Intelligence 91 (2022) 101617

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Thirty years of research on general and specific abilities: Still not much more than *g*

Malcolm James Ree [a,*], Thomas R. Carretta [b]

[a] Our Lady of the Lake University, San Antonio, TX 78207, United States of America
[b] Air Force Research Laboratory, Wright – Patterson AFB, OH 45433, United States of America

A R T I C L E I N F O

Keywords:
General cognitive ability
Specific cognitive abilities
Construct validity
Incremental validity

A B S T R A C T

Results of a 30-year research program indicate that specific cognitive abilities (*s*) provide little or no incremental validity beyond general cognitive ability (*g*). Definitions of *g* and *s* are provided and examples from training and job performance are presented. All samples are large adding to confidence in the results. On average, the increased validity for multiple regressions between using *g* versus *g* plus *s* was 0.02. The weight of the evidence suggests that the increment of 0.02 is an artifact of measurement error. An alternative ability model that fails to separate *g* and *s* is presented.

One of the longest controversies in educational and industrial-organizational psychology is the measurement and roles of general (*g*) and specific abilities (*s*). Although there are many ways to think about abilities, division into general and specific abilities makes sense, especially in structuring research, applications, and communication. Resurgence of interest in general and specific abilities in the last few decades makes it likely that the controversy will continue. In this paper, we define general and specific abilities and review our program of research and that of others.

1. Defining general and specific abilities

General cognitive ability (*g*) is a source of reliable variance that is common across multiple measures. For example, the average correlation for the subtests of the Armed Services Vocational Aptitude Battery (ASVAB Forms 11, 12, and 13), a multiple aptitude test, is about 0.60 and is an indicator of a large first factor, *g*, accounting for about 64% of the total variance. The ASVAB is highly *g*-saturated (Ree & Carretta, 1994). The average correlation of the Air Force Officer Qualifying Test (AFOQT; Carretta & Ree, 1996) subtests is about 0.43, where *g* accounts for only about 41% of the total variance. The difference is the consequence of the AFOQT aviation knowledge subtests. The correlations for both the ASVAB and AFOQT display "positive manifold" (Spearman, 1927) as found in all multiple-ability cognitive batteries.

A specific ability (*s*) is a source of reliable variance found in each measure but specific to that measure and independent of *g*. It must be independent so that its effect can be differentiated from that of *g*. There could be 9 specific abilities in a battery of 9 tests or two or more subtests could measure the same *s*.

A test of arithmetic reasoning (AR) has at least three sources of variance. The dominant source is *g*, one or more sources of *s*, and random error. The variance of test scores is the sum of these variances; $AR = g + s + e$ without covariances among these independent sources. The goal is to isolate these sources for use in theory building or application. To conflate *g* and *s* and identify the sum as a specific ability (see, for example, Kell & Lang, 2017; Krumm, Schmidt-Atzert, & Lipnevich, 2014) is erroneous.

The model of general and specific sources of variance has been applied to other domains such as personality, emotional intelligence, cognitive components, job performance, and psychomotor (Carretta, Perry Jr., & Ree, 1996; Carretta & Ree, 1997; Carretta, Ree, & Teachout, 2012; Ree, Carretta, & Teachout, 2015; Stauffer, Ree, & Carretta, 1996).

2. Correction for range restriction

In most studies of abilities samples subject to prior selection must be used, creating the condition known as range restriction (Thorndike, 1949). If the sample has been selected on one or more variables any correlations among the variables will be downwardly biased, although rarely in extreme conditions the correlations may increase (Levin, 1972). For more than a century methods to correct range restricted estimates have been available. Some researchers (Richardson & Norgate,

* Corresponding author.
E-mail addresses: Malcolmree@att.net (M.J. Ree), Thomas.Carretta@US.AF.MIL (T.R. Carretta).

https://doi.org/10.1016/j.intell.2021.101617
Received 22 November 2021; Received in revised form 14 December 2021; Accepted 14 December 2021
Available online 24 January 2022

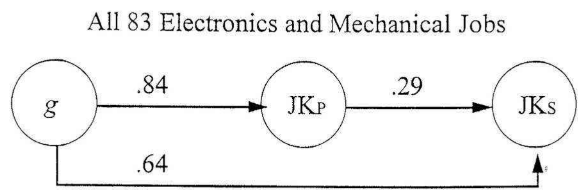

Fig. 1. Structural equation model for the role of g and prior job knowledge (JK_P) in the acquisition of subsequent job knowledge (JK_S) during training for 83 US Air Force Electronics and Mechanical training specialties (Ree, Carretta, & Doub, 1998/1999).

