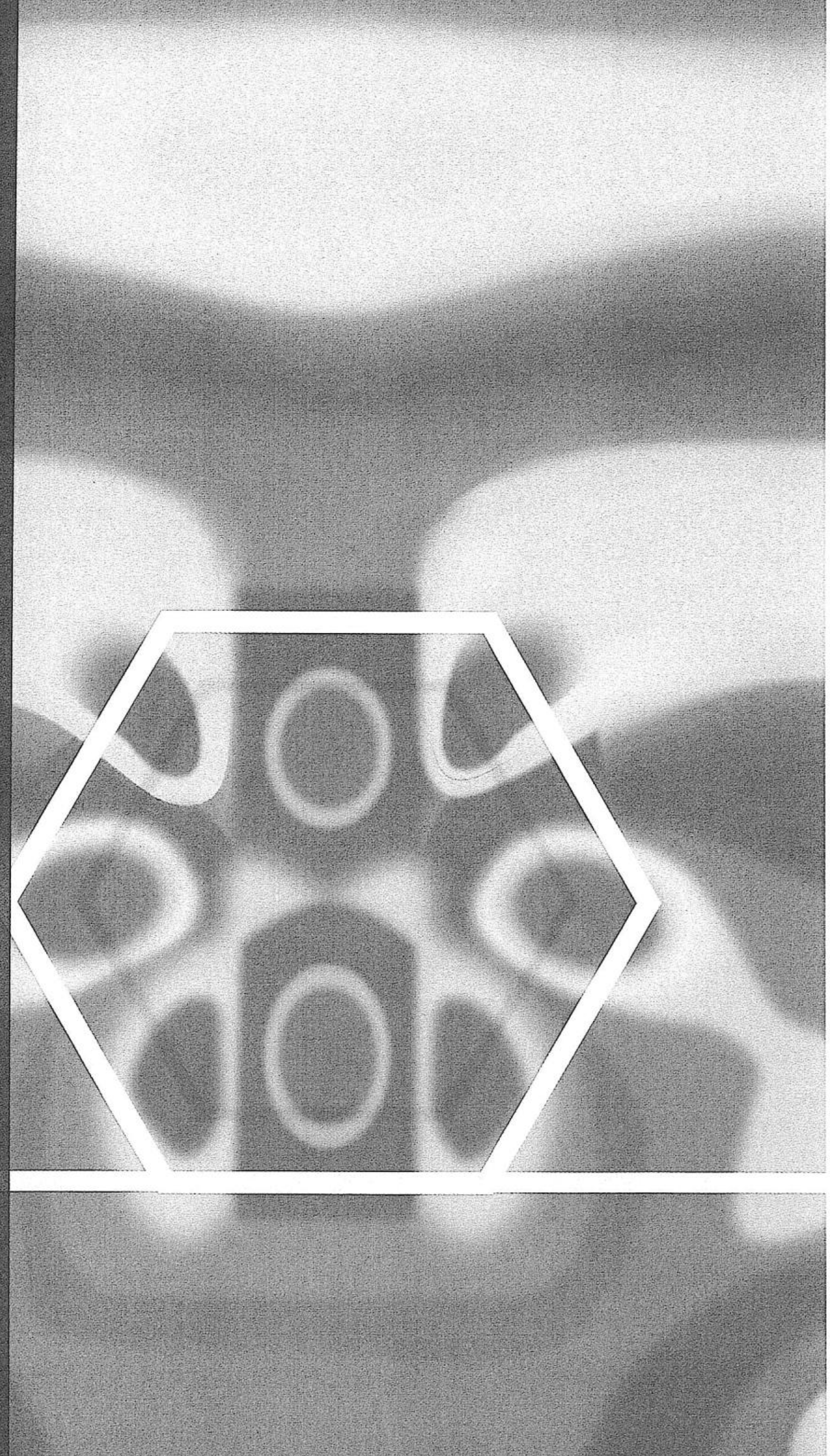
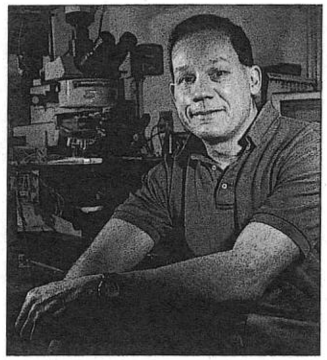
F&M SCIENTIST

WINTER 2013



About the author



Charles M. Lieber graduated with honors in chemistry from Franklin & Marshall College. After doctoral studies at Stanford University and postdoctoral research at the California Institute of Technology, he assumed an assistant professor position at Columbia University in 1987. He moved to Harvard University in 1991 where he is now Mark Hyman Professor of Chemistry in the Department of Chemistry and Chemical Biology. His work has been recognized through numerous awards, including the Wolf Prize in Chemistry (2012), ACS Inorganic Nanoscience Award (2009), NIH Pioneer Award (2009), ACS Award in the Chemistry of Materials (2004), APS McGroddy Prize for New Materials (2003), MRS Medal (2002), and Feynman Prize in Nanotechnology (2001). Lieber is an elected member of the National Academy of Sciences and the American Academy of Arts and Sciences, and an elected Fellow of the Materials Research Society, American Physical Society, American Chemical Society, and American Association for the Advancement of Science. He is co-editor of *Nano Letters*, and serves on the editorial and advisory boards of many science and technology journals. Lieber has published more than 340 papers, which have been cited more than 64,000 times, and is the principal inventor on more than 35 patents. In his spare time, Lieber has founded several nanotechnology companies, coaches high-school wrestling, and grows giant pumpkins.

Effect of F&M education on my career

Professors Fred Snavely and Claude Yoder had a defining role in my career. After originally entering F&M as a premed student, and even applying to and being accepted by medical schools, I realized through independent research opportunities at F&M that my true love was basic research and discovery. Their support and encouragement of students in research (and willingness to look past my many faults) were nothing short of amazing, and have helped to guide my own training of students over the past two decades. Thank you!

Testing the Idea of General Intelligence

John J. McArdle '73

Senior Professor of Psychology
Head, Quantitative Methods Area
University of Southern California, Los Angeles, CA 90089

Abstract

Fundamental questions are raised about the existence of a general intellectual ability, and these are framed as the "*g* hypothesis." The research and views of several prominent scientists are considered and it is concluded that there is still some debate about this issue, especially when comparing classic ideas, such as those of Charles Spearman, to more contemporary ideas, such as those of John L. Horn. To resolve this debate, or at least bring it into focus, Spearman's (1904) classic method of "common factor analysis" is presented as a way to test almost any formal hypothesis. This technique is illustrated using data from the first occasion of eight scales from the Wechsler Intelligence Scale for Children (WISC; N=204; from Osborne & Suddick, 1972). Several alternative common factor models are fitted and evaluated for misfit and by contemporary standards there seems to be more than one common factor present. For these reasons, the theory of "Fluid and Crystallized Intelligence" (g_f/g_c from Horn & Cattell, 1969) is now given more support than the theory of "General Intelligence" (*g* by Spearman, 1904). There is no doubt that most research and practicing psychologists will find this result surprising, but using Spearman's (1904) own methods, we conclude that a *g* factor of intelligence does not exist.

Introduction

Some of the early history of General Intelligence

Sir Francis Galton's work in the late 1800s started much of the research and theory on the assessment of individual differences in intelligence. Galton seemed to believe that all intelligent behavior was related to innate sensory abilities but his attempts to empirically validate that key assumption were largely unsuccessful. A key turning point was around 1904 when Alfred Binet & Théodore Simon were commissioned by the French Minister of Public Instruction to develop a procedure to select children unable to benefit from regular public school instruction for placement in special educational programs because they required special services. As it turns out, this was the beginning (in Western cultures) of a placement system for educational purposes based on measurable test scores.

Binet & Simon (1905) subsequently published what has been termed an objective and standardized intelligence test that consisted of thirty subtests of mental ability, including tests of digit span, object and body part identification, sentence memory, and so on. In 1908 the subtests were organized according to the age level at which they could be successfully performed by most children of normal intelligence. Children were characterized and compared to each other in terms of their intellectual or mental age. In fact, children who were considered to be at lower mental ages were never even administered some of the test items. The Binet and Simon intelligence test was widely adopted in Europe and in the United States, and many of these subtests, with minor modifications, are included in the Stanford-Binet intelligence test (see Terman, 1925). Soon afterwards David Wechsler (1949) created his own series of tests to be used for a wide range of adults (see McArdle et al., 2009) and these were subsequently made for children as well.

In these early days we can also find that Charles A. Spearman from the University of London (1904) presented an influential theory of a unitary intelligence, termed *g-theory*. Less well known, it seems, is that over the next few decades he (Spearman, 1927) also developed a way to estimate and test his *g-theory* as a mathematical and statistical hypothesis. The latter development was a most important scientific discovery because it provided a formal way to reject the single factor hypothesis he himself had created. This was the special contribution of a great scientist, and this is a lesson that we must all learn better. Spearman developed a rudimentary form of the techniques of what is now called *factor analysis* and he applied these techniques to real data (see McArdle, 2007a). It seems now that this concept and this approach are once again embroiled in one of the greatest controversies in all of psychological research, no doubt due to its important racial implications (e.g., Herrnstein & Murray, 1994; Jensen, 1998).

Current sentiments about General Intelligence

My close friend, colleague, and mentor Dr. John L. Horn (University of Denver, and later the University of Southern California; see McArdle, 2007b) when asked about General Intelligence often used to give intentionally provocative answers like, *"What do I conceive invisible green spiders to be? For current knowledge suggests to me that intelligence is not a unitary entity of any kind. Attempts to describe it are bound to be futile"* (Horn, 1986, p. 91). There are probably many others who said such things, but I doubt they said it in exactly this way.

This opening statement by Horn seems rather different than what other equally prominent psychologists seem to say. For example, the major proponent of *g* theory nowadays seems to be the well-known advocate Arthur Jensen (1980, 1998). Among many books and articles on the very same topic, he writes:

"The relationship of the g factor to a number of biological variables and its relationship to the size of the white-black differences on various cognitive tests (i.e. Spearman's hypothesis), suggests that the average white-black difference in g has a biological component. Human races are viewed not as discrete or Platonic categories but, rather, as breeding populations that, as a result of natural selection, have come to differ statistically in the relative frequencies of many polymorphic genes. The genetic distances between various populations form a continuous variable that can be measured in terms of differences in gene frequencies. Racial populations differ in many genetic characteristics, some of which, such as brain size, have behavioral and psychometric correlates, particularly g." (from Jensen, 1998).

That is, Jensen assumes that Spearman's hypothesis was about the nature of *g* not the existence of *g*. While the purported racial differences on *g* seems to be driving the debate, at the heart of the matter is the existence of *g* at all. That is, if *g* did not exist, this debate would certainly be moot. Nowadays, more support for Spearman's first hypothesis can also be found in the work of Dr. Douglas Detterman (Professor at Case Western University and Editor of the Journal *Intelligence*). Dr. Detterman's research suggests that *g* becomes increasingly dominant at the lower extreme of the IQ continuum; that is, correlations among scores on cognitive tests are the highest in individuals with the lowest IQs. This suggests that people with mental retardation have a deficit in something that powers all areas of cognition, rather than discrete deficits in specific information-processing capabilities (Detterman & Daniel, 1989). Once again, Detterman seems to assume the existence of *g*, and does not seem to evaluate his results as a tendency of any distribution of scores, but this leaves us to wonder if this should be considered new evidence of a *g* factor.

