Cognitive ability in everyday life: the utility of open source measures

William Revelle, Elizabeth Dworak, and David Condon
Northwestern University and University of Oregon

Abstract

The measurement of individual differences in cognitive ability has a long and important history in psychology, but has been impeded by the proprietary nature of most assessment measures. With the development of validated open source measures of ability (available from the International Cognitive Ability Resource at ICAR-project.com) it is now possible for many researchers to assess ability in large surveys or small lab based studies without the expenses associated with proprietary measures. We review the history of ability measurement and discuss how the growing set of items included in ICAR allow ability assessments to be more generally available for all researchers.

Introduction

Ever since antiquity, people have used measures of cognitive ability for selection and prediction. The story is told in the Hebrew Bible (*Judges 7) of Gideon who* rejected potential soldiers for showing fear and not having battle wisdom; Plato, in *The Republic; VII: 534, 537* thought that leaders should show exceptional ability and discussed principals of assessment; Theophrastus in his *Characters* depicts the 'stupid man' as slow in speech and action. Given the belief that "Never before in the history of civilization was brain, as contrasted with brawn, so important; never before, the proper placement and utilization of brain power so essential to success" (Yoakum & Yerkes, 1920, p vii.), U.S. Army recruits in the First World War were screened for levels of intelligence deemed necessary to complete their training. An emphasis on cognitive performance continues to this day in the form of

---

contact: William Revelle revelle@northwestern.edu
Final version of February 14, 2020. Not quite the final version as submitted to *Current Directions in Psychology*. (Four more references included in this version.)

standardized testing, such as the SAT for admission to college and the Graduate Record Examination (and several similar tests) for selection to graduate and professional schools (Kuncel & Hezlett, 2007). Of course, successful outcomes have been shown to depend upon much more than cognitive ability. Success in graduate training in clinical psychology requires a mix of ability, stability, and interests (Kelly & Fiske, 1950) and graduate school performance is predicted better by the subject test than either the verbal or quantitative, suggesting some combination of ability and motivation (Kuncel & Hezlett, 2007).

Although initially designed to study "inferior states of intelligence" in children (Binet & Simon, 1916, p 9), early test administrators began assessing "normal" children in terms of their mental age using test items ordered by average performance as a function of chronological age. This practice grew out of efforts to ensure that students received a level of education that was appropriate for their intellectual development (Binet, 1908)[1]. Introduction of the "Intelligence Quotient" led to an explosion of research examining its validity. Terman (1916), for example, demonstrated that children who scored at levels typical of older children were also rated by teachers as smarter or more intelligent. A test that had been developed to assess low levels of ability thus became one that could assess the entire range of cognitive ability.

Early research on intelligence also contributed to advances in measurement and theory. While still a graduate student, Charles Spearman published a fundamentally important paper (Spearman, 1904) establishing the tradition of measuring "general" intelligence (g) that continues to this day (de la Fuente et al., 2019). Although Spearman's samples were tiny by today's standards, his correlations of psychophysical sensitivity to pitch, weight, and light with teacher ratings of "common sense" and cleverness in 24 village children and with school performance in the Classics, French, English and Mathematics in the upper class of a preparatory school (N = 22) showed, when correcting for reliability, a "general function" which he labelled "general intelligence". (In 1904, Spearman also developed the fundamentals of reliability theory as well as the basis of factor analysis.) Students' performance in the Classics correlated highly with performance in other subjects well as their psychophysical sensitivities.

There were several prominent applications of early intelligence research. For example, the notions of item difficulty and deviations from mean performance led to the creation of an index of competence used in the Army Alpha exam for placing US army recruits in the first World War (Yoakum & Yerkes, 1920). In 1932, every 11 year old school child in Scotland was assessed, laying the foundation for a remarkable followup study 69 years later showing the stability of ability measures (r = .66, Deary et al., 2004) as well as their use in predicting important life outcomes such as mortality (Deary, 2008). Indeed, despite ongoing controversies about their use (Hunt & Carlson, 2007; Rindermann et al., 2020), ability measures are associated not just with living longer, but also with success in school,

[1]Binet and Simon's articles from 1905, 1908 and 1911 were translated into English and released as one volume in 1916 (Binet & Simon, 1916); they are still well worth reading.

in job performance, marital stability, and social mobility (Gottfredson, 1997, 2004).

