attempt to change the opinions of what seems to be the majority opinion of those reading the book. Who is the audience or who is intended to read the book? That is not clear to me, even after having read the book carefully.

I began reading this book and agreeing with the authors that two of the most influential books in test theory were Harold Gulliksen's *Theory of Mental Tests* (1960) and Lord and Novick's *Statistical Theories of Mental Test Scores* (1968). I wondered initially whether Wainer and Robinson's book might join that group of legendary scholarly works. It does not, if only because it appears less of a psychometric book and more of an instructional book, attempting to educate the reader about the empirical facts of some studies on controversial testing topics, all based on science and the authors' orientations. So, who should read the book? It might parallel, for example, Bok and Bowen's *The shape of the river*.

First, it should be required reading for those considering entering the fields of psychometrics, educational testing, college admissions, and university leadership. It would provide them with a useful perspective, whether they adhere to that orientation or not. It could be usefully read by those in the field as well, especially if they have not focused their work on admissions or licensing testing. The focal group who are to learn from the book are an assembly of the intelligencia, university administrators, admissions professionals, those concerned about licensing testing, and policy wonks. I respect the points of the book, but I also fear that it may fall into the hands of far-right policy gurus who will take some of the recommendations to an illogical conclusion, much as the Nixon White House did in attempting to kill Head Start after Arthur Jensen (1969) wrote his seminal *Harvard Educational Review* article.

The authors address several issues related to testing, with the primary focus on admissions and licensing tests. The chapters include brief statements on the history of testing, why tests are disliked, the origins of group testing in the U.S. military, testing in K–12 settings, licensing examinations, admission testing in higher education, tests for awarding scholarships and other prizes, tests for evaluating public school teachers, tests and subtest scores, three controversial issues in testing—which they call Zombies—and a brief summary chapter. While the listing of chapters makes the book appear to cover the range of what has traditionally been educational testing, this reviewer believes that the focus is primarily admissions testing and related uses of admissions test scores, and to a lesser extent, licensing testing. This pairing is not surprising, because Wainer spent most of his professional life on those two topics.

The authors take some positions that are controversial. One of these is related to statewide and federal requirements to test for accountability, wherein the authors provide evidence that such measures do not work as intended. However, they also hold that value-added modeling, as is used in many states to evaluate teachers by the extent to which their students improve scores on tests, is a bad practice, something on which most testing professionals would agree. They argue that value-added approaches encourage those schools and those working in them to cheat to beat the system, but Wainer and Robinson argue that the expected improvements on which the system is purportedly based do not accomplish what was intended. Their argument is compelling.

The authors describe quite well the logic underlying professional educational testing (not classroom testing). The chapters in many ways could stand alone; that is, one could read a single chapter and make sense of it without reading the rest of the book. They may be expecting such use because there is much duplication of examples throughout the book. I was disappointed in only one chapter: why people do not like tests. I do not think that they really enumerate those reasons; rather, they explain why students and

others should like tests, at least that was my reading of the chapter. I also note that although I agree with most of the points in this thesis, I will quibble with a few. There are other books on why tests are disliked (e.g., Gould, 1981; Hoffman, 1962; Kamenetz, 2015; Kohn, 2000; Koretz, 2017; Lemann, 1999; Sacks, 1999; and Soares, 2020). Wainer and Robinson's book is the anti-antitesting book that takes on points made by many of the above cited books. Few of the antitesting books provide balance in recognizing the strengths of testing as well as the faults. Koretz (2017) and perhaps Lemann (1999) are exceptions. Rather than refuting many complaints about test use in admissions, licensure, and other educational settings, Wainer and Robinson make the case that test scores are and can be useful.

The initial chapters provide a nice, brief history of the advent of group testing in admissions, licensure, and of candidates for the military. They spend considerable time on the advantages of objective scoring. I have two concerns about their strong statements on the benefits of objective scoring, which to my mind are nevertheless clear. First, there are licensure companies that charge extra for hand scoring when it is requested by an individual who failed by a narrow margin. I know of some cases where minor adjustments have been made, and I was told by one individual at a licensing agency that almost 50% of hand scoring leads to minor changes to their scores. Optical scanning machines are very good, but not perfect, what with erasures, light pencil marks, and the like. There are also types of individual difference variables where assessments can be only poorly evaluated objectively.

