Understanding Human Intelligence: Where Have We Come Since Spearman?

John L. Horn[i]

University of Southern California

[Note. Dr. John Horn has given permission for this pre-publication draft of this paper to be made available by Kevin McGrew via IAP's blog/listserv/web page. The chapter will be published in a book titled Factor Analysis at 100*, edited by R. Cudeck and R. MacCallum, and published by Lawrence Erlbaum, 100 Mahwah, NJ.]*

One hundred years have passed since Charles Spearman (1904a, 1904b) published papers in which --it has been said --nearly all the basic formulas that are particularly useful in test theory are found (Gulliksen, 1950, p.1), and g, "a general mental ability that enters into every kind of activity requiring mental effort, was discovered" (Jensen, 1998, p.18). Such thoughts led to a question that led to organizing a conference, which led to writings and the idea of putting the writings together in a book. The question was: "What have we learned about test theory and human intelligence since Spearman wrote those articles that seemed to say it all?" In this chapter I'll talk about the part of the question that pertains to human intelligence.

To deal with this part of the question we will lay out knowledge acquired from Spearman's work. In doing this we find that what often is assumed to be known is not known. We find, also, that we learned a lot that did not come directly from Spearman's work even as it did come from following leads that Spearman laid down. We will consider those leads first, and then look at major substantive contributions of Spearman's work.

Spearman's Theory of General Intelligence

A scientific theory must be falsifiable: that's what distinguishes scientific theory from other theory. In his1904a paper, Spearman put forth a scientific theory of general intelligence. Others of his time had theories about intelligence, but they had no test that could falsify their theory. Binet and Simon (1905) were prominent among these others. They said intelligence is a collection of cognitive abilities. That collection-of-abilities theory became the accepted theory, the standard used throughout the world in psychology and education.

Still today it is the theory of the large majority of people who talk and write about human cognitive abilities. Jenson (1998), for example, one of the best known members of that majority, argues that intelligence is the first principal factor in any large collection of diverse mental ability tests. The first principal factor is simply a weighted sum of the test scores. The theory thus argues that intelligence is a sum of the scores obtained with a diverse collection of mental ability tests –just what Binet and Simon (1905) said.

Such theories are not really scientific theories because there is no test that can possibly falsify them: one collection-of-abilities theory of intelligence is no better –or worse— than another. Collection-of-abilities theories spawn commercial enterprise --indeed, many marketed intelligence tests have derived from such theories—but they do not spawn a science.

Spearman’s theory was different. It described what the results of doable experiments would be if the theory was correct –and, just as important, what the results would be if the theory was not correct. It required that one identify intelligence –whatever it was—and distinguish it from what was not intelligence. Spearman’s theory thus directed the fledgling field of psychology toward research that could build a science to describe that which people referred to when they used the term “human intelligence.”

The essence

It may seem on first consideration that Spearman’s theory is no better than collection-of-abilities theories, for it argues that intelligence is a common factor manifested in every kind of mental effort. Thus one might think that this would be the first principal factor in any large collection of diverse mental ability tests –just what

Jensen argued. But it isn't. Spearman's theory requires that that factor be the only common factor, not just the first principal factor.

The theory requires an experiment in which people are measured with separate devices[ii] for each kind of activity requiring mental effort. It argues that the measures obtained with such devices involve two things – general intelligence and a specific factor. Persons scoring well on a device may do so in virtue of having and using of general intelligence, but also in virtue of having and using ability, or other abilities or by luck. Such other ability or abilities or luck is referred to with the singular term "specific factor." In an experiment required to test Spearman's theory it is necessary that there be no duplication of any specific factor in the devices selected to measure different kinds of mental effort. If this condition is met, and the theory of general intelligence is correct, then every one of the off-diagonal 2-by-2 determinants of the matrix of intercorrelations among the measurement devices will be zero (to within chance variation). On the other hand, if the theory is not correct, or an experiment is not adequately designed (to meet the conditions of the theory), then the 2-by-2 determinants will not be zero. The 2-by-2 determinants came to be called tetrad differences. The tetrad differences will all be zero if each of the sampled measurement devices measure one (and only one) factor in common. The tetrad differences will not be zero --even if the theory is correct-- if any specific factor is reliably measured in more than one measurement device of an experiment.

Again it might seem on first consideration that Spearman's theory is not scientific because it requires that intelligence be general –operate in every activity requiring mental effort. This necessitates representative sampling of all such activities, which is impossible, so experiments to test the theory are not doable; therefore, the theory is not testable, and thus is not scientific. But this reasoning misconstrues what is required of a scientific theory. A scientific theory need not be testable in a limiting statement of the theory. Most scientific theories are not testable in this sense.

For example, the kinetic theory of molecules states that all molecular motion stops at absolute zero (-273 centigrade). Such a condition can never be sampled – Brownian motion can never be stopped entirely. This does make the theory unscientific. It merely indicates the impossibility of testing the absolute limiting condition of the theory. The theory specifies doable experiments that could falsify it. That's what makes it a scientific theory. Similarly, the ideal of obtaining a representative sample of all indicators of intelligence need not be realizable in order for Spearman's theory to be testable[iii].

For any scientific theory there are hypotheses that are not testable. This does not render a theory unscientific; it renders it not fully tested, never proven for sure, which is a hallmark of scientific theories (in contrast to non-science theories, which claim to be fully and absolutely proven). The requirement for a theory to be scientific is that it be falsifiable in doable experiments, and Spearman's theory meets that requirement

The Origins

Spearman had studied for the doctoral degree with Wilhelm Wundt, usually regarded as the founder of experimental psychology –that is, research based on experimental-versus-control-group designs. But Spearman also read the works of Francis Galton (1869; 1883), often regarded as the founder of differential psychology –that is, research based on covariation designs (studies of individual differences). Spearman's theory of general intelligence, and his specifications for testing it, derived primarily from the Galton school of psychology.

Galton had theorized that individual differences in intellectual achievements – particularly the differences between the achievements of eminent people and ordinary people-- reflect hereditary differences in a power of the mind. He thought this power would be indicated by keenness of sensory discrimination (acuity in seeing, hearing, tasting, smelling, sensing touch). He reasoned that thinking is required in the intellectual achievements that indicate power of the mind, and since all thinking must depend on the 5 senses (a premise that Locke had advanced), individual differences in sensory discrimination should indicate individual differences in power of the mind. He sought evidence to support this idea.

He had established an Anthropometric Laboratory in London's South Kensington Science Museum. Visitors to the museum were allowed to pay 3 pence to have their sensory discriminations measured (and provide other information). Thousands of people paid the 3 pence and thus provided Galton with data. He had developed a way of calculating the correlation between variables. He reasoned that if keenness of sensory discrimination indicated intellectual achievement, then the intercorrelations among the discrimination measures for the different senses and the correlations of these measures with intellectual achievements should be large. This is not what he found: the correlations he calculated were very small— near zero.

By the time Spearman undertook his 1904 studies, he had created methods for estimating reliability of measurement and had proved that correlations can be low primarily because the reliabilities of the measures are low. He had invented a method for correcting correlations for attenuation due to unreliability. Had such corrections been used with the correlations Galton obtained, the estimated correlations would have been larger, so the results would have presented a more favorable case than Galton presented for the hypothesis that measures of sensory discrimination are inter-related and related to intellectual achievements. Spearman's 1904a article is largely a presentation of this more favorable case[iv].

First Studies

The Spearman (1904a) article is a 93-page report of four studies of measures of sensory discrimination and measures of "the intelligences." In these studies he did more than merely correct correlations for attenuation due to unreliability. To obtain what he regarded as proper data for analyses, he eliminated subjects from initial samples to control for experience differences; he rescaled variables to remove possible practice effects; he adjusted for age and gender differences; he partialled out a variety of what he called irrelevant influences; he "arbitrarily assumed observational errors" to estimate reliabilities. Then, after all this, he corrected the correlations for attenuation due to measurement unreliability.

With the data adjusted in these ways Spearman directed analyses at showing that there is a "correspondence" –by which he meant a single common factor-- among the different forms of sensory discrimination, a comparable correspondence among different measures of intelligence, and a correspondence among both "the discriminations and the intelligences". Surveying the outcomes of these analyses, he concluded that "…these results …and other analogous observed facts indicate that all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all others" (Spearman, 1904a, p.284).

The calculations for the critical test of Spearman's hypothesis in these earliest studies were not literally those I described previously: Spearman did not calculate all the tetrad differences and compare their distribution around zero with the standard error of the distribution. Those calculations --in the days before computers—would have taken a great amount of time. Spearman worked out calculations that approximated results that would obtain if all the tetrad differences were zero, and that could be done in a reasonable amount of time. I can illustrate such calculations with Spearman's analyses to show "hierarchical order." Such an order occurs if the tetrad differences are zero. Correlations Spearman (1904a) presented to show this hierarchy are shown in Table 1.

Insert Table 1 About here

If the tetrad differences are zero, the partial correlation between any two variables will be zero when the general factor[v] is partialled out; then the correlation between the two variables will be equal to the product of the correlation of each variable with the general factor. Spearman recognized that it follows directly from this condition that the correlations in the same rows of any two columns of the matrix of intercorrelations will be proportional. Such proportionality indicates, he said, "a very remarkable uniformity" –what he called "The hierarchy of the intelligences:" If all the correlations are due to one common factor and they can be arranged in a table in order of magnitude --from the top left corner to the bottom right corner of the table, as seen in Table 1—then it will be seen that the correlations decrease in the same proportion in both vertical and horizontal directions throughout the table. This is what Spearman called a "perfectly constant Hierarchy. [vi]"

The subjects in the study on which table 1 is based were 22 boys[vii] "…in a preparatory school of the highest class, which principally trained boys for Harrow." There were six kinds of measurements. Four were said to be measures of "the intelligences." These were rank-orders of the sums of percentage-grades the boys received within courses in Classics (Latin and Greek), French, English, and Mathematics. There were two kinds of measures of sensory discrimination, one a measure of pitch discrimination, the other –said to be a measure of discrimination in music-- was the rank-order of the sum of percentage-grades the boys received in their music course.

Spearman described the hierarchy in Table 1 in this way:

> "…if we consider the correspondences between the four branches of school study, a very remarkable uniformity may be observed. English and French, for instance, agree with one another in having a higher correlation with Classics than with Mathematics. Quite similarly, French and Mathematics agree in both having a higher correlation with Classics than with English. And the same will be found to be the case when any other pair is compared with the remainder. The whole thus forms a perfectly constant Hierarchy in the following order: Classics, French, English, Mathematics. This unbroken regularity becomes especially astonishing when we regard the minuteness of the variations involved, for the four branches have average correlations of 0.77, 0.72, 0.70, and 0.67 respectively.

> When…we turn to the Discrimination of Pitch, we find its correlations to be of slightly less magnitude (raw) but in precisely the same relative rank, being: 0.66 with Classics, 0.65 with French, 0.54 with English, and 0.45 with Mathematics.…
>
> Just the same principle governs even Musical Talent…For it is not only correlated with all the other functions, but once again in precisely the same order: with Classics 0.63, with French 0.57, with English 0.51, with Mathematics 0.51, and with Discrimination 0.40.
>
> Altogether, we have a uniformity that is very nearly perfect and far surpasses the conceivable limits of chance coincidence …when we consider that the probable error varies between about 0.01 for the ordinary studies to about 0.03 for music…(Spearman, 1904a, pp. 274-275).

Spearman estimated the correlation of each variable with the common factor by dividing the sum of the correlations in each column by the square-root of the sum of all the correlations. He then ordered (from high to low) the variables in accordance with the squares of these correlations to show "the full absolute saturation of each variable with General Intelligence." (Spearman, 1904a, p. 276).

These, then, are the kinds of analyses, results and statements of proof that are referred to when the Spearman (1904a) paper is cited to support a claim that he did the first factor analysis and discovered general intelligence. They are not the crisp calculations I outlined when I argued that Spearman was the first to put forth a testable theory of intelligence, but in his arguments for "correspondences", displays of "very remarkable uniformity" and marshalling of results Spearman is, indeed, putting forth a form of common-factor analysis[viii] and presenting evidence in support of the principal hypothesis of his theory of general intelligence.

In a later paper Spearman (1914) would present a more mathematically rigorous argument for the proportionality of columns and hierarchical order. He would then argue, first, that if the column proportionality holds good throughout the correlation matrix, the correlation of a column of that matrix with any other column will be 1.0,[ix] and, second, that the "… theory now…possesses one of the most valuable characteristics in highest degree: the capability of being readily submitted to crucial quantitative verification." (Spearman, 1914; p. 108). He would use this criterion of near-perfect correlations among the columns of the correlation matrix in later studies in which he made the claim that one common factor described the data (Spearman, 1923). It his 1927 book Spearman applied the tetrad difference test in the manner I described in the first part of this chapter.

