General and Specific Mental Abilities

General and Specific Mental Abilities

Edited by

Dennis J. McFarland

Cambridge Scholars Publishing

General and Specific Mental Abilities

Edited by Dennis J. McFarland

This book first published 2019

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-3310-7
ISBN (13): 978-1-5275-3310-3

To my wife, Loretta and my three sons, John, Christopher, and Michael.

CONTENTS

PREFACE

The history of testing mental abilities has seen two contrasting approaches, that of psychometrics and that of neuropsychology. These two traditions have different theories and methodologies but overlap considerably in the tests they use. Historically, psychometrics has emphasized the primacy of a general factor while neuropsychology has emphasized specific abilities that are dissociable. Other disciplines have recently become interested in this issue. This book includes the opinion of experts from several fields including psychometrics, neuropsychology, speech language and hearing, and applied psychology. These experts have diverse opinions on the relative importance of general and specific abilities.

There is not a consensus about the nature of human mental abilities despite over one hundred years of study. Yet this issue is of importance for many practical concerns. Questions such as gender, ethnic, and age-related differences in mental abilities are relatively easy to address if these are due to a single dominate trait. Presumably such a trait can be measured with any collection of complex cognitive tests. If there are many specific mental abilities, these would be much harder to measure and associated social issues are more difficult to resolve. The relative importance of general and specific abilities also has implications for educational practices. For example, are there specific learning disabilities amendable to remediation and do certain instructional approaches benefit some students more than others?

In deciding on a title for this book I at first considered using "general and specific cognitive abilities". However, I chose mental abilities instead. This is due to my feeling that the term "cognitive" is applied much too broadly. For example, perceptual and motor functions are described as cognitive. I once asked an eminent expert in cognitive psychology where cognition began and he replied "at the retina". To me, this implies that all mental events are cognitive! But if a term refers to every aspect of mental life it loses meaning. Perceptual phenomena should be those processes that are sensory-modality specific. Motor processes should likewise be those that involve specific output modalities. The term cognitive should be reserved for describing "thought" rather than "automatic" processes.

The authors in this text have a diversity of views on issues concerning the relative importance of general and specific mental abilities as well as other issues. For example, are abilities discovered or created? That is, do our constructs represent real entities or are they useful constructions? My own thinking about such matters continues to evolve over time.

LIST OF CONTRIBUTORS

Alexander Beaujean
Baylor University

Jens F. Beckmann
Durham University

Nadin Beckmann
School of Education
Durham University

Damian Birney
School of Psychology
University of Sydney

Moritz Breit
Department of Psychology
University of Trier, Germany
,
Martin Brunner
Faculty of Human Sciences
University of Potsdam, Germany

Vivian Chou
Baruch College and the Graduate Center
City University of New York

Roberto Colom
Universidad Autónoma de Madrid, Spain

J. P. Das
University of Alberta, Edmonton, Canada

José María Ruíz Sánchez de León
Universidad Complutense, Spain

George K. Georgiou
University of Alberta, Edmonton, Canada

Harold Goldstein
Baruch College and the Graduate Center
City University of New York

Abe Hofman
University of Amsterdam

Kees-Jan Kan
University of Amsterdam

Harrison Kell
Educational Testing Service
Princeton, NJ

Kristof Kovacs
Maarten Marsman
University of Amsterdam

Dennis McFarland
National Center for Adaptive Neurotechnologies
Albany,NY

Kevin R. Murphy
University of Limerick
Ireland

Rachel Omansky
Baruch College and the Graduate Center
City University of New York

Franzis Preckel
Department of Psychology
University of Trier, Germany

M. Ángeles Quiroga
Universidad Autónoma de Madrid, Spain

Alexander O. Savi
University of Amsterdam

Charles Scherbaum
Baruch College and the Graduate Center
City University of New York

Lazar Stankov
School of Psychology
The University of Sydney, Australia

Kara M. Styck
Department of Psychology
Northern Illinois University

Irina Trofimova
McMaster University
Hamilton, ON, Canada

Han L. J. van der Maas
University of Amsterdam

Grant M. Walker
University of California, Irvine

Kenneth Yusko
University of Maryland

CHAPTER ONE

A BRIEF HISTORY OF THEORY AND TESTING OF GENERAL AND SPECIFIC MENTAL ABILITIES

DENNIS MCFARLAND

The present review summarizes the early development of theories and testing of mental abilities with a particular emphasis on issues related to general and specific abilities. Any such review necessarily omits a vast amount of material. Contributions of certain individuals are covered but it should be kept in mind that their work was in the context of that of their contemporaries. The reader may want to consult some of the original sources to get a flavor for the language and thinking of each period. This review is largely chronological rather than by topic so as to maintain the order of the unfolding of events.