The authors use writing as an example of a situation where objectively scored items that assess editing skill predict written essay scores better than other essay scores do. They are correct in this conclusion (see, for example, Godshalk et al., 1966). However, while such assessments are suggestive of writing ability, they are not writing per se. This argument is a point that many writing scholars express. It is the equivalent of flying with a pilot or choosing a surgeon based on their performance on a paper-and-pencil test. Again, the evidence that they cite is correct; however, there are some assessments that need to be based, at least in part, on actual performance. I question whether Hemingway (whose sentences were too short) and Faulkner or Thomas Mann would do well (because their sentences were simply too long) on such editing measures. I doubt that a concert pianist could be selected to perform based on paper-and-pencil success. Similarly, the attempts to measure creativity objectively have failed miserably. The authors mostly cite reliability evidence, but reliability is a prerequisite for validity not a substitute for it, and there are times when prediction is not as good as sampling the kinds of behaviors sought on a measure. Moreover, this point moves us into the current hot topic of artificial intelligence (AI), which Wainer and Robinson do not directly address. Imagine that a person is falsely accused of a crime, and the judge is a supercomputer that has digested thousands of crimes and uses AI to resolve the case. Given the exceedingly high base rate of convictions after one is charged, would anyone accept such a judge? Nevertheless, they are correct, objective scoring is more reliable, if perhaps less authentic than actual behavior.

Wainer and Robinson correctly report that the stakes of licensure are more critical than those of college admissions. False positive licensure decisions permit a less qualified individual to provide services and thus cause potential harm to the public. The primary goal of all licensure decisions is to avoid harming the public, so that is a critical concern.

A false negative is hurtful to the recipient of the licensure decision, but they may take the examination again and demonstrate their abilities and skills; the harm is temporary, although hurtful to the recipient, which is probably better than harming the public. Meanwhile, a false negative in college admissions means that a student does not go to their first-choice university but can go to another. Their career plans are not thwarted. A false positive decision may have worse consequences; it means that it is more likely that a student will not graduate, which harms both the student and the university. For this latter reason, as Wainer and Robinson argue, false positive admissions decisions may be considered worse than false negative ones, at least to the university and long term to the student. The authors spend considerable time reviewing this kind of thinking and resultant decision-making.

Their chapter on licensing tests and the need for them is very strong and compelling. They trace the history of a couple of licensing tests in the United States, primarily for physicians and airplane pilots. The inclusion of testing in the licensing process has improved both practices and thereby meets the primary requirement of licenses: to protect the public. The authors also report a change in the United States Medical Licensing Examination (USMLE), the licensing test for physicians, from previous practices. Presently, the USMLE is scored as a pass–fail decision; one does not get actual scores. Actually, only the first step of the USMLE is so scored. Those who fail receive information on their performance including content feedback on strengths and weaknesses. The latter two steps remain scored. Moreover, such a system only works, as it does in medicine and psychology, for example, if there is one passing score across the country. In some fields, individual states can set different passing scores, and there is difficulty in many aspects of the transportability of licenses. In medicine and psychology, former state-set standards have been replaced by a single national one.

Throughout the book and especially in one chapter (Chapter 11—Evidence in Science: What Data can we Trust?), they argue that all decisions should be based on data, not based on anecdotes, and the authors go as far as to report that many anecdotes taken together do not form a data set. They decry the attempt by some to stop gathering data and to make decisions based on beliefs; no scientist should disagree with that strategy. This approach has been termed “evidence-based practice” in many fields. While I might agree with that statement in many circumstances, I also can imagine situations where that point does not hold. Imagine a school that gets numerous reports of sexual harassment by a staff member, for example. Are such reports anecdotes or data upon which a decision can be made?

