

**REFERENCES ON TESTING OF  
PEOPLE OF COLOR, WOMEN, LANGUAGE MINORITIES, AND  
PEOPLE WITH DISABILITIES**

**COMMITTEE ON PSYCHOLOGICAL TESTS AND ASSESSMENT (CPTA)  
2002**

- Anthony, L. C., Harris, V. F., & Pashley, P. J. (1999). Predictive validity of the LSAT: A national summary of the 1995-1996 correlation studies. (LSAT Technical Report 97-01). Newtown, PA: Law School Admission Council.
- Bennett, R.E. (1999). Computer-based testing for examinees with disabilities: On the road to generalized accommodation. In S. J. Messick (Ed.), Assessment in higher education: Issues of access, quality, student development, and public policy (pp. 181-191). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bleistein, C. A., & Wright, D. J. (1987). Assessment of unexpected differential item difficulty for Asian-American examinees on the Scholastic Aptitude Test. In A. P. Schmitt & N. J. Dorans (Eds.), Differential item functioning on the Scholastic Aptitude Test (ETS Research Memorandum No. 87-1). Princeton, NJ: Educational Testing Service.
- Bond, L. (1995). Unintended consequences of performance assessment: Issues of bias and fairness. Educational Measurement: Issues and Practice, 14 (4), 21-24.
- Braun, H. T., & Jones, D., H. (1981). The Graduate Management Admission Test prediction bias study (GMAC Report 81-4). Princeton, NJ: Educational Testing Service.
- Braun, H. T., & Jones, D. H. (1985). Use of empirical Bayes methods in the study of the validity of academic predictors of graduate school performance. (ETS Research Report 84-34). Princeton, NJ: Educational Testing Service.
- Breland, H. M., & Durán, R. P. (1985). Assessing English composition skills in Spanish-speaking populations. *Educational and Psychological Measurement*, 45, 309-318.
- Bridgeman, R. L. (1998). Fairness in computer-based testing: What we know and what we need to know. (The GRE FAME Report Series Vol. 2) (pp.4-11). Princeton, NJ: Educational Testing Service.
- Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning Test*. (Research Report No. 2000-01). New York: College Entrance Examination Board.
- Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). Prediction of freshman grade-point average from the revised and recentered SAT I: Reasoning Test (College Board Report 2000-1). New York: College Entrance Examination Board.

- Bridgeman, B., & McHale, F. (1996). Gender and ethnic group differences on the GMAT Analytical Writing Assessment (ETS Research Report 96-2). Princeton, NJ: Educational Testing Service.
- Bridgeman, B., & Schmitt, A. (1997). Fairness issues in test development and administration. In W. W. Willingham & N. Cole, Gender and fair assessment (pp. 185-226). Mahwah, NJ: Lawrence Erlbaum Associates.
- Burton, E., & Burton, N. W. (1993). The effect of item screening on test scores and test characteristics. In P. W. Holland & H. Wainer (Eds.), Differential item functioning (pp. 321-336). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Burton, N. W., & Ramist, L. (2001). Predicting success in college: SAT studies of classes graduating since 1980 (Research Report 2001-2). New York: College Entrance Examination Board.
- Burton, N. W., & Turner, N. (1983). Effectiveness of Graduate Record Examinations for predicting first year grades: 1981-82 summary report of the Graduate Record Examinations Validity Study Service. Princeton, NJ: Educational Testing Service.
- California State Auditor (2000, November). Standardized tests: Although some students may receive extra time on standardized tests that is not deserved, others may not be getting the assistance they need (Summary of Report 2000-108). Sacramento, CA: Bureau of State Audits.
- Camara, W. J., & Schmidt, A. E. (1999). Group differences in standardized testing and social stratification (College Board Report 99-5). New York: College Entrance Examination Board.
- Camara, W. J., Copeland, T., & Rothchild, B. (1998). Effects of extended time on the SAT I: Reasoning test score growth for students with learning disabilities (College Board Report 98-7). New York: College Entrance Examination Board.
- Clark, M. J., & Grandy, J. (1984). Sex differences in the academic performance of Scholastic Aptitude Test-takers (College Board Report 84-8). New York: College Entrance Examination Board.
- Cleary, T. A. (1968). Test bias: Prediction of Negro and White students in integrated colleges. Journal of Educational Measurement, 5 (2), 115-124.
- Cole, N. S. (1999). Merit and opportunity: Testing and higher education at the vortex. In A. L. Nettles & M. T. Nettles (Eds.), Measuring up: Challenges minorities face in educational assessment (pp. 215-224). Boston: Kluwer Academic.
- Coley, R. J. (2001). Differences in the gender gap: Comparisons across racial/ethnic groups in education and work (ETS Policy Information Report). Princeton, NJ: Educational Testing Service.
- Curley, W. E., & Schmitt, A. P. (1992, April). Revising SAT-Verbal items to eliminate DIF: Does it work? When? How? Paper presented at the American Educational Research Association/ National Council on Measurement in Education, San Francisco.