Speculation about the nature of human mental abilities has a long history. For example, Aristotle described analytical, practical, and creative varieties of intelligence (Tigner & Tigner, 2000). Testing of mental abilities was practiced by the ancient Chinese (Bowman, 1989) and Greeks (Doyle, 1974). The Chinese exams were part of an elaborate system for evaluating potential civil servants that evolved over a considerable period of time. Bowman (1989) notes that controversy over testing practices in ancient China foreshowed many current issues in mental abilities testing. These included the relative importance of memory and expert knowledge, effects of social class on test performance, and the use of geographical quotas.

The psychometric approach

Most accounts of testing of human mental abilities begin at the turn of the twentieth century. This effort was part of the new discipline of experimental psychology that had recently begun with the work of Wundt

(Mackintosh, 2011). J. McKeen Cattell, a student of Wundt, devised a test of association in which individuals wrote down as many words as they could in 20 seconds in response to a single spoken word (Cattell & Bryant, 1889). This procedure is reminiscent of modern tests of word fluency. Cattell and Galton (1890) proposed assessing mental abilities with a series of tests measuring simple perceptual abilities, reaction time, and a letter span task. They suggested that collection of data with these tasks on a large number of individuals would provide insight about the constancy of mental processes and their interdependence. Cattell and Farrand (1896) subsequently reported observations on such tests from students at Columbia University.

Wissler (1901) examined Cattell and Farrand's tests using Pearson's recently developed method of correlation. Wissler (1901, page 1) asserted that "If a test is general, then its results should correlate with many other special tests, and, in turn, if there is an integral relation between general and specific ability, the results of the latter should correlate with the former." He reported that the correlation between various scores was not "significant", but the criterion he used for what he called "significant" was not based on probability but rather on the colloquial usage of this term. My own calculation of the probability for his findings of a correlation of 0.21 between speed of letter cancellation and speed of naming colors with an n of 159 yields a $p < 0.01$. Likewise, Wissler described a correlation between auditory and visual digit span of 0.39 as significant but small. Correlations between academic standing and mental tests were generally small, the largest being between logical memory at $r = 0.19$. In contrast, correlations between relative standing in different classes were much higher, ranging between 0.60 and 0.75. Wissler concluded that the mental tests used by Cattell & Farrand (1896) had little interdependence and that they were not of practical value. This investigation was influential in suggesting that tests of perception and speeded responding were not useful measures of intellect. However Wissler's evaluation of the magnitude of correlations was much different than what is practiced today.

Spearman (1904) advocated that a "correlational psychology" be applied to mental tests. Spearman's correlation differed from that of Pearson in that it involved rank ordering the data. He applied this method to "selected laboratory psychics" in young students that included tests of sensory discrimination in three senses and three estimates of intelligence by their teachers. Spearman (1903) found that all nine correlations between these measures were positive and concluded that they all reflected a common element. He offered a theory of "intellective unity" and suggested a hierarchy of the specific intelligences. This hierarchy was ordered according to the "saturation" of each item with general intelligence. Thus, each measure was viewed as composed of general intelligence and a specific factor which was uncorrelated with other specific factors. Spearman (1903, page 284) concluded:

> 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.

Jensen (2000) states that Spearman's interests were predominately in theory and the nature of cognition.

Spearman (1914) later offered an improved method for verifying his two factor theory that was applied to a larger series of mental tests that were collected by Thorndike. Spearman's method involved the computation of tetrad differences between the ratios of pairs of correlations among four mental tests. If correlations are due to a single common factor the difference between these ratios should be within sampling error. Spearman (1914, page 105) stated:

> the two factors in success are quite distinct; firstly, there is the state of the particular group of neurons, their development and organization; and secondly, there is the whole cortex. The former may be called the 'specific' factor, as it is specific to that particular performance. The latter constitutes the 'general' factor, since it is required for all performances.

Thus, Spearman offered a primitive hypothesis of how his two factors depended on the physiology of the brain. Spearman also devised several approaches for extracting the loadings of test scores on a single factor (Vincent, 1953). Thus, Spearman could be considered the founder of single-factor analysis.