In addition, there is also Dr. Ian Deary (prominent Head of Psychology at the University of Edinburgh) who in his recent book "Looking Down On Human Intelligence" (2000) points out that some people are more mentally able than others, and examines the reasons why this is true. While there are no real differences of opinion expressed, the author inquires about the cognitive and biological foundations of human mental ability differences and discusses important ideas from antiquity through the renaissance and enlightenment to the late 19th century beginnings of scientific psychology. The main body of his work charts the progress of modern research down through the main reductionist efforts to account for variance in human intelligence. Concepts from the fields of psychometrics, cognitive-experimental psychology, psychophysics, and biology that correlate with psychometric intelligence differences are described and discussed. It is noted that such correlations are numerous, often replicated, though usually modest in effect size. Once again, this work leaves much to be desired in terms of the existence of *g*.

We can also find Dr. David Lubinski (Professor of Psychology at Vanderbilt University, and Head of Study of Mathematically Precocious Youth) saying: *"The study of individual differences in cognitive abilities is one of the few branches of psychological science to amass a coherent body of empirical knowledge withstanding the test of time. There is wide consensus that cognitive abilities are organized hierarchically, and C. Spearman's (1904) general intelligence occupies the vertex of this hierarchy. In addition, specific abilities beyond general intelligence refine longitudinal forecasts of important social phenomena and paint a rich portrait of this important domain of psychological diversity..."* (Lubinski, 2004, p. 96). This specific article identifies and then reviews five major areas concerning the significance of cognitive abilities and the methods used to study them. Lubinski suggests that, in models of human behavior and important life outcomes, cognitive abilities are critical in more ways than most social scientists tend to realize. Because multiple abilities are considered in the hierarchy, there is no doubt actually raised here that *g* is comprised of more than one concept.

As a final example, we can point to Dr. Linda Gottfredson (Professor at the University of Delaware) stating *"The g factor is a universal and reliably measured distinction among humans in their ability to learn, reason, and solve problems."* (Gottfredson, 2004). Dr. Gottfredson is also a leader in the public policy debate about this topic (see "Mainstream Science on Intelligence," *Wall Street Journal*, 12/13/94). The signers of the Mainstream document were/are indeed prominent in psychology, but very little effort was made to tell us who was asked to sign it and did not sign it, and it is doubtful that some of the signers had completely understood the gravity of their message (see Hauser, 2010).

But the list of notable researchers goes on and on, including many other areas of research (e.g., see Schmidt & Hunter, 1999, on *g* and job performance). What I think this should be taken to mean is that there seems to be broad agreement that a construct like *g* does exist and it is a useful construct to have around and use in psychological research.

Additional scientific evidence

Other important work was carried out on the topic of a general factor of intelligence. In my own department at USC, for example, J.P. Guilford (1956) proposed the first version of the "Structure Of Intellect" (SOI) model, including a Content dimension, Products dimension, and Operations dimension. SOI was most easily represented as a cube with each of the three dimensions occupying one side. Each mental ability was then defined by a conjunction of the three categories, occupying one cell in the three-dimensional figure. In the SOI, there were five categories of Content (i.e., visual, auditory, symbolic, semantic, and behavioral), six categories of Products (i.e., units, classes, relations, systems, transformation, and implications), and five kinds of Operations (i.e., cognition, memory, divergent production, convergent production, and evaluation), and this led to at least 150 measures of abilities. The SOI theory was allegedly an open system that allowed for newly discovered categories to be added in any of three directions. Although orthogonal factor models were always used to understand the data from many collated lab experiments, the abilities were believed to be correlated with each other, and the SOI model suggested where new abilities may be discovered. It follows that in SOI theory and practice, intelligence was thought to be incredibly complex. This is not a single *g* theory. If the SOI was to be believed, then no longer was intelligence a single innate aspect of human functioning.

A set of debates, published in *Psychological Bulletin*, pitted a young psychologist named John L. Horn (then at the University of Denver) up against the famous USC psychometrist, J. P. Guilford. The seeds of this debate can be found in Horn's classic works "A Rationale and Test for the Number of Factors in Factor Analysis" (1965) and "Subjectivity in Factor Analysis" (1967). But the debate became more focused with Horn & Knapp's (1974) complete refutation of the factorial evidence of Guilford's popular SOI model of cognitive abilities. Using statistical simulations they criticized the evidence in favor of SOI theory by showing how easy it is to use "Procrustean" rotation to obtain virtually any factorial solution desired, especially one with as many dimensions as SOI theory (i.e., 120). The response by Guilford (1974, 1980) was heated, to say the least, but it is no coincidence that the popular SOI model soon lost its elevated status in psychometric practice. So *g* theory lived on.

In more recent work, the specific topic of hierarchies of cognitive abilities has been discussed by many others, with continued agreement. We now have the seemingly new assertions of Carroll (1998): *"The three-stratum theory of cognitive abilities...proposes that there are a fairly large number of distinct individual differences in cognitive ability, and that the relationships among them can be derived by classifying them into three different strata: stratum I, 'narrow' abilities; stratum II, 'broad abilities; and stratum III, consisting of a single 'general' ability (Carroll, 1998, p. 122)."*

In terms of more recent methods with policy implications, Wendy Johnson et al. (2004) in a recent analysis provocatively titled "Just One *g*: Recent Results From Three Test Batteries" start by suggesting reasons why the concept of a general intelligence factor or *g* is controversial in psychology. They point out that one of the most important issues involves *g*'s identification and measurement in a group of individuals. That is, if *g* is actually predictive of a range of intellectual performances, the factor identified in one battery of mental ability tests should be closely related to that identified in another dissimilar aggregation of abilities. This seems quite reasonable as a method, but it really seems to require an alternative model. In any case, these authors addressed the extent to which this prediction was true using three mental ability batteries administered to a sample of (n=463) adults. Though the particular tasks used in the batteries reflected varying conceptions of the range of human intellectual performance, the *g* factors identified by the batteries were completely correlated (with correlations approaching or at unity). They suggested that this result provides further evidence for the existence of a higher-level *g* factor and suggests that its measurement is not dependent on the use of specific mental ability tasks.

Is g-Theory a question to be answered?

So maybe the outrageous sentiment expressed by Horn (1986) is simply isolated. After all, there do seem to be a lot of people who favor *g*. But I think not, not really, and more evidence should be considered. In fact, there are many other credible scientists who have tried to consider this specific topic in detail. Among the first was the well-known scientist Sir Cyril Burt (1909, see 1949), a colleague of Spearman, who, although recently discredited for falsifying genetic estimates, used Spearman's own data to detect more isolated "group" factors than found before. This seems to have been the beginning of the idea of "multiple common factors." Some time afterwards, the important work of Thurstone (1938, 1947) made many important contributions to multiple factor analysis. Thurstone, for example, argued that multiple primary factors were needed to represent discrete intellectual abilities, and he developed distinct tests to measure them. Among the most important of Thurstone's (1938) "primary mental abilities" are verbal comprehension, word fluency, numerical ability, spatial relations, memory, reasoning, and perceptual speed. He suggested these could be found using a "simple structure" form of factor rotation (see Thurstone, 1947). But he also was among the first to point out that correlations among multiple common factors could themselves be factor analyzed, and this could lead to a new concept of higher order *g*. This seeming compromise offered by a hierarchical *g* position helped lay the groundwork for future researchers who proposed hierarchical theories and theories of multiple intelligences (Ruzgis, 1994). Of course, this influential work on higher order factors clearly predated Carroll's later statement by about fifty years.

It is also useful to point out that a graduate student of Spearman's, R.B. Cattell (1941, 1998), had already discovered sixteen factors underlying human personality using the then most advanced and available techniques of factor analysis. In his effort to bridge the gap between Spearman and Thurstone's view, Cattell collected new cognitive data. In Cattell's analysis of this new cognitive data, he thought he had found the existence of two general kinds of intelligence, fluid and crystallized intelligence (termed g_f/g_c to honor Spearman; see Cattell, 1943, 1978, 1989). Cattell argued that the data suggested that g_f factor represents an individual's basic biological capacity whereas the g_c represented the types of abilities required for most school activities. This is provocative, and basically asserts that g_f is equivalent to g, but was presented without further evidence. This was broadly taken as evidence suggesting the utility of a g_f factor of mental reasoning or thinking abilities and a separation of g_c mental facilities related to knowledge acquisition from the dominant culture. This seemed a long way from the basic tests of Binet and Simon. Cattell also labeled three minor general factors as visual abilities (g_v), memory retrieval (g_r), and performance speed (g_s). It is probably important to point out that subsequently John L. Horn was a graduate student of R.B. Cattell, and a major developer of this g_f/g_c theory with adults (see Horn & Cattell, 1966a, 1966b, 1967, 1998). Perhaps this partly explains why Horn did not approve of Spearman's general factor concept?

Richard E. Snow, (from Stanford University) himself a prominent educational psychologist who worked with Carroll, but who did not work with Horn or Cattell, dedicated much of his own work toward studying human aptitudes and learning environments. He aligned himself with g_f/g_c theory in many ways, but he *"stressed the importance of looking at individual differences in cognitive processing and analyzing these processes in relation to variations in environmental affordances to develop a person-situated interaction theory of intellect"* (see Snow, 1998).

But, obviously, Horn's statement that g is a fictional concept is not believed by everyone in the know nowadays, and certainly not by most psychologists either. I now think it is in our best interests to get to the bottom of this argument. In most general terms, we can ask "What evidence do we require for the existence of any important construct in behavioral science?" But specifically, we want to examine g as a theory. So what are we as new to this area expected to believe? I, probably just like you, find this whole debate somewhat surprising and would like to see this matter settled in a fair and honest way. The purpose of this article is to do just this. We will examine this basic idea of g-theory using contemporary statistical models. Unfortunately, the story also tells us something about the rather wide-ranging human uses of modern statistical tools. It seems that many people only like and trust the techniques that tell them what they already know or want to hear.

Factor analytic techniques

Factor analytic techniques have been developed and used by psychologists for most of the 20th century, and these methods have become especially useful in studies of individual differences. The key concept in any factor analysis is that a large number of observed behaviors are a direct result of a smaller number of unobserved or latent sources termed *factors*. This theoretical principle was used by Spearman (1904) in his early studies of the concept of general intelligence, and it has been applied in a wide variety of empirical studies to isolate and identify latent factors representing parsimonious and reliable sources of differences between individuals and groups. In this sense, common factors in Psychology share much with ideas from other areas of science: e.g., the quarks and atoms of Physics, the molecules and elements of Chemistry, the genes and viruses of Biology, and the unobserved planets and black holes of Astronomy. These key scientific concepts are not directly observed, but we imply their existence from a large set of observed measurements.