Today, the kind of calculations Spearman so arduously worked through to imperfectly estimate the tetrad-difference condition would be compressed into fitting -- and estimating the goodness of fit of-- a one-common-factor model. Once the raw scores were in the computer, this would take only a small fraction of a second using a computer with a program such as Mx (Neale, 1993). Results of such analysis for the correlations of Table 1 are shown in Figure 1. Here we see that an estimate of the probability that the model fits the one-common-factor model is 0.99 and the root mean square error of approximation (RMSEA, Browne & Cudeck, 1993; Steiger, 1990), which estimates departure from fit, rounds to zero. These two statistics indicate that one –and only one-- common factor very well accounts for the intercorrelations among the measures of "sensory discrimination and the intelligences".

Insert Figure 1 About Here

Criticisms and Reponses

Today the adjustments Spearman applied to the data –for example, estimating reliability from "arbitrarily assumed observational errors," adjusting correlations for attenuation due to unreliability-- would render his study unacceptable for publication in a scientific journal. But such procedures were not a principal concern in those early days in psychology. The concerns were with the ideas and the logic of the methods for garnering support for ideas.

Early Problems.

Burt (1909, 1911) appears to have been the first (in publication) to question the ideas and logic of Spearman's (1904a) study. His principal point was that the Spearman's results did not indicate that intelligence was general because important activities requiring mental effort were not considered in Spearman's analyses. Burt brought together and analyzed broader samples of cognitive abilities extracted a general factor -- in accordance with Spearman's stipulations—and found that one common factor would not account for the correlations: he had to calculate a second (numerical) factor and third (verbal) factor to do this.

In his answer to these criticisms Spearman (1914; 1923; Hart & Spearman, 1912) presented results that showed that the one-common factor hypothesis was retainable for correlation matrices on different particular sets of abilities. Burt (1924) replied with analyses of even broader samples of "indicators of intelligence" that showed that one common factor was not sufficient and that group factors of memory span, scholastic aptitude and manual skills, as well as the verbal and numerical factors were needed to account for the correlational data. In the years immediately following other studies by other investigators (Carter, 1928; Cox, 1928; Kelley, 1928; Patterson & Elliot, 1930) were presented to show that when samples of what were well-regarded as indicators of intelligence were analyzed, more than one common factor was indicated. Cognitive speed and visualization were added to the list of replicable group factors.

A general factor was always extracted in these early studies. Some reasoned that perhaps that general factor was the factor of Spearman's theory. But a flaw in that reasoning was spotted: the general factors of different studies were different. The factor had a different composition in each study. It came to be realized that a general factor could always be calculated, but this factor was simply a summary of the correlations among the collection of variables sampled in a particular study. There was no test of whether the general factor in one study represented the same phenomenon as the general factors in other studies. Simply calculating the general factor of each particular battery did test Spearman's hypothesis that one –the same one-- common factor was required in all mental activities.

Spearman's Major Response

In his much-cited book "The Abilities of Man: Their Nature and Measurement" Spearman (1927) marshaled a comprehensive response to criticisms of his theory. He brought together the empirical findings of his previous studies, but he also called in evidence and theory from different branches of psychology and a variety of other sources. He aimed: "… to set forth the conditions under which every ability is divisible into two factors, one universal and the other specific"(Spearman, 1927; p.87). He also described the nature, origins, development and correlates of the universal factor.

Comprehensiveness was a principal feature of this statement of his theory. Spearman tried to bring together all that was known in psychology to help us understand human intelligence. This 1927 statement, more than Spearman (1904a) or any of his other early works, most influenced subsequent research on human cognitive capabilities (a.k.a. human intelligence).

The theory Spearman produced owes much to Galton's thinking. But Spearman, much more than Galton, appealed to evidence of the field of psychology, generally, to give credence to his principal arguments. For example, what Galton had described as power of the mind, Spearman described as mental energy: but where Galton had left power of the mind largely as only a metaphor, Spearman used two chapters and major sections of other chapters to bring together evidence and arguments to show that the mental energy concept was well-based on findings and explanations of general psychology – findings that supported concepts of mental competition, fatigue, retentivity, conative control, and primordial potencies. His claim was that these various lines of evidence indicate, first, that there is a form of mental energy that "…is wanted to explain the general factor," and, second, that there is a system of "…mental engines that might go far toward explaining the specific factors" (Spearman, 1927, p. 135).

The mental energy of the general factor was said to be manifested in cognitive behavior, but Spearman suggested that underlying this mental energy is a neural energy that flows from throughout the brain and affects all abilities. He suggested that different neural systems serve the specific-factor engines.

This thinking is similar to that of a prominent neurologist of the time, Lashley (1929) --although Spearman (1927) makes no reference to Lashley, and Lashley (1929) makes no reference to Spearman. Lashley argued that although specific, somewhat different cognitive deficits are produced by injuries in different parts of the brain, the brain functions as a whole, so that an injury anywhere in the brain produces general cognitive decline. He spoke of a mass action of the neural system that –to some extent-- determines all intellectual capabilities.

Spearman's theory also specified that individual differences in the general factor were –to some extent-- innate. Again his thinking is similar to Galton's, but much more based on appeal to evidence. To arrive at the main points of this argument, Spearman reviewed findings from some 28 studies of mean ability differences for groupings of people classified in accordance with nationality, occupation, race, familial relationship, social class, gender, and educational training. He concluded that "…education has a dominant influence on specific abilities, but normally it has little if any influence in respect to the general factor" (Spearman, 1927, p. 32). On this point, he acknowledged that "… the question is, no doubt, in great need of further more exact investigation."

Emergence of the theory of g. Spearman's early theory was about intelligence – general intelligence. But increasingly in his writings over the years preceding his 1927 book he had expressed concerns about using the word "intelligence." It referred to too many things. Then, in 1927, he characterized it "as "cankered with equivocality…Chaos itself can go no farther! Disagreement …has reached its apogee…In truth, 'Intelligence' has become a mere vocal sound, a word with so many meanings that finally it has none" (Spearman, 1927, p. 14). He had used the letter g to stand for general intelligence in some of his earlier writings. Now he used the letter *g*, always italicized, to replace "general intelligence," and sought to define *g* in a way that would eradicate equivocality.

This *g,* he said, "…is primarily not any concrete thing but only a value or magnitude. Further, that which this magnitude measures has not been defined by declaring what it is like, but only by pointing out where it can be found. It consists in just that constituent –whatever it may be—that is common to all abilities that are inter-connected by the tetrad equation" (Spearman, 1927, p.75-76).

This did not immediately clean up the definitional muddle Spearman so deplored in the concept intelligence. To find *g*, one had to find "the abilities that are interconnected by the tetrad equation" and identify "that constituent --whatever it may be-- common to all" those abilities. Most important, one had to distinguish the *g* common factor from the specific factors that also account for individual differences in mental abilities. The specific factor of each measure of mental ability had to be identified because the tetrad-difference test of Spearman's theory required that there be no duplication of such factors in a battery of measures designed to test the theory. The muddle remained because investigators had different ideas about the abilities that should be interconnected by the tetrad equation; they had different ideas about the essential constituents of *g;* and they had different ideas about what constituted a specific factor. There was still much wrangling about where to point to find *g*.

Nevertheless, the Spearman definition of *g* provided a way clean up the definitional mess. It required experiments. Measurement devices had to be assembled in accordance with hypotheses that they inter-correlated in a manner such that the tetrad-differences to be zero. To meet Burt's criticism the devices would have to measure abilities that were accepted has indicating human intelligence. A single experiment, such as Spearman's (1904a) study, would not prove the point, but it would be evidence in support of the point, and if more studies answering to Burt-like criticisms, also proved the point, the theory of *g* would gain acceptance in the scientific community. In the process, the constituents of *g* would become clear. Spearman's definition of *g* was thus a call to action that could end the wrangling about the meaning of intelligence. Experiments would indicate *g,* the constituents of *g* and the nature of specific abilities that are independent of *g*.

Laying out the structural evidence. Spearman (1927) answered to that call. And he presented "proof that *g* and *s* (specific abilities) exist." "To the question, whether the divisibility of abilities into g and s (with s throughout independent) really occurs to any large extent …our evidence appears to have answered convincingly in the affirmative. Such two independent factors have been demonstrated for at any rate a great number of sets of tests commonly used for 'general intelligence'" (Spearman, 1927. p. 150).

The evidence to which Spearman referred was, first, inter-columnar correlations calculated on the correlation matrices of 14 studies conducted between 1889 and 1914: these correlations ranged between 1.16 and 0.89 and averaged 0.99. "All this evidence," he said, "lies beyond reasonable doubt…the inter-columnar correlation shows itself to be excellently satisfied. On the other hand, we must remember that this criterion itself leaves much to be desired…Let us turn to the genuine criterion, the tetrad difference… " (Spearman, 1927, p. 140). He then presented evidence indicated by the distributions of tetrad differences calculated for correlation matrices originally obtained by Simpson (14 measurement devices, 37 persons); Brown (8 devices, 66 persons); Bonser (5 devices; 757 persons); Holzinger (9 devices; 50 persons) and Magson (7 devices; 149 persons). In each case the median observed tetrad difference was nearly as small as, or smaller than, the probable error of the tetrad differences. The agreement of observed with theoretical errors of sampling the tetrad-differences was, Spearman said, "nearly perfect" and, moreover, the results obtained by in applying this correct criterion have fully corroborated the results obtained with the earlier inter-columnar correlation criterion.

So he concluded, first with a caution: "Science knows no finality…all conclusions drawn in the present work are subject to 'inevitable eventual corrections and limitations.'" But then with confidence: given " ...the degree of exactness attained already, the agreement of the observed values with those required by theory must be admitted by any unbiased person to have been surprisingly close. In general, it seems quite as good as, if not better than, that usually reached in determining the mechanical equivalent of heat and thus establishing the law of conservation of physical energy." (Spearman, 1927, pp. 159-160).

Dealing with discrepancies at hand. There were problems, however. Some of these were in the data Spearman presented. For example, the Simpson matrix was not positive definite: it had two negative roots. Given only the computing capability available to Spearman, this probably could not have been noticed. There were two common factors, not one, in the Holzinger data. The reported correlation matrix of Magson was asymmetrical. Three of the 8 factor coefficients of the Brown data were too small to reach statistical significance; the near-zero correlations for the variables would have contributed small tetrad differences, but these would mainly indicate the smallness of the correlations.

But again these were procedures problems --not attended to much at the time. Most of the problems with Spearman's claims resulted because other investigators carefully applying his test did not find what he found. He faced some of these problems head-on in his 1927 presentation of "proof that *g* and *s* exist." To explain why others testing his hypothesis did not get results consistent with his theory, he identified a number of conditions that could produce specific factor overlap, and thus yield results that would falsely falsify his hypothesis.

For results obtained on 1,000 army recruits with the 8 measurement devices of the Army Alpha, he argued that because the measures were obtained at 9 different camps, then if "…in any camp the testing or marking happened to be more generous than in others with respect to any of these tests, the men here would tend to shine in these particular tests; the result must be to generate additional correlation between these tests quite independently of any psychological connection between them." (Spearman, 1927; p. 157).

For results obtained on 2,599 members of the British Civil Service with 7 devices intended to measure "general intelligence," he argued that because some of the devices required selecting an answer from among several choices and some required "inventing" an answer, these two response forms produced specific factor overlap; thus test of the tetrad-differences seemed to invalidate the *g*-factor hypotheses when it was valid. He calculated the tetrad-differences separately for the devices that had the same response form and found that "…the agreement of observation with theory at once becomes admirable." (Spearman, 1927; p. 154).

For results obtained with six successive measurements with the Binet test-series at intervals of six months, he argued the "…tests change in nature as the age of the testee increases. Consequently, the tests for any two neighboring ages will have much in common that does not extend to ages farther apart." (Spearman, 1927, p.151).

In these ways Spearman mounted a general argument that whenever there were differences in test-battery administers or scorers, or diversity in age or training or sex or other such factors within the samples of testees, or diversity in measurement devices in respect to item form (as the above-mentioned selective or inventive response requirement), then specific factor overlap would occur, the distribution of the calculated tetrad-differences would not match the theoretical distribution, and the one-common factor hypothesis would be falsely rejected.