Wainer and Robinson also state that decisions should be made consistent with values. While they provide many examples documenting the need for data-based decision making, the examples of integrating values into those decisions are much less present. One example that they present, however, is that to improve the teachers in the country, higher rewards (i.e., salaries) are needed to attract more academically able teachers to the profession. In the field, many of the values (seen as consequences by Messick, 1994, 1998) such as those related to race, gender, and ethnic group differences are not discussed thoroughly. Industrial-organizational psychologists have worked in recent decades to combine tests of general mental ability with personality, integrity, and other measures to balance workforce quality and diversity, for example. Such a strategy combines data and values in a manner to optimize both quality and diversity. Wainer and

Robinson do suggest that one can collect data on values that should be considered in making various kinds of decisions and that the results of such data collections should be made public. While that seems like a good strategy, where majority opinions do not protect the underserved minorities, there can be serious problems.

The authors appear to believe that merit should rule the day in admissions, scholarships, and, obviously, licensure. A testing day is much like a day when a race is run. The authors seem to believe that the only thing that matters is who crosses the line first (at least in admissions and scholarships). It does not matter if some runners faced obstacles to get there as opposed to the other runners who ran with advantages rather than hindrances. The position that Wainer and Robinson appear to be taking is that universities should accept the very best students as determined by test scores that they earn, and because of the grade inflation in high schools, universities get much better information from tests than other information including high school grades, letters of reference, and other information. That statement is clearly accurate. It is more economical for institutions to use admissions tests, they argue, because it holds the educational costs down and perhaps maximizes institutional income because fewer students withdraw or fail out. While their facts are true and their financial analyses appear appropriate and accurate, they take what this reader considers is a short-term view. If we want all groups in society to have a chance for their members to succeed, then we must accommodate those who have faced more challenges in their development, perhaps not because they will succeed as do the more advantaged candidates for admission, but because in the long run, these groups will all receive society's rewards and the disadvantages that some runners face will be reduced. While Wainer and Robinson are correct that the "Strivers" research did not achieve positive results, admissions research at the University of California system (Academic Senate, University of California, 2020) concluded that admissions counselors were well able to combine such information in making admissions decisions.

I believe that the book would have been enhanced with a more thorough description of what fairness is. They seem on one hand to argue that not accepting, licensing, or providing scholarships to the very best tested candidates is unfairness. And yet they also argue that one should infuse decisions with one's values, but they provide few examples of this being done. I also believe that most institutions want both the best candidates and a diverse student body or pool of licensees. They do not discuss methods to compromise the two, which is a limitation of the book.

It is also worth asking that if colleges and universities are the groups that are advantaged by test scores, why are they not paying for the tests? While in the U.S. context this question may appear extreme, in the late 1960s and early 1970s, Canada, working with help from the College Entrance Examination Board and Educational Testing Service, developed its own college admissions test, to be administered once each year (Traub & Elliott, 1973). However, the test never really became operational. Traub (2002) informed me that given the test would benefit universities, the Canadian government decided that universities should pay the costs of the testing rather than students. The universities choose not to do so, and the vision of a Canadian SAT died, although they do use some specialized tests, such as for medical school admissions.

Obviously, the above discussion has racial, ethnic, and gender implications. Wainer and Robinson state "Although admissions tests are not biased against minority students,

there still remains the problem of group differences in test scores” (p. 93). Based on test fairness research, many psychometricians and testing experts would probably agree with that statement, but many others in society will not. I note that in several places in the book, while the authors encourage decision makers to base their decisions on test scores, they also encourage such users to know their values. As stated above, they do not provide much guidance, however, in terms of how values should be applied to balance validity and equity. Zwick (2017, 2023a, 2023b), Zwick et al. (2019), Gaertner and Barnes (2023), Gaertner and Hart (2005), and McClarty, Mattern, and Gaertner (2017) have provided examples of how there can be successful compromises of test data and characteristics that may affect student performance in one way or another. These scholars have determined ways to increase diversity in college classes with an exceedingly limited reduction in the average quality of undergraduate students in overall quality. Moreover, many test publishers use differential item functioning techniques to eliminate or improve items that differentially impact members of specific groups. However, if a test uniformly adversely affects a minority group, no items will be identified as functioning differentially, and it is easy to imagine such a result when the test is built to be highly unidimensional. While the kinds of compromise solutions Zwick, Gaertner, and their colleagues have sought to help universities select fine students, they also substantially diversify the campuses and improve the entire learning environment.