Theories of intelligence

Ever since Spearman it has been routinely noticed that all cognitive measures form a "positive manifold" (the correlations are all positive) which has been taken as an indication of a unified general factor of ability. The correlations of almost all cognitive ability measures are not just positive, but may be arranged in a replicable three or four level hierarchy of specific tests of narrow abilities, groups of tests of broad abilities (e.g., fluid, crystallized, memory) and a higher factor known as g (Carroll, 1993). Alternatively, it has been proposed that the third level is better represented with factors for verbal, perceptual and rotation ability below the higher order g (Bouchard, 2014; Johnson & Bouchard Jr., 2005).

However, it has been recognized for more than 100 years (e.g., Thomson, 1916) that the existence of such a positive manifold is a descriptive finding and should not be taken as having any necessary causal meaning, as there are several ways that such a positive manifold might be produced (Bartholomew et al., 2009; Kovacs & Conway, 2019; Van Der Maas et al., 2006). Sampling independent "bonds" (Bartholomew et al., 2009), dynamic mutualism (Van Der Maas et al., 2006), and overlapping processes (Kovacs & Conway, 2019) all result in the same set of positive correlations without a causal general factor. This can be seen via simulation of a genetic factor model of independent genes with pleiotropic effects (simulated as cross loadings) which yields a positive manifold and a "g" factor even though the underlying casual mechanisms are independent (for demonstration, see the sim.bonds function in the *psych* package (Revelle, 2020)).

By analogy, an equivalent positive manifold may be found in measures of body size. Whether measuring weight, height, chest circumference or hundreds of more precise measures, adult humans differ in a general factor of size (e.g., the USAF data set in *psych*). Even among a homogenous group of male Air Force personnel, there is a clear general factor of size, with positive correlations across many anatomical features. The utility of this analogy to g can be extended further, for both general factors show: (1) clear hierarchical structure; (2) additive effects among (and across) many genes; (3) high sensitivity to environmental effects (e.g., nutrition) ; and (4) robust age trends. Regrettably, changes in body size and g tend to drift in the opposite direction with age, though both reliably change with greater variability in more specific domains.

Developmentally, cognitive ability can be thought of as a *propensity* to acquire new information and new reasoning skills. It is analogous to differences in stickiness as snowballs roll down hill. Just as sticky snowballs become larger than those less sticky, so do high ability individuals acquire more information as they experience life.

Classic longitudinal studies

The question of causality does not diminish the usefulness of the general factor as a predictor of real world outcomes. Terman & Oden (1959) reported on the life time accomplishments of his 1,528 "termites" — these were very bright 3rd to 8th grade Californians with Stanford Binet scores mainly above 140 (roughly, the top 1% of the student population). Contrary to the prevalent hypothesis when the study began that high ability was related to psychological fragility, the participants were psychologically healthy and showed impressive levels of accomplishment over their lifetimes (see Lubinski, 2016) . In more recent longitudinal study based upon the representative sample of 440,000 US high school students in Project Talent (Flanagan et al., 1962), 50 year follow-ups of 1,952 9th to 12th graders demonstrated the predictive validity of cognitive performance tests. Ability measures taken 50 years earlier correlated .50, .35, and .35 respectively with (subsequent) educational attainment levels, occupational level, and estimated income (Spengler et al., 2018), and the effects remained robust even when controlling for parental social status (partial correlations were .40, .29, and .28).

The often stated claim that differences in ability do not make much difference for the outcomes of the top 1-2% in ability is contradicted by differences in the achievement of participants in another 50 year longitudinal study of mathematically precocious youth (Lubinski & Benbow, 2006). Even among students identified by their SAT scores at age 14 to be among the top 1%, those in the top 1 in 10,000 (.01%) had even more accomplishments in the next 35-50 years than did those who were "merely" exceptional. Lubinski reminds us that there are six standard deviations of ability *above* the mean level and that one third of the total range is observed within the top 1%.