2015) misunderstand the assumptions of range restriction correction procedures. They are two of the three assumptions of correlation (linearity, and homoscedasticity); normality is not assumed for the correction procedures. If you can estimate a correlation you can correct it for range restriction.

There are those who consider correction for range restriction improper. For example, Landy (2003 p. 155) called it the "psychometric equivalent of alchemy." However, Millsap (1989) showed correcting correlations usually under corrects and the corrected values are closer to the population values. Held and Foley (1994) found the same result. Corrected correlations were always closer to the population correlations than the uncorrected correlations.

Sackett and Yang (2000) published a valuable taxonomy of methods of correcting for range restriction. This taxonomy should be consulted when performing research on general and specific cognitive abilities.

3. Knowledge versus ability

There are measures of knowledge such as the ASVAB Auto/Shop and Electronics Information subtests or AFOQT Aviation Information and Instrument Comprehension subtests. Knowledge should never be identified as ability. Acquisition of knowledge is dependent on *g*. Hunter (1986) provides a good example as do Ree, Carretta, and Doub (1998/1999). Ree et al. investigated the role of *g* and prior job knowledge (JK_P) on the acquisition of subsequent job knowledge (JK_S) during training using structural equation models. Data were from 83 independent jobs with a total sample of 42,399 trainees. Three models were examined: 1) a role for *g* only, 2) a role for JK_P only, and 3) a role for both. Results supported a model where JK_P mediated the relations of ability (*g*) and acquisition of JK_S (see Fig. 1). The total direct and indirect causal effect of *g* on the acquisition of subsequent job knowledge was 0.88 for the model that included all 83 jobs. The causal effect for prior job knowledge was 0.29.

4. Alternate models of general and specific abilities

Recently some researchers (Kell & Lang, 2017) have conceptualized specific ability as a combination of *g* and *s* like second-order factors in hierarchical factor analysis or that *g* is solely comprised of independent specific abilities (Krumm et al., 2014). The inclusion of *g* with *s* makes comparing their predictive efficiencies impossible.

Frequently the name of the test suggests a specific ability. A test of space perception is deemed specific due to its name and questions. This is sometimes called content-aligned specific ability. The AFOQT Arithmetic Reasoning and Word Knowledge subtests correlate at about $r = 0.46$ despite radically different content (Carretta & Ree, 1996). This correlation is the consequence of *g*. Content is an unreliable indicator of factors leading to the "topological fallacy" (Walters, Miller, & Ree, 1993).

Finally, some researchers argue that a particular method of estimating *g* is preferable. Possible methods include unrotated principal components (PC), unrotated principal factors (PF), and bi-factor (BIF) (Jensen, 1998; Ree, Carretta, & Earles, 1998; Ree & Earles, 1994). Wilks (1938) demonstrated almost perfect correlation between composites of variables when the mean value of the correlations is large. Ree, Carretta, and Earles (1998)) illustrated this using ASVAB subtest scores with positive weights (as would be found in PC, PF, and BIF) some of which were randomly determined. With as few as 5 variables, and low subtest intercorrelations of $r = 0.10$, the expected correlation of linear composites was 0.57 (see Ree, Carretta, & Earles, 1998, p. 415, Table 6). With a greater number of variables or greater average intercorrelation, the composite correlation approaches 1.

Kell and Lang (2017) proposed a model in which the common variance that defines *g* is associated with the variance of $s_1 \ldots s_n$. As this common variance is found in each factor, ordinary regression fails to handle it well due to colinearity. Therefore, they conducted analyses using regression weight relative importance analysis. This model and this procedure will fail to embody characteristics attributable to *g*.