Thompson (1916) countered Spearman's argument with a dice-throwing experiment that showed that the hierarchy of intelligences could be produced without a general factor (*g*). Subsequently, Thompson (1919) conducted additional simulations using playing cards. These simulations were done by assigning values drawn from dice or cards to artificial variables that were used to generate simulated test scores. Thompson's experiments are probably the first Monte Carlo simulations of mental test performance. Thompson concluded that Spearman's evidence in favor of a general factor was by no means crucial and could be accounted for by multiple independent factors. This work may be seen as the first demonstration that Spearman's *g* may be a statistical artifact.

Thompson and Spearman continued to debate the issue of a general factor, and it is interesting that their positions evolved. In a series of meetings described by Deary et. al. (2008), Spearman appears to have accepted that there might be group level factors (i.e., factors common to a subset of tests) in addition to general and specific factors. Likewise, Thompson is described as leaning toward Spearman's *g* and is quoted as saying that "Surely the real defense of *g* is simply that it has proved useful" (Deary et. al., page 129). Although historical accounts may present the positions of researchers as static, the views of these two individuals, like those of many other scientists, changed over time.

Binet & Simon (1916) sought to develop a means of determining whether a young child was suitable for normal instruction or needed to be sent to a "special class". Their intent was to describe the current condition of the child and not to speculate about etiology or prognosis. Their approach differed from the work of earlier researchers such as Cattell & Galton (1890) in that of Binet and Simon used more complex tasks involving judgment and reasoning rather than simple laboratory tasks. Binet & Simon (1916, page 41) stated that "The scale that we shall describe is not a theoretical work; it is the result of long investigations…", and "all the tests which we proposed have been repeatedly tried, and have been retained from among many, which after trial have been discarded." Thus, Binet & Simon's scale was the result of an empirical investigation. Boake (2002) states that validity of their scale was based on the increase of scores with age and by the scale's ability to identify mentally impaired children. Binet & Simon (1916) did have thoughts about the nature of the characteristics that distinguished "mental defectives" from other children. They thought that judgment was primary, along with comprehending and reasoning well. On the other hand, they thought that memory was not important. As an example of the relative unimportance of memory, they described the case of a "backward" girl with an exceptional memory, what today might be called a savant. According to Boake (2002), Binet emphasized that a particular test in isolation is of little value and that the important information was in the subject's average performance over various tests. On the other hand, in describing what their tests measured, Binet & Simon (1916, page 52) stated that they "find it difficult to define which mental functions are being exercised because they are very numerous".

The Binet & Simon scale was subsequently translated and revised by Terman (1916) who added new items and standardized the test on a large group of children. This scale came to be known as the Stanford-Binet. Terman included items that discriminated between children of different ages and which correlated with the scale as a whole. While the practice of including items that correlated with the total score would seem to imply that the Stanford-Binet scale measured a single ability, Terman stated that a single test alone was not accurate since intelligence has many aspects. Terman (1916) suggested a number of applications for intelligence testing beyond identifying mental defectives. Terman (1918, page 165) asserted that his method "probes beneath the veneer of education and gives an index of raw brain power". Terman (1924) also spent considerable time emphasizing that the use of mental tests was a methodology that is equally appropriate for experimental psychology as is the use of controlled experiments. He noted that many of his contemporary experimental psychologists were in agreement with this position.

Yerkes (1917) criticized both the Binet scale and the Stamford-Binet scale based on differences from his “point scale”. Yerkes suggested that tests should be selected to evaluate basic psychological functions rather than being selected to discriminate between age groups. Yerkes’s tests of individual psychological functions were designed to be applied in isolation or as a group of tests. In addition, Yerkes asserted that a given test item should be applicable to all age groups and produce a continuous score. However Yerkes point scale ultimately did not enjoy the success of Binet and Terman’s scales.

Terman (1918) was also involved with the construction of tests used by the US army during the First World War. These scales were designed by a group of prominent experimental psychologists of that period led by Yerkes (Kevles, 1968). The tests developed by this group included the Alpha, which made use of verbal material and was designed for men who could read and write English, and the Beta, which was designed for those who could not. Both the army Alpha and Beta were group tests that involved multiple choice questions. According to Spring (1972), the criterion used to validate these tests was the ability to be a good soldier. This ability was assessed by correlating test scores with officer's ratings of their men in terms of practical soldier value. As such, Spring suggested that this form of validation paralleled Binet’s attempt to determining the suitability of children for normal school instruction. Spring suggested that both the army testing project and that of Binet and Simon evaluated the ability of individuals to function in a highly organized institutional structure. Kevles (1968) states that intelligence testing was not accorded much respect prior to the war but gained a considerable following as a result of this project.