Genetics of cognitive ability

Classic behavioral genetics work comparing the similarities of identical twins to fraternal twins as well as the lack of similarity of adopted siblings shows that roughly 70-80% of the variance in ability as measured by conventional intelligence tests (among those with a middle class background) is under genetic influence (Bouchard, 2014). These findings show systematic increases with age. Siblings pairs, whether adopted, dizygotic or monozygotic twins are all very similar when 5-7 years old but the adopted sibs become less similar while the mono-zygotic twins more similar as they age (Bouchard, 2014). Much lower estimates of heritability come from Genome Wide Association Studies which exam common polymorphisms. Analyses of more than 1 million participants in the UK Biobank have shown that years of education (a proxy for cognitive ability and motivation) may be associated with 1,271 independent Single Nucleotide Polymorphisms (Lee et al., 2018). The implications of these findings is that ability and subsequent outcomes are substantially heritable, but this does not imply that environmental influences are not important. It also underscores the fact that heritability is a hodge-podge ratio of genetic variance to total variance (genetic plus environmental) for a particular sample, leaving many unanswered questions about the extent to which changes in the environment can affect phenotypic scores. Psychological and physical differences can be be highly heritable but also highly malleable by the environment (e.g., height). Furthermore, in the US, heritability of ability estimates vary as a function of social class (Giangrande et al., 2019), but this effect is not observed in Europe or Australia which may be taken as a sign of greater socioeconomic inequality in the US (Tucker-Drob & Bates, 2016).

Cognitive ability and cognitive processes

Although faced with problems of small samples and restriction of range when studying college students, individual differences in g may be related the basic cognitive processes studied in experimental psychology (Engle, 2018). Structural equation modeling of a variety of cognitive tasks along with more conventional psychometric tasks shows remarkable agreement between the higher order factors of each, with some evidence of moderation of loadings of basic cognitive tasks depending upon the level of the higher order g factor (Kovacs et al., 2019). Some lower level processes (e.g. object recognition) show much smaller correlations with measures of g (Richler et al., 2017) than do measures of working memory.

Measurement: the development of ICAR

Even though clearly important, the study of individual differences in cognitive ability has been limited by several constraints, including the related issues of cost, sample size, and "scalability." The high costs of ability testing stems from the field's reliance mainly on proprietary licensed measures. The expense of licensing tends to severely constrain researchers' budgets, leading to the collection of smaller sample sizes than might otherwise be possible. Even the ETS "French Kit" (Ekstrom et al., 1976) is $.15 per copy for graduate students and is not suitable for web based administration. It is also the case that the most widely-used ("high stakes") measures tend to require one-on-one or proctored, small-group administration. These problems are compounded by the tradition of relying on undergraduate samples as this often leads to restriction of range and concerns about generalizability.

To alleviate these problems we developed and validated an open source ability test that is well-suited for administration on the web (the International Cognitive Ability Resource, Condon & Revelle, 2014). Although the original instrument had just 60 items spanning four constructs, with the help of an international consortium (Condon et al., 2014) we have expanded the total item pool to more than 1,000 items and 19 lower level constructs. Additional measures are currently under development for an increasingly broad range of constructs. For the sake of cross-validation against other ICAR measures, subset of each type are administered to large online samples using a "Massively Missing Completely at Random" design (Revelle et al., 2016). The original form (Condon & Revelle, 2014) was based on four sub-factors (three-dimensional rotation, matrix reasoning, letter/number series, and verbal reasoning) with a clear hierarchcal factor structure. The newer measures include a forced choice remote associates test, two dimensional rotations, propositional reasoning, figural analogies, numeracy, map use, and more complex matrix reasoning problems. Computer generated number series have been validated against the original items and added to ICAR (Loe et al., 2018).