There are many reasons why diversity is important. I spent 14 years of my career as a dean and vice president of academics, and increasing diversity in students and faculty was one of my institution’s and my goals. In a job interview I had once, given my own scholarly record, one faculty member asked me in a group setting whether I wanted all faculty to have records like mine. My response was “No,” and I used the example that no one wants to listen to an orchestra if all its members are playing the same instrument. I also read in a higher education journal once that personnel directors of companies look to hire individuals who can perform the required job functions. CEOs, on the other hand, look for those who will eventually accede to leadership positions, and incumbent on such positions is the need to deal effectively with diverse employees. Without that experience in education, graduates will not have the skills that they need.

I would have liked to see Wainer and Robinson address the potential problem of bias in criteria. For example, they cite that minority students in medical school who have lower admissions test scores also do worse in terms of grades in clinical as well as more academic courses. One wonders the extent to which that such grades are free of any biases, whether intentional or unintentional. In several industries, I have seen clear rater-race-by-rater interactions indicating a higher rating going to the person with the same racial identity. That kind of bias could easily exist in higher education too.

My favorite chapter, Chapter 12 (Testing Zombies), relates to what the authors call the three zombies. By zombies, they mean widely held beliefs that are not consistent with data. These three beliefs are test coaching, strivers, and test-optional college admission programs. They provide clear and compelling evidence that each of these beliefs are in large measure mistaken if not outright wrong. Many studies of test coaching have found these training programs to be ineffective, leading to changes that are likely to be matched simply by retesting. Not only are the coaching programs ineffective, but they are also often expensive. However, I do have a couple of quibbles here. First, there has been some—admittedly anecdotal—evidence that students in education and similar

fields do not take quantitative courses while in college. A training program that refreshes the mathematics they learned in high school can help them remember what they once learned and perform better on graduate and professional school admissions tests. In fact, the founder of the Kaplan nationwide test coaching schools was a New York City mathematics teacher who asked students about the math questions that were asked on the SAT. He identified certain types of mathematics questions that were often asked and taught students about the concepts involved and gave them considerable practice rather than training them on specific questions likely to be on the test. He was a true teacher, and his 2001 book, *Test Pilot* (Kaplan, 2001), is an informative read. There is also much evidence that taking many tests—a regular practice in test coaching programs—reduces test anxiety, whether it improves their scores or not. In this world of such increased use of tests in the schools, one might wonder if more testing practice is needed.

A research program at the Educational Testing Service and the College Board attempted to identify “strivers,” defined as students who were overachievers based on their school achievement being higher than their tested abilities. Such a student group was expected to have a disproportionate number of minority students, so the program was also expected to be one to help increase diversity on higher education campuses. Unfortunately, the test scores earned by these seeming overachievers ably predicted their performance in college, meaning they did not appear to overachieve in college. For this or other reasons, the program was halted.

Most undergraduate schools in the United States have dropped the college admissions test requirements. These schools either do not require test scores or will not accept them, and these latter schools make all decisions wholistically without test scores. Wainer and Robinson demonstrate that the outcomes of this change in procedures are well known to admissions test professionals and scholars. First, students who do not submit test scores—a minority of the applicants at most schools—are those who have scored less well than their record might suggest, or lower than what they believe is the average for accepted candidates. Wainer and Robinson demonstrate using data from a single test-optional school that the test scores not sent in (but which were available elsewhere) nevertheless did accurately predict the students’ academic success in college. Second, in Geisinger (2023a), I reported that there were three reasons why schools made this admissions test policy change: (1) the pandemic made it difficult to test every student during their high school senior years, (2) the so-called Varsity Blues scandal—in which a number of rich and highly visible parents paid funds to individuals who had connections at universities to get their children admitted—made the country question the legitimacy of college admission tests, and (3) the Black Lives Matter movement accentuated the degree to which the higher education community failed to understand that this important societal group had experienced discrimination. In this review, I identify a fourth. In the past decade, approximately a dozen small colleges have closed for enrollment declines or other fiscal issues. In addition, for example, even a major institution like Penn State University has been affected. It is currently in the process of closing several of its branch campuses, and the Pennsylvania State College system (formerly the State Teachers Colleges) has been combining campuses. Much of this development is because of the upcoming reduction in the number of college-age students coming through the school systems. Some schools believe that adopting test-optional admissions policies is a draw for some students, and many institutions are even advertising their test-optional status. Moreover, many schools accept a high proportion