- Dalessandro, S. P., Anthony, L. C., & Reese, L. M. (2001). LSAT performance with regional, gender, and ethnic breakdowns: 1993-1994 through 1999-2000 testing years. (LSAT Technical Report 00-01). Law School Admission Council.
- Durán, R. P. (1983). *Hispanics' Education and Background: Predictors of College Achievement*. New York: College Entrance Examination Board.
- Durán, R. P. (1988). Validity and language skills assessment: Non-English background students. In H. Wainer & H. I. Braun (eds.), *Test Validity*, pp. 105-127. Hillsdale, NJ: Erlbaum.
- Dwyer, C. A., & Johnson, L. M. (1997). Grades, accomplishments, and correlates. In W. W. Willingham & N. Cole, Gender and Fair Assessment (127-156). Mahwah, NJ: Lawrence Erlbaum Associates.
- Educational Testing Service (1999). Overview: ETS fairness review. Author. Retrieved August 30, 1999 from the World Wide Web: <http://www.ets.org>.
- Educational Testing Service (2000). Sex, race, ethnicity, and performance on the GRE General Test (Technical Report). Princeton, NJ: Author.
- Fuertes, J. N., Sedlacek, W. E., & Liu, W. M. (1994). Using the SAT and noncognitive variables to predict the grades and retention of Asian American university students. *Measurement and Evaluation in Counseling and Development*, 27, 74-84.
- Gallagher, A., Morley, M. E., & Levin, J. (1999). Cognitive patterns of gender differences on mathematics admissions tests (The GRE, FAME Report Series 3) (pp. 4-11). Princeton, NJ: Educational Testing Service.
- Gándara, P., & López, E. (1998). Latino students and college entrance exams: How much do they really matter? *Hispanic Journal of Behavioral Sciences*, 20, 17-38.
- García, E. E., Jorgensen, R. E., & Ormsby, C. (1999). How can public universities still admit a diverse freshman class? *Journal of College Admission*, No. 164, pp. 5-11.
- Geisinger, K. F. (1998). Testing accommodations for the new millennium: Computer-administered testing in a changing society (The GRE FAME Report Series Vol. 2, pp.12-20). Princeton, NJ: Educational Testing Service.
- Gould, S. J. (1996). The mismeasure of man (Revised & expanded edition). New York: Norton.
- Gray, W. H., III, Nettles, M. T., & Millett, C. M. (1999). Tests as barriers to access. In A. L. Nettles & M. T. Nettles (Eds.), Measuring up: Challenges minorities face in educational assessment (pp. 197-207). Boston: Kluwer Academic.

- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science, 269, 41-45.
- Hoover, H. D., & Han, L (1995, April). The effect of differential selection on gender differences in college admission test scores. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Jencks, C. (1998). Racial bias in testing. In C. Jencks & M. Phillips (Eds.), The Black-White test score gap (pp. 55-85). Washington, DC: Brookings Institution Press.
- Jolly, P. (1992). Academic achievement and acceptance rates of underrepresented-minority applicants to medical school. Academic Medicine, 67 (11), 765-769.
- Koenig, J. A., & Leger, K. F. (1997). Test-taking behaviors and their impact on performance: A comparison of retest performance and test-preparation methods for MCAT examinees grouped by gender and race-ethnicity. Academic Medicine, 72 (10), S100-S102.
- Koenig, J. A., Sireci, S. G., & Wiley, A. (1998). Evaluating the predictive validity of MCAT scores across diverse applicant groups. Academic Medicine, 73 (10), 1095-1106.
- Legg, S. M., & Buhr, D. C. (1992). Computerized adaptive testing with different groups. Educational Measurement: Issues and Practice, 11 (2), 23-27.
- Leonard, D., & Jiang, J. (1999). Gender bias and the college prediction of the SATs: A cry of despair. Research in Higher Education, 40 (4), 375-408.
- Lewis, C., & Willingham, W. W. (1995). The effects of sample restriction on gender differences (ETS Research Report 95-13). Princeton, NJ: Educational Testing Service.
- Linn, R. L. (1983). Predictive bias as an artifact of selection procedures. In H. Wainer & S. Messick (Eds.), Principals of modern psychological measurement: A Festschrift for Frederic M. Lord (pp. 27-40). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Linn, R. L. (1990). Admissions testing: Recommended uses, validity, differential prediction, and coaching. Applied Measurement in Education, 3 (4), 297-318.
- Loveless, P. (2000). Does self-stereotyping affect performance on the ACT assessment? (ACT Research Report 2000-6). Iowa city, IA: ACT, Inc.
- Madaus, G. F., Raczek, A. E., & Thomas, S. (1999). Performance assessment and issues of differential impact: The British experience--lessons for America. In A. L. Nettles & M. T. Nettles (Eds.), Measuring up: Challenges minorities face in educational assessment (pp. 137-165). Boston: Kluwer Academic.
- Mandinach, E. B. (2000, April). Flagging: Policies, perceptions, and practices. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