To describe this general factor theory in the first paper, Spearman first created a technique for testing hypotheses about latent variables using only measured variables. He applied the first factor analysis model to a matrix of correlations from data collected on the school grades and mental tests of children and, indeed, he suggested that the single factor model fit well. Over the next 20 years, Spearman posed more technical problems and invented many solutions for these difficult psychometric issues. He wrote a basic algebraic model for multiple observed variables based on a single common latent variable, he considered different ways that the parameters in this model could be uniquely calculated, and he considered ways to evaluate the statistical goodness-of-fit of this model to empirical data (see Spearman, 1927).

In order to raise this general factor concept to a high level of a scientifically respectable theory, Spearman created a strong and rejectable method for the understanding of individual differences. In my own view, most of the research that has followed Spearman's work on this topic has soundly rejected the simple idea of a single common source of intellectual differences among people, although I recognize others have different views on this matter (see Methods section). More importantly, in this early work on factor analysis, Spearman laid the basic foundation for all modern work in *structural equation modeling* (SEM).

This paper uses factor analysis as a generic term for explaining the processes of describing, testing hypotheses, and making scientific inferences about unobserved variables by examining the internal structure of multiple variables (as in McArdle, 1994; McArdle & Nessleroade, 1994). This is obviously the technique we will use to evaluate the Spearman hypothesis. We start with a description of the data we will use, and we then describe several basic factor models for cross-sectional data, The analyses presented here allow us to address other contemporary issues, including model fitting and factor rotation, and future directions.

Method

Participants

The set of real data we use here have been presented in several other publications on developmental data analysis (e.g, McArdle & Epstein, 1987; McArdle & Nessleroade, 1994, 2012). We have been working with a selected set of longitudinal WISC (Wechsler Intelligence Scale for Children) data for about thirty-five years now, and we have shared these data with many other researchers (see our public website). The original sample of participants for a longitudinal study of the WISC data were described as: *"The original group consisted of 163 white and 110 Negro preschool children selected from three counties in Georgia, representative of small rural and medium and large industrial urban populations. Of the original base group of 273 children who were tested in the spring before they entered the first grade, 204 were retested at the end of the first grade, at the end of the third grade, and also upon completion of the fifth grade. The children who were not retested had moved out of the state or were no longer in public schools." Children who transferred to other school systems within the state were located and reexamined. At the time of the initial testing. Spring 1961, the mean WISC full scale IQ was 96.16; SD, 15.6. The age range of the group was narrow; mean six years one month; SD two months. One hundred eighteen children were Caucasian, 86 were Negro; there were 95 boys and 109 girls. [Osborne & Suddick, 1972, p. 84]*

We have already tried out many different kinds of analyses with these longitudinal data, including the beginnings of latent curve analysis (McArdle & Epstein, 1987), multivariate latent curve analysis (McArdle, 1988), multiple group factor analysis of latent curves (McArdle, 1989), auto-regression versus latent curves (McArdle & Aber, 1990), MANOVA (multivariate analysis of variance) and the beginnings of latent change score analysis (McArdle & Nesselroade, 1994), and multivariate latent change score analysis (McArdle, 2001). Perhaps the best or most proper way to analyze these data has yet to be found, but these data have certainly been subjected to many forms of data analysis (see references listed above). The factor analyses presented now are new and have not been published elsewhere.

Measures in the WISC

The variables in any study of the *Wechsler Intelligence Scale for Children* (WISC; Wechsler, 1949) can be numerous (see Osborne & Suddick, 1972), and much of the available data were never used. But here we select some of these data—we use only the first grade data on eight of the WISC sub-scales, each with its own maximum score. The selected WISC tasks are:

1. *Information* (IN, 30 points possible) — recall a series of facts that are known to be correct.
2. *Comprehension* (CO, 28 max.) — understand what to do correctly in various situations.
4. *Similarities* (SI, 28 max.) — suggest the best reasons two things are alike.
4. *Vocabulary* (VO, 80 max.) — provide proper definitions of specific words
5. *Picture Arrangement* (PA, 20 max.) — arrange a series of pictures so they tell a coherent story.
6. *Picture Completion* (PC, 55 max.) — isolate the missing element in a single picture.
7. *Block Design* (BD, 57 max.) — reorganize a series of physical blocks into a pattern that matches a printed picture (originally from the set of Blocks by Koh, 1923).

8. *Object Assembly* (OA, 34 max.) — reorganize cutouts into a pattern that makes a picture.

Although we realize there are many alternative ways to weight and transform these scores, we will rely on the simplest method of "proportion correct." This is used simply to avoid confusion about the different number of points actually given to each scale. That is, each raw sub-scale score is first divided by its maximum possible scores (ranging from 20 to 80), and then multiplied by 100, so each new raw scale and each weighted composite can be interpreted as the proportion correct (between 0 and 100) for that scale. This simple POM (percentage of maximum) transformation does not alter the distribution of any score, nor does it alter the correlation with other any scores, but it provides a simple way to make an interpretation in POM terms. (Yes, we agree this is an easy transformation for any data; see McArdle & Epstein, 1987).

The correlations of all eight WISC scale scores at the first occasion of testing are presented as a scatter plot matrix in **Figure 1** (created with the R-psych program *pairs.panels*). This kind of plot shows the generally symmetric and normal distribution of all scale scores, and the generally positive relationships among all sub-scales. **Table 1** gives all the summary statistics on these proportion correct transformed scores at the first occasion of measurement

Figure 1: Time 1 (Grade 1, Age 6) Summary Statistics for 8 WISC subscales

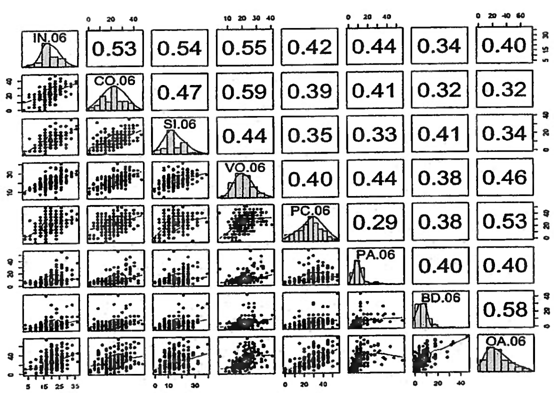


Table 1: Summary Statistics for Eight Measures from the Wechsler Intelligence Scale for Children (WISC) from Time 1 (Age 6)

	WISC Measure							
	Information IN	Comprehension CO	Similarities SI	Vocabulary VO	Picture Completion PC	Picture Arrangement PA	Block Design BD	Object Assembly OA
Means								
Raw	6.05	6.20	4.28	16.58	5.72	5.66	4.13	8.80
POM	20.16	22.15	15.30	20.73	28.58	10.29	7.24	25.89
SD								
Raw	1.75	2.67	2.06	4.96	2.40	4.50	3.66	5.51
POM	5.83	9.55	7.37	6.20	12.14	8.18	6.43	16.21
Correlations								
IN	1.000							
CO	0.530	1.000						
SI	0.539	0.475	1.000					
VO	0.549	0.589	0.440	1.000				
PC	0.421	0.393	0.351	0.405	1.000			
PA	0.438	0.408	0.338	0.444	0.289	1.000		
BD	0.343	0.323	0.408	0.380	0.385	0.404	1.000	
OA	0.403	0.322	0.338	0.462	0.534	0.399	0.583	1.000
Reliabilities	0.66	0.59	0.66	0.77	0.59	0.72	0.84	0.63

Note: N=204. Values presented as raw scores (items correct) and as a percentage of the maximum score (POM). Data from Osborne & Sudick, 1972. See McArdle & Epstein 1987, McArdle & Aber 1990, McArdle & Nesselroade 1994 for details of scaling and analyses. Internal consistency reliability values from Wechsler, 1949.

(Grade 1, Age 6). Of most importance, these eight sub-scales of the WISC were selected to represent one or two separate constructs (see McArdle & Prescott, 1992).

In several earlier reports we have compared our analyses over a single grouping variable, "Mother's Level of Education," before the children entered the first grade. For convenience, Mother's Education was initially coded into three groups (termed MOED)—(0, n=76) Mothers with little or no High School Education, (1, n=82) Mothers who were High School Graduates, and (2, n=46) Mothers with Some Post-High School Education. Of course this grouping and the observational variable it represents was not intended to be of the same inferential status as a full randomization to group, because we did not randomly assign anyone to these groups (see McArdle & Epstein, 1987).

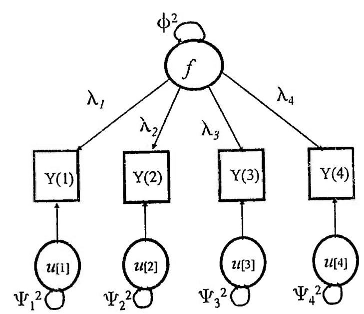
Nevertheless, when we put everyone together—that is, when we ignore what group they are in—we are basically assuming the same model applies to each group. If we find any evidence that one big grouping may not apply, we will study this specific grouping question in various ways.

The statistical basis of common factor models

This section is intended to be technical but it is presented without equations on purpose. Here we provide some details on the statistical model of factor analysis. This is included mainly to establish factor analysis as a statistical method, comparable to the analysis of variance (e.g., ANOVA), among many other well-known techniques. But the non-technical reader may wish to skip this section (and go directly to the RESULTS) without too much loss of context, and then return to this section when deemed useful (or even to try McDonald, 1985).

Figure 2 is a SEM path diagram representing a single common factor as a latent variable (LV). The common factor score (the circle termed *F*) is not directly measured. But this common factor is thought to have its own variability (ϕ^2) and to produce the variation in each of the 4 observed variables (Y_m, m=1 to 4) through the common factor loadings (λ_m). The common factor loadings have all the properties of traditional regression coefficients, except here the predictor is unobserved (McArdle, 1991). Each of the measured variables is also thought to have a unique variance (φ_m^2) that is assumed to be uncorrelated with other unique factors and with the common factor as

Figure 2: A path diagram of a one factor model



well. So, although we have many measured variables (Y_m), we only have one unobserved common factor score (*F*). As in traditional regression analysis we presume there is a score for each person and that is multiplied by a group weight (λ_m). So the reason we observe variation in any outcome score is partly due to its common factor component ($\lambda_m F_n$) and partly due to the uncorrelated unique component ($u_{m,n}$).

This set of assumptions imply that the expectation we have for the covariance terms among measures includes only common factor variance while the expected variance terms includes both common and specific variance. That is, if this single factor model is true, then each variance and covariance must have a very simple structure.

