Ways to Address Flawed Assumptions in Testing: Annotated Bibliography

Joel P. Wiesen, Ph.D.
April 27, 2026

References for the Master Tutorial and a few annotations are provided below, organized by slide number.

Slide 5: Historical Perspective

Brigham, C. C. (1922). *A study of American intelligence*. Princeton: Princeton University Press; London: Oxford University Press, 1922 to 1923.
pages 190, 197, 245
pg 190 "Our figures, then, would rather tend to disprove the popular belief that the Jew is highly intelligent."
pg 197 "… the Alpine and Mediterranean races in our immigration are intellectually inferior to the representatives of the Nordic race…"
pg 245 "The really important steps are those looking toward the prevention of the continued propagation of defective strains in the present population."

Goddard, H. H. (1917). Mental tests and the immigrant. *Journal of Delinquency. 2(5)*, 243-277.
page 252, Table 2
10% of Jews and 7% of Italians are in the normal range of intelligence
76% and 79% are in the Moron range (IQ of 50-70 or mental age of 8-12 years)

Eysenck, H. J., & Kamin, L. J. (1981). *Intelligence: The battle for the mind*. (No Title). London : The Macmillan Press Ltd
page: 156
"Over the years, Pearson suggested, good Jews, brave Jews, clean Jews would have rebelled against the Tsars, and would have been exterminated. The genes for goodness, bravery and cleanliness would thus have died out in the Jewish race; only the dregs would have survived, clamouring for admission to England."

Slide 6: Historical Perspective

Many others: Galton, Cattell, Munsterberg, Woodworth, Washburn

Army Beta was “language free” because no reading or writing required. But some verbal instructions in English, such as “Start there” Find a way through the maze to the end. Do it as quickly as possible.”

Slide 11: Number Correct Predicts Best

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior research methods, 52(6),* 2287-2305.

Slide 13: Number Correct Predicts Best

Guion, R. M. (2011). Assessment, measurement, and prediction for personnel decisions. Routledge.

Guion cites Lord article, 50 years old, with no critique. Lord was respected but article has logical flaw: focuses on risk of false negatives rather than risk of false positives

Lord, F. M. (1962). Cutting scores and errors of measurement. Psychometrika, 27, 19-30.
Lord, F. M. (1963). Cutting scores and errors of measurement: A second case. Educational and Psychological Measurement, 23, 63-68.

Page 283-284, Guion 2011
"Objections to cut scores in bivariate prediction apply even more to multiple cut scores, where even very low cut scores can result in rejecting too many candidates. A cut score about 1.5 SD below the mean of a normal distribution will reject about 7% of the applicants. If a similar cut score is set on another, uncorrelated measure, 7% will be rejected by it also. Some people might be in the low 7% on both tests, but the percentage of the total group being rejected will approach, even if it does not reach, 14%. More hurdles mean more rejections. Many of those passing all of the hurdles will do so with scores too low to suggest any genuinely useful qualifications at all. Cut scores high enough to assure people qualified on each trait may mean that no applicant qualifies on all of them."

(Lord, 1962 cited in Guion, 2011, pg 284)

Slide 20: Use z-scores to Equate

Sizes are based on a small Monte Carlo study

Slide 24: Use z-scores to Equate

Elliott, M. R., West, B. T., Zhang, X., & Coffey, S. (2022). The anchoring method: Estimation of interviewer effects in the absence of interpenetrated sample assignment. *Survey methodology, 48(1),* 25.
(Discusses multiple panels rating at least some common subset of candidates.)

Slide 32: Weight by Number of Items

Tatum v. Commonwealth of Massachusetts, C.A. No. 0984CV00576 (Mass. Super. Ct., Suffolk Cnty, Oct 27, 2022).

Judge was incredulous that applicants were assigned a 600 page book in administration and the exam had only ONE question.

Slide 42: High Rater Reliability = Valid
Wiesen, J. P. (2025, July 28). Reliability of BARS: New Data from Two-Board Grading. *International Personnel Assessment Council Annual Conference*, Atlanta, GA, United States.