Thorndike (1918) thought that psychologists tend to reduce the infinitude of tendencies to think and feel and act in certain ways in response to varied situations to a few tendencies called traits, abilities, and interests. Thorndike (1918, page 149) stated that "if the scale by which individuals are measured is very coarsely divided, their differences may be hidden". He also stated (Thorndike, 1918, page 158) that there might be "one single type or as many types as there are individuals" depending on whether one wanted to emphasize commonalities or differences between individuals. Thorndike (1921) viewed intellect in general as the power of good responses. He thought that it would not be wise to spend too much time trying to separate intelligence from emotional and vocational abilities. Thorndike also thought that ability varied according to the particular task. He viewed the value of test scores not in terms of their capacity to identify some general power residing in the individual but in terms of their ability to predict future performance. Thorndike believed that this prophesy was less accurate the more the test content differed from the skills that are to be predicted. Thorndike noted that tests current to his time "favored words, numbers, space-forms, and pictures, neglecting three-dimensional objects and situations involving other human beings." He suggested that this might be due in part to convenience. Thorndike developed a number of tests of specific abilities and achievement for domains such as reading (Thorndike,1914a) and mathematics (Thorndike, 1914b).

Thurstone (1934) noted that Spearman's method of tetrad differences often failed to show that there was only a single common factor that accounted for the correlations between tests in a battery. According to Thurstone (1934), the proponent of *g* would consider the tests as inadequate and discard them while the opponent of g would consider Spearman's theory as inadequate. Thurstone believed that neither conclusion was correct and that more than one general factor was necessary to account for observed correlations. Thurstone (1934) stated that "The multi-dimensionality of mind must be recognized before we can make progress toward the isolation and description of separate abilities." Thurstone devised the centroid method to extract these multiple factors (Vincent, 1953). Thurstone (1935) elaborated on methods for extracting multiple factors, noting that this problem has at least two parts. The first of these concerns the minimum number of factors that will account for the observe intercorrelations of test scores. The second concerns the minimum number of factors for each trait that will account for the correlations between test scores (i.e., each trait can be described by of the smallest possible number of factors). This second issue involves the rotation of factors. Thurstone (1935) believed that the solution to the problem of an infinite number of possible factor rotations was to be found in simple structure. Simple structure results when there are many factor loadings that are vanishingly small (i.e., a sparse matrix of factor loadings). Thurstone also developed methods for higher-order factor analysis.

At this time multiple factor analysis was very labor intensive and Thurstone & Thurstone (1941) acknowledged the efforts of several assistants in performing the manual calculations. Using these methods, Thurstone & Thurstone (1941) characterized six primary mental abilities they describe as clearly indicated: verbal comprehension; word fluency; space; number; memorizing; and inductive reasoning. In addition, they indicated that two more mental abilities were not as clearly defined: deductive reasoning; and perceptual speed. However Thurstone & Thurstone (1941, page 8) stated that:

> No one knows how many primary mental abilities there may be. We know about one memory factor now, but several new memory factors may be found.

In addition, they stated that "It should not be assumed that the primary mental abilities are elemental or indivisible." Thurstone (1940) thought that some of his primary mental abilities might be associated with physiology while others might result from experience and education.

Wechsler (1939) developed a battery of tests, the revised versions of which are currently the most popular scales of human mental abilities (Rabin et al., 2016). The Army group examinations were a major source of sub-tests and items used in the Wechsler-Bellevue scale. According to Tulsky et. al. (2003a), Wechsler's scale was unusual for the time in that he combined both verbal and non-verbal (performance) sub-scales. However, Frank (1983) describes Wechsler's initial scale as almost identical to the scales of several previous authors. Frank's table 1 shows five other scales with both verbal and performance tests. Wechsler (1958) recommended considering possible discrepancies between these verbal and performance tests. Wechsler (1958) describes the final selection of these tests as primarily based on three considerations: (1) that previous studies showed that the tests correlated with composite measures of intelligence; (2) that the tests as a group encompassed a sufficiently diverse content and (3) that the nature of the subjects' failures on the tests had diagnostic implications. Wechsler (1958, page 7) described his views of intelligence as follows:

> Intelligence, operationally defined, is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment. It is aggregate or global because it is composed of elements or abilities which, though not entirely independent, are qualitatively

differentiable. By measurement of these abilities, we ultimately evaluate intelligence.