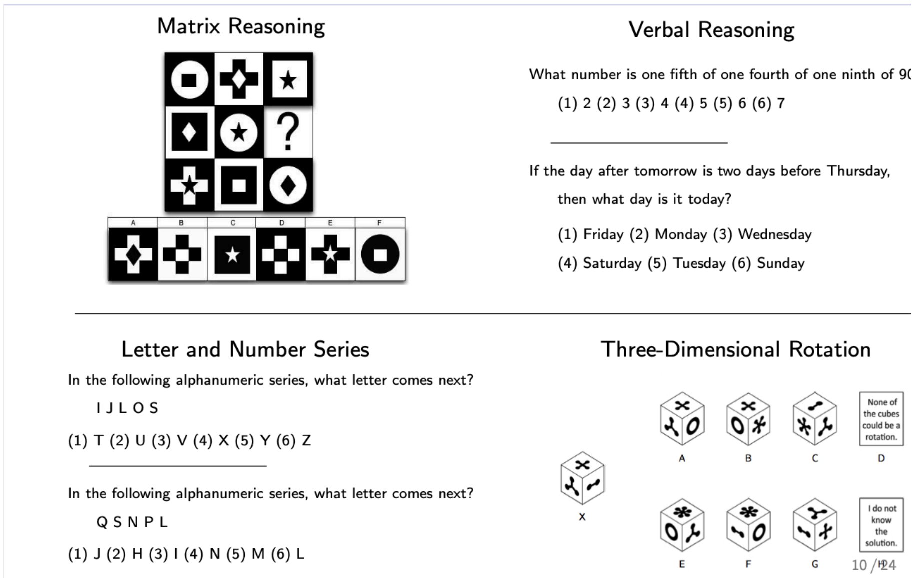

Figure 1. The original 60 item ICAR was composed of four item types (examples of which are shown here) with a clear hierarchical factor structure. See Condon & Revelle (2014) for more example items and join the ICAR project at ICAR-Project.com for access to all of the items.

Applications of ICAR

Although one reviewer suggested that to compare the ICAR to the Stanford Binet is analogous to comparing a cheap ripoff to a Versace handbag, we view the utility of ICAR in terms of its wide range of applications in just the past few years. The use of the ICAR measures of cognitive ability have already been seen across many studies and publications with various real world criteria and different item types (e.g., the 79 studies reviewed in Dworak et al., 2020). Such projects include an online survey that utilized 35 verbal reasoning and three-dimensional rotation items to provide participant feedback and evaluate individual differences in a nationwide sample (Krieke et al., 2016). Other studies assessed how 46 verbal reasoning and matrix reasoning items related to genetic scores of education attainment and showed that large scale genetic studies can rely on online collection of cognitive ability measures (Liu et al., 2017). ICAR items have also been utilized with experience sampling methods to test the relationship between cognitive ability and creativity. Cognitive ability was also found to moderate the relationship between everyday positive affect and everyday creativity (Karwowski et al., 2017). Using 16 items, one cross-sectional study found higher cognitive ability related to greater aptitude in discriminating between "pseudo-profound bullshit" and profound statements (Bainbridge et al., 2019). Research has used as few as 4 items to find that cognitive ability relates negatively to the political ideologies of right-wing authoritarianism and social dominance orientation and attitudes towards Trump (Choma & Hanoch, 2017).

Future Directions

We have received requests for the use of ICAR items with younger subjects (less than 14) and as potential measures of cognitive decline in the elderly. The factor structure of the original 60 items of the ICAR was based on the responses of 96,958 participants with a median age of 22 but which ranged from 14-90 years of age. A subsequent validation against self reported SAT and ACT was done for those 34,229 participants between 18 to 22 years of age. Thus, there is a need to further validate the items with younger and older participants. Although some researchers have used as few as four items in their studies, and many have used just the 16 items from the sample test, we encourage users to go beyond these 16, and even the 60 described in Condon & Revelle (2014) and use items sampled from the larger ($> 1,000$) pool of items that are available at the ICAR-project web site.

Recommended Readings

Deary, I. J. (2000). Looking down on human intelligence: From psychometrics to the brain. Oxford: Oxford University Press. A thoughtful and well integrated series of essays on the history, measurement and correlates of intelligence.

Deary, I. J. (2001). Intelligence: A very short introduction. Oxford: Oxford University Press One of the Oxford "Short introduction" series, this is a delightful and informative review of the meaning and importance of intelligence meant for the general reader.