Downfall of the theory. In describing conditions under which the tetrad-difference test might fail when the one-common-factor hypothesis was valid, Spearman's did not deal with the results of Burt (1909, 1911, 1924) and others who were presenting similar findings at about the same time (Carter, 1928; Cox, 1928; Kelley, 1928). Prominent critics of his theory did not accept his post hoc arguments that the theory was correct even when the results of well-conducted tests did not support it. Faced with this dissention and recognizing the many, many ways in which the tetrad difference test might fail, Spearman (1927) set forth a different theory. He argued that *g* entered into every kind of ability measure and this was indicated by positive intercorrelations among these measures "…for the purpose of indicating the amount of g possessed by a person, any test will do …the most ridiculous "stunts" will measure the self-same *g* as will the highest exploits of logic or flights of imagination." (Spearman, 1927; p. 197). He referred to this argument for the "universality of *g*" as "the theorem of indifference of the indicator." In accordance with this theorem, evidence of positive correlations among abilities was sufficient to support his theory. This evidence was (and is) prevalent.

Spearman's argument from the "theorem of indifference of indicator" has been embraced by a preponderance of prominent scientists studying human abilities. For example, Jenson (1998) in his widely praised (Beaujean, 2002; Detterman, 1998; Miele, 2003; Nyborg, 2003) book made the "theorem" the centerpiece of his claim that Spearman discovered general intelligence. He used it to explain why different ability tests are positively correlated, why positive intercorrelations among ability tests are prevalent, why a g factor must be found at the top in higher-order multiple-factor analysis, and why, therefore, the theory of general intelligence is basically correct. These arguments have been put forth by other prominent investigators (e.g., Carroll, 1993; Eysenck, 1982; Gottfredson, 1997).

But this theory is a step down from the scientific theory of *g*. Indeed, it is, at best, barely a scientific theory. It is true that in broad samples of people, almost all tests regarded as measuring any aspect of human intelligence are positively correlated, and the rare exceptions --for samples of very young children (Bayley, 1969) and for the correlations between highly speeded tests at low levels of difficulty and unspeeded tests at high levels of difficulty (Guilford, 1964)—might be written off as unimportant or explained as in some way compatible with the hypothesis. But it is true, also, that practically every variable that in any sense indicates the good things in life –health, money, education, etc., and athletic and artistic abilities, as well as cognitive abilities—correlate positively with every other such thing[x]. Granted that some (Herrnstein & Murray, 1994) might take such findings evidence that general intelligence is truly universal, there is no falsifiable test that distinguishes this theory of the "good things" from the theory of *g*. There is no way to identify the constituents of *g* and distinguish them from what is not *g* –the specific factors. The theory has the same character as the theories Spearman had earlier so roundly (and soundly) criticized: *g* has so many meanings that finally it has none.

Indeed, the so-called "theorem of indifference of indicator" is not a theorem at all. If it were true that when all intercorrelations among variables are positive, there must be one common factor –a factor that might have to be found the top of the order in higher-order multiple factor analysis— it could be said that this was proof of the theorem. But this is not true. This was demonstrated by Thomson (1919) a few years after Spearman's theory came into prominence.

The Thomson-Thorndike theory. Positive intercorrelations among all of a set of variables can be indicative of two or three or four or more common factors; it need not be indicative a single higher-order common factor. Indeed, Thomson argued that positive intercorrelations among abilities is just as compatible with a theory of many common factors as it is with a theory of one common factor. He showed that if a measurement device --Device 1, say-- involves elementary processes a, b, c, and d, for example, then if Device 2 involves processes a, e, f, and h it can correlate with Device 1 in virtue of sharing process a (but not b, c, d, e, f, or h). Similarly, if Device 3 involves processes b, e, i, j, and k, it can correlate with Device 1 in virtue of sharing process b, and it can correlate with Device 2 in virtue of sharing process e, (and not sharing any other processes with Tests 1 or 2). Device 4 can share process c with Device 1, process f with Device 2, and process h with Device 3 and thus be positively correlated with all three tests without sharing any processes common to all four tests. Continuing in this way Thomson showed that all the devices of a battery of measurement devices can be positively correlated and not involve a single common factor.

Thomson's demonstration was prompted by a theory of many factors of intelligence that was put forth by E.L. Thorndike in 1903, the year before Spearman's theory was first presented. In line with Thorndike's thinking, Thomson argued that performance on any cognitive measure can be seen to involve many processes of perception, apprehension, retention, association, reasoning, reflection, retrieval, etc. and that such processes can be configured --organized, called forth, applied, and expressed-- in a great variety of different ways, and these different configurations can overlap and be shared in an even larger number of ways to produce the performances on the different measures that make up a battery of positively correlated cognitive devices.

Two conditions observed in many studies of cognitive abilities --positive manifold of the intercorrelations among samples of ability variables and the varying composition a general factor among these variables -- are consistent with Thomson's theory of human abilities, but not with Spearman's theory of one common factor. Thomson's theory does not claim that there is one general factor that pervades all cognitive abilities[xi]. Several investigators of the last century, Humphreys (1971) most prominent among them, favored this kind of theory of human cognitive capabilities.

Attempts to Retain g-theory

g-theory did not go quietly into the night. There were many efforts to retain it. Recognizing the validity of Spearman's argument that overlapping specifics would spoil tests of the one common-factor hypothesis, Alexander (1935), Brown (1933), Brown and

Stephenson, (1933), El Koussy (1935), and Rimoldi (1948) designed studies in which the assembled tests were thought to at once measure the important abilities of intelligence and not introduce specific factor overlap. These efforts failed. In each case several common factors were required to fit the data. This was the verdict of almost all factor analytic studies conducted from the 1930's onward (summarized in Carroll, 1993; Ekstrom, French & Harmon, 1979; Hakstian & Cattell, 1974).

Such evidence did not end the matter in Spearman's day and it has not ended the matter even today. A theory that humans differ in an innately-determined general intelligence is widely and strongly believed. Belief in the theory is entrenched in our culture and language. It is not simply Spearman's theory; it is the theory of many. Such belief in the basic correctness of the theory has charged efforts to retain it even in the face of mounting evidence of its inadequacy. It is instructive to consider these salvage attempts.

The high correlations argument. It is sometimes argued that the correlations among different devices said to measure intelligence are large and this indicates that *g* is operating throughout the devices.

There is no merit to this argument. It is merely a version of the-correlations-are-positive argument. It is consistent with Thomson's theory, not *g*-theory. The test of g-theory does not require that correlations be large. What constitutes a "large" correlation is a question for debate or for practical need to predict, not a question required of *g*-theory. By almost any reasonable criterion, correlations between devices thought to measure intelligence are not always large –even when adjusted to estimate attenuation due to unreliability: they are in the neighborhood of from .15 to .80. The only refutable hypothesis of the argument –that a correlation can not be smaller than, say, x—is arbitrary. Different sum-of-abilities measures of intelligence –the Binet and Wechsler devices, for example—correlate highly because they measure the same things, not because they measure the same thing. The "high correlation" argument is baseless.

The higher-order common factor argument. It has been argued that if positively correlated ability measures are factored to yield common factors at a primary level, and the primary common factors are factored to yield second-order factors, and these are factored to yield third-order factors and factoring continues in this manner until only one factor is indicated at the kth-order, that one factor is evidence of *g*.

This again is merely a version of the-correlations-are-positive argument. There is no refutable hypothesis: a higher-order factor can always be computed. In practice, the one factor determined in this manner in one study is not the same as the one factor found in other studies. If a battery contains many reliable tests measuring spatial abilities, for example, the one factor calculated at the highest order has its highest correlations with spatial abilities. If the battery is comprised largely of reliable measures requiring verbal comprehension, the one factor at the higher-order is a verbal comprehension factor.

This is well-illustrated in Carroll's (1993) tour de force re-analysis of some 477 batteries of ability measuring devices. In Table 15.4 of his monograph Carroll identified "…53 factors, in 146 datasets, classified as measuring "general intelligence" or possibly Spearman's factor *g*." (p. 591). Inspection of these results indicates that often the general factors of different analyses contained no measurement devices (no abilities) that are the same; the factors were thus "general" in respect to entirely different sets of variables. When the batteries contained some of the same variables, the "general" factors were different: one was a “general” visualization factor; another was a “general” verbal comprehension factor; and so on. Even when batteries were made up of the same measurement devices, the order of the correlations with the general factor were notably different: the factors did not pass the test of metric invariance (Meredith and Horn, 2001). To argue that these kinds of results indicate general intelligence, one must appeal to the discredited “theorem of indifference of indicator” – essentially, assume what one is trying to prove. Carroll’s findings do not support a theory of *g*.

Hierarchical analysis with test for one common factor. A higher-order factor analysis converging on one factor at the top would provide evidence in support of Spearman’s theory if three conditions were met: (1) the factor intercorrelation at the next-to-highest-order satisfied the one-common-factor (rank-one) requirements of the Spearman model, (2) the factor at the highest order was invariant across different samples of people and different occasions of measurement, and (3) the set of analyzed variables – hence the factors at the various levels— included a good complement of the abilities indicating human intelligence. If these conditions were met, there could still be questions about whether important indicators of intelligence were left out of the sample of abilities, but one would need to identify such indicators to give credence to the claim and discount the evidence. Until then, the results would be supportive of a general-factor theory.

A study reported by Thurstone and Thurstone (1941) seemed to have met these conditions. This study was based on research in which L. L. Thurstone (1935; 1938) put forth a multiple factor theory to describe the cognitive abilities of intelligence. Rather than specifying a single common ability (Spearman) or overlapping bondings of many, many abilities (Thomson), Thurstone specified a relatively small number of primary abilities, each of which could be identified as a common factor among several (three or four) different exemplar measures of the ability. A test of one major hypothesis of the theory was in principle the same as the test of Spearman's theory: the rank of the matrix of intercorrelations had to equal the number of hypothesized common-factors. In his landmark studies Thurstone (1935; 1938) gathered broad samples of the abilities of intelligence and, in multiple-factor analyses, identified primary abilities of verbal comprehension (V), word fluency (W), number facility (N), spatial thinking (S), associative memory (M), perceptual speed (P), general reasoning (R), inductive reasoning (I), and deductive reasoning (D).

In the study that seemed to support Spearman's theory Thurstone and Thurstone (1941) tested the one-common-factor hypothesis at the second-order among primary abilities. The rank of the matrix of intercorrelations among the V, W, N, S, M and I primary factors was found to be very close to unity. This was consistent with Spearman's theory. Unfortunately, the results were obtained for a trimmed battery –only 6 of the 9 primary abilities had been analyzed. When another --or other-- primary abilities were included in the battery, the rank-one condition no longer obtained. Thus, the finding was that the general factor was not truly general.

Another excellent higher-order study (Gustafsson & Undheim, 1996) is often cited as indicating support for Spearman's theory of *g*. The study was well-designed to represent a broad sampling of the abilities accepted as indicating human intelligence. The findings indicated that at the second-order in analysis among primary abilities there were several broad factors representing different forms of intelligence. Gustafsson & Undheim interpreted one of these second-order factors as indicating fluid intelligence, Gf, an ability I will describe in some detail later sections of this chapter. Gustafsson and Undheim allowed the second-order factors to be correlated, and calculated a single factor at what they regarded as the third-order. They found, however, that they could rotate this "third-order" factor into perfect alignment (correlation = 1.0) with the second-order Gf factor. The third-order factor was interpreted as *g*.

Thus, the finding seemed to be that *g* is equivalent to Gf and, because the third-order factor accounts for the correlations among the second-order factors, the factor interpreted as *g* accounts for the components held in common by all the abilities and thus is a general factor consistent with Spearman's theory. There are several problems with this interpretation of the findings:

First, because the third-order factor correlates perfectly with a second-order factor, there really is no need for a third-order factor. The second-order factors account for the primary factor intercorrelations just as well without, as with, the third-order factor. Rotation of the second-order factors could just as well have been orthogonal, for example. The third-order factor is simply a different way of summarizing the second-order findings.

Second, granted that it is informative to describe the second-order factors with an oblique solution and to condense the intercorrelations thereby introduced into a third-order factor, that third-order factor could be aligned with any one of the second-order factors, not simply Gf. It could just as well be aligned with a second-order factor Gustafsson & Undheim interpreted as crystallized intelligence, Gc, for example[xii].

Third, any second-order factor chosen to absorb (summarize) the intercorrelations would still be independent of the other second-order factors. Such a factor is identical to the third-order factor, but not identical to the other second-order factors: it does not account for them. The evidence thus indicates that several second-order factors are required to account for the primary factor intercorrelations. One factor will not do it. The findings indicate that Gf (or *g* if we prefer that label) is not a general factor: it is but one among several second-order factors required to describe the correlations among primary ability indicators human intelligence.