Krumm et al. (2014) suggested use of a variant of the Brunswick Lens model. It requires:

"A basic assumption of this model is that several cues, which are more or less indicative of the criterion, are used for the prediction." (p 118). They also hypothesize that general cognitive ability is made up only of a group of specific abilities. Further, they stated that general cognitive ability is a good predictor if the underlying specific abilities are "indicative" of the criterion.

The model proposed by Krumm et al. (2014) does not allow for *g* and *s* to coexist and that they can be separated in analyses through residualization (Schmid & Leiman, 1957). They also call for *g* to be assembled from a set of cognitive components. This suggests that measuring the right cognitive component could increment *g*. Much of the evidence is otherwise. Stauffer et al. (1996), in a reanalysis of Kyllonen (1993), demonstrated that cognitive components are highly *g*-saturated and can only increment prediction by adding to the total *g* variance available for prediction (Ree & Carretta, 2011). Stauffer et al. conducted separate confirmatory factor analyses for the ASVAB and the cognitive components battery. Both batteries exhibited a hierarchical structure with a general cognitive ability factor at each apex, The two general factors correlated 0.994, supporting Spearman's (1927) concept of the indifference of the indicator.

5. The specificity doctrine

The specificity doctrine, sometimes called "situational specificity," posits that specific abilities or combinations of specific abilities are at least equal to or superior to *g* for prediction and understanding. In the third decade of the 20th century, Hull (1928) asserted that test scores should be differentially weighted for predicting performance in each job. He believed that specific abilities could replace or compensate for a lack of *g*. This means that validity would not generalize across jobs, locations, or selection techniques. A validation would be required for every job and selection situation. Meta-analyses of multiple jobs (e.g., Hunter, 1986) do not support the specificity doctrine.

Also, the specificity doctrine says that *s* should be incremental to *g* in prediction. Nye et al. (2020) explored the specificity doctrine with cognitive ability measures, knowledge measures, simple cognitive components, and psychomotor tests. In a sample of 310 Navy student pilots, they reported incremental validity beyond *g* for psychomotor tests for training outcomes. The results were ambiguous because the sample data were not corrected for range restriction.

This doctrine holds that there would be a weighted composite of ability that is unique for job tasks that should moderate the relationship of *g* and *s* for jobs and job families. Jones and Ree (1998, see their Fig. 1) tested this aspect of the specificity doctrine using the ASVAB multiple abilities battery with 24,482 Air Force enlistees in 37 jobs. Building on Schmidt, Hunter, and Pearlman (1981) and substituting job ability requirements for job task requirements the prediction of differences among and between jobs was tested. The sample was subjected to direct and indirect selection creating range restriction which was corrected using the multivariate method (Lawley, 1943). Job families were based on hierarchical clustering using Ward' method Alley, Treat, & Black, 1988. In neither jobs nor job families was moderation found. The specificity doctrine was not supported.

6. Training and job performance

Ree and Earles (1991) studied training outcomes in a sample of 78,041 Air Force enlistees in 82 jobs using the 10 ASVAB subtests. The ASVAB subtests are attuned to high school curricula and measure verbal, quantitative, perceptual speed, and technical knowledge. All enlistees had completed basic military training and job-specific technical training. ASVAB scores were corrected for multivariate range restriction (Lawley, 1943). The first unrotated principal component of the ASVAB served as the measure of *g* and the other 9 unrotated principal components as measures of *s*. The most complex regression model allowed each job to have its own regression slope and intercept. There were 902 variables from the 82 job identifying binaries and 820 product variables for the interaction of principal components and jobs. A less complex model allowed common slopes for principal components of *g* and *s*. The least complex model used only the first principal component (*g*) with a common slope for all jobs. See Ree and Earles (1991, p. 327, Table 3) for a detailed description of the models. Differences between regression models were tested via the *F* test at $p < .01$. Again, only trivial validity gains were found for specific abilities which averaged about 0.02.