of their applicants. In such situations, test scores are certainly less critical, except for counseling purposes.

Because Wainer and Robinson seem to focus on elite, highly selective or selective academic institutions, the use of tests in admissions can affect the proportion of students entering who are able to graduate. Many institutions, however, are not selective. Many colleges and universities in the country accept most of their applicants. For this reason, I think that the book would benefit from the inclusion of a discussion of the concepts underlying the Taylor-Russell tables (1939). The Taylor-Russell tables were developed in a personnel selection context, but provide information on the proportion of candidates likely to be successful given three factors: the validity of the test, the base rate of successful candidates for admissions, and the selection ratio (the proportion of applicants who are accepted). Such a discussion could show how the proportion of applicants to be accepted and the proportion of the applicant pool who would likely succeed clearly can limit the utility of even valid admissions measures.

There is yet another unmentioned reason that this book is an important read. In recent years, doctoral programs in psychometrics and quantitative psychology have focused virtually all their time and energies on advancing quantitative techniques. Many graduates of such programs, perhaps especially as they are increasingly in schools of education, lack a firm understanding of differential psychology, the traditional psychological foundation of psychometrics (e.g., Anastasi, 1958; Minton & Schneider, 1980; Stern, 2024; and Tyler, 1965,1974). The Wainer-Robinson volume discusses general mental ability in ways that will help such students begin to comprehend the complexity of individual differences.

In closing, this book is not a book on fairness, which is what I expected it to be from the title. It is rather one on the validity and, to some extent, utility of tests and test scores. The authors' argument is simple: if a test is valid, it is ipso facto fair. They are correct that few studies have found differential predictions due to ethnic or racial group differences, and meta-analyses of these studies confirm the general conclusion. That is, differential prediction studies have generally shown that test scores are equally predictive for all test takers, regardless of their group affiliation. The authors focus primarily on college, graduate, and professional school admissions and licensing, but address a few other topics less thoroughly. Throughout the book there is a focus on efficacy, which they evaluate primarily in terms of the percentage of students who graduate. It is a decent criterion from one point of view, but it is also one that likely has negative implications for diversity. At the same time, Wainer and Robinson recommend integrating one's values into the decision-making process, which I found somewhat contradictory. Efficacy is also a goal that affects most colleges, but the authors concentrate primarily on elite institutions like those in the Ivy League and public institutions such as the University of Michigan and the University of California, Berkeley. Moreover, while most schools wish to graduate as many entering students as possible, they also desire diverse student bodies.

Again, this is a book that should be read—by testing specialists, education leaders, policy makers, and those interested in such matters. It does not require much statistical knowledge to be read; the authors explain these concepts in a manner that should make even difficult concepts understandable. It is an important book, one that rebuts beliefs and decisions made in different traditional educational testing contexts. But while the

book is based on data-based research, it also espouses a particular orientation, one that runs counter to pro-diversity perspectives now widely held in higher education. Even in licensure settings, many professions seek to increase their minority licensees, in part to better provide services to all of society. One should, as Wainer and Robinson suggest, read this book with their own values in mind and reassess them from the perspectives shared in the book. Had the authors shared approaches where one applied test data in conjunction with the increasingly common values of those who administer higher education, the book would be even more useful.