Thus, we learn from the Gustafsson & Undheim (1996) study what we learned from the Thurstone & Thurstone (1941) study –namely, that one-common-factor does not describe the intercorrelations among primary abilities of human intelligence.

There is another lesson we can learn from Gustafsson & Undheim that we do not learn from Thurstone & Thurstone. This is that there are several intelligence-like factors at the second-order among primary factors. Only two such factors were required in Thurstone & Thurstone, but the Gustafsson & Undheim results indicate that more than two “intelligences” are needed to describe broad samples of the human abilities. I will look into this matter in a later section, after we consider other major parts of Spearman’s theory.

Spearman’s Theory of Processes

Although he railed against doing it in some of his writings, Spearman developed several important ideas about processes of intelligence. Indeed, second only to his development of a test for a common-factor hypothesis, these ideas have most influenced research on human cognitive capabilities. It is interesting, too, that these ideas contradict some of Spearman’s most forceful criticisms of the theorizing of his contemporaries.

In his work on processes Spearman went well beyond merely pointing to where *g* "can be found." He put thought into identifying the particular behavior that indicates *g* and distinguishes it from specific factors. He tried to define *g* by declaring what it is like, contrary to what he advised when he criticized others' theories of intelligence.

Spearman described "what *g* is like" with what he called "laws of behavior." By "laws" he meant regularities established by experiments. The laws he was concerned with were intended to describe "the entire range of possible operations of knowing." Although the laws were presented as statements of fact, we would see them today as hypotheses about processes such as are analyzed in studies of cognitive psychology. The laws called for operational definitions of constructs which should, if *g* theory is correct, relate to each other in the manner described in the theory.

The "noegenetic[xiii] laws" were at the core of the theory. With the term "noegenetic" Spearman referred to capacity for creating understanding and building knowledge out of what is sensed, perceived and comprehended. Three noegenetic laws were said to account for this capacity: the law of apprehension of experience, the law of eduction of relations, and the law of eduction of correlates.

In the law of apprehension of experience Spearman argued that in order to think in a way that would indicate g –and show it in the behavior of attempting to solve a problem-- one must first sense and perceive the fundamental features –the fundaments-- of the problem. With the law of eduction of relations Spearman argued that there are relations among the fundaments of a problem, and one must comprehend these in order to make a response (in attempt to solve the problem) that can indicate a magnitude of g. With the law of eduction of correlates, he argued that for a person to make a response that indicates a quantity of *g*, that person must extrapolate or interpolate or generalize to infer a not-immediately-educed relation from the evidence of the extant relations.

The laws of eduction of relations and correlates were said to operate with all of several different kinds of relations –relations of conjunction, space, and time; relations of causation, constitution, attribution, and identity; relations of evidence, likeness and conjunction; and relations among psychological concepts. To illustrate operations of measurement that would indicate g, test items were constructed --or taken from the constructions of others-- to show each of the relations of eduction and to indicate how the relations operated in eduction of correlates.

An important departure from earlier theory. The law of apprehension of experience fell by the wayside in Spearman's definitive statements of his theory: individual differences in *g* were said to be primarily a function of the two eductive processes. De-emphasizing the apprehension law in his definitive theory is a major departure from Spearman's original theory. In this latter, as I mentioned earlier, he argued strenuously for Galton's hypothesis that keenness of sensory discrimination indicates quality of intellect. He had reported in 1904, for example, that "…we arrive at the remarkable result that the common and essential element in the Intelligences wholly coincides with the common and essential element in the Sensory Functions," (Spearman, 1904a, p.269). But this "remarkable result" was obtained only after many adjustments of the data and corrections of correlations for attenuation due to unreliability of the measures. The raw, uncorrected correlations were small, as they had been in Galton's study, and as they were usually found to be in the studies of others. In his corrections of correlations Spearman "arbitrarily assumed observational errors" (his words) that gave him low reliabilities, which in the denominator of corrections for unreliability, gave him unrealistically large and mathematically impossible “corrected” correlations. Spearman probably would have noticed such unbelievable results. This, plus the persistent finding of near-zero correlations for the apprehension-of-experience measures, very possibly led him away from his earlier conclusion that sensory discrimination is a major feature of intelligence.

In any case, in his later writings Spearman relegated apprehension of experience and keenness of sensory discrimination to minor roles in the theory of *g*. Apprehension was still seen to be part of the processing of *g* --it provided a foundation for educing relations and correlates-- but the level of apprehension achieved by most people –excluding only the retarded-- was sufficient for exercise of the other noegenetic processes; these other processes were seen as the essential processes that indicated individual differences in magnitudes of *g* in studies of normal people.

Further contributions. This account of how apprehension of experience relates to abilities regarded as central to human intelligence is accepted in contemporary theory. Elementary sensory processes relate at only a low level to the reasoning, acquisition and retention processes of human intellect. Other features of Spearman’s theory also have become part of contemporary scientific theory of human cognitive capability.

Notable among these “other features” are Spearman’s ideas about speed of thinking. Spearman had found that quickness in educing relations and correlates --when measured in a particular person and problem-- is in competition with the quality of that thinking: slower (more thoughtful) thinking usually is associated with better, more nearly correct thinking. But Spearman found, also, that in analyses of between-person differences, quickness in educing relations and correlates was often positively –though only lowly-- correlated with quality of that thinking. "On the whole", he said, "…*g* has shown itself to measure a factor both in goodness and in speed of cognitive process. The connection between the goodness and the speed is that of being inter-changeable. If the conditions …eliminate the influence of speed, then *g* measures goodness, and vice versa. When –as is most usual—both influences are at play, then *g* measures the efficiency compounded of both." (Spearman, 1927, p.138). It is now generally recognized that a speed-accuracy trade-off operates within a person and that individual differences in speed and quality of thinking are usually positively –though lowly-- correlated (e.g., Salthouse, 1985, 1991 for review).

Also imbedded in contemporary theory is Spearman's account of a relationship between what he called "intensity of thinking" --which today is referred to as the level of difficulty of problems solved-- and "extensity of thinking," which today is called working memory span –that is, the ability to hold information in awareness while doing other things such as solving a problem that requires the retained information. According to Spearman, "both the intensity and extensity of cognitive operations depend on *g*… the two constitute alternative dimensions of the same constant cognitive output characterizing each individual." (Spearman, 1927; p. 269). It is now generally recognized that working memory span is closely related to ability to solve reasoning problems of a high level of difficulty (e.g., Baddeley, 1994).

Spearman's findings and theory in regards to recall memory have also become a part of contemporary thinking about short-term memory –the ability to remember for a few seconds items for which one has no organizational scheme. The human can retain for a short time (less than 60 seconds without rehearsal) only about 7 unrelated items – with individual differences ranging generally from about 9 items to 5 about items. If there is a momentary distraction, the memory is lost. Usually, for example, we can retain a 7-digit telephone number long enough to dial it, provided no one asks us a question when we are trying to do the dialing.

Spearman recognized in his studies that such memory is a lower-order process relative to the eductive processes of *g* –in his words "…memory correlates with measures of *g* to an amount close upon .30…the memorizing even of sentences and passages has only a medium correlation with *g.* And in proportion as the material to be learnt becomes either unrelated or sensory –so that the influence of eduction whilst learning diminishes—the correlation with *g* dwindles down towards the point of disappearance." (Spearman, 1927; p. 280). This is essentially the modern-day view of the way short-term memory is related to what is known as fluid reasoning (Flanagan et al 1997; McArdle & Woodcock, 1998; McGrew, Werder & Woodcock, 1991).

Restructuring the Theory

The Thurstone & Thurstone (1941) study showed that a set of important abilities of intelligence can be found to fit the one-common-factor model. This was indicated, also, in the Spearman studies to which we have referred. In these studies Spearman and his coworkers carefully selected variables that would fit a one-common-factor model, and they sometimes explicitly dropped from their analyses variables that spoiled such a fit (Brown and Stephenson, 1938; Hart & Spearman, 1912; Spearman, 1927). The findings thus suggested that for some sets of abilities, Spearman's test of his theory applies, but for other sets of abilities, it does not.

Spearman's unintentional two common-factor theory. The sets abilities found to fit the model in Spearman's work were well accepted as indicating human intelligence. But some of the abilities that didn't fit were also abilities that at least some who were studying human abilities regarded as indicative of human intelligence. For example, Spearman found that "…general information turns out to measure intelligence very badly indeed…[and] …scholastic tests do not appear to have manifested any correlation with *g* except insofar as they involve eduction, either at the actual testing, or during the antecedent learning. …there is nothing to indicate that *g* has any correlation with pure retentivity. ..the available evidence indicates that g is exclusively involved in eduction and not at all in bare retention…" Spearman, 1927; pp. 277,278; 290).

Thus, Spearman, in effect, reduced his claims for *g*: it was not truly general, for it did not account for "pure retentivity," "bare retention" and "general information" – surely activities requiring mental effort. Indeed, activities that some investigators regarded as among the best indicators of human intelligence. For example, general information is a principal component of the Stanford-Binet, Wechsler and other measures of intelligence.

Thus, with these kinds of observations, Spearman introduced a theory of two intelligences, although, granted, the introduction was rather oblique and unintentional. A theory of this form was later developed by Raymond Cattell, who, as a student, had worked closely with Spearman.

Other evidence of a two- common-factor theory. At about the time Spearman was noticing a distinction between *g* and "pure retentivity" other investigators in entirely different lines of research were noticing that some abilities of intelligence --what I will now call Gf abilities-- declined irreversibly with brain damage and with aging in adulthood, but other abilities of intelligence --which I will call Gc abilities—did not show this decline pattern. Interestingly, the Gf abilities were described in much the same way as Spearman described the processes of *g*; and the Gc abilities were those of retained general information and scholastic knowledge, very like the "pure retentivity" abilities that Spearman said were not good indicators of *g*. In studies of brain damage, for example, Bianchi (1922), Feuchtwanger, (1923), Kubitschek (1928), Dandy (1933), Weisenberg & McBride (1935) and Rowe (1937) reported that pathology (stroke) and surgery in the adult brain produced very little or no enduring loss of abilities of knowledge, verbal facility, fluency, and everyday judgment –even after an entire hemisphere of the brain had been removed (Rowe, 1937)-- but in the same person the pathology or surgery produced profound and lasting, seemingly irreversible, loss of ability to understand and reason with complex novel relationships. Similarly, in studies of aging, Willoughby (1927), Jones, Conrad & Horn (1928), Babcock (1930), Miles (1934), Christian & Paterson (1936) reported that older adults did as well as, or better than, younger adults on tests that measured knowledge and verbal facility, but performed more poorly than younger adults on measures of logical reasoning when the relations that had to be comprehended were not such that the person could have studied and used them at prior times.

These findings led Cattell (1941), looking primarily at the age-differences research, and Hebb (1941), reviewing the brain-injury findings, to propose that there must be two broad intelligences --neither a general intelligence. Cattell coined the terms "fluid intelligence" to describe the Gf abilities that declined with age and brain damage, and "crystallized intelligence" to describe the Gc abilities that did not irreversibly decline with brain damage and improved with age in adulthood. He developed a comprehensive theory to describe the development and effects of these two intelligences (Cattell, 1963).

Gf and Spearman's process theory of g. Cattell's concept of fluid intelligence borrowed heavily from Spearman's process theory of *g*. The close similarity between these two concepts can be seen in the following descriptions of variables sampled to indicate Gf in a series of factor analytic and developmental studies (Horn, Donaldson & Engstrom, 1981):

Variable Description	Spearman Process
1. Span of apprehension. Measured with an adaptation of the Sperling (1960) paradigm.	Immediate apprehension span -- awareness of fundaments.
2. Primacy Memory . Measured as recall of first two elements of a series of elements.	Retaining fundaments in awareness.
3. Working Memory. Measured as recall in reverse order of a series of elements.	Maintaining awareness. Required to educe relations.
4. Comprehension of conjunctions. Measured with power letter series.	Eduction of relations among fundaments.
5. Drawing inferences. Measured with remote associations.	Eduction of correlates.
6. Focused attention. Measured with slow tracing.	??? Capacity for concentration.
7. *Carefulness,* Measured with few incorrect answers on several tests.	Pervasiveness of capacity for Apprehending experience

The intercorrelations among these indicators of Gf satisfy the requirements of the one-common-factor model (RMSEA[xiv] = .067). We have found comparable approximations to the model with other combinations of tasks designed to indicate Gf.