Haier, R. J. (2016). The neuroscience of intelligence. Cambridge University Press. The current status of biological models of intelligence.

Johnson, W. (2010). Understanding the genetics of intelligence: Can height help? can corn oil. Current Directions in Psychological Science, 19(3), 177-182. A very clear discussion of why heritability within groups is irrelevant when discussing between group differences.

Mackintosh, N.J. (2011) IQ and human intelligence. Oxford University Press. Oxford. A very useful review of the history of intelligence testing.

Lubinski, D. (2016). From Terman to today: A century of findings on intellectual precocity. Review of Educational Research. doi: 10.3102/0034654316675476 A very thoughtful review of intellectual precocity featuring the Terman and Stanley/Benbow/Lubinski longitudinal studies.

Sackett, P. R., & Kuncel, N. R. (2018). Eight myths about standardized admissions testing. In Buckley, J, Letukas, L. and Wildavsky, B. (Eds) Measuring Success: Testing, Grades, and the Future of College Admissions, Baltimore: Johns Hopkins University Press pp13-38. Addresses many false claims about the use of ability tests for college admissions.

References

Bainbridge, T. F., Quinlan, J. A., Mar, R. A., & Smillie, L. D. (2019). Openness/intellect and susceptibility to pseudo-profound bullshit: A replication and extension. *European Journal of Personality*, *33*(1), 72-88. doi: 10.1002/per.2176

Bartholomew, D., Deary, I., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological review*, *116*(3), 567–579. doi: 10.1037/a0016262

Binet, A., & Simon, T. (1916). *The development of intelligence in children: Translated by Elizabeth S. Kite* (H. H. Goddard, Ed.). Baltimore, Md.: William and Wilkens Company.

Bouchard, T. J., Jr. (2014). Genes, evolution and intelligence. *Behavior Genetics*, *44*(6), 549-577. doi: 10.1007/s10519-014-9646-x

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY, US: Cambridge University Press. doi: 10.1017/CBO9780511571312

Choma, B. L., & Hanoch, Y. (2017). Cognitive ability and authoritarianism: Understanding support for Trump and Clinton. *Personality and Individual Differences*, *106*, 287 - 291. doi: 10.1016/j.paid.2016.10.054

Condon, D. M., Doebler, P., Holling, H., Gühne, D., Rust, J., Stillwell, D., . . . Revelle, W. (2014). International Cognitive Ability Resource. *https://icar-project.com.*

Condon, D. M., & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52-64. doi: 10.1016/j.intell.2014.01.004

Deary, I. J. (2008). Why do intelligent people live longer? *Nature*, *456*(7219), 175–176. doi: 10.1038/456175a

Deary, I. J., Whiteman, M., Starr, J., Whalley, L., & Fox, H. (2004). The impact of childhood intelligence on later life: Following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*, 130–147. doi: 10.1037/0022-3514.86.1.130

de la Fuente, J., Davies, G., Grotzinger, A. D., Tucker-Drob, E. M., & Deary, I. J. (2019). Genetic "general intelligence," objectively determined and measured. *bioRxiv*. doi: 10.1101/766600

Dworak, E. M., Revelle, W., Doebler, P., & Condon, D. M. (2020). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences (in press).*

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Manual for kit of reference tests for cognitive factors:.* Educational Testing Service.

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, *13*(2), 190-193. doi: 10.1177/1745691617720478

Flanagan, J. C., Dailey, J. T., Shaycoft, M. F., Gorham, W. A., Orr, D. B., & Goldberg, I. (1962). *The talents of American youth: Design for a study of American youth* (Vol. 1). Boston, MA: Houghton Mifflin.

Giangrande, E. J., Beam, C. R., Carroll, S., Matthews, L. J., Davis, D. W., Finkel, D., & Turkheimer, E. (2019). Multivariate analysis of the scarr-rowe interaction across middle childhood and early adolescence. *Intelligence*, *77*, 101400. doi: 10.1016/j.intell.2019.101400

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79-132. doi: 10.1016/S0160-2896(97)90014-3

Gottfredson, L. S. (2004). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology.*, *86*(1), 174–199. doi: 10.1037/0022-3514.86.1.174

Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science*, *2*(2), 194–213.

