How to Build a Cognitive Ability Test with Reduced Mean Group Differences

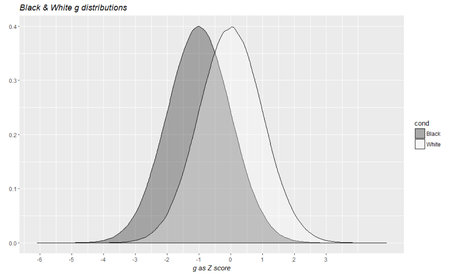
Michael A. McDaniel, Work Skills First, Inc. & Virginia Commonwealth University (Emeritus)

Abstract

- One can build cognitive ability tests with lower mean group differences using items with low *g* saturation and reducing the reliability of the test. Such a diminished *g* test predicts *g* related criteria worse than a *g* test with high *g* saturation and high reliability. Assertions about specific item types causing reduced mean differences are likely incorrect.

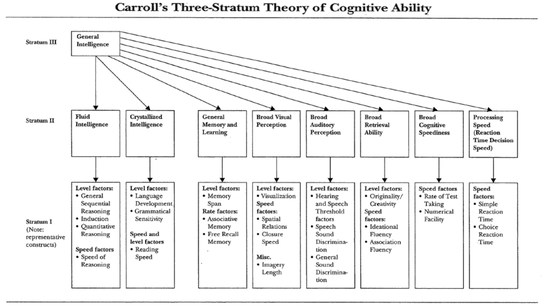
Challenges and Literature

- There are two challenges in building a general cognitive ability (*g)* test with low mean group differences.
- Challenge #1 is that there are large Black-White mean differences in *g*.



- Challenge #2. Spearman (1927) noted that the magnitude of mean Black-White differences co-varied with the extent to which a test was "saturated with *g*" (Spearman, 1927, p. 379). By "saturated with *g*," he meant the extent to which the test measures *g*. Challenge #2 is that if Spearman's Hypothesis is correct, then one cannot build a *g* test that measures *g* well (has high *g* saturation) and has low mean group differences.
- However, one can build a *g* test that does not measure *g* well (has low *g* saturation) that has smaller mean group differences.
- So what *g*-ish measures assess *g* poorly? Carroll's (1993) three stratum model suggests constructs that are related to *g* but which have lower *g* saturation. A good source of item types corresponding to some of Carroll's (1993) constructs is Ekstrom et al. (1976).

This research was funded by the U. S. Army Research Institute for the Behavioral and Social Sciences through Grant W911NF1310353 awarded to Work Skills First, Inc. The views, opinions, and/or findings contained in this article are solely those of the authors and should not be construed as an official Department of the Army or Department of Defense position, policy, or decision, unless so designated by other documentation.



- Yusko et al. (2010) argued that the Siena Reasoning Test shows reduced mean racial differences because it seeks to reduce reliance on prior knowledge, reduce the use of language, and incorporate graphical stimuli.
- Past research does not support the assertions about such item types being responsible for reduced mean group differences (Jenson, 1980, Raven et.al. 1994, 1998).
- Test *g* saturation and test reliability are likely better explanations.

Sample-Size-Weighted Correlations with *g* Tests	
Sienna Reasoning Test (Yusko et al., 2012, slide 24)	.32
Situational Judgment Tests (McDaniel et al. 2007)	.32
Employment Interviews (Huffcutt, et al., 1996)	.40

- If situational judgment tests and employment interviews correlate with *g* as well as or better than the Siena Reasoning Test, should they also be called *g* measures?

Empirical Demonstration

- Data from 12 cognitive scales administered to 927 respondents were used to estimate the *g* saturation of items.
- Scales comparing use of nonsense and real words were written to have the same logical structure which gives them similar *g* saturation.

  A logic-based measurement scale that used real words. ("John likes all dogs")

  A logic-based measurement scale that used some fake words. ("John likes all doferts")

- These scales did not show statistically significant differences in the magnitude of Black-White means.
- Scales comparing graphics- and text-based items were written to have the same logical structure which gives them similar *g* saturation.

  Comparison items using graphics:  is less friendly than

  Comparison items using words: Muffin is a less friendly cat than Fuzzy.

- These scales did not show statistically significant differences in the magnitude of Black-White means.
- Across all items, Black-White item level group differences correlated .60 with item level *g* saturation.
- Mean Black-White mean differences were smaller for tests composed of lower *g* saturation items than for tests composed of higher *g* saturation items. Shorter (30 items), less reliable tests, yielded lower mean Black-White differences than longer (40 items) more reliable tests.

Means of the standardized mean differences for 100 30-item tests vs. 100 40-item tests				
	30 item tests		40 item tests	
	Low *g*	High *g*	Low *g*	High *g*
White-Black *d*	0.45	0.57	0.47	0.58

- Mean reliability of 30-item low *g* tests was smaller (.60) than mean reliability of 30-item high *g* tests (.84). Similar findings (.66 vs. .88) were found for 40-item tests.
- For predicting educational attainment, mean correlations for 30-item low *g* tests were lower than for 30-item high *g* tests (.08 vs. .13). Similar findings (.09 vs. .13) were found for 40 item tests.

References

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Ekstrom, R.B., French, J.W., Harmen, H.H., & Dermen, D. (1976a). *Kit of factor-referenced cognitive tests*. ETS: Princeton, NJ. Available at https://www.ets.org/Media/Research/pdf/Kit_of_Factor-Referenced_Cognitive_Tests.pdf

Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology*, *81*, 459-473.

Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.

McDaniel, M.A., Hartman, N.S., Whetzel, D.L. & Grubb. W.L., III (2007). Situational judgment tests, response instructions and validity: A meta-analysis. *Personnel Psychology, 60*, 63-91.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices*. Oxford: Oxford Psychologists Press Ltd.

Raven, J. C., Court, J. H., & Raven, J. (1994). *Advanced progressive matrices: Sets I and II. Manual for Raven's progressive matrices and vocabulary scales*. Oxford, England: Oxford Psychologists Press.

Yusko, K.P., Goldstein, H.W., Oliver, L.O., and Hanges, P.J. (2010, April). *Building cognitive ability tests with reduced adverse impact: Lowering reliance on prior knowledge.* Paper presented at the 25th annual meeting of the Society of Industrial and Organizational Psychology. Atlanta.

Yusko, K.P., Goldstein, H.W., Scherbaum, C.A., and Hanges, P.J. (2012, April). *Siena Reasoning Test: Measuring intelligence with reduced adverse impact*. Paper presented at the 27th annual meeting of the Society of Industrial and Organizational Psychology. San Diego.