Wiesen, J. P. (2026, April 30 – May 2). The BARS Inter-rater Reliability Illusion: a Within Versus Between-panel Comparison [Poster]. *Society for Industrial and Organizational Psychology Annual Conference,* New Orleans, LA, United States.

Forthcoming, July 2026:
Wiesen, J. P. & Gunn, J. J. (2026). Two-Panel Grading Using BARS: Attempts to Understand Disturbing Reliability Findings. *International Personnel Assessment Council Annual Conference,* Charleston, SC, United States.

Slide 51: Quote sources ➜ Defensibility
For example:
Tatum v. Commonwealth of Massachusetts, C.A. No. 0984CV00576 (Mass. Super. Ct., Suffolk Cnty, Oct 27, 2022).

Smith v. City of Boston, Civil Action No. 12-10291-WGY (United States District Court District of Massachusetts, Oct 26, 2020) 496 F. Supp. 3d 590 (D. Mass. 2020)

Slide 59: Top Hires Will Perform Well
Wiesen, J. P. (2021, July 26). Select Tests Based on Utility to Maintain Job Performance and Reduce Adverse Impact. **International Personnel Assessment Council Annual Conference** (virtual).
(Esp. Slide #21)

Slide 60: Top Hires Will Perform Well
Source: My Monte Carlo research

Slide 61: Top Hires Will Perform Well
Aamodt, M. G. (2004). *Research in Law Enforcement Selection*. Boca Raton: BrownWalker

Press.

Hirsh, H. R., Northrop, L. C., & Schmidt, F. L. (1986). Validity generalization results for law enforcement occupations. *Personnel Psychology, 39(2)*, 399-420.

Wiesen, J. P. (2018). Tools to Increase Diversity, Utility, and Validity in Hiring Police Officers [Master Tutorial]. *Society for Industrial and Organizational Psychology Annual Conference*, Chicago, IL, United States.

Slide 67: Tests Have Sufficient Validity
"a given selection score …will often result in proportionately more false negative decisions in groups with lower mean test scores" (AERA, APA, & NCME, 1999, page 79, col 2, par 2).

AERA, APA, & NCME (1999) *Standards for Educational and Psychological Testing (2nd ed.)* Washington, DC: American Psychological Association.

Project A found statistically significant different prediction equations for almost all pairs of the 9 jobs considered.

Wise, L. L., McHenry, J., & Campbell, J. P. (1990). Identifying optimal predictor composites and testing for generalizability across jobs and performance factors. *Personnel Psychology, 43(2),* 355-366. (See Tables 6 and 7.)

Slide 68: Tests Have Sufficient Validity
Kuechler, W. L., & Simkin, M. G. (2010). Why is performance on multiple-choice tests and constructed-response tests not more closely related? Theory and an empirical test. Decision Sciences *Journal of Innovative Education, 8(1),* 55-73. (M/C and constructed response, r=.45-.63, Tables 2 and 3)

Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology, 58(4)*, 981-1007. (SJT, r=0.66, on page 993)

Schleicher, D. J., Van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2010). If at first you don't succeed, try, try again: Understanding race, age, and gender differences in retesting score improvement. *Journal of Applied Psychology, 95(4)*, 603. (performance tests, r=.23-.30, on page 608)

Van Iddekinge, C. H., Morgeson, F. P., Schleicher, D. J., & Campion, M. A. (2011). Can I retake it? Exploring subgroup differences and criterion-related validity in promotion retesting. *Journal*

of Applied Psychology, 96(5), 941.
(carefully constructed job knowledge test, r=.48, on page 948, Table 3, last entry)

Slide 69: Tests Have Sufficient Validity
Wise, L. L., McHenry, J., & Campbell, J. P. (1990). Identifying optimal predictor composites and testing for generalizability across jobs and performance factors. *Personnel Psychology, 43(2),* 355-366. (See Table 6.)