Thus, Gf and *g* are similar constructs. This is suggested in other work. For example, in citing evidence in support of *g*-theory, researchers often refer to evidence indicating Gf (e.g., Jensen, 1998). Devices that are referred to as providing good measures of *g* --matrices and topology— are found to be good marker variables for identifying the Gf factor (but not good indicators of the Gc factor).

Emergence of multiple-abilities theory.

Thus, a theory of two common-factors of intelligence emerged, partly as a result of Spearman's work. But the accumulating evidence often suggested even more than two factors.

We have seen that from Burt's (1909) efforts onward, results from attempts to validate Spearman's theory repeatedly indicated several common factors among indicators of human intelligence. In this early work there was first an attempt to fit Spearman's model and when that failed, group factors were calculated to account for the residuals left after the general factor was partialled out. As investigators learned about centroid analysis and principal component analysis, these methods were used in place of Spearman's methods. In any case, the method of calculating factors begged the question of a general ability factor by simply calculating it; and the method required that other abilities –group factors-- be left-overs, residuals of what was not accounted for by a general factor.

The results of such studies were highly unstable. Not only was the general factor of one study not the same as, and often very different from, the general factor of other studies, the residual factors were contrasts between different sets of abilities and the contrasts were often very different from one study to another.

The contrasts, necessitated by the methods[xv], sometimes seemed to make pretty good sense. For example, the variables correlating positively with the second factor could be verbal abilities when the variables correlating negatively were mathematical abilities, and one could reason that in school, after the general factor was taken into account, students tended to get sorted into those who liked and studied literature (history, etc.) – hence showed well on measures of verbal abilities-- and those who liked and studied mathematics (the physical sciences. etc.), and thus evinced good mathematical abilities. The problem was that the factoring methods that were produced results were not replicable. In studies in which there were somewhat different samplings of variables, instead of getting the interpretable result of verbal abilities contrasted with mathematical abilities, for example, the results would indicate a second factor in which verbal abilities were contrasted with spatial abilities or a factor in which mathematics abilities were contrasted with speeded abilities. There was little stability in the results of different studies.

There were attempts to rectify these problems, while retaining the idea that a general factor had to be calculated. The bi-factor theory and method of Holzinger (1934), particularly, in which the group factors were specified and calculated, tended to remedy the problem of the instability of the group factors, but it compounded the problem of the instability of first factor.

Tryon (1932; with cluster analysis) and Thurstone (1931; 1947; with simple-structure analysis) developed methods that did deal with the instability problem. Their methods simply did not require a first, general factor. Thurstone's method became the method of choice of most researchers. Results from use of these methods were often stable from study to study under conditions in which the results from using unrotated centroid and truncated principal components analyses were quite unstable.

Thurstone's method served a new theory of human intelligence –a theory of multiple primary mental abilities. This theory had two principal tests. It required, first, as mentioned before, that the rank of the matrix of intercorrelations be equal to the number of common factors specified in theory, and, second, that a particular structure be specified for the factor coefficients, and that this structure be found for the factors rotated to meet the separately defined criteria of simple structure.

The concept of simple structure stemmed from reasoning that no influence in nature affects everything; hence no factor should affect all variables. And no variable is affected by all the influences in nature; indeed, most variables should be affected by only a few factors. As with Spearman's theory, Thurstone's theory required experiments that were well-designed to pass the two tests.

The simple-structure test was particularly difficult to formulate. In any particular study there could be a number of common factors each expected (by hypothesis) to influence particular specified sets of variables. The sets –i.e. the "small" number of variables affected by a factor-- could vary; communalities of the variables could vary; the number of factors could vary; and the factors need not be orthogonal. This made it very difficult –perhaps impossible-- to write general mathematical requirements for simple structure for all cases. Several mathematical/statistical procedures for attaining approximations to simple structure were developed (often sounding like nostrums from a pharmacy –Varimax, Oblimax, Equamax, Promax).

The beauty of these methods (as seen from a scientific perspective) was that they defined particular structures, each different from the other, quite independently of any substantive theory. They specified mathematical models –meta-theories— a scientist could attempt to fit by appropriate sampling of variables and subjects. For example, the Varimax model required one fairly broad factor (but not a factor as broad as the general factor of a centroid and principal axis models) along with several less-broad factors, whereas the Equamax model required that all factors be equally broad.

The main problem with the models was that they did not restrict just how broad was broad or how equal was equal with a statistical tests that indicated departures from chance. For this reason they came to be called "exploratory" methods, although, as noted, they were distinctly not exploratory in requiring experimenters to design studies in accordance with an objectively-defined model.

With the advent of structural equation modeling, the hypotheses of a specified simple structure (or any other structure) could be tested and regarded as either tenable or not at a designated alpha level.

Thurstone's theory of cognitive abilities required specifying how each sampled variable was affected by each primary ability, and how each primary ability was distinguished from other primary abilities, from specific factors and from error. Ideally in most cases, variables were expected (by hypothesis) to relate primarily to only one primary ability factor, although in some well designed cases a variable was selected to indicate the influence of two primary abilities. Thus, the factor structure would be simple because most of the correlations between variables and factors would zero (chance-like): variables would correlate primarily only with one factor, and factors would correlate primarily with only the few variables.

Chaos rendered by a plethora of primary abilities. The promise when Thurstone's theory first appeared was that experiments would establish a boundary for the number of primary abilities needed to describe and understand human intellectual capacities. As we have noted, Thurstone's (1935; 1938) initial experiments suggested that this number might be about nine. To the dismay of many, however, further applications of Thurstone's logic and methods with different batteries of variables turned up many more than 9 primary abilities. From the 1940's onward dozens of studies of the common factors among tests thought to measure important features of human intelligence produced dozens of factors regarded as indicating primary mental abilities. Summary studies identified replications of, first, over 40 such abilities (e.g. Ekstrom et al, 1979; Hakstian & Cattell, 1974), and then, as results from more studies rolled in, over 120 primary mental abilities (Carroll, 1993; Guilford, 1967). It came to be recognized (Humphreys, Ilgen, McGrath & Montanelli, 1969) that, depending only on the number and ingenuity of scientists who might construct mental ability tests --and most could make up a new test every morning before breakfast-- the number of primary abilities is arbitrarily large. Thus, it seemed that research on the nature of human intellectual abilities had come a full circle, back to where it was before Spearman (1927) moaned that "chaos itself can go no farther".

Salvation through construct validation at the second-order. Thurstone (1947) had pointed out that identifying a common factor is only an initial form of the evidence that is needed to build scientific understanding: it is necessary also to establish a network of lawful relationships a factor has with other variables. These relationships form the basis for the explanatory framework that defines a scientific theory. This network of relationships prescribes the construct validity of a factor.

There appeared to be little hope of building construct validities for the multitude of primary ability factors that grew up in the aftermath of Thurstone's research. But the distinctly different relationships to brain damage and aging in adulthood that had been found for Gf and Gc factors suggested that construct validities might be established for factors identified at this broader, second-order level. Evidence has gradually accumulated to support this supposition. It indicates that second-order factors among primary abilities relate not only to the Gf and Gc distinctions, but also to distinctions between visual and hearing processes, and to concepts of immediate memory and consolidation in memory for which relationships have been established through the controlled-manipulative experimental research of cognitive psychology. Second-order factors thus appear to provide at least a beginning basis of operational definitions of constructs for building a scientific theory of human cognitive functioning.

The second-order factors were first identified in studies largely aimed at indicating the nature of the Gf and Gc factors. However, as I indicated earlier in discussing the Gustafsson & Undheim (1996), the findings from these studies suggested that the two-common-factor theory of intelligence needed to be extended to a several-common-factor theory. The results from such studies came to be referred to as indicating an extended Gf-Gc theory. Usually 8 or 9 such factors were indicated. In analyses based on practically all the factor analytic studies done up to about 1990 –some 477 batteries of cognitive ability measures-- Carroll (1993) identified 8 factors at the second order. These factors have now been described, and their construct validities discussed, in wide variety of articles and books (for example, Carroll, 1993; Flanagan et al, 1997; Flanagan & Harrison (2005); Horn, 1968; McArdle & Woodcock, 1998; McGrew et al, 1991; Perfect & Maylor, 2000; Schaie, 1996; Woodcock, 1995). Here I'll describe the factors only briefly in what follows immediately, and then, in later sections, skim-off some major indications of their distinct construct validities.

Fluid reasoning (Gf), the factor that most resembles Spearman's concept of *g*. It indicates capacities for identifying relationships, comprehending implications and drawing inferences in novel content.

Acculturation knowledge (Gc), the factor that represents Spearman's ideas about variables that do not provide good indications of g, particularly "general information" and "bare retention." It indicates breadth of knowledge.

Fluency of retrieval from long-term storage (TSR), a factor that indicates consolidation in memory and association memory over long periods of time.

Short-term apprehension and retrieval (SAR). This factor indicates a capacity for maintaining awareness stimulus elements for a span of a minute or so.

Visual processing (Gv), abilities of visual closure, maintaining visual constancy and fluency in recognizing the way objects appear in space as they are rotated and flip-flopped in various ways.

Auditory processing (Ga), abilities of perception of sound patterns under distraction or distortion, maintaining awareness of order and rhythm among sounds, and comprehending groupings of sounds.

Processing speed (Gs), although involved in almost all intellectual tasks, this factor is Indicated most purely in rapid scanning and comparisons in intellectually simple tasks in which almost all people would get the right answer if the task were not highly speeded.

Correct decision speed (CDS), measured in quickness in providing answers in tasks that are not of trivial difficulty.

Also, although broad quantitative knowledge was not identified as a separate second-order factor in Carroll's (1993) mega-analysis, such a factor has construct-validity relationships that indicate that it is quite distinct from Gc and other second-order factors.

These broad factors, including quantitative knowledge, although positively correlated, are operationally independent and have predictive independence, as well as independence in virtue of having distinct construct validities. Predictive independence is indicated by evidence that a best-weighted linear combination of any set of eight of the factors does not account for the reliable covariance among the elements of the ninth factor. This means that the beta-weight for each of the nine factors can be significant in the prediction of complex criteria.

Although there are suggestions that some of the second-order factors are more related to genetic influences than others, the broad patterns do not represent a clean distinction between genetic and environmental determinants. (Such distinctions appear to be better made at the primary level or below).

Evidence Pertaining to Construct Validity

In previous sections of this chapter I considered details of the factor analytic evidence pertaining to Spearman's *g*-theory. I did this because claims for support for that theory have appealed primarily to structural –i.e., correlational, factor analytic— evidence. But the evidence that most surely indicates the there is not one general factor that pervades all cognitive abilities is really a lack of construct validity for a general factor and the presence of at least some construct validity for the second-order factors. That is, studies of development throughout the life span, of neural function and brain damage, of behavior genetics and environmental influences, and of prediction of criteria most often indicate that clumping abilities together in a single general composite intended to represent IQ or *g* obscures distinct relationships that can be seen if second-order factor measurements are used in place of general factor composites. There is space here to provide only a bird's eye view of this evidence.

Developmental evidence. The evidence I referred to earlier --indicating that Gf and Gc abilities have quite different relationships to age in adulthood— is dramatically indicative of how considering only a general factor obscures relationships that can be made clear : Gf declines while Gc increases over age; if the two are clumped together in a general factor, these distinct relationships are not seen.

Studies of childhood development also indicate a distinction between the two forms of intelligence. Gf and Gc can be identified as distinctly different abilities as early as in the third year of life, at which age the correlation between the two is approximately .65 when the internal consistency reliabilities of the factors are approximately .90. The correlation between the two becomes smaller at successively later stages of development.

In adulthood the correlation is found to be in a range of about .40 to .50 (with factor internal consistency reliabilities in a range of .80 to .90). See Horn (1991) for reviews.

Gc correlates in a range of .40 to .60 with the educational or economic level of one's parents and with one's own educational or economic level at later ages. It correlates also with other indicators of social class. Gf, in contrast, correlates in a range of .20 to .45 with these same indicators of social class (Cattell, 1971). Such findings suggest that the development of Gc abilities is promoted by acculturation.

Individual differences in the Gc abilities are associated with individual differences in the quality and amount of formal education –explicit acculturation. This, in turn, is related to child-rearing that promotes a valuing of formal education and the attainment of the knowledge of the dominant culture. These educational and child-rearing conditions are positively associated with socioeconomic level. They are also correlated with a cluster of factors that point to secure home, neighborhood and school environments (Bohlin, Hagekull, and Rydell, 2000).

Other second-order factors have trajectories over age that are different from those of either Gf or Gc. The averages for the abilities of SAR memory decline at a different rate than the averages for the abilities of Gf; the averages for the abilities of TSR consolidation and retrieval from stored knowledge increase at a different rate than the averages for the abilities of Gc.