Using data from the Air Forces' job performance measurement project, Ree, Earles, and Teachout (1994) evaluated the contributions of *g* and s for eight jobs using the methodology as described in Ree and Earles (1991). There were 1545 Air Force enlistees tested on the ASVAB. The job performance criteria were developed as part of the Joint-Services Job Performance Measurement project (Wigdor & Green Jr., 1987). The criterion measures were hands-on work samples, technical interviews in which participants explained how to perform technical tasks, and a combination of these criteria. The data were corrected for multivariate range restriction (Lawley, 1943) with regressions comparing prediction of the criteria by *g* with prediction by *g* and s. The average increment to prediction for *g* and *s* versus only *g* was about 0.02 for all three criteria.

Olea and Ree (1994) used Air Force samples ranging from 1856 to 3942 pilot and 1176 to 1411 navigator students to investigate the role of *g* and *s* in predicting training outcomes. The 16 Air Force Officer Qualification Test (AFOQT) subtests measure verbal, quantitative, spatial, perceptual speed, and pilot job knowledge; however, the AFOQT has no job knowledge tests for navigators. AFOQT composite scores were used to select the trainees so they constituted a range restricted sample. The sample was corrected for range restriction using the multivariate method (Lawley, 1943). General and specific abilities were represented by the AFOQT unrotated principal components. The criteria included academic grades and ratings of flying maneuvers (e.g., landings, loops, and rolls) for pilot and airborne navigation tasks (e.g., day and night celestial fixes and locations) for navigators. *g* was the best predictor of all criteria and *s* contributed little beyond *g*, despite the wide variety of criteria. The incremental validity of specific measures beyond the prediction provided by *g* averaged 0.08 for pilot and 0.02 for navigator criteria. Results showed that the incremental validity of specific measures for pilots was due to *specific knowledge* about aviation principles, aviation instruments, and aircraft controls, not a specific ability. No analogous navigator-specific knowledge subtests were in the AFOQT.

Brown, Le, and Schmidt (2006) hypothesized that the small differences between prediction by *g* versus by *g* and *s* were a reliability artifact. In a study of 26,097 Navy recruits in 10 jobs, they conducted regressions of training performance on six ASVAB subtests (Arithmetic Reasoning, Mathematics Knowledge, Word Knowledge, Paragraph Comprehension, Electronics Information, and Mechanical Comprehension). Three composites of two subtests each were created by summation to represent quantitative, verbal, and technical measures. Noting that the most informative test of *g* versus *s* would come from true scores regressions, the scores were corrected for unreliability based on ASVAB subtest reliabilities from the 1987 applicant population. Regressions were used to determine whether individual weighting of the subtests was superior to a *g* composite created by summing the six subtests. Results indicated that true score regressions erased the small incremental validity differences found by Ree and colleagues.

7. Conclusions

The controversy over *g* versus *s* cognitive abilities is still active despite numerous comparative studies. In this paper general cognitive ability and specific abilities have been defined as separate and orthogonal. Other theories propose a chimera of general and specific abilities. 'Abilities' developed by the joining of general and specific abilities variances are usually not tested with linear regression but by some form of dominance analysis. This analytic method neither supports nor refutes accumulated results.

Many of the proposed alternate models do not separate *g* from *s* making tests of relative predictive efficiency impossible. Other models propose that general cognitive ability is made up of separate specific abilities indicating that general cognitive ability cannot exist along with specific abilities. Models where *g* and *s* are not orthogonal typically use non-regression statistical analyses. These analyses with variables that include a portion of *g* with *s* show *s* to dominate *g*. A flaw in these studies is that the sample correlations are never corrected for range restriction and results cannot generalize beyond the sample.

The need to correct observed correlations for range restriction is vital to understanding the relations between *g* and *s* and generalizability. Bryant and Gokhale (1972) wrote that "…to infer beyond the sample a correction for restriction in range is necessary" (p. 305).