Slide 70: Tests Have Sufficient Validity
Sackett, P. R., Demeke, S., Bazian, I. M., Griebie, A. M., Priest, R., & Kuncel, N. R. (2024). A contemporary look at the relationship between general cognitive ability and job performance. *Journal of Applied Psychology, 109(5)*, 687. (See Tables 1, 2)

Sackett, P. R., Lievens, F., & Landers, R. N. (2026). Hiring People in Organizations: The State and Future of the Science. *Annual Review of Organizational Psychology and Organizational Behavior, 13,* 49-75. (See page 52, last paragraph.)

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting Meta-Analytic Estimates of Validity in Personnel Selection: Addressing Systematic Overcorrection for Restriction of Range. *Journal of Applied Psychology, 107(11)*, 2040–2068.

Wise, L. L., McHenry, J., & Campbell, J. P. (1990). Identifying optimal predictor composites and testing for generalizability across jobs and performance factors. *Personnel Psychology, 43(2),* 355-366. (See Table 4 for validity as high as .65.)

Slide 75: Our Tests Are Fair
This is a derivative of the fact that the B-W difference in test performance is twice the size of the B-W difference in job performance.

Slide 79: Our Tests Are Fair
Aguinis, H., & Culpepper, S. A. (2024). Improving our understanding of predictive bias in testing. *Journal of Applied Psychology, 109(3)*, 402-414.

Slide 82: Our Tests Are Fair
Ferrara, E. (2024). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci, 6(1),* 3.
Esp. paragraph 5.1. Definition of Fairness in AI and Its Different Types

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv* preprint arXiv:1609.05807. https://arxiv.org/pdf/1609.05807

Slide 101: SMEs Can Brainstorm KSAPs

For grit:
Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92(6)*, 1087-1101. (See abstract and pg 1093, col 2, par 5, r=.25 with educational attainment among elite students)

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100(3)*, 363-406. (See abstract)

For peripheral vision:
Vater, C., Wolfe, B., & Rosenholtz, R. (2022). Peripheral vision in real-world tasks: A systematic review. *Psychonomic bulletin & review, 29(5)*, 1531-1557.

Slide 102: SMEs Can Brainstorm KSAPs

Reyes, G. (2023). Cognitive Endurance, Talent Selection, and the Labor Market Returns to Human Capital. *arXiv.Org*. (https://arxiv.org/abs/2301.02575)

Jones, K. S., Newman, D. A., Su, R., & Rounds, J. (2022). Vocational interests and adverse impact: How attraction and selection on vocational interests relate to adverse impact potential. *Journal of Applied Psychology, 107(4)*, 604â€“627. https://doi.org/10.1037/apl0000893

Sternberg, R. J. (2015). Successful intelligence: A model for testing intelligence beyond IQ tests. *European Journal of Education and Psychology, 8(2)*, 76-84.

Slide 103: SMEs Can Brainstorm KSAPs

Call, M. L., Nyberg, A. J., & Thatcher, S. (2015). Stargazing: an integrative conceptual review, theoretical reconciliation, and extension for star employee research. *Journal of Applied Psychology, 100(3),* 623-640. (Page 628, col 2)
(For Striving for competence and mastery, learning goal orientation)

Sternberg, R. J. (1996). *Successful intelligence: How practical and creative intelligence determine success in life.* New York: Simon & Schuster.
(Delay of gratification, page 161)

Gallagher, C., & Burke, T. (2007). Age, gender and IQ effects on the Rey-Osterrieth complex figure test. *British Journal of Clinical Psychology, 46(1)*, 35-45. (For visiospacial ability)

Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology, 96(6)*, 1167.

Woehr, D. J., Sheehan, M. K., & Bennett Jr, W. (2005). Assessing measurement equivalence across rating sources: A multitrait-multirater approach. *Journal of Applied Psychology, 90(3)*, 592-600.
(For self-development and self-control, e.g., pages 595-596)

Slide 104: SMEs Can Brainstorm KSAPs

Irons, J. L., & Leber, A. B. (2020). Developing an individual profile of attentional control strategy. *Current Directions in Psychological Science, 29(4)*, 364-371.

Zavlis, O., Bentall, R. P., Fonagy, P. & Rigoli, F. (2025). A Formal Theory of Mood Instability. *Clinical Psychological Science, 13*, 871-892.