In longitudinal studies McArdle and his coworkers found repeated-measures age differences of the same general form as have been found in cross-sectional studies.

Brain function and malfunction. A good candidate for a construct validity link between neural function and a general factor is the spike potential of the neuron, for this operates in all neurons and thus all the neural functioning that underlies human behavior. It has not (yet) been shown that individual differences in any aspect of the spike potential is related to cognitive ability differences in humans. Similarly, the evidence does not indicate any other single neural function --neural speed or neural energy or neural mass action— accounts for a part of the variability in all of the cognitive abilities regarded as indicating human intelligence.

Indeed, the findings suggest the opposite. Different abilities relate to different brain functions. With respect the build-ups of memories, for example (which build-ups become parts of cognitive abilities), there appear to be at least three --perhaps four— distinct neural functions. One system is centered in the cerebellum; one is largely associated with the function of the hippocampus; one is related to function of the amygdala; and one is characterized by protein synthesis, perhaps mainly in the frontal lobes (Thompson, 1998).

There is direct evidence linking classical-conditioning learning to neural function in the cerebellum. It is shown, specifically, that selectively disabling and enabling the interpositus nucleus with freezing techniques first wipes out and then restores classical conditioning association learning (Lavond & Kanzawa, 2000). Function in this area of the brain thus affects an elementary form of learning. Some elementary language learning, for example, is clang association –i.e., classical conditioning-- learning. Language acquisition is, of course, part of what is regarded as human intelligence.

The hippocampus plays a different role than the cerebellum in memory, learning and consolidation. The hippocampus is in quite a different part of the brain than the interpositus nucleus, and there are no direct pathways between the two. The hippocampus is essential for retaining the outcomes of instrumental learning in long-term memory. It seems not to be heavily involved in long-term memory storage, as such. Several lines of evidence lead to these conclusions.

The most dramatic of this kind of evidence has derived from studies of HM, a person whose hippocampus became entirely nonfunctional in consequence of surgical removal of approximately two-thirds of the tissue (Corkin, 2002; Skovile & Milner, 1957). After the surgery some of HM's abilities appeared to be quite normal, but his ability to consolidate in learning –to commit what he learned to memory—was completely gone. In conversation, he could recall information learned and stored in the past. This indicated that Gc knowledge was intact. HM could remember a telephone number long enough to dial it, which indicated that SAR apprehension/retrieval was normal He could learn a complex motor skill as well as most people, which suggested that he normal functioning of classical conditioning learning mediated through the interpositus nucleus.

But HM's intermediate association memory was lost. He could carry on an intelligent conversation, but he couldn't remember that he had the conversation. He could remember a telephone number long enough to dial it, but he couldn't commit the number to memory. He was quite unable to remember experiences he had just a few minutes prior to a test for recall. Removal of the hippocampus removed this ability to consolidate learning.

This relationship between consolidation in learning and hippocampus function has been documented in studies of other people who have had lesions in the hippocampus area, and analogous effects have been found in controlled experiments with monkeys

(Thompson, 1998). The relationship has also been indicated in studies of injuries other than ablation. For example, heavy use of alcohol –drinking to the point of passing out— has been found to result in loss of neural function in the mammilary bodies and nearby hippocampus, and this loss, too, is associated with loss of intermediate-term memory and consolidation in learning --a condition known as Korsakoff's syndrome.

Although checks and balances work to ensure that all areas of the brain receive an adequate supply of blood, still conditions such illnesses, high fever, exhaustion, and poisoning bring about decreases in blood flow. The hippocampus area of the brain is particularly vulnerable to such diminutions (Hachinski, 1980). The arteries that supply blood to the hippocampus branch at right angles from the main trunks and terminate as end-arteries in the area – unlike the Y-branches from the main trunks that supply other parts of the brain. This means that any drop in blood flow can result in "drying-up" at the end-arteries, death of neurons, and loss of neural function. Infarcts occur more frequently in the hippocampus than in comparison structures (Corsellis, 1976). Again, what is lost in cognitive ability is consolidation in learning, recorded as poor memory over time periods of more than a few minutes.

Thus, consolidation in learning, an ability that surely is part of what is referred to as intelligence, is associated primarily with neural function operating in and around the hippocampus, and this function is not much related to Gc and SAR. The hippocampus function is different from the cerebellum function of the interpositus nucleus, which supports another form of learning. There is little reason to suppose that these functions of different parts of the brain are merely parts of a unitary brain function such as proposed in the theory of *g*.

These are simply examples. However, such examples add up to suggest that there is no single, unitary function that could represent a principle of mass action process of the kind hypothesized by Lashley (1938) and Spearman (1927) in his theory of g. The evidence indicates instead that different neural functions support different cognitive abilities. The neural functions are affected by events such as those illustrated in the extreme in the case of HM, a case of Korsakoff's syndrome, and diminutions of blood flow. Less extreme versions of these extreme conditions likely occur for different people at different points of development -- before birth, at birth, in infancy, childhood, adolescence, adulthood, old age—and result in individual differences in measures of the various abilities that depend on different neural functions.

Genetic influences. I'm aware of very little direct evidence linking particular genes and genetic factors to particular abilities of reasoning, learning, retention, and retrieval. Such evidence as there is suggests that different sets of genes produce different brain structures and functions, which in turn support and help determine different cognitive abilities. But it is clear that brain structure and function are determined in part by genes; so, since different neural structures and functions are associated with different cognitive abilities of human intelligence, different genes most likely are as well.

Individual differences in brain structures are rather like individual differences in the faces of people: when you see different faces, you can visualize different brain structures behind those faces. Just as there are separate genetic factors affecting ear shape and nose length, so there are separate genetic factors affecting the size and shape and function of the right and left hemispheres of the brain, the cerebellum, the limbic brain, the hippocampus, and so on. Different sets of genes also affect neurotransmitter systems and the pathways joining regions of the brain. Different structures, determined by different combinations of genes, enable different reactions to environmental stimulations. Different environmental stimulations affect separate brain functions differently. Such combinations of genetic and environmental influences –huge in number-- produce great variety in the patterns of abilities we can measure.

It is possible, of course, that a single set of genes operate in unison throughout the separate neural systems, and thus operate throughout all human abilities, but on the face of it, this seems unlikely. If it is true, it would seem that it must be through a gene of set of genes that affect some elementary function of all neurons, such as the spike potential function of each individual neuron. If such a feature does indeed affect all human abilities, that influence is very likely small relative to the influences of the separate functions of different parts of the neural system.

Emerging Theory

The results I have reviewed in previous sections of this chapter provide glimmerings of the nature of adult human intelligence, but they are dim glimmerings,. Some of the information is incorrect. The big picture is not in focus.

Problems With Current Theory of Cognitive Capabilities.

What we see as intelligence in the theory and research findings I reviewed in previous sections is not consistent with what we see when see adults doing the jobs they do in our society. The current theory points to adulthood aging declines Gf, SAR and Gs—major abilities of intelligence. But decline does not characterize what we see in every-day observations of adults. In the research, we see adolescents and young adults more intelligent than older adults, but in life we don't see increasing deficits of reasoning and memory (at least not through the main period of adulthood, from the 30's into the 70's); we see advanced-age adults doing most of the intellectual work of maintaining and advancing the culture; we see older people who are the intellectual leaders in science, politics, business, and academics, people who are in their positions of responsibility largely because (we think) they are –in some sense we need to define—more intelligent than younger adults and adolescents.

So, there's something out of kilter here. Are we measuring the wrong things in the research thus far done? The answer appears to be yes. It may be yes both in regards to abilities that are regarded as not declining in adulthood –Gc and TSR—as well as in regards to the abilities for which the research does indicate decline –Gf, SAR, and Gs.

Consider Gc first. As defined, this is supposed to indicate the depth of the knowledge of the culture, as well as breadth of this knowledge. In the practice of measurement, however, we get only breadth, no depth. The primary abilities that indicate the factor are surface-like –the beginning course—not what one who seriously studies a domain of knowledge comes ultimately to understand. Instead of measuring understanding of science or literature, for example, we measure vocabulary, information and reasoning sampled only at the elementary level of introduction to these domains of knowledge. Our measures provide only dilettante indications of knowledge of a culture. A person flitting over many areas of knowledge in his or her study will score higher on these measures of Gc than a person who has devoted intensive study to develop truly profound understanding in an area of knowledge. But we recognize this latter, not the dilettante, as the most intelligent –the one most likely to make significant contributions to the culture, the one who becomes CEO, the one to whom we award the Nobel Prize. Such persons are referred to as expert. Experts best exemplify the capabilities that indicate the nature and limits of human intelligence.

Consider next the reasoning we measure in the primary abilities that define Gf, fluid intelligence (and equate with Spearman's *g*, the *sine qua non* of intelligence). This is a measure of reasoning with a made-up problem that really is no problem. The reasoning is inductive, requiring –in the device used to measure it-- as little knowledge as possible.

In contrast, the reasoning we regard as indicating intelligence (in the CEO, for example) is reasoning with relevant information. It is deductive reasoning employing knowledge. It is the kind of reasoning identified in descriptions of the thinking of experts –in chess, financial planning, and medical diagnosis (Charness, 1991; de Groot, 1978; Ericsson, 1996; Walsh & Hershey, 1993).

The expert is able to construct a framework within which to organize and effectively evaluate presented information, while novices, with no expertise basis for constructing a framework, search for patterns and do reasoning by trial-and-error evaluations –inductive reasoning. The expert apprehends large amounts of organized information, comprehends many relationships among elements of this information, infers possible continuations and extrapolations, and, as a result, is able to select the best from among many possibilities in deciding on the most likely outcome, consequence or extension of relationships. The expert goes from the general comprehension of relations and knowledge of principles to most likely specifics.

The person able to do this kind of reasoning is regarded as intelligent. This is what is recognized as intelligence in adults.

This is not to argue that the inductive reasoning of Gf is not an aspect of what is also called intelligence. However, it is to argue that the reasoning of Gf may not be a central characteristic of intelligence, but expertise reasoning may be.

A Glimpse at Findings From Research on Expertise.

To summarize, the this: (1) abilities that come to fruition in adulthood best represent the quintessential expression of human intellectual capacity (2) the measures currently used to estimate intelligence do not assess these abilities; they do not assess all the important abilities of human intelligence, (3) when measures currently used do assess these abilities, they don't assess them at a depth sufficient to indicate the essential character of intelligence, (3) the abilities not measured and not among the abilities currently used to estimate intelligence are in-depth abilities of expertise.

The principal problems of design of research for describing adult human intelligence are problems of circumscribing domains of expertise, and constructing measures of the abilities of the highest levels of expertise in these domains. The research should identify the relationships between these measures and other measures thought to indicate human intelligence. The research should show the relationships these measures have with variables that help indicate causes and effects –i.e., variables that indicate the validity of the expertise-ability constructs.

Hiromi Masunaga and I have studied expertise in playing the game of GO. We have designed research in accordance with the stipulations I outline above. I do not have the space here to provide much of a summary of what we have discovered so far. For better summaries, I refer you to Horn & Masunaga (2000), Masunaga & Horn (2000) and, particularly, Masunaga & Horn (2001).

One principal conclusion we draw from review of the results of research by others is that intensive practice is required to attain and maintain high levels of expertise. This practice must extend into adulthood, which means that there can be increases in expertise abilities of intelligence in adulthood. If there is such practice, expertise abilities do not decline. Intensive practice is not simply practice; it is practice focused on attaining ever-higher levels of expertise.

Also important for our research, we found that research on expertise pointed to some of the abilities that, it appeared, would be important features of adult intelligence. Charness (1991) had identified expertise deductive reasoning as one such ability. Ericsson & Kintsch (1995) had identified a wide-span working memory that also appeared to be important. These two abilities, it seemed, were not sampled in the research on primary mental abilities and the second-order factors of extended Gf-Gc theory, but the abilities characterized high levels expertise that indicated high levels of human intelligence.

Following these leads, we found in our studies that a form of expertise deductive reasoning (EDR) that is quite separate from Gf reasoning, and a wide-span working memory (WSWM) that is quite separate from short-term working memory (SAR). The span for WSWM is several times as wide as the span of SAR. An hypothesis stipulating that a factor of expertise speed is separate from cognitive speed (Gs) was not supported: expertise speed and cognitive speed were collinear. Gf, SAR Gs were found to decline with age in adulthood. EDR and WSWM were found to decline as a function lack of intensive practice and lack of expertise. Advanced levels of EDR and WSWM were maintained throughout adulthood in people who worked to advance their level of expertise. Expertise abilities of older high-level experts exceeded the abilities of younger persons at lower levels of expertise.