The expected parameters of this factor model all include the factor variance term (ϕ^2), so at least one additional restriction will be needed to make the parameters "uniquely identified." This is typically done by either restricting the unknown factor variance (ϕ^2 = a positive value, such as 1) or by restricting one of the factor loadings (ϕ_1 = a positive value, such as 1). The specific choice of the identification "constraint" (also referred to as "setting the metric") should not really alter the estimation, fit, or interpretation of the result. That is, because the predictor variable is unobserved the only parameters that are really identified are the ratios of the loadings (e.g., λ_j / λ_k) and these need to remain the same under any scaling of the metric of the unobserved variable. But, of course, this is all true if the model is largely correct and not if it is largely incorrect. So, after adding the appropriate identification constraint ($\phi^2 = 1$), the new covariance expectations are now more "restricted" and can be written in the much simpler form. As it turns out, this provides exactly enough information for us to estimate all factor loadings, and this is used to tell which measured variables are most closely aligned with the unobserved common factor and which are not—i.e., the ratio of the loadings are also used to provide a label for the unobserved common factor. This also means the covariance expressions can be examined for fit by comparison to real data (using SEM software), and when this model does not seem to fit, we usually assume the score model of observable variables is not based on a single common factor.

We typically examine whether or not the one-factor model fits the data by comparing the observed variances and covariance (**S**) to the model expected variances and covariance (Σ). The index of fit is summarized as a single number, a likelihood (L_1^2), that can be compared to the fit of other alternatives (L_2^2). Under the assumption of normally distributed residuals, this difference can be compared to a chi-square distribution (χ^2) with degrees-of-freedom (*df*) based on the differences in the number of model parameters. In this way, the likelihood difference test is an index of "statistical misfit" (see Lawley & Maxwell, 1971). There are several other variations of this basic idea of model misfit (see Browne & Cudeck, 1993, and their RMSEA or ε_a).

Incidentally, our choice of using M=4 indicators per common factor should not be considered arbitrary—this is the smallest number of indicators that offers positive degrees-of-freedom for model evaluation (i.e., df=1). Of course with only three indicators we could look at the resulting factor loadings to make sure they are all large and positive. But this is not often done. Instead, many researchers seem to treat this "extraction" as if it were based on the same technical procedures used to form the indicators. Unfortunately, while it is possible to fit a one-factor model with only three indicators, this model always fits the observed data perfectly (e.g., because there are as many unknowns as knowns), so we have no way to tell if it is incorrect. However, a perfect fit is generally not possible with four indicators, so here we do have a way to examine the model fit.

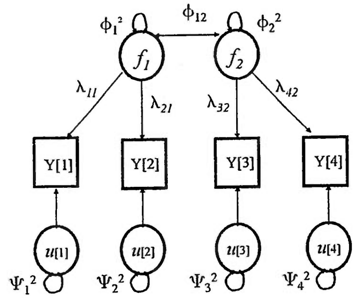
One additional note may be useful—the "principal components" (PC) of any set of variables is a weighted linear combination of the observed scores. We can always calculate the optimal PC by asking for a first weighted linear combination that has the largest variance. The second PC can be extracted as an orthogonal projection by calculating the linear combination with the largest variance of what remains after the first has been removed, and so on. In this sense the PCs are not latent variables so weighted linear composite scores can be created for them. In general, the PC approach seems very much like the common factor approach, and this is the good news. The bad news is that it is not really possible to test the concept of a PC as a formal hypothesis and the PC loadings seem to be biased upward (i.e, the values are too large) when we have a small number of indicators per construct (see McArdle, 1991). Thus, we can calculate as many PCs as we like, but we do not have any standard statistical tests of their utility or their potential bias.

If the factor model expectations seems to match the observed data, there are many ways to restrict this model even further. As a first very general alternative, we could simply say there is no common factor at all, so all covariances and correlations are collectively zero. This is a typical "baseline" or "null" model that, without any doubt, we really must be able to say does not fit our data. This simpler model could be examined for goodness-of-fit. Of course we know that if this model fits our data we are effectively finished. That is, we really must have enough statistical power to reject this no-common factor hypothesis. If the null baseline model does fit we might as well stop here because this forms a very simple set of relationships. In another simpler alternative, we could say that all the factor loadings are equal (as done in a Rasch-type model; see McDonald, 1999; McArdle, Grimm, Bowles, Hamagami & Meredith, 2009), and this simpler model could be examined for goodness-of-fit. If this model fits the observed data we can say the common factor has a lot of simple properties (e.g., the factor scores can be well estimated by simply summing up the items). In any case, both the null factors baselines and the equal loading model are simpler and testable alternatives to the one common factor concept.

If the simple one common factor model does not seem to provide a good fit, we can go the other direction and relax some of the model constraints. For example, we can posit the existence of a second common factor (F_2), and a simple version of this model is drawn as **Figure 3**. In this simple model we posit that each common factor is related to a specific set of observed variables. Perhaps it is not clear that this new two-factor model is decidedly more complicated than the single factor model. But the only difference between this two-factor model and the one-factor model is the covariance among the common factors (ϕ_{12}). That is, if this value turns out to be the same as the variance terms (or the factor inter-correlation ρ_{12}=1), then this model reduces to become the one factor model. This holds true even when we have more variables measured, as in **Figure 4**. This subtle difference could be important in the model fit where there is now one degree of freedom difference between these models, testing the covariance hypothesis, and we can evaluate the gain in fit (or the loss of misfit). We illustrate this point with the real example later.

The factor analysis approach is more rigid than most people think, but it is not without its own problems. For example, it appears that simple kinds of SEMs are the only ones that can be fitted, and this is not true. We can start with the simple two-factor model of **Figure 4** and add one additional factor

Figure 3: A path diagram of a restricted two common factor model



loading where the first factor is indicated by variable 4, and the second factor is not. This is drawn as a path diagram in **Figure 5**, and it could be a correct representation of the data, even though variable 4 now loads on factor 1 and not factor 2. We would fit both models as alternatives, but we would quickly see they cannot be compared using our standard logic because they have the same number of parameters and are not proper subsets of one another.

In a different example we can start with the simple two-factor model (of **Figure 5**) and add two additional factor loadings where common factor 1 is indicated by variables 4 and 5, and two additional factor loadings

Figure 4: A two common factor model (with id constraints)

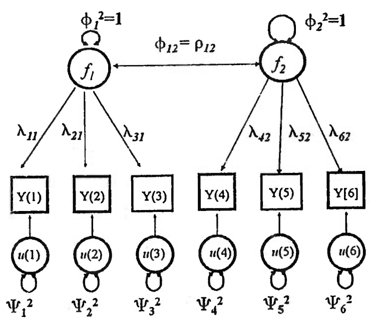
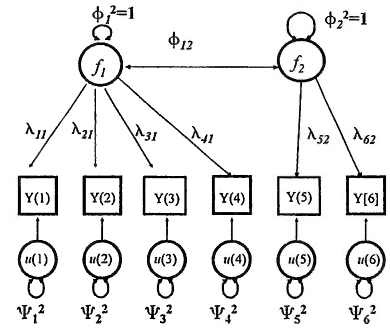


Figure 5: A "non-nested" two common factors model with unbalanced constraints



where common factor 2 is indicated by variables 2 and 3. These additional parameters reduce the number of degrees of freedom (by 4) but the overall two-factor model is still identified, and unique estimates can be found for all parameters. We know if we fit this kind of a model we could trim out the non-salient loadings and, hopefully, have a reasonable fitting model. So here our data analysis strategy is put to an important test. Do we use our *a priori* logic and fit the hypothesized model (i.e., **Figure 4**)? One advantage to this *a priori* approach is that we could use all listed probability calculations. Or do we fit the exactly identified solution to help us understand our data and gain a good fitting model? Since we typically do not define everything we could do here, we cannot correctly use all SEM probability calculations. Of course, in practice, we often do both anyway! This specific issue is discussed in many other presentations (e.g., McArdle & Cattell, 1994). Because the overall two-factor model can be represented with different diagrams (i.e., "rotated" into a different position) without any change in fit, this means that we have possibly different parameter values and factorial interpretations.

In general, it is well-known that a comparison of alternative factor models cannot be judged by goodness-of-fit alone (see Kaiser, 1976; McArdle & Cattell, 1994). Perhaps more importantly, this also illustrates that the highly restricted two-factor model (of **Figures 3, 4, 5**) is a specific factorial hypothesis that can have a "unique solution" and "cannot be rotated" any further. This is the essential benefit of what is usually termed "confirmatory factor analysis" (see Tucker & Lewis, 1973; Jöreskog, 1971; Lawley & Maxwell, 1963; McDonald, 1985). Just as a clarifying note, in our view, if the factor loading values were all specified in advance, then this would really be a "confirmatory" approach to model evaluation. This would mean the entire set of covariances was expected rather than just the model pattern. But, of course, this very rigid form of confirmatory factor analysis is hardly ever used.

In the typical unrestricted factor model, we initially estimate a K-factor orthogonal solution using some "convenient restrictions" (e.g., Lawley & Maxwell, 1963; Jöreskog, 1971, p.23). In this model the parameters are "exactly-identified" and require further rotation for interpretation. Exactly the same solution can also be obtained from a restricted SEM when we place the minimal constraints necessary for a unique solution. The required constraints have been previously described by Jöreskog (1971), and are repeated in McArdle & Cattell (1994), to achieve an *exactly identified* rotation with same index of fit.

These generic statements mean that even more factor loadings can be estimated than we typically think. Most specifically, if we want to estimate 2 common factors from 8 observed variables we require only 4 constraints, and this leaves a combination of 12 free elements in the Λ and Φ matrices. One common way to do this in SEM would be to define a fixed unity for each factor (e.g., $\lambda(m,k)=1$) for scaling purposes, and this largely defines what the factor "is." To this we add a fixed zero (e.g., $\lambda(m,k)=0$) in a different location for each factor, to separate the first and second factor, and to designate what the factor "is not." This approach allows the estimation of the six other loadings for each factor as well as the correlation between the factors. This is termed an "exactly identified" two-factor solution which means it can be rotated. That is, many alternative solutions (i.e. with K^2 constraints) would yield the same goodness-of-fit (for details, see McArdle & Cattell, 1994).

The range of possibilities for factorial structures for multiple variables is so vast that it is rare to consider "all" possible alternatives. Instead we consider the goodness-of-fit of the models to be an indicator of whether a specific model does not fit the data, and we never explicitly know if any specific model is the best one to fit. As with most models, using the classical arguments of Popper (1970), "we can reject models with data, but we can never know we have the best model for any set of data." Another way to say this is that we can use the data to tell us which models are "false," but we cannot use the data to tell us which models are "true."

One of the classical problems in factor analysis with multiple factors is the definition of the position of the factors with respect to the variables. In one of the most important developments in factor analysis, Thurstone (1947) proposed an important meta-theoretical solution to the optimal choice of these constraints: *"One of the turning-points in the solution of the multiple factor problem is the concept of 'simple structure.' It will be shown that this concept enables us to obtain an invariance of factorial description that has not, so far, been available by other means. . . . " "When a factor matrix reveals one or more zeros in each row, we can infer that each of the tests does not involve all the common factors that are required to account for the intercorrelations of the battery as a whole. This is the principle characteristic of a simple structure." (1947, p.181). "The factorial description of a test must remain invariant when the test is moved from one battery to another which involves the same common factors. . . . The factorial composition of a set of primary factors that have been found in a complete and over-determined simple structure remains invariant when the test is moved to another battery involving the same common factors and in which there are enough tests to make the simple structure complete and over-determined."* (Thurstone, 1947, p.365).