## References

- Academic Senate, University of California. (2020). *Report of the UC Academic Council: Standardized Test Task Force*. Systemwide Academic Senate, University of California.
- Anastasi, A. (1958). *Differential psychology: Individual and group differences in behavior*. Macmillan.
- Bok, D., & Bowen, W. G. (2000). *The shape of the river*. Princeton University Press.
- Gaertner, M. N., & Barnes, M. (2023). Colorado's affirmative action experiment: Disadvantage, overachievement, and new routes to racial diversity. In K. F. Geisinger (Ed.), *College admissions and admissions testing in a time of transformation* (pp. 165–189). Routledge. <https://doi.org/10.4324/9781003326823-10>
- Gaertner, M. N., & Hart, M. (2005). Considering class: College access and diversity. *Harvard Law & Policy Review*, 7(2), 367–403.
- Geisinger, K. F. (2023a). College admissions: What our goals should be and what works to get us there. In K. F. Geisinger (Ed.), *College admissions and admissions testing in a time of transformation* (pp. 1–9). Routledge. <https://doi.org/10.4324/9781003326823-1>
- Geisinger, K. F. (2023b). Some final thoughts on college admissions: International perspectives, summary thoughts, and a look to the future. In K. F. Geisinger (Ed.), *College admissions and admissions testing in a time of transformation* (pp. 190–219). Routledge. <https://doi.org/10.4324/9781003326823-11>
- Godshalk, F. I., Swineford, W. F., & Coffman, W. E. *The measurement of writing ability*. College Board.
- Gould, S. J. (1981). *The mismeasure of man*. Norton.
- Gulliksen, H. (1950). *Theory of mental tests*. Routledge. <https://doi.org/10.1037/13240-000>
- Hoffman, B. (1962). *The tyranny of testing*. Crowell-Collier.
- Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39(1), 1–123. <https://doi.org/10.17763/haer.39.1.13u15956627424k7>
- Kamenetz, A. (2015). *The test: Why our schools are obsessed with standardized testing, but you don't have to be*. Public Affairs.
- Kaplan, S. H. (2001). *Test pilot*. Kaplan. <https://doi.org/10.1145/381344.381355>
- Kohn, A. (2000). *The case against standardized tests*. Heinemann.

- Kolkata, M., Silva, M., & Sireci, S. G. (2019). College admissions and social responsibility. *Educational Measurement: Issues and Practice*, 40(4), 22–27. <https://doi.org/10.1111/emip.12425>
- Koretz, D. (2017). *The testing charade*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226408859.001.0001>
- Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. Farrar, Straus, and Giroux.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- McClarty, K. L., Mattern, K. D., & Gaertner, M. N. (2017). *Preparing students for college and careers: Theory, measurement, and educational practice*. Routledge. <https://doi.org/10.4324/9781315621975>
- Messick, S. (1994). The interplay of evidence and consequences in the validity of performance assessments. *Educational Researcher*, 23(2), 13–23. <https://doi.org/10.3102/0013189X023002013>
- Messick, S. (1998). Test validity: A matter of consequences. *Social Indicators Research*, 45(1), 35–44. <https://doi.org/10.1023/A:1006964925094>
- Minton, F., & Schneider, F. (1980). *Differential psychology*. Brooks/Cole.
- Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Perseus.
- Soares, J. A. (Ed.). (2020). *The scandal of standardized tests: Why we need to drop the SAT and ACT*. Teachers College Press.
- Stern, L. W. (2024). On the psychology of individual differences: Toward a “differential psychology.” Palgrave Macmillan. <https://doi.org/10.1007/978-3-031-69638-1>
- Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology*, 23(5), 565–578. <https://doi.org/10.1037/h0057079>
- Traub, R. E. (2002). Personal communication.
- Traub, R. E., & Elliott, H. A. (1973). Development of the Canadian Scholastic Aptitude Test. <https://www.thecommonwealth-library.org/index.php/comsec/catalog/download/262/259/1973?inline=1>.
- Tyler, L. E. (1974). *Individual differences: Abilities and motivation differences*. Appleton-Century-Crofts.
- Tyler, L. E. (1965). *The psychology of human differences*. Appleton-Century-Crofts.
- Zwick, R. (2023a). Using mathematical models to improve access to postsecondary Education. In M.E. Oliveri & C. Wendler (Eds.), *Higher education admissions practices: An international perspective* (pp. 333–346). Cambridge University Press. <https://doi.org/10.1017/9781108559607.018>
- Zwick, R. (2023b). Using quantitative techniques to promote diversity. In K. F. Geisinger (Ed.), *College admissions and admissions testing in a time of transformation* (pp. 151–164). Routledge. <https://doi.org/10.4324/9781003326823-9>