Johnson, W., & Bouchard Jr., T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*(4), 393 - 416. doi: 10.1016/j.intell.2004.12.002

Karwowski, M., Lebuda, I., Szumski, G., & A.Firkowska-Mankiewicz. (2017). From moment-to-moment to day-to-day: Experience sampling and diary investigations in adults' everyday creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(3), 309-324. doi: 10.1037/aca0000127

Kelly, E. L., & Fiske, D. W. (1950). The prediction of success in the VA training program in clinical psychology. *American Psychologist*, *5*(8), 395 - 406. doi: 10.1037/h0062436

Kovacs, K., & Conway, A. R. A. (2019). What is iq? life beyond "general intelligence". *Current Directions in Psychological Science*, *28*(2), 189-194. doi: 10.1177/0963721419827275

Kovacs, K., Molenaar, D., & Conway, A. R. (2019). The domain specificity of working memory is a matter of ability. *Journal of Memory and Language*, *109*, 104048. doi: https://doi.org/10.1016/j.jml.2019.104048

Krieke, L. V. D., Jeronimus, B. F., Blaauw, F. J., Wanders, R. B., Emerencia, A. C., Schenk, H. M., ... others (2016). HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International journal of methods in psychiatric research*, *25*(2), 123–144. doi: 10.1002/mpr.1495

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*(5815), 1080-1081. doi: 10.1126/science.1136618

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... Consortium, S. S. G. A. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. doi: 10.1038/s41588-018-0147-3

Liu, M., Rea-Sandin, G., Foerster, J., Fritsche, L., Brieger, K., Clark, C., ... others (2017). Validating online measures of cognitive ability in genes for good, a genetic study of health and behavior. *Assessment*. doi: 10.1177/1073191117744048

Loe, B., Sun, L., Simonfy, F., & Doebler, P. (2018). Evaluating an automated number series item generator using linear logistic test models. *Journal of Intelligence*, *6*(2), 20. doi: 10.3390/jintelligence6020020

Lubinski, D., & Benbow, C. P. (2006). Study of Mathematically Precocious Youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, *1*(4), 316–345. doi: 10.1111/j.1745-6916.2006.00019.x

Revelle, W. (2020, January). psych: Procedures for personality and psychological research [Computer software manual]. https://CRAN.r-project.org/package=psych. (R package version 2.0.1)

Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G. (2016). Web and phone based data collection using planned missing designs. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *Sage handbook of online research methods* (2nd ed., p. 578-595). Sage Publications, Inc.

Richler, J. J., Wilmer, J. B., & Gauthier, I. (2017). General object recognition is specific: Evidence from novel and familiar objects. *Cognition*, *166*, 42 - 55. doi: 10.1016/j.cognition.2017.05.019

Rindermann, H., Becker, D., & Coyle, T. R. (2020). Survey of expert opinion on intelligence: Intelligence research, experts' background, controversial issues, and the media. *Intelligence*, *78*. doi: 10.1016/j.intell.2019.101406

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology*, *15*(2), 201-292. doi: 10.2307/1412107

Spengler, M., Damian, R. I., & Roberts, B. W. (2018). How you behave in school predicts life success above and beyond family background, broad traits, and cognitive ability. *Journal of Personality and Social Psychology*, *114*(4), 600-636. doi: 10.1037/pspp0000185

Terman, L. M. (1916). *The measurement of intelligence*. Houghton Mifflin.

Terman, L. M., & Oden, M. (1959). *The gifted group at mid-life: Thirty-five years' follow-up of the superior child* (Vol. 5). Stanford Univ Pr.

Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, *8*, 271-281.

Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene x socioeconomic status interaction on intelligence. *Psychological Science*, *27*(2), 138–149. doi: 10.1177/0956797615612727

Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006, 10.1037/0033-295X.113.4.842). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842 - 861.

Yoakum, C. S., & Yerkes, R. M. (Eds.). (1920). *Army mental tests.* Henry Holt and Company.