This concept of simple structure as a goal of exploratory factor rotation has been very influential in factor analysis because it suggested both rules for selecting an optimal test battery (i.e., selecting the edges of the conic structure) and a clear set of scientific and technical restrictions on the resulting loadings and the transformations needed (e.g, as in Promax, see Browne, 2001). As we have seen earlier, in contemporary SEM we can impose more than the required constraints (K^2) and obtain an *over-identified* and, therefore, un-rotatable solution, but the result model Σ might not fit the observed S as well. In SEM the choice of these constraints is used to test the critical hypotheses, and the cost due to these restrictions is judged in the same way as described before (i.e., using L^2, df, and ε_a). Often these hypotheses are based on simple factor patterns, especially those where each measured variable loads on one and only one factor. Although this simple pattern is not a necessary requirement for a good SEM, most current SEM applications try to achieve this level of "very simple structure" (after Revelle, 1983). This procedure probably should be questioned more rigorously.

Results

Overall WISC results

With these possibilities in mind, we now begin a first measurement analysis by taking the summary statistics for the eight measured WISC variables at the first occasion (Time 1, Age 6, Grade 1, presented earlier in **Figure 1** and **Table 1**) and look for a reasonable number of common factors. We should point out that the eigenvalues of this Time 1 (or Grade 1, about age 6) correlation matrix can be easily calculated ([Λ[1]=[3.98, 0.96, 0.72, 0.65, 0.52, 0.46, 0.38], and from this we can see that it is unlikely that we will ever find more than two common factors (due to the smallish size of the third component). In any case, **Table 2** presents a simple example of a contemporary SEM-based factor analysis of the initial data from the WISC study. Here six alternative SEMs are presented with the eight WISC variables listed in the rows, and the six models across the columns. **Table 3** is a summary of the goodness-of-fit of these models.

The first model fit (*M0*) assumes no common factors are apparent, with an expected value for every correlation as zero. This model, like all those that follow in this section, does not make any restriction about the variable variances (σ_m^2) or variable means (μ_m), so they can be estimated in any way needed. Typically, these parameters will be estimated at their sample values but this is not required. Not surprisingly, with N=204 and the current summary statistics, this simple zero-factors model does not fit the data very well (χ^2=598. on df=28, ε_a=0.32). This model fit is poor because the chi-square is relatively large compared to the degrees of freedom. What we are usually looking for is a model where the chi-square is less than two times the size of the degrees-of-freedom and the model parameters all make good sense. But his model is not intended to fit because it proposes all correlations in **Table 1** and **Figure 1** are collectively zero. It is mainly used as a baseline against which we can judge the fits of other alternatives.

The next model fit (*M1*) is a one common factor model with equal loadings (see Rasch, 1960; McDonald, 1999). The common factor and its loadings are identified by requiring the common factor variance to be fixed (at ϕ^2=1) and

Table 2: Results for Alternative Factor Models of WISC Data from First Occasion

	Model								
	(M0) Zero Common Factors	(M1) One-Factor Rasch Model	(M2) One Common Factor	(M3) Two Common Factors– Verbal & Performance		(M4) Two Common Factors– Relaxed		(M5) Two Common Factors– Restricted	
Group Effects									
λ(IN)	=0	4.77 (17) [0.78]	4.21 (11) [0.72]	4.36 (12) [0.75]	=0	4.29 (11) [0.74]	=0	4.39 (12) [0.75]	=0
λ(CO)	=0	4.77 (*) [0.55]	6.55 (10) [0.69]	6.89 (11) [0.72]	=0	8.57 (7) [0.90]	-2.03 (1.7) [-0.21]	6.94 (11) [0.73]	=0
λ(SI)	=0	4.77 (*) [0.65]	4.69 (10) [0.64]	4.81 (10) [0.66]	=0	4.63 (6) [0.63]	0.10 (0.1) [0.01]	4.76 (10) [0.65]	=0
λ(VO)	=0	4.77 (*) [0.75]	4.57 (12) [0.74]	4.69 (12) [0.78]	=0	4.68 (12) [0.76]	=0	4.69 (12) [0.76]	=0
λ(PC)	=0	4.77 (*) [0.42]	7.24 (9) [0.60]	=0	7.75 (9) [0.64]	3.17 (2.6) [0.26]	5.09 (4) [0.42]	2.91 (2.4) [0.24]	5.42 (4) [0.45]
λ(PA)	=0	4.77 (*) [0.59]	4.82 (9) [0.59]	=0	4.68 (8) [0.57]	3.58 (4) [0.44]	1.52 (1.7) [0.19]	3.43 (4) [0.42]	1.75 (2.1) [0.22]
λ(BD)	=0	4.77 (*) [0.67]	3.79 (9) [0.59]	=0	4.42 (10) [0.69]	=0	4.48 (10) [0.67]	=0	4.49 (10) [0.70]
λ(OA)	=0	4.77 (*) [0.32]	10.37 (9) [0.64]	=0	12.48 (12) [0.77]	=0	13.62 (12) [0.84]	=0	13.55 (12) [0.84]
Individual Effects									
$\varphi\{g\}^2$	--	=1.0	=1.0	=1.0	=1.0	=1.0	=1.0	=1.0	=1.0
$\varphi\{V,P\}$	--	--	--	0.772 (14) [0.77]		0.692 (11) [0.69]		0.657 (10) [0.66]	
Model Fit Indices									
χ^2 *(df)*	598 (28)	139 (27)	78 (20)	45 (19)		24 (15)		27 (17)	
$d\chi^2$ *(ddf)*	--	459 (1)	61 (7)	33 (1)		21 (4)		3 (2)	
P(perfect fit)	p<.01	p<.01	p<.01	p<.001		p<.07		p<.052	
ε_a	0.32	0.14	0.12	0.08		0.05		0.06	
P(close fit)	p<.01	p<.01	p<.01	p<.046		p<.041		p<.38	

Note: the symbol "=" indicates a parameter fixed at the indicated value; in parentheses is the *z* value for the parameter estimate divided by its standard error; [ ] indicates standardized coefficient; – indicates value not relevant to the model; * indicates parameter equated across time intervals; $d\chi^2$ – difference in fit, and *ddf* – difference in degrees of freedom relative to prior model.

Table 3: The Alternative Fits of Several Common Factor Models of the WISC

Model	Parms (M+V+R) =T	Chi-Square/dfs	Prob. {Perfect Fit}	Change in Fit/ Change in dfs	RMSEA	Prob. {Close Fit}
0 Common Factors	8+8+0= 16	598 / 28	p < .01	-- / --	0.316	p < .01
1 Common Factor Rasch	8+8+1= 17	190 / 27	p < .01	408 / 1	0.172	p < .01
1 Common Factor	8+8+8= 24	78 / 20	p < .01	112 / 7	0.119	p < .01
2 Factors V & P	8+8+9= 25	45 / 19	p > .01	34 / 1	0.082	0.046
Maximum 2 Common Factors	8+8+15= 31	19 / 13	0.0124	26 / 6	0.047	0.490
Improved Restricted 2 Common Factors	8+8+13= 29	24 / 16	-- / --	-- / --	0.051	--

the common factor mean to be zero (at ν=0). Incidentally, this scaling is not an effort at standardization although it has that effect; it is used to simplify the expected values (as seen above). Now the resulting factor loading (λ=4.77 (17)) forms the expected value for all pairs of covariances ($\sigma(j,k)$= 4.77^2). This specific value seems a bit odd at first, so we need to explain its meaning. This is the estimated linear regression coefficient from factor one to each variable, so it is termed a factor loading here. In typical terms it means that a one unit shift in people, or a difference between people, in the factor score distribution would lead to almost a five point shift in the WISC scale score distribution. But we must remember that these two variables are not on the same scale of measurement. In theory the standardized factor score ranges from about -3 to +3 and is centered at zero, while the WISC raw score ranges from 0 to 100 and is centered at its mean value (see **Table 1**). This makes a shift of almost 5 points seem more reasonable, and it will happen in this way for each variable we consider.

Since this factor model allows some covariances to exist, this model fits the WISC data much better (χ^2=139. on df=27, ε_a=0.14; $\Delta\chi^2$=459. on Δdf=1). The fact that the raw loading is the same across all measures does not imply the standardized loadings (i.e., when both the common factor and the observed variable have unit length) are all the same size, they are not ([Λ]=[.8, .6, .7, .8, .4, .6, .7, .3]). It is fairly easy to see that the first four variables, and maybe variable 7 as well, have the highest standardized loadings. The use of a common factor with a single loading is a reasonable idea if we want to create equally weighted composite scores for the common factor. In fact this is a necessary feature of a simple composite scale (see Rasch, 1960; McDonald, 1999) and, unfortunately, it does not seem to fit well here.

The next model fit (*M2*) is a more traditional one common factor model with possibly unequal loadings (see McDonald, 1985, 1999). Since these are cognitive data, this model represents the idea that we have a General (*g*) factor (among many others, see Jensen, 1980). The common factor and its loadings are once again identified by requiring the common factor variance to be fixed (at φ_g^2=1) and the common factor mean to be zero (at ν_G=0). Now the resulting factor loadings (Λ_g=[4.2, 6.6, 4.7, 4.6, 7.2, 4.9, 3.8, 10.4]) form the expected value for all pairs of covariances ($\sigma(j,k)$= $\lambda(j)*\lambda(k)$). Since this model allows all covariances to exist, it fits the WISC data even better (χ^2=78. on df=20, ε_a=0.12) and the sequential test of the models is a test of whether the factor loadings are the same ($\Delta\chi^2$=61. on Δdf=7). The fact that the raw loadings are unequal implies that the standardized loadings can be unequal as well ([Λ_g]= [.7, .7, .6, .7, .6, .6, .6, .6]). The use of a common factor with unequal loadings now seems to be a reasonable idea for the WISC, especially if we believe in the existence of a single *g* construct of cognition (i.e., Jensen, 1980). The model is not a perfect fit to these data however, and **Table 4** gives a listing of the model misfits for each covariance we have tried to fit. As we can see here, some of the misfits are relative large (i.e., BD–OA is 2.517).