These conclusions flow from few studies. There have been no longitudinal follow-up studies of broad samples of adults that could help indicate the extent to which the findings apply to people in general. There is particular need for this kind research.

Closing Comments and Conclusions

One can question whether a study ever has been adequately designed to test Spearman's model of one common factor pervading all expressions of human intelligence: the possibility of specific factor overlap has never been entirely ruled out. But granting that uncertainty, most of the factor analytic evidence does not support the theory. A wide array of evidence from research on development, education, neurology and genetics suggests that it is unlikely that a factor general to all abilities produces individual-differences in all of what are regarded as indicators of human intelligence. There have been many efforts to discredit and counteract this evidence: they have not altered the conclusion: no general factor has been found. The evidence suggests that if there is such a factor --a behavioral concomitant of neural spike potential, for example--it accounts for no more than a miniscule part of the variance in human intellectual abilities.

But while his theory of a general factor has not found support, Spearman's model for testing that theory, and his ideas about processes that indicate important features of a general factor, have been very important for the development of scientific theory of human cognitive capabilities over the last century. His idea of testing an hypothesis stipulating that a latent variable may underlie many manifest indicators, and his mathematical and statistical methods for accomplishing this test, are the core concepts of the theory of common factors. That theory is a fundamental part of the current theory of multiple factors and structural equation modeling, which theory, in turn, is not only the basis for modern theory of human capabilities, but also the basis for a very large amount of other substantive theory in psychology. Spearman's concept of *g* is in major respects the concept of Gf, fluid reasoning, in modern theory. His ideas about what *g* is not became –in modern theory-- ideas about what Gc, crystallize knowledge, is. He described apprehension of experience, cognitive speed-power trade-off and individual differences, and short-term memory in much the same way as these concepts are described in modern theory. Spearman's influence on present-day scientific thinking has been large. His ideas are well embedded in that thinking.

Thurstone's theory of simple structure for multiple common factors of cognitive abilities also was very important. Current theory of human intellectual capacities derives directly and mainly from the work of Spearman and Thurstone.

That current theory is described under the heading of extended Gf-Gc theory. This is an account of evidence indicating that individual differences in behaviors characterizing human intelligence can, to a considerable extent, be described in terms of broad factors of capacity to consolidate new information and retrieve it (TSR), capacity to retain information in accessible storage (Gc), capacity for reasoning in novel situations (Gf) capacity for holding unorganized information in immediate awareness over short periods of time (SAR), capacity for organizing and retaining visual information (Gv), capacity for organizing and retaining auditory information (Ga), capacity for quantitative thinking (Gq), and capacity for speedy thinking (Gs). The theory organizes evidence of the relationships these broad constructs have with other variables, principally variables pertaining to development and neurological functioning, but also variables pointing to achievements in school and work and variables representing genetic variations. Gf, SAR and Gs are found to decline with age and brain damage in the same people in which Gc and TSR increase with age and are less affected by brain damage. Evidence from several sources indicates links between the separate broad factors and separate brain functions. .

Extended Gf-Gc theory does not adequately describe abilities that appear to be quintessential expressions of human intelligence –in particular, abilities that reach their peaks of development in adulthood. Research on expertise has pointed to two of these abilities, one a factor expertise deductive reasoning (EDR) that is distinct from Gf reasoning, the other a factor of wide-span working memory (WSWM) for which the span is considerably larger and more flexible than for SAR. It appears that these two expertise abilities do not decline over the period of adulthood if there is continued intensive practice to improve --unlike the findings for the comparable Gf and SAR abilities.

Footnotes

---

[i] I particularly thank Jack McArdle for editorial suggestions and many good ideas he contributed to this chapter. I also thank Hiromi Masunaga, Robert Sternberg and Kevin McGrew for their very helpful comments.

[ii] In this context to keep “test” in the sense of testing a theory separate from “test” in the sense of a psychological test to measure an ability, I will use the term “measurement device” or simply “device” to refer to a psychological test.

[iii] More specifically to the point, Meredith (1964), citing Lawley (1943-44) as the originator of the theorem, has shown that a common factor of a population can be identified when neither the sample of indicator variables nor the sample of subjects is representative of their respective populations.

---

iv The other 1904 article –Spearman, 1904b-- is a presentation of test theory, not a study of intelligence. Here I refer to some results from that article –for example, the idea of correcting correlations for attenuation due to unreliability— but I do not review it here.

v There was no mention of *g* in this early article. Reference to *g* appears in Spearman (1923), Theory of Two Factors, and is very prominent in Spearman (1927).

vi In Table 1 compare the elements in the same rows of any 1 two columns --say, columns 1 and 2. Perfect fit of the model requires $r_{31}/r_{32} = r_{41}/r_{42} = r_{51}/r_{52} = r_{61}/r_{62}$. In Table 1 the proportions are .78/.67 = 1.16; .70/.67 = 1.04; .66/.65 = 1.02; .63/.57 = 1.10. Throughout the table the average of the 90 independent proportions for off-diagonal elements is 1.07.

vii To control for practice effects in music and pitch discrimination, only the boys who were taking music lessons were selected from the entire class of 33 boys; the 11 boys who were not taking music lessons were excluded from the sample.

viii The claim that Spearman invented factor analysis needs to be understood in the context of mathematical inventions that preceded him (much of which history is discussed elsewhere in this book). Spearman did not invent or prove the principles of determinants and matrices that are the crux of his test. These principles were developed by others well before Spearman's time. Grattan-Guinness & Ledermann (1994) provide a good history indicating how and when these ideas came into use. They report that as early as the 17th century --in the writings of Leibniz in Europe and Seki in Japan, both in the 1680's-- it was known that in solving linear equations, determinants comprised of the coefficient multipliers of the unknowns indicate the number of consistent equations needed to solve for –determine— a solution for the unknowns. As early as 1812 in the work of Cauchy the concepts of matrix, characteristic equation, and characteristic values (eigenvalues) were in use. In the writings of Frobenius, Sylvester and Caley in the 1870's and 1880's there were proofs that a matrix satisfies its own characteristic equation, that the number of linearly independent (basis) variables among a set of variables is the rank of the product-moment matrix (Gram product) of these variables, that rank is equal the number of nonzero eigenvalues, and is one less than the largest non-zero minor of that Gram-product matrix. In the 1890's Pearson developed the product moment correlation and Yule described partial and semi-partial correlation.

So these ideas on which factor analysis is based were known and in available publications in the days when Spearman read mathematics and studied for his degrees in engineering and psychology. They indicate that if measurements can be regarded as real numbers (at the interval level of measurement) in parametric equations for which there are unknowns, then mathematical analyses can be applied to solve for the unknowns. Spearman's invention of factor analysis involved applying these principles to the idea that, indeed, measurements of persons might be assumed to be real numbers in equations and therefore the mathematical principles of determinants and matrices can be applied to solve for the unknowns. His idea –the basic idea of common factor analysis-- was that there may be both common and unique basis variables and that only the off-diagonal elements (the covariances) of the Gram product indicate the common basis variables. This is different from the way the principles of determinants and matrices had been applied in eigenvalue-eigenvector decomposition analysis, and in the method of scaling the eigenvectors that is called principal components analysis (Pearson, 1901). In stating and applying the principles of determinants and matrices in this novel way Spearman invented the one-common-factor form of factor analysis.

[ix] Spearman (1914) gives the product moment correlation formula for this. Kendall's (1962) tau had not yet been invented. Although Spearman (1904b) had developed the product moment correlation for ranks, he did not argue for use of that formula here.

[x] The same positive correlation obtains when things are measured at the opposite pole of the not-so-good things in life –poor health, little money, lousy education, and lack of abilities, as discussed at some length in Herrnstein & Murray (1994).

[xi] Nor does it call for a theory of mass action of neural activity underlying all cognitive behavior.

[xii] It is reasonable that Gustafsson & Undheim chose to align their *g*-factor with Gf – because the indicator variables of Gf are most similar to the process variables that Spearman described as indicating *g* (as I will point out in another section of this paper). But such alignment should not be taken as proof of the equivalence of *g* and

Gf: it is merely a statement that the abilities that define Gf appear to be the same as the processes Spearman identified as indicating *g.* ,

[xiii] As from noesis, purely intellectual apprehension

[xiv] Root mean square error of approximation RMSEA, Browne & Cudeck, 1993; Steiger, 1990

[xv] In centroid analysis the positive and negative correlations had to sum to zero, and these were approximately the conditions imposed by Spearman's method; in principal components analysis, the sum of squares of the positive correlations had to equal the sum of squares of the negative correlations.

References

Alexander, W. P. (1935). Intelligence, Concrete and Abstract. British Journal of Psychology: Monograph Supplement.

Babcock, E. (1930). An experiment in the measurement of mental deterioration. Archives of Psychology, 18, No.117.

Baddeley, A. (1994). Memory. In A.M. Colman (Ed.), *Companion encyclopedia of psychology* (Vol. 1, pp. 281-301). London: Routledge.

Bayley, N. (1969). Manual for the Bayley Scales of Infant Development. New York: Psychological Corporation.

Beaujean, A.A. (2002). On the life and scholarship of Arthur R. Jensen. Proceedings the SASP Convention. London.

Bianchi, L. (1922). The Mechanism of the Brain and the Function of the Frontal Lobes. Edinburgh: Oliver & Boyd, 1922.

Binet, A. & Simon, T (1905). Sur le necessite d'etablit un diagnostic scientifique des etats inferieurs de l'intelligence. (On the necessity of establishing a scientific diagnosis of inferior states of intelligence.) L'Annee Psychologique, 11, 163-190; also 191-244; 245-366.

Bohlin, G. Hagekul, B. and Rydell, A. M. (2000). Attachment and social functioning: a longitudinal study from infancy to middle childhood. Social Development, 9, 24-39.

Brown, W. (1933). The mathematical and experimental evidence for the existence of a central intellective factor. British Journal of Psychology, 23, 171-179.

Brown, W. A. & Stephenson, W. (1933). A test of the theory of the two factors. British Journal of Psychology, 23, 352-370.

Browne, M.A. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.). Testing Structural Equation Models (Chapter 6, pp. 136-162). Newbury Park, CA: Sage Publications.

Burt, C. (1909). Experimental tests of general intelligence. British Journal of Psychology, 3, 94-177.

Burt, C. (1911). Experimental tests of higher mental processes and their relation to general intelligence. Journal of Experimental Pedagogy and Training, 1, 93-112.

Burt, C. (1924). Report of consultive committee on psychological tests of educable capacity. London: H. M. Stationery Office.

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

Carter, H.D. (1928). The organization of mechanical intelligence. Journal of Genetic Psychology, 35, 270-285.

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

Cattell, R. B. (1963). Theory for fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 54, 1-22.

Cattell, R. B. (1971) *Abilities: Their structure, growth and action.* Boston: Houghton-Mifflin.

Charness, N. (1991). Expertise in chess: the balance between knowledge and search. In K. A. Ericsson and J. Smith (Eds.), *Toward a general theory of expertise* (pp. 39-63). New York: Cambridge University Press.

Christian, A.M. & Paterson, D.G. (1936). Growth of vocabulary in later maturity. Journal of Psychology, 1, 167-169.

Corkin, S. (2002). What's new with the amnesic patient H.M. *NatureReviews Neuroscience*, 3(2), 153-60.

Corsellis, J. A. N. (1976). Aging and the dementias. In W. Blackwood & J. A. N. Corsellis (Eds.): Greenfield's Neuropathology. London: Arnold.

Cox, G.W. (1928). Mechanical aptitude. London: Methuen. Craik, F. I. M. (1977). Age differences in human memory. In J. E. Birren & K. W. Schaie (Eds.): *Handbook of Psychology of Aging*. New York: Van Nostrand-Reinhold.

Dandy, W.S. (1933). Physiologic studies following extirpation of the right cerebral hemisphere in man. Bulletin: Johns Hopkins Hospital, 53, 31-51.

de Groot, A.D. (1978). *Thought and choice and chess*. The Hague: Mouton.

Detterman, D. K. (Ed.). (1998). Kings of men: A special issue of the journal Intelligence about Arthur Jensen. [Special Issue]. *Intelligence, 26*(3).

El Koussy, A.A.H. (1935). The visual perception of space. British Journal of Psychology. Monograph Supplement, No. 20.

Ekstrom, R. B., French, J. W. & Harman, M. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs, 79*.