- Mandinach, E. B. & Cahalan, C., & Camara, W. J. (2001, April). The impact of flagging on the admissions process: Policies, practices, and implications. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Marco, G. L. (1988). Does the use of test assembly procedures proposed in legislation make any difference in test properties and in the test performance of Black and White test takers? Applied Measurement in Education, 1 (2), 109-133.
- Nettles, A. L., & Nettles, M. T. (1999). Introduction: Issuing the challenge. In A. L. Nettles & M. T. Nettles (Eds.), Measuring up: Challenges minorities face in educational assessment (pp. 1-11). Boston: Kluwer Academic.
- Noble, J. P. (2000, October 16). Non-race-based admissions decisions using ACT composite scores and/or high school grade averages. (ACT Draft Report). Iowa city, IA: ACT, Inc.
- O'Neill, K. A., & McPeck, W. M. (1993). Item and test characteristics that are associated with differential item functioning. In P. W. Holland & H. Wainer (Eds.), Differential item functioning (pp. 255-276). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pearson, B. Z. (1993). Predictive validity of the Scholastic Aptitude Test (SAT) for Hispanic Bilingual Students. *Hispanic Journal of Behavioral Sciences*, 15, 342-356.
- Pennock-Román, M. (1990). Test validity and language background: A study of Hispanic American students at six universities. New York: College Entrance Examination Board.
- Pennock-Román, M. (1994). College major and gender differences in the prediction of college grades (College Board Report 94-2). New York: College Entrance Examination Board.
- Pennock-Román, M. (1993). The status of research on the Scholastic Aptitude Test (SAT) and Hispanic Students in postsecondary education. In Gifford, B. R. (ed.). *Policy Perspectives on Educational Testing*, pp. 75-115. Boston: Kluwer.
- Powers, D. E. (1986b). Test preparation for the GRE analytical ability measure: Differential effects for subgroups of GRE test takers (ETS Research Report 86-40). Princeton, NJ: Educational Testing Service.
- Powers, D., & O'Neill, K. (1993). Inexperienced and anxious computer users: Coping with a computer-administered test of academic skills. Educational Assessment, 1 (2), 153-173.
- Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). Student group differences in predicting college grades: Sex, language, and ethnic groups (College Board Report 93-1). New York: College Entrance Examination Board.

- Ramist, L., Lewis, C., & McCamley-Jenkins, L. (2002). Using achievement tests/SAT II Subject Tests to demonstrate achievement and predict college grades: Sex, language, ethnic, and parental education groups. College Board Report.
- Rigol, G. W. (1997, June). Common sense about SAT score differences and test validity (College Board Research Notes RN-01). New York: College Entrance Examination Board.
- Rogers, H. J., & Kulick, E. (1987). An investigation of unexpected differences in item performance between Blacks and Whites taking the SAT. In A. P. Schmitt & N. J. Dorans (Eds.), Differential item functioning on the Scholastic Aptitude Test (ETS Research Memorandum No. 87-1). Princeton, NJ: Educational Testing Service.
- Rooney, C. (1998). Test scores do not equal merit: Enhancing equity & excellence in college admissions by deemphasizing SAT and ACT results. Cambridge, MA: FairTest.
- Rosser, P. (1989). The SAT gender gap: Identifying the causes. Washington, DC: Center for Women Policy Studies.
- Schmitt, A. P. (1987). Unexpected differential item performance of Hispanic examinees. In A. P. Schmitt & N. J. Dorans (Eds.), Differential item functioning on the Scholastic Aptitude Test (ETS Research Memorandum No. 87-1). Princeton, NJ: Educational Testing Service.
- Schmitt, A. P. & Dorans, N. J. (1988). Differential item functioning for minority examinees on the SAT (ETS Research Report 88-32). Princeton, NJ: Educational Testing Service.
- Schneider, L. M., & Briel, J. B. (1990). Validity of the GRE: 1988-89 summary report. Princeton, NJ: Educational Testing Service.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Identity salience and shifts in quantitative performance. Psychological Science, *10*, 80-83.
- Steele, C. M. (1997). A threat in thin air: How stereotypes shape intellectual identity and performance. American Psychologist, *52* (6), 613-629.
- Steele, C. M. (1999, August). Thin ice: "Stereotype threat" and Black college students. The Atlantic Monthly. Retrieved September 19, 1999 from the World Wide Web: <http://www.theatlantic.com>
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual performance of African Americans. Journal of Personality and Social Psychology, *69*, 797-811.
- Steele, C. M., & Aronson, J. (1998). Stereotype threat and the test performance of academically successful African Americans. In C. Jencks & M. Phillips (Eds.), The Black-White test score gap (pp. 401-427). Washington, DC: Brookings Institution Press.