Moving on to two common factors

The next common factor model fit (*M3*) is a restricted two common factor model with possibly unequal loadings. In this model we presume a first factor produces variation in the first four observed WISC variables (IN, CO, SI, VO), while a second possibly correlated factor produces variation in

Table 4: Covariance Model Misfits from One Factor g Solution

	IN. 6	CO. 6	SI. 6	VO. 6	PC. 6	PA. 6	BD. 6	OA. 6
IN. 6	0.000							
CO. 6	0.405	0.000						
SI. 6	0.976	0.473	0.000					
VO. 6	0.170	1.000	-0.407	0.000				
PC. 6	-0.161	-0.238	-0.410	-0.491	0.000			
PA. 6	0.134	0.031	-0.561	0.101	-0.872	0.000		
BD. 6	-1.160	-1.137	0.407	-0.768	0.413	0.720	0.000	
OA. 6	-0.815	-1.619	-0.963	-0.157	1.900	0.277	2.517	0.000

the last four variables (PC, PA, BD, OA). We have used this pattern to mimic Wechsler's (1949) theoretical concept of a Verbal (*V*) factor that is separable from a Performance (*P*) or Non-Verbal factor (see McArdle & Prescott, 1992). The two common factors, and their respective loadings, are here identified by requiring each common factor variance to be fixed (at $\varphi_v^2=1$, and $\varphi_p^2=1$) and both common factor means to be zero (at $\nu_v=0$, and $\nu_p=0$). Now the resulting MLE factor loadings are mostly higher (Λ_v=[4.4, 6.9, 4.8, 4.7]; Λ_P=[7.8, 4.7, 4.4, 12.5]) as are the standardized versions ([Λ_v]=[.8, .7, .7, .8]; [Λ_p] =[.6, .6, .7, .8]). One key feature of this model is the estimation of the two factor inter-correlation (ρ_{vp} =[.77]). Since this model also allows all covariances to exist, but in an even more relaxed pattern, this model fits these WISC data even better (χ^2=45. on *df*=19, ε_a=0.08). In this specific case, the sequential test of the models is a test of whether the factor intercorrelation is unity ($\Delta\chi^2$=33. on Δdf=1), and it does not seem to be. The use of this *V–P* common factor model with unequal loadings but correlated factors is presented in **Figure 6**. This model now seems to be a more reasonable idea for the WISC, even though it argues against the existence of a single construct of cognition (i.e., Jensen, 1980). Of course this model is not a perfect fit to these data, however, and **Table 5** gives a listing of the model misfits for each covariance we have tried to fit. As we can see here, some of the misfits are still relative large (i.e., IN–PA is 1.401, but not BD–OA which is reduced to .631).

A lot has been made about the relatively high correlation estimate found among these two new factors in our V&P solution. What we have done in the comparison of the previous two models is test whether or not this factor correlation is unity (when it would be a single factor), and this analysis suggests that it does not seem to be unity. But this is not a test of whether the correlation is zero, and it most certainly is not this zero value either. So we are stuck with a dilemma of modern data analysis: How can independent constructs be so highly correlated? There are many models for this, including the use of higher order factor analysis (see Thurstone, 1947) discussed earlier. This approach is not viable here because we only have two indicators (*V* and *P*) or a higher order factor (*g*) so there are no degrees of freedom and this means no rejectable test is possible. Some researchers ignore this problem and fit a higher order model with as few as three lower order factors, but it is a mistake to think this is a test of anything at all. Thus we are basically stuck with a correlated factors model at this level.

Figure 6: Two common factors (*M3*) for WISC scales (*N*=204) using Eight Variables at Age 6

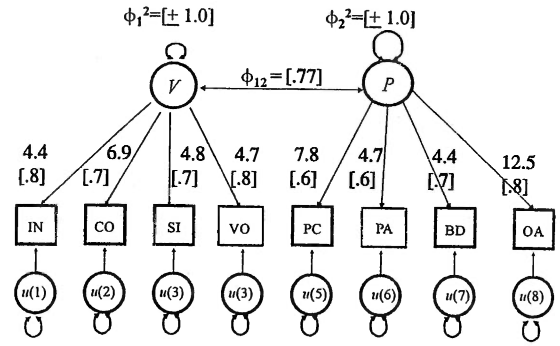


Table 5: Covariance Model Misfits from Two Factor V&P Solution

	IN. 6	CO. 6	SI. 6	VO. 6	PC. 6	PA. 6	BD. 6	OA. 6
IN. 6	0.000							
CO. 6	-0.153	0.000						
SI. 6	0.611	0.017	0.000					
VO. 6	-0.247	0.501	-0.740	0.000				
PC. 6	0.678	0.489	0.379	0.412	0.000			
PA. 6	1.401	1.178	0.623	1.427	-1.060	0.000		
BD. 6	-0.749	-0.829	0.797	-0.308	-0.755	0.111	0.000	
OA. 6	-0.561	-1.467	-0.694	0.146	0.511	-0.567	0.631	0.000

Now perhaps we should not be so disturbed that this journey seems to end here. After all, this is certainly true for other independently and well-measured variables such as height and weight. These are scores which are highly positively correlated (r>.60; see Sargent, 1963) but certainly have independent predictors and independent outcomes and may be correlated due to being based on living samples (where height and weight must be correlated).

What we usually require is that these factors, if there are two, behave in different ways from one another. This behavior is external to the current analysis and is yet to be determined (for details, see McArdle & Prescott, 1992). But we would be very skeptical if subsequent models for both variables led to the same conclusions (they usually do not). Incidentally, Thurstone (1947) also showed that part of the correlation among factors can be the result of a non-random sampling of persons. What seems a bit odd about this cognitive literature is that researchers in this domain seem to ignore these external features at all and simply continue to advocate *g* at a higher level where no test is possible (e.g., Jensen, 1980; Carroll, 1993; Salthouse, 2010).

Exploring the two factor solution further

The previous model was based on a specific hypothesis about the WISC, and we know it is not the only two common factor model that could be useful. To consider this possibility we fit the next model (*M4*). In order to retain a nested sequence of models, we simply added four factor loadings to the prior model. Incidentally, we could have added up to three loadings on each factor and be uniquely estimated (see McArdle & Cattell, 1994). In this way we allow the first factor, possibly *V*, to load on PC and PA, and the second factor *P*, to load on CO and SI. All other loadings and the factor inter-correlation were allowed, and the resulting inter-factor correlation is slightly lower (ρ_{VP} =[.69]). This model improves the fit a bit more (χ^2=25. on df=15, ε_a=0.05) and the overall contribution of the four loadings is evident ($\Delta\chi^2$=21. on Δdf=4). When we look at the estimates we see that while the additional loadings on the first factor *V* seem like they could be useful (i.e., it seems possible that PA is better off loading on the first factor, or on both factors), those on the second factor *P* yield nothing at all. In the next model (*M5*), we simply removed the two small loadings, and the result is a very clean model. Of course, these exploratory attempts to reduce or trim the SEM do not have a true probability value. If we did want to go to the maximum extraction for two common factors, we could have added two more factor loadings (and obtained χ^2=19. on df=13, ε_a=0.05, p\{perfect\}=0.12, p\{close\}=0.50), but we already have most of the fit. So, in any case, what we now conclude is that a single factor *g* is not such a good idea for these WISC data (c.f., Jensen, 1980), and the basic ideas of *V* and *P*, as stated by Wechsler (1949), may be much better in the long run. We will keep this idea in mind.

Discussion

Conclusion—so where do we stand now?

The result presented here now suggests that there is more than one common factor in the WISC scores of first-graders. Nevertheless some researchers would say that a single factor exists among these variables, while others would say that at least two factors are needed, probably more. But the latter is not really controversial. As I have tried to demonstrate here, the first common factor by itself could account for a lot of the covariance among the measures, almost 50% for that matter, so it must be useful for something. The large impact of a one factor model is not denied by this analysis. But just to say that a global composite is useful because it accounts for a large proportion of observed variance is not typically how we judge the validity of constructs in contemporary analyses. In contemporary terms, the single *g* model is overly simplistic and fails to meet the minimal conditions of the original test (from Spearman, 1904; see McArdle, 2007a). As an aside, this is one clear place where the author(s) of the "Bell Curve" went wrong as well. They simply asserted that there was a *g* factor and this led to their further confusion about race differences.

So, to be clear, using the statistical basis of factor analysis we now conclude "There is not a *g* factor, there never was a *g* factor, and we can demonstrate this with anyone's data" (see McArdle & Woodcock, 1977; McArdle & Prescott, 1992). It appears the reason other prominent researchers do not come to this conclusion is because (a) this basic logic of statistical model fitting is ignored, or (b) there is nothing much better to say (the alternatives do not make sense), or (c) there is a practical need for a strict rank ordering of persons. I suggest all of these alternatives are correct to some degree.

Perhaps we need a new definition of intelligence

Given this starting point it is often useful to ask "what else can it be?" That is, what can we say if we do not have a single dimension to talk about how we can organize the mental ability differences we observe among people. One answer has already been provided.

"*... [I]ntellectual abilities are organized at a general level into two general intelligences, viz., fluid intelligence and crystallized intelligence and in terms of visual, auditory, memory and speed-of-thinking kinds of intelligence.... there are those influences which directly affect the physiological structure upon which intellectual processes must be constructed—influences operating through the agencies of heredity and injury: in adulthood development these are most accurately reflected in measures of fluid intelligence. In early (at birth, infancy and childhood) these influences affect both fluid and crystallized abilities. And on the other hand there are those influences which affect physiological structure only indirectly through agencies of learnings and acculturations: crystallized intelligence is the most direct resultant of individual differences in these influences*" (Horn & Cattell, 1967).

John Horn obviously believed that the weight of the evidence argues against a general factor (*g*) as being responsible for all intelligent behavior. The Cattell-Horn theory of fluid and crystallized intelligence (R. B. Cattell, 1941, 1971, 1998; Horn, 1965; Horn & Cattell, 1966a, 1966b) proposed that general intelligence is actually a conglomeration of a large number of different abilities working together in various and different ways in different people to bring out different intelligences. Their g_f-g_c theory separates these abilities broadly into, first, two different sets of abilities that have quite different trajectories over the course of development from childhood through adulthood.

To quote other sources, "Fluid abilities (g_f) drive the individual's ability to think and act quickly, solve novel problems, and encode short-term memories. They have been described as the source of intelligence that an individual uses when he or she doesn't already know what to do. Fluid intelligence is grounded in physiological efficiency, and is thus relatively independent of education and acculturation (Horn, 1967). The other factor, encompassing crystallized abilities (g_c), "stems from learning and acculturation and is reflected in tests of knowledge, general information, use of language (vocabulary) and a wide variety of acquired skills (Horn & Cattell, 1967). Personality factors, motivation and educational and cultural opportunity are central to its development, and it is only indirectly dependent on the physiological influences that mainly affect fluid abilities. Many studies have demonstrated that fluid intelligence peaks in early adulthood and then declines, gradually at first and then more rapidly as old age sets in after about 70. Crystallized abilities continue to improve as individuals age (Horn & Cattell, 1967)." "Horn's most recent work, done primarily with Hiromi Masunaga suggests that in adulthood people funnel their abilities into areas of expertise (http://www.indiana.edu/~intell/horn.shtml)." These more recent extensions to this g_f-g_c theory are important in a statistical sense as well.

More than a hundred years ago Spearman made a clear methodological suggestion—a model dealing with latent variables was possible, to put it to a rigorous empirical test by measuring multiple realizations of it in the form of indicators (see McArdle & Prescott, 1992, 2010). This also implies that we should not confuse Crystallized abilities with Verbal materials and Fluid abilities with Non-Verbal abilities. Nevertheless, due to the fact that decisions about goodness-of-fit are relative to the data at hand, comparisons *within* a data set can be informative, and absolute rules for goodness-of-fit indices are useful but typically misplaced. This implies the popular "non-significant χ^2" or "$\varepsilon_a < .05$" are not justified in all cases, and it may also be important to remember Kaiser's (1976) comment on the statistical basis of factor analysis: *"Delight in its elegant algebra and prose ...but for God's sake, don't take it seriously!"* However, even given all these technical caveats there seems to be no need to drop the whole idea of testability. In fitting any factor model it is most useful to recall: "Factors in a factor analysis are not *things*, but they are our evidence for the *existence of things*" (R.B. Cattell, 1998). Substantive knowledge is always needed in the interpretation of a factor analysis.