Zwick, R. (2017). *Who gets in? Strategies for fair and effective college admission*. Harvard University Press. <https://doi.org/10.4159/9780674977648>

Zwick, R., Ye, L., & Islam, S. (2019). Using constrained optimization organization to increase representation of students from low-income neighborhoods. *Applied Measurement in Education*, 32(4), 281–297. <https://doi.org/10.1080/08957347.2019.1660346>

### About the Reviewer

**Kurt F. Geisinger** is Director of the Buros Center for Testing and the Meierhenry Distinguished University Professor at the University of Nebraska–Lincoln. He is a past president of American Psychological Association’s (APA) Quantitative and Qualitative Methodology and International Psychology divisions, the International Test Commission (ITC), and the Assessment and Evaluation division of the International Association of Applied Psychology (IAAP); and a past board member of APA. He is currently Treasurer of the IAAP. A fellow of APA (seven divisions), the Association for Psychological Science, the AERA, Eastern Psychological Association, and ITC, he has chaired the Graduate Record Examination Board, chaired the College Board’s Research Committee, the Joint Committee on Testing Practices, and served on the SAT Committee. He edited/co-edited more than 20 volumes including the *APA Handbook of Testing and Assessment in Psychology*, *Psychological Testing of Hispanics* (two editions), seven *Mental Measurements Yearbooks*, *Fairness in Educational and Psychological Testing*, *College Admissions and Admissions Testing in a Time of Transformational Change*, several monographs and instructors’ manuals and written about 170 chapters and journal articles. He chaired the GRE Board, the GRE Technical Advisory Committee, the College Board’s Research Advisory Committee, and served on the SAT Committee. He has been an expert witness in more than two dozen cases, primarily defending and critiquing civil service tests and arguing for the rights of those with disabilities in admission and licensure tests. He served as chair of the Fordham University Psychology department, Dean of Arts and Sciences at SUNY-Oswego, and Academic Vice President/Provost at Le Moyne College and the University of St. Thomas.



### About the Book Authors

**Howard Wainer** is an award-winning American statistician and research scientist. His areas of work include testing, graphical methods for data analysis and communication, and robust statistical methodology. He has served on the faculty of the University of Chicago, at the Bureau of Social Science Research during the Carter Administration, and as Principal Research Scientist in the Research Statistics Group at Educational Testing Service for 21 years. In 2016, he retired after 15 years as Distinguished Research Scientist at the National Board of Medical Examiners. This is his 26<sup>th</sup> book.





**Daniel H. Robinson** is the K-16 Mind, Brain, and Education Endowed Chair at the University of Texas at Arlington. He received his Ph.D. in Educational Psychology in 1993 from the University of Nebraska where he majored in both learning/cognition and statistics/research. He previously taught at Mississippi State University, the University of South Dakota, the University of Louisville, the University of Texas at Austin, and Colorado State University. He has served as department chair, director, and associate dean of research.



**Education Review      Reseñas Educativas**  
**Resenhas Educativas**



*Education Review/Reseñas Educativas/Resenhas Educativas* is supported by the Scholarly Communications Group at the Mary Lou Fulton College for Teaching and Learning Innovation, Arizona State University. Copyright is retained by the first or sole author, who grants the right to first publication to the *Education Review*. Readers are free to copy, display, distribute, and adapt this article, as long as the work is attributed to the author(s) and *Education Review*, the changes are identified, and the same license applies to the derivative work. More details of this Creative Commons license are available at <https://creativecommons.org/licenses/by-sa/4.0/>.



**Disclaimer:** The views or opinions presented in book reviews are solely those of the author(s) and do not necessarily represent those of *Education Review*.