The military has provided large samples for studies. Those researchers using hierarchical linear regression with orthogonal measures of *g* and *s* have consistently found little incremental validity for specific abilities. An average result is an increment of 0.02. This has been true for both training and job performance criteria. This finding applies to simpler jobs and to very complex jobs. Brown, Le and Schmidt (2006) applied a method using true scores and structural equation models and reduced the 0.02 difference to nearly zero in a sample with 10 jobs. Replication with larger samples and additional jobs is needed.

Constructs that have been shown to increment *g* are personality, psychomotor, integrity, and job knowledge. Personality, integrity, and job knowledge are measurable with a computer or paper–and–pencil, but psychomotor tests require computers and specialized control mechanisms (e.g. joysticks, foot pedals). These create another source of error as control mechanisms must be calibrated.

Another issue is the lack of thorough evaluation of new constructs. Here are five important questions. Have construct and empirical validity been assessed in reference to established constructs? Are the same constructs being measured for various groups? What are the mean differences among groups (Carretta & Ree, 1995), how large are practice effects, and does the new specific ability pass a cost-benefit analysis? Finally, there should be no proliferation of new constructs when these questions remain unanswered.

References

Alley, W., Treat, B., & Black, D. (1988). *Classification of Air Force jobs into aptitude clusters. (AFHRL-TR-88-14)*. Brooks AFB, TX: Air Force Human Resources Laboratory.

Brown, K. G., Le, H., & Schmidt, F. L. (2006). Specific ability theory revisited: Is there incremental validity for training performance? *International Journal of Selection and Assessment, 14*, 87–100. https://doi.org/10.1111/j.1468-2389.2006.00336.x

Bryant, N. D., & Gokhale, S. (1972). Correcting correlations for restrictions in range due to selection on an unmeasured variable. *Educational and Psychological Measurement, 32*, 305–310. https://doi.org/10.1177/001316447203200207

Carretta, T. R., Perry, D. C., Jr., & Ree, M. J. (1996). Prediction of situational awareness in F-15 pilots. *International Journal of Aviation Psychology, 6*, 21–41. https://doi.org/10.1207/s15327108ijap0601_2

Carretta, T. R., & Ree, M. J. (1995). Near identity of cognitive structure in sex and ethnic groups. *Personality and Individual Differences, 19*, 145–155. https://doi.org/10.1016/0191-8869(95)00031-Z

Carretta, T. R., & Ree, M. J. (1996). Factor structure of the Air Force Officer Qualifying Test: Analysis and comparison. *Military Psychology, 8*, 29–42. https://doi.org/10.1207/s15327876mp0801_3

Carretta, T. R., & Ree, M. J. (1997). Expanding the nexus of cognitive and psychomotor abilities. *International Journal of Selection and Assessment, 5*, 149–158. https://doi.org/10.1111/1468-2389.00055

Carretta, T. R., Ree, M. J., & Teachout, M. S. (2012). Measuring personality in business: The general personality factor in the mini-IPIP. *American Journal of Management, 12*, 81–87.

Held, J. D., & Foley, P. P. (1994). Explanations for accuracy of the general multivariate formulas in correcting for range restriction. *Applied Psychological Measurement, 18*, 355–367. https://doi.org/10.1177/014662169401800406

Hull, C. L. (1928). *Aptitude testing. Great Britain: World Book*. https://doi.org/10.1037/11019 - 0000

Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior, 29*(3), 340–362. https://doi.org/10.1016/0001-8791(86)90013-8

Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.

Jones, G. E., & Ree, M. J. (1998). Aptitude test score validity: No moderating effect due to job ability requirement differences. *Educational and Psychological Measurement, 58*, 284–294. https://doi.org/10.1177/0013164498058002011

Kell, H. J., & Lang, J. W. B. (2017). Specific abilities in the workplace: More important than *g*? *Journal of Intelligence, 5*, 13. https://doi.org/10.3390/jintelligence5020013

Krumm, S., Schmidt-Atzert, & Lipnevich, A. (2014). Specific cognitive abilities at work: A brief summary from two perspectives. Journal of Personnel Psychology, 13*,* 117–122. doi: 10. 1027\1866–5888\a000117.

Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: A test of the four-sources model. *The Journal of General Psychology, 120*, 375–405. https://doi.org/10.1080/00221309.1993.9711154

Landy, F. J. (2003). Validity generalization: Then and now. In K. R. Murphy (Ed.), *Validity generalization: A critical review*. Hove, England: Erlbaum.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, Section A, 62*(Part 1), 28–30. https://doi.org/10.1017/S0080454100006385

Levin, J (1972). The occurrence of an increase in correlation by restriction of range. *Psychometrika, 37*, 93–97. https://doi.org/10.1007/BF02291414

Millsap, R. E. (1989). Sampling variance in the correlation coefficient under range restriction: A Monte Carlo study. *Journal of Applied Psychology, 74*, 456–461. https://psycnet.apa.org/record/1989-31731-001.

Nye, C., Chernyshenko, O., Stark, S., Drasgow, F., Philips, H., Phillips, J., & Campbell, J. (2020). More than *g*: Evidence for the incremental validity of performance-based assessments for predicting training performance. *Applied Psychology, 69*, 302–324. https://doi.org/10.1111/apps.12171

Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology, 79*, 845–851. https://doi.org/10.1037/0021-9010.79.6.845

Ree, M. J., & Carretta, T. R. (1994). Factor analysis of ASVAB: Confirming a Vernon-like model. *Educational and Psychological Measurement, 54*, 459–463. https://doi.org/10.1177/0013164494054002020

Ree, M. J., & Carretta, T. R. (2011). The observation of incremental validity does not always mean unique contribution to prediction. *International Journal of Selection and Assessment, 19*, 276–279. https://doi.org/10.1111/j.1468-2389.2011.00556.x

Ree, M. J., Carretta, T. R., & Doub, T. W. (1998/1999). A test of three models of the role of *g* and prior job knowledge in the acquisition of subsequent job knowledge in training. *Training Research Journal, 4*, 135–150.

Ree, M. J., Carretta, T. R., & Earles, J. A. (1998). In top-down decisions, weighting variables does not matter: A consequence of Wilks' theorem. *Organizational Research Methods, 1*, 407–420. https://doi.org/10.1177/109442819814003

Ree, M. J., Carretta, T. R., & Teachout, M. S. (2015). Pervasiveness of dominant general factors in organizational measurement. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8*, 409–427. https://doi.org/10.1017/iop.2015.16

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than *g*. *Personnel Psychology, 44*, 321–332. https://doi.org/10.1111/j.1744-6570.1991.tb00961.x

Ree, M. J., & Earles, J. A. (1994). *Estimating the general cognitive component of the Armed Services Vocational Aptitude Battery (ASVAB): Three faces of g, AFHRL-TR-(0-38)*. Brooks AFB, TX: Air Force Human Resources Laboratory.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology, 79*, 518–524. https://doi.org/10.1037/0021-9010.79.4.518

Richardson, K., & Norgate, S. H. (2015). Does IQ really predict job performance? *Applied Developmental Science, 19*, 153–169. https://doi.org/10.1080/10888691.2014.983635

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112–118. https://doi.org/10.1037/0021-9010.85.1.112

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor situations. *Psychometrika, 22*, 53–61. https://doi.org/10.1007/BF02289209

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. *Journal of Applied Psychology, 66*, 166–185. https://doi.org/10.1037/0021-9010.66.2.166

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. London: Macmillan.

Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-components tests are not much more than *g*: An extension of Kyllonen's analyses. *The Journal of General Psychology, 123*, 193–205. https://doi.org/10.1080/00221309.1996.9921272

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. NY: Wiley. https://psycnet.apa.org/record/1949-05074-000.

Walters, L. C., Miller, M., & Ree, M. J. (1993). Structured interviews for pilot selection: No incremental validity. *International Journal of Aviation Psychology, 3*, 25–38. https://doi.org/10.1207/s15327108ijap0301_2

Wigdor, A., & Green, B. F., Jr. (1987). *Assessing the performance of enlisted personnel; evaluation of a joint-service project*. Washington, DC: National Academy Press.

Wilks, S. S. (1938). Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika, 3*, 23–40. https://doi.org/10.1007/BF02287917