The second contribution made by Spearman 100 years ago was the substantive theory of a single underlying and unifying force behind all intellectual activity—the factor he labeled *g*. In my view, Spearman's *g-theory has been soundly rejected* by the very factor analysis methods he created. I do not want to seem naïve about the ongoing debate where many researchers and educators still strongly support some form of a *g*-theory (e.g., among many others, see the references in Jensen, 1998; Herrnstein & Murray, 1994; Lubinski, 2004). However, it may be wise to focus on the cases where we can use Spearman's factor methods to examine the strength of the structural model (*dfs*) and the evidence of goodness-of-fit (ε_a) to evaluate *g*-theory. In the first example presented in Spearman (1904) the original *g*-factor argument of multiple cognitive tests fits nicely (ε_a<.05). However, in subsequent research the single factor model uniformly failed at the first-order level. A first attempt to fix this solution was offered in Burt's (1909, as reported by Burt, 1949) "orthogonal group factors" solution. Thurstone (1938, 1947) created the multiple oblique factors solution widely used today and suggested, again due to constant lack of fit at the 1st-order, *g*-theory was reconsidered at the 2nd-order level. Schmidt & Leiman (1957) showed how the multiplication of the 1st and 2nd order factor loadings could be used to show the orthogonal impacts of the 2nd-order factor on the observed variables (i.e., Burt's group factor solution with constraints).

In contrast to this other research, Cattell (1941), one of Spearman's best-known students, suggested the need for at least two general factors at the 2nd-order level to adequately represent the available cognitive data; in deference to Spearman, he labeled these factors as g_f and g_c. When a positive correlation was found between these factors (ρ=.5-.7), some naïve researchers again suggested this correlation was supportive evidence for *g* at a higher level—Spearman's factor method points out that any explanation of a single correlation is not rejectable (i.e., has negative *df*). More recent SFA research suggested evidence of *g* at the 3rd-order in cross-sectional data, but also that this *g*-factor is almost entirely related to the 2nd-order g_f. For example, Gustaffson (1989) fit a higher-order factor model with three factors at the 2nd-level and still made an interpretation of *g* as g_f (based on zero *df* but with deficient rank, *c.f.*, McArdle, 1991). As stated earlier, and more recently, Carroll (1993, 1998) used the traditional Schmid-Leiman multiplication on a wide variety of correlation matrices from cross-sectional data and made a similar statement about the *g* at the 3rd order being close to g_f. But other researchers have been fairly critical of this slippery movement of *g* upwards:

"The Humphrey's lesson about factor analysis supports the argument that first principal component and IQ measures of intelligence should be banished from most scientific research... The lessons of Humphrey also teach that stepping up to higher order solutions, as in Schmid-Leiman transformations, do not solve the problems stemming from the arbitrariness of collections of tests... But one can calculate this general factor for any mixture of abilities, and there is no assurance that the factor thus calculated in one arbitrary battery is at all equivalent to a factor calculated in the same way in another such battery." (J.L. Horn, 1989, p.38).

In contrast with this support for hierarchical *g*-theories, Cattell (1971) postulated the "investment theory" which suggested any contemporaneous correlation among g_f and g_c was because g_f (fluid reasoning) led to the development (hardening) of the g_c (crystallized-knowledge) especially early in individual development. It is clear that this kind of a developmental theory is not testable using cross-sectional data alone, so more recent research has used longitudinal data to model these dynamic hypotheses directly (see McArdle et al, 2001; McArdle & Hamagami, 2006; Ferrer & McArdle, 2004).

If we are to continue to make progress in the next century in intelligence research we need to carefully consider the powerful lessons of this *g*-theory debate. This is hardly ever a debate about the models of factor analysis, and the much heralded statistical evidence available in the factor analysis is often overlooked. The well-known critique by Gould (1981) seems to attack the factor analysis methods, but it is actually an attack on the misuses of factor analysis by factor analysts, supposedly due to personal prejudices and policy goals. If we are to have any hope to create an "objectively determined" set of scientific results in the future we need to reverse this trend and pay more attention to the basic evidence that actually emerges from our factor analysis. In this ironic sense, Spearman's factor analysis methods should now be considered an outstanding scientific achievement largely because his more well-known *g*-theory should not. Yes, even though I fought it for a very long time, I guess I have decided that John L. Horn was essentially correct!

References

Binet, A. & Henri, V. (1896). "La psychologie individuelle." *Année Psychologie, 2*, 411–465.

Binet, A. & Simon, S. (1905). Methods nouvelles pour le disgnostic de niveau intellectuel des anormaus. *L'Annee Psychologique, 11*, 191–244.

Block, N.J. & Dworkin, G. (1976). *The IQ Controversy*. New York: Pantheon Books.

Brody, N. (1992). *Intelligence*. San Diego, CA: Academic Press.

Browne, M.W., (2001) An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–150.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. and Long, J.S. (Eds.), *Testing Structural Equation Models*. Beverly Hills, CA: Sage, pp 136–162.

Burt, C.L. (1949). The structure of the mind: A review of the results of factor analysis. *British Journal of Educational Psychology, 19*, 100–111, 176–199.

Carroll, J.B. (1993). *Human Cognitive Abilities: A Survey of Factor Analytic Studies*. Cambridge, Eng.: Cambridge University Press.

Carroll, J.B. (1998). Human cognitive abilities: A critique. In McArdle, J.J. and Woodcock, R.W. (Eds.), *Human Abilities in Theory and Practice*. Mahwah, NJ: Erlbaum, pp 5–24.

Cattell, R.B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38*, 592.

Cattell, R.B. (1971). *Abilities: Their Structure, Growth and Action*. Boston: Houghton-Mifflin.

Cattell, R.B. (1998). Where is intelligence? Some answers from the triadic theory. In McArdle, J.J. and Woodcock, R.W. (Eds.), *Human Abilities in Theory and Practice*. Mahwah, NJ: Erlbaum, pp 29–37.

Cattell, R.B. & Horn, J.L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15*, 139–164.

Deary, I.J. (2000). *Looking Down on Intelligence: From Psychometrics to the Brain*. New York: Oxford University Press.

Detterman, D. (1987). Theoretical notions of intelligence and mental retardation. *American Journal of Mental Deficiency, 92*, 2–11.

Detterman, D.K. & Daniel, M.K. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence, 13*, 349–359.

Ferrer, E. & McArdle, J.J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology, 40*, 935–952.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191.

Gardner, H. (1999). *Intelligence Reframed: Multiple Intelligences for the 21st Century*. New York: Basic Books.

Gottfredson, L.S. (2004). Intelligence: Is it the epidemiologists' elusive "Fundamental Cause" of social class inequalities in h*ealth? *Journal of Personality and Social Psychology, 86*, 174–199.

Gould, S.J. (1981). *The Mismeasure of Man*. New York: Norton.

Guilford, J.P. (1956). The structure of intellect. *Psychological Bulletin, 53*, 267–293.

Guilford, J.P. (1967). *The Nature of Human Intelligence*. New York: McGraw-Hill.

Guilford, J.P. (1974). Rotation problems in factor analysis. *Psychological Bulletin, 81*, 498–501.

Guilford, J.P. (1980). Fluid and crystallized intelligences: Two fanciful concepts. *Psychological Bulletin, 88*, 406–412.

Gustavson, J.E. (1989). Broad and narrow abilities in research on learning and instruction. In Ackerman, P., Kanfer, R. and Cudeck R. (Eds.), *Learning and Individual Differences: Abilities, Motivation, and Methodology*. NJ: Erlbaum, pp 203–238

Hauser, R.M. (2010). Causes and consequences of cognitive functioning across the life course, *Educational Researcher, 9*, 95–109.

Herrnstein, R.J. & Murray, C.A. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. NY: The Free Press.

Horn, J.L. (1965). *Fluid and Crystallized Intelligence: A Factor Analytic Study of the Structure Among Primary Mental Abilities*. Ph.D. Thesis. University of Illinois.

Horn, J.L. (1967). Intelligence: Why it grows, why it declines. *Transaction*, 23–31.

Horn, J.L. (1979). The rise and fall of human abilities. *Journal of Research on Developmental Education, 12*, 59–78.

Horn, J.L. (1985). Remodeling old models of intelligence. In B.B. Wolman (Ed.), *Handbook of Intelligence: Theories, Measurements, and Applications*. New York: Wiley, pp 267–300.

Horn, J.L. (1986). Some thoughts about intelligence. In Sternberg, R.S. and Detterman, D.K. (Eds.), *What is Intelligence? Contemporary viewpoints on its nature and definition*. Norwood, NJ: Ablex, pp 91–96.

Horn, J.L. (1989). Models of intelligence. In R. Linn (Ed.), *Intelligence, Measurement & Public Policy*. Urbana, IL: University of Illinois Press, pp 29–73.

Horn, J.L. & Cattell, R.B. (1966a). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*, 253–270.

Horn, J.L. & Cattell, R.B. (1966b). Age differences in primary mental ability factors. *Journal of Gerontology, 21*, 210–220.

Horn, J.L. & Cattell, R.B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*, 107–129.

Horn, J.L. & Goldsmith, H.H. (1999). Reader beware: Review of Arthur Jensen's book: "The *g* factor."

Horn, J.L. & Knapp, J.R. (1974). Thirty wrongs do not make a right: Reply to Guilford. *Psychological Bulletin, 81*, 502–504.

Horn, J.L. & McArdle, J.J. (1992). A practical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117–144.

Horn, J.L. & McArdle, J.J. (1980). Perspectives on mathematical and statistical model building (MASMOB) in research on aging. In L. Poon (Ed.), *Aging in the 1980s: Psychological Issues*. Washington, D.C.: American Psychological Association, 503–541.

Horn, J.L. & McArdle, J.J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R. MacCallum (Eds.), *Factor Analysis at 100 years*. Mahwah, NJ: Erlbaum, pp 205–247.

Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Indiana University. Human Intelligence: John L. Horn. http://www.indiana.edu/~intell/horn.shtml (accessed September 3, 2012).

Jensen, A.R. 1980. *Bias in Mental Testing*. New York: Free Press.

Jensen, A.R. 1998. *The g Factor: The Science of Mental Ability*. Westport, CT: Praeger.

Johnson, W., Bouchard, T.J., Kruger, R.F., McGue, M. & Gottesman. I.I. (2004). Just one g: Consistent results from three test batteries. *Intelligence, 32*, 95–107.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426.

Kaiser, H. (1976). Book Review: D.N. Lawley and A.E. Maxwell. Factor analysis as a statistical method (2nd ed.). New York: American Elsevier, 1971. *Educational and Psychological Measurement, 36*: 586–589.