Ericsson, K.A. (1996). The acquisition of expert performance. In K.A. Ericsson (Ed.), *The road to excellence* (pp. 1-50). Mahwah, NJ: Lawrence Erlbaum Associates.

Ericsson, K. A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 105*, 211-245.

Eysenck, H. J. (Ed.) (1982). *A model for intelligence*. Berlin: Springer-Verlag.

Feuchtwanger, E. (1923). Funktionen des stiruhirns ihre Pathologie und Psycholgic. Monograph Gesalk Neurologia Psychiatri, 38, 194-206/

Flanagan, D.P., Genshaft, P. I. & Harrison, P. L. (Eds.) (1997). *Contemporary Intellectual Assessment* (pp. 53-91). New York: Guilford Press.

Flanagan, D.P. & Harrison, P. L. (Eds.), (2005) Contemporary Intellectual Assessment: Theories, Tests and Issues. New York: Guilford Press.

Galton, F. (1869). Hereditary Genius: An Inquiry into Its Laws and Consequences. New York: Horizon Press, 1952. (Original work published 1869.)

Galton, F. (1883). Inquiries into Human Faculty and Its Development. London: Dent,

Gottfredson, L. S. (1997). Why g matters: the complexity of everyday life. Intelligence, 24, 79-132.

Grattan-Guinness, I. & Ledermann, W. (1994). Matrix theory. In Grattan-Guinness, I. (ed.), *Companion Encyclopedia of the History and Philosophy of the*

Mathematical Sciences. London: 775-786.

Guilford, J. P. (1964). Zero intercorrelations among tests of intellectual abilities. Psychological Bulletin, 61, 401-404.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley

Gustafsson, J. E. & Undheim, J.O. (1996). Individual differences in cognitive functions. In D.C. Berliner & R.C. Calfee (Eds.). Handbook of educational psychology (pp. 186-242. New York: Simon & Schuster Macmillan.

Hachinski, V. (1980). Relevance of cerebrovascular changes in mental function. Mechanisms of Aging and Development, 10, 1-11.

Hakstian, A. R., & Cattell, R. B. (1974). The checking of primary ability structure on a broader basis of performances. British Journal of Educational Psychology, 44, 140-154.

Hart, B. & Spearman, (1912). General ability, its existence and nature. British Journal of Psychology, 5, 51-81.

Hebb, D. O. (1941). The clinical evidence concerning the nature of normal adult test performance. Psychological Bulletin, 38, 593.

Herrnstein, R. & Murray, C. (1994). The bell curve: Intelligence and class structure in American life. New York: Free Press.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, 75*,* 242-259.

Horn, J. L. (1991). Measurement of intellectual capabilities: a review of theory. Chapter 7 in K. S. McGrew, J. K. Werder & R.W. Woodcock (Eds.), *Woodcock-Johnson Technical Manual*. Allen, TX: DLM Teaching.

Horn, J. L., Donaldson, G., & Engstrom, R. (1981). Apprehension, memory, and fluid intelligence decline in adulthood. *Research in Aging, 3*(1), 33-84.

Horn, J.L. & Masunaga, H. (2000). New Directions for research on aging and intelligence. In T. J. Perfect & E. A. Maylor (Eds.), Models of Cognitive Aging. (pp. 125 -159). Oxford: Oxford University Press.

Holzinger, K. J. (1934). Preliminary reports on the Spearman-Holzinger unitary trait study. Nos. 1-9. Chicago: University of Chicago, Statistical Laboratory, Department of Education.

Humphreys, L. G. (1971). Theory of intelligence. In R. Cancro (Ed.), Intelligence: Genetic and environmental influences (pp. 31-42). New York: Grune and Stratton.

Humphreys, L. G., Ilgen, D., McGrath, D. & Montanelli, R. (1969). Capitalization on chance in rotation of factors. Educational and Psychological Measurement, 29, 259-271.

Jensen, A. R. (1998). The g factor: the science of mental ability. London: Praeger.

Jones, H.H., Conrad, A.S. & Horn, A. (1928). Psychological studies of motion pictures. II. Observation and recall as a function of age. University of California Publications in Psychology, 3, 225-243.

Kelley, T. L. (1928). Crossroads in the mind of man. Stanford, CA: Stanford University Press.

Kendall, M. G. (1961). Rank Correlation Methods. New York: Hafner Publishing Company.

Kubitscheck, P.S. (1928). The symptomatology of tumors of the frontal lobe based on a series of twenty-two cases. Archives of Neurology in Psychiatry, 20, 559-579.

Lashley, K. S. (1929). Brain Mechanisms and Intelligence: a quantitative study of injuries to the brain. Chicago: University of Chicago Press.

Lashley, K. S. (1938). Factors limiting recovery after central nervous system lesion. *Journal of Nervous and Mental Diseases, 78*, 733-755.

Lawley, D. N. (1943-44). A note on Karl Person's selection formula. *Proceedings of the Royal Society of Edinburgh (Section A)*, 28-30.

Lavond, D.G. & Kanzawa, S.A. (2000). Inside the black box. In J.E. Steinmetz, M. Gluck & P. Solomon (Eds.) Model systems and the neurobiology associative learning: a festschrift in honor of Richard F. Thompson. (pp. 245-269) Hillsdale, N.J. Lawrence Erlbaum Associates.

Masunaga, H. & Horn, J. L. (2000). Characterizing mature human intelligence: expertise development. Learning and Individual Differences, 12, 5-33.

Masunaga, H. & Horn, J. L. (2001). Expertise and age-related changes in components of intelligence. Psychology and Aging, 16, 2, 293-311.

McArdle, J.J. & Woodcock, R. (Eds.) (1998), *Human cognitive abilities in theory and practice*. Chicago, IL: Riverside press.

McDougall, W. (1932). The Energies of Man. London: Methuen.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *Woodcock-Johnson technical manual*. Allen, Texas: One DLM Park.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika, 29*, 177-185.

Meredith, W., & Horn, J. L. (2001). The role of factorial invariance in modeling growth and change. In A. G. Sayer & L. M. Collins (Eds.), *New methods for the analysis of change* (pp.203-240). Washington, D.C.: American Psychological Association.

Miele, F. (2003). *Intelligence, race and genetics: Conversations with Arthur R. Jensen*. Boulder, CO: Westview.

Miles, C.C. (1934). The influence of speed and age on intelligence scores of adults. Journal Genetic Psychology, 10, 208-210.

Neale, M. C. (1993). *Mx structural equation modeling*. Richmond, VA: Medical College of Virginia.

Nyborg, H. (Ed.). (2003). *The scientific study of general intelligence -- Tribute to Arthur R. Jensen.* New York: Pergamon.

Patterson, D. G., & Elliot, R. N. (1930). Minnesota Mechanical Ability Tests. Minneapolis, MN: University of Minnesota Press.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 6, 559-572.

Perfect, T. J. & Maylor, E. A. (Eds.), (2000). Models of Cognitive Aging. Oxford: Oxford University Press.

Rimoldi, H. J. (1948). Study of some factors related to intelligence. Psychometrika, 13, 27-46.

Rowe, S.N. (1937). Mental changes following the removal of the right cerebral hemisphere for brain tumor. American Journal of Psychiatry, 94, 605-614.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed.) New York: Van Nostrand Reinhold.

Salthouse, T. A. (1991). Theoretical perspectives on cognitive aging. NJ: Lawrence Erlbaum.

Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle longitudinal study.* Cambridge: Cambridge Univ. Press

Scoville, W. & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. Journal of Neurology, Neurosurgery and Psychiatry, 20, 11.

Spearman, C. E. (1904a). ``General intelligence,'' objectively determined and measured. American Journal of Psychology, 15, 201-293.

Spearman, C. E. (1904b). ``The proof and measurement of association between two things. American Journal of Psychology, 15, 72-101/

Spearman, C. (1914) Theory of two factors. Psychological Review, 21,101-115.

Spearman, C. The Nature of Intelligence and the Principles of Cognition. London: Macmillan, 1923.

Spearman, C. E. (1927). The abilities of man: Their nature and measurement. New York: Macmillan.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*, 498.

Steiger, J.H. (1990). Structural model evaluation and modification: an interval estimation approach. Multivariate Behavioral Research, 25, 173-180.

Thompson, R. F. (1998) The brain: a neuroscience primer. New York: Freeman.

Thomson, G. A. (1919). On the cause of hierarchical order among correlation coefficients. Proceedings of the Royal Society, A, 95, 400-408.

Thorndike, E. L. (1903). Intelligence. Educational Psychology.

Thurstone, L. L. (1931). Multiple factor analysis. Psychological Review, 38, 406-427.

Thurstone, L. L. (1935). The vectors of mind. Chicago: University of Chicago Press.

Thurstone, L. L. (1938). Primary mental abilities. Psychometric Monographs, (1). Chicago: University of Chicago Press.

Thurstone, L. L. (1947). Multiple factor analysis. Chicago: University of Chicago Press.

Thurstone, L. L. & Thurstone, T. G. (1941). Factorial studies of intelligence. Psychometric Monographs.

Tryon, R.C. (1932). Multiple factors vs. two factors as determiners of abilities. Psychological Review, 39, 324-351.

Walsh, D.A. & Hershey, D.A. (1993). Mental models and the maintenance of complex problem solving skills in old age. In J. Cerella, J. & Rybash, W. Hoyer, & M. Commons (Eds.), *Adult information processing: Limits on loss* (pp. 553-584). San Diego: Academic Press.

Weisenberg, T. & McBride, K. E. (1935). Aphasia: a Clinical and Psychological Study. New York: Commonwealth Fund.

Willoughby, R.R. (1927). Family similarities in mental –test abilities. Genetic Psychology Monograph, 2, 235-277.

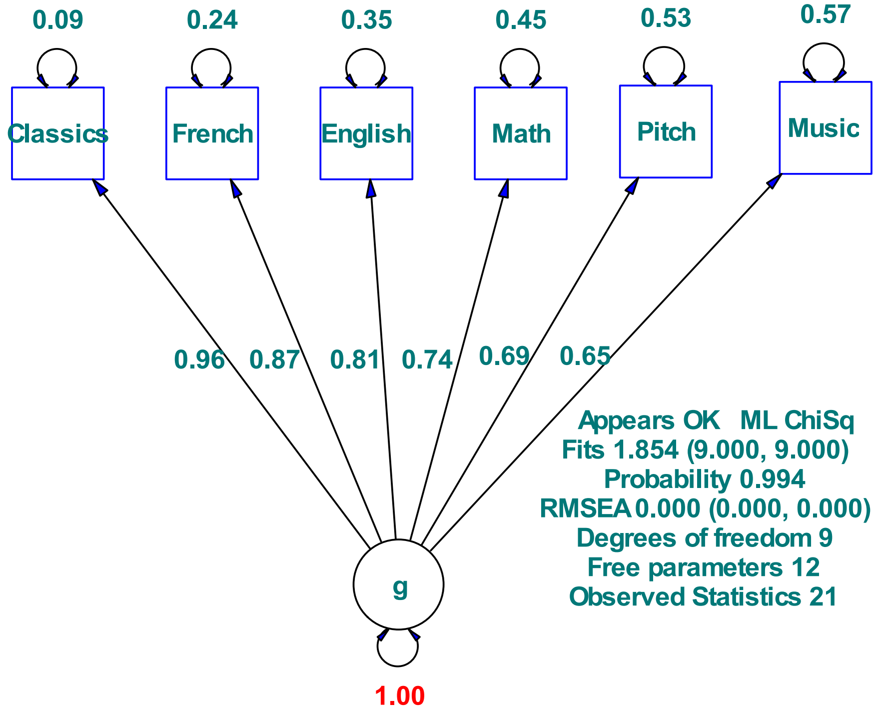
Woodcock, R.W. (1995). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231-258.

Table 1

Intercorrelations On Which Spearman Based His Analysis of the Hierarchy of the Intelligences, as Reported in Spearman (1904a)

	Classics	French	English	Math	Discrim.	Music
Classics	1.00					
French	.83	1.00				
English	.78	.67	1.00			
Math	.70	.67	.64	1.00		
Discrim.*	.66	.65	.54	.45	1.00	
Music	.63	.57	.51	.51	.40	1.00

* Pitch discrimination



0.09
0.24
0.35
0.45
0.53
0.57
Classics
French
English
Math
Pitch
Music
0.96
0.87
0.81
0.74
0.69
0.65
Appears OK ML ChiSq
Fits 1.854 (9.000, 9.000)
Probability 0.994
RMSEA 0.000 (0.000, 0.000)
Degrees of freedom 9
Free parameters 12
Observed Statistics 21
g
1.00