- Stricker, L. J. (1998). Inquiring about examinees' ethnicity and sex: Effects on AP Calculus AB examination performance (College Board Report 98-1). New York: College Entrance Examination Board.
- Stricker, L. J., & Ward, W. C. (1998). Inquiring about examinees' ethnicity and sex: Effects on Computerized Placement Tests performance (College Board Report 98-2). New York: College Entrance Examination Board.
- Stricker, L. J., Rock, D. A., & Burton, N. W. (1993). Sex differences in predictions of college grades from Scholastic Aptitude Test scores. Journal of Educational Psychology, 85 (4), 710-718.
- Sue, S., & Abe, J. (1988). *Predictors of academic achievement among Asian American and White students*. (Report No. 88-11). New York: College Entrance Examination Board
- Tracey, T. J., & Sedlacek, W. E. (1984). Noncognitive variables in predicting academic success by race. Measurement and Evaluation in Guidance, 16 (4), 171-178.
- Vars, F. E., & Bowen, W. G. (1998). Scholastic Aptitude Test scores, race, and academic performance in selective colleges and universities. In C. Jencks & M. Phillips (Eds.), The Black-White test score gap (pp. 457-479). Washington, DC: Brookings Institution Press.
- Wightman, L. F. (1993). Test takers with disabilities: A summary of data from special administrations of the LSAT (LSAC Research Report 93-03). Newtown, PA: Law School Admission Council.
- Wightman, L. F. (1994). Analysis of LSAT performance and patterns of application for male and female law applicants (LSAC Research Report 94-02). Newtown, PA: Law School Admission Council.
- Wightman, L. F., & Muller, D. G. (1990). Comparison of LSAT performance among selected subgroups (Law School Admission Council Statistical Report 90-01). Newton, PA: Law School Admission Council.
- Wiley, A., & Koenig, J. A. (1996). Undergraduate performance assessment and prediction: The validity of the medical college admission test for predicting performance in the first two years of medical school. Academic Medicine, 71 (10), S83-S85.
- Willingham, W. W. (1989). Standard testing conditions and standard score meaning for handicapped examinees. Applied Measurement in Education, 2 (2), 97-103.
- Willingham, W. W. (1999). A systemic view of test fairness. In S. J. Messick (Ed.), Assessment in higher education: Issues of access, quality, student development, and public policy (pp. 213-242). Mahwah, NJ: Lawrence Erlbaum Associates.
- Willingham, W., & Cole, N. (1997). Gender and fair assessment. Mahwah, NJ: Lawrence Erlbaum Associates.

- Willingham, W. W., Cole, N. S., Lewis, C. & Leung, S. W. (1997). Test performance. In W. W. Willingham & N. Cole (Eds.), Gender and Fair Assessment (55-126). Mahwah, NJ: Lawrence Erlbaum Associates.
- Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, H., Rock, D. A., & Powers, D. E. (1988). Testing handicapped people. Boston: Allyn and Bacon, Inc
- Young, J. W. (1991). Improving the prediction of college performance of ethnic minorities using the IRT-based GPA. Applied Measurement in Education, 4 (3), 229-239.
- Young, J. W. (2001). *Differential validity, differential prediction, and college admissions testing: A comprehensive review and analysis* (Research Report No. 2001-6). New York: The College Board.
- Ziomek, R. L., & Andrews, K. M. (1996). Predicting the college grade point averages of special-tested students from their ACT assessment scores and high school grades. ( ACT Research Report 96-7). Iowa City, IA: ACT, Inc.
- Zwick, R. (1991). Differences in graduate school attainment patterns across academic programs and demographic groups: (ETS Research Report 91-17). Princeton, NJ: Educational Testing Service.
- Zwick, R. (2001, March/April). What causes the test-score gap in higher education? Perspectives on the Office for Civil Rights resource guide on high-stakes testing. Change, pp.33-37.
- Zwick, R., & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. Journal of Educational Measurement, 26, 55-66.
- Zwick, R., & Schlemmer, L (2002). How Well Does the SAT Predict College Performance for Language Minority Students? A Study of Two Freshman Cohorts at UC Santa Barbara . To appear, University of California Linguistic Minority Institute report series.
- Zwick, R. (2002). *Fair game? The use of standardized admissions tests in higher education*. New York: RoutledgeFalmer.