Lawley, D.N. & Maxwell, A.E. (1971). *Factor analysis as a statistical method*. New York: American Elsevier.

Lemann, N. (1999). *The Big Test: The Secret History of the American Meritocracy*. New York: Farrar, Strauss, and Giroux.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General Intelligence,' Objectively Determined and Measured." *Journal of Personality and Social Psychology, Vol 86*, 96–111.

Maydeau, A. & McArdle, J.J. (Eds., 2005). *Contemporary Advances in Psychometrics*. Mahwah, NJ: Erlbaum.

McArdle, J.J. (1984). On the madness in his method: R. B. Cattell's contributions to structural equation modeling. *Multivariate Behavioral Research, 19*, 245–267.

McArdle, J.J. (1988). Dynamic but structural equation modeling of repeated measures data. In J.R. Nesselroade and R.B. Cattell (Eds.), *The Handbook of Multivariate Experimental Psychology, Volume 2*. New York, Plenum Press, pp 561–614.

McArdle, J.J. (1989). Structural modeling experiments using multiple growth functions. In P. Ackerman, R. Kanfer and R. Cudeck (Eds.), *Learning and Individual Differences: Abilities, Motivation, and Methodology*. Hillsdale, NJ: Erlbaum, pp 71–117.

McArdle, J.J. (1991). Structural models of developmental theory in psychology. In P. Van Geert and L.P. Mos (Eds.), *Annals of Theoretical Psychology, Volume VII*, pp 139–160.

McArdle, J.J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29*, 409–454.

McArdle, J.J. (1996). Current directions in structural factor analysis. *Current Directions in Psychological Science, 5*, 11–18.

McArdle, J.J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. du Toit and D. Sorbom (Eds.), *Structural Equation Modeling: Present and future*. Lincolnwood, IL: Scientific Software International, pp 342–380.

McArdle, J.J. (2007a). Five steps in the structural factor analysis of longitudinal data. In R. Cudeck & R. MacCallum, (Eds.). *Factor Analysis at 100 years*. Mahwah, NJ: Erlbaum, pp 99–130.

McArdle, J.J. (2007b). Obituary for John L. Horn (1928–2006). *American Psychologist, 62*(6), 596–597 (Reprinted in *Intelligence*, 2008, and in Kyllonen, Roberts & Stankov (2008), pp 219–222.)

McArdle, J.J. (2009). Latent variable modeling of longitudinal data. *Annual Review of Psychology, 60*, 577–605.

McArdle, J.J. & Aber, M.S. (1990). Patterns of change within latent variable structural equation modeling. In A. von Eye (Ed.). *New Statistical Methods in Developmental Research*, New York: Academic Press, 151–224.

McArdle, J.J. & Cattell, R.B. (1994). Structural equation models of factorial invariance in parallel proportional profiles and oblique confactor problems. *Multivariate Behavioral Research, 29*, 63–113.

McArdle, J.J. & Epstein, D.B. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110–133.

McArdle, J.J., Ferrer-Caja, E., Hamagami, F. & Woodcock, R.W. (2002). Comparative longitudinal multilevel structural analyses of the growth and decline of multiple intellectual abilities over the life-span. *Developmental Psychology, 38*, 115–142.

McArdle, J.J., Fisher, G.G. & Kadlec, K.M. (2007). Latent variable analysis of age trends in tests of cognitive ability in the health and retirement survey, 1992–2004. *Psychology and Aging, 22*, 525–545.

McArdle, J.J., Grimm, K., Hamagami, F., Bowles, R. & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods, 14*, 126–149.

McArdle, J.J. & Goldsmith, H.H. (1990). Alternative common factor models for multivariate biometric analyses. *Behavior Genetics, 20*, 569–608.

McArdle, J.J. & Hamagami, F. (2006). Longitudinal tests of dynamic hypotheses on intellectual abilities measured over sixty years. In C. Bergeman & S.M. Boker (Eds.), *Methodological Issues in Aging Research*. Mahwah: Erlbaum, 43–98.

McArdle, J.J., Hamagami, F., Meredith, W. & Bradway, K.P. (2000). Modeling the dynamic hypotheses of g_f-g_c theory using longitudinal life-span data. *Learning and Individual Differences, 12*, 53–79.

McArdle, J.J. & Lehman, R.S. (1992). A functionalist view of factor analysis. In D.F. Owens & M. Wagner (Eds.), *Progress in Modern Psychology: The Contributions of Functionalism to Modern Psychology*, Hillsdale, N.J.: Erlbaum, 167–187.

McArdle, J.J. & Nesselroade, J.R. (1994). Using multivariate data to structure developmental change. In S.H. Cohen and H.W. Reese (Eds.), *LifeSpan Developmental Psychology: Methodological Innovations*. Hillsdale, N.J.: Erlbaum, 223–267.

McArdle, J.J. & Nesselroade, J.R. (2003). Growth curve analyses in contemporary psychological research. In J. Schinka & W. Velicer (Eds.), *Comprehensive Handbook of Psychology, Volume Two: Research Methods in Psychology*. NY: Pergamon Press, 447–480.

McArdle, J.J. & Nesselroade, J.R. (2012). *A Workshop in Longitudinal Structural Equation Modeling*. Washington, D.C.: APA Books.

McArdle, J.J. & Prescott, C.A. (1992). Age-based construct validation using structural equation modeling. *Experimental Aging Research, 18*, 87–115.

McArdle, J.J. & Prescott, C.A. (1996). Contemporary models for the biometric genetic analysis of intellectual abilities. In D.P. Flanagan, J.L. Genshaft and P.L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests and Issues*. New York: Guilford Press, pp 403–436.

McArdle, J.J. & Prescott, C.A. (2010). Contemporary modeling of Gene-by-Environment effects in randomized multivariate longitudinal studies. *Perspectives on Psychological Science, 5*, 606–621.

McArdle, J.J. & Woodcock, J.R. (1997). Expanding test-retest designs to include developmental time-lag components. *Psychological Methods, 2*, 403–435.

McArdle, J.J. & Woodcock, R.W. (Eds., 1998). *Human Abilities in Theory and Practice*. Mahwah, NJ: Erlbaum.

McDonald, R.P. (1985). *Factor analysis and related methods*. Mahwah, NJ: Erlbaum.

McDonald, R.P. (1999) *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

Osborne, R.T. & Suddick, D.E. (1972). A longitudinal investigation of the intellectual differentiation hypothesis. *Journal of Genetic Psychology, 121*, 83–89.

Popper, K.R. (1970) Normal Science and its Dangers. In I. Lakatos and A. Musgrave (Eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, pp 51–58.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research, expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.

Revelle, W. (1983). Factors are fictions and other comments on individuality theory. *Journal of Personality, 51*, 707–714.

Ruzgis, P. (1994). Thurstone, L.L. (1887–1955). In R.J. Sternberg (Ed.). *Encyclopedia of human intelligence*. New York: Macmillan, pp 1081–1084.

Salthouse, T.A. (2010). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology, 24*, 563–572.

Sargent, D. (1963). Weight-height relationship of young men and women, *American Journal of Clinical Nutrition, 13*, 318–325.

Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

Snow, R. (1998). Abilities as aptitudes and achievements in learning situations. In McArdle, J.J. & Woodcock, R.W. (Eds.), *Human Abilities in Theory and Practice*. Mahwah, NJ: Erlbaum. pp 93–112.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology 15*, 201–293.

Spearman, C. (1927). *The Abilities of Man: Their Nature and Measurement*. New York: Macmillan.

Terman, L.M. (1925). *Genetic Studies of Genius*. Stanford, CA: Stanford University Press.

Thurstone, L.L. (1938). *Primary Mental Abilities*. Psychometric Monographs No. 1. Chicago: University of Chicago Press.

Thurstone, L.L. (1947). *Multiple Factor Analysis*. Chicago: University of Chicago Press.

Tucker, L.L & Lewis, C.A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10.

Wechsler, D. (1949). *Manual of the Wechsler Intelligence Scale for Children*. New York: Psychological Corp.

About the author

John J. (Jack) McArdle, Ph.D., is now senior professor of psychology at the University of Southern California (USC) where he heads the Quantitative Methods training program. He received a B.A. in psychology from Franklin & Marshall College in 1973, his Ph.D. from Hofstra University (in Hempstead, N.Y.) in 1977, then went to the University of Denver as an NIH Postdoctoral Fellow to work with Dr. John L. Horn. In 1984 he moved to the University of Virginia to start a quantitative methods program, and in 2005 he moved to USC (with John L. Horn) to do the same. He now mentors seven seriously committed graduate students. He has both written a book, called A *Theory of Longitudinal Structural Equation Modeling* (with J.R. Nesselroade, APA Books, 2013), and has edited a book called *Contemporary Issues in Exploratory Data Mining* (with G. Ritschard, Routledge Press). McArdle was recently awarded an NIH-MERIT grant from the National Institute on Aging for his work on "Longitudinal and Adaptive Testing of Adult Cognition." (2005–2016). He has also been heavily involved with research on the Academic Skills of College Student Athletes with the National Collegiate Athletics Association (NCAA).

Effect of F&M education on my career

There were at least two influential scientific experiences that I had at F&M. The first was failing chemistry in my freshman year as a result of choosing to play freshman football. This was my own idea and it was obviously a bad one from a professional point of view. The second experience was doing my senior thesis in psychology with Dr. Richard S. Lehman. From Prof. Lehman, I learned not only about the use of statistics in psychology, but also about computer programming and simulation methods of psychological research. I knew nothing about these issues before this wonderful experience. Ultimately, this interest in both statistical methods and computer programming proved very valuable to me, and I have encouraged many other students along the same lines. Interestingly, this was exactly the path that was recommended to me by Prof. Lehman during my senior year at F&M.

Tracing the Paths from Basic Research to Economic Impact

Francis Narin '55

Retired President,
CHI Research, Inc.

Abstract

A series of studies throughout my research career have illuminated many of the steps by which basic research leads ultimately to economic rewards. This sequence started with the TRACES study in the late 1960s (Technology in Retrospect and Critical Events in Science), which identified the non-mission and mission oriented research events leading to 5 important technical advances. The TRACES study was performed essentially by hand, with knowledgeable experts identifying the events. However, quantitative analysis of the events did reveal key characteristics of the process, such as a 20–30 year lag between the peak of non-mission research and the eventual innovations. Following TRACES, more systematic bibliometric (publication and citation) analyses were initiated to analyze the process more objectively, leading to quantitative indicators of research performance incorporated in the U.S. National Science Board's biennial Science and Engineering Indicators reports, from the first 1972 report to the most recent 2012 report. These indicators delineated the citation dependence of more applied research on high impact basic research, and of industrial developments on publically funded basic research. In the 1980s these citation techniques were extended to patents, first patent-to-patent citation and then to citation links between patents and papers. Key findings in the 1990s were a rapid increase in the dependence of U.S. technology on basic and applied science, and that 73% of the science base of U.S. industrial patents came from publically supported science. In the economic realm it was shown in 1999 that companies whose patents were highly linked to science, and