Raymond Cattell (1943) described the literature of his time as containing a host of divergent definitions of intelligence. He proposed a scheme where adult mental capacity could be viewed as consisting of fluid and crystalized abilities. Cattell's fluid intelligence was a general ability to discriminate and perceive relations. Crystalized intelligence consisted of the habits formed previously through the use of fluid intelligence. However, once formed, crystalized intelligence no longer required the use of its fluid counterpart. This theory was subsequently refined to include nine factors accounting for both mental abilities and general personality dimensions (Horn & Cattell, 1966). Of these nine factors, those related to intellectual performance were fluid intelligence, crystalized intelligence, visualization, speediness, use of concept labels, and carefulness The nine factors in Horn & Cattell's solution were correlated, which these authors interpreted as being due to interactions during the individual's development.

Cronbach & Meehl (1955) summarized the conclusions of a committee of the American Psychological Association on the topic of test validation. As discussed previously, test developers had used a variety of procedures to select test items and rationalize their use. Various criteria included: predictive validity, the ability of a test or battery to predict (i.e., correlate with) some future performance of interest; concurrent validity, the correlation of a test with similar measures taken at the same time; content validity, the extent to which test items are a sample of the trait or abilitiy the investigator is interested in, as determined deductively; and construct validity, which includes many forms of the evidence that supports interpretation of a test, including theory. An important feature of construct validity was an emphasis on theory in the validation process. Cronbach and Meehl (1955) described construct validation as a continuing process, where tests are not validated but rather supported by evidence of their validity. Cronbach (1957) also suggested that the process of construct validation should include experimental research. He noted that investigators using experimental methods had become quite separate from those using correlational methods. This situation is quite distinct from the time when mental tests were initially being devised by experimentalists such as Terman. Cronbach (1957) felt that the two disciplines could mutually benefit from interacting. However, it is not apparent that this interaction occurred at that time.

Vincent (1953) asserted that factor analysis was not a statistical technique, owing to guesswork involved in its implementation. This was a common belief among statisticians at the time. Factor analysis developed within the field of psychology, quite apart from statistics. The guesswork involved in factor analysis included the necessity of estimating commonalities and the fact that Thurstone's centroid method did not produce unique solutions. These problems were overcome by Lawley & Maxwell (1962) who showed that the factor problem could be solved by use of maximum likelihood. This approach involved finding the matrix of factor loadings that minimized the difference between the common factor model and the observed correlations. Lawley & Maxwell's method also provided a statistical method of determining the number of factors to retain based on the difference between the predicted and observed correlation matrices. The structural equation modeling approach of Joreskog & van Thillo (1972) can be regarded as an extension of Lawley & Maxwell's method that allows for evaluation of many alternative models that are specified by the analysist. These methods became practical for widespread use with the advent of modern computing hardware and software.

One of the more prolific collections of mental abilities was proposed by Guilford (1956) who suggested that there were at least 40. Guilford's approach involved first postulating the existence of some unitary ability and then selecting or designing tests to measure that ability. Guilford believed that these separate abilities could be grouped into a number of classes. Major divisions were between cognitive factors (i.e., discovery), production factors (i.e., convergent and divergent thinking), evaluation factors, and memory. Within each of these factors Guilford conceptualized matrices formed by the intersection of content and processes. Guilford noted that empty cells in his matrices suggested the existence of additional mental abilities. Guilford (1956, page 285) stated that "parsimony has led us in the past to the extreme of one intellectual dimension, which everyone should now regard as going too far in that direction".

Kirk & McCarthy (1961) devised the Illinois Test of Psycholinguistic Abilities (ITPA) to provide a means of assessing children with specific disabilities in areas such as language, perception, or behavior. Children with learning disabilities were conceptualized as having wide discrepancies among abilities that result in a failure in academic subjects. Kirk (1968) felt that diagnostic tools such as the Stanford-Binet were inadequate for this purpose. The ITPA consisted of 9 subtests evaluating three processes (decoding, association, and encoding) in each of two channels (auditory-vocal and visual-motor) at two levels of organization (symbolic and non-symbolic). As such the ITPA represented an initial attempt to assess specific abilities in children with specific learning problems.

Elwood (1969) discussed the possibility of automating psychological testing. He noted that in addition to saving time, automation could potentially standardized presentation of test items, improve accuracy in recording and scoring responses, allow measurement of new dimensions of the response process, and increase the reliability and validity of test results. He described an automated system that could administer most of the WAIS sub-scales. This automated system produced results similar to manual testing. Elwood's apparatus was based on technology that was primitive by today's standards. Yet despite the immense progress in technology, automated testing of mental abilities has yet to become the norm.

Daneman and Carpenter (1980) used individual differences in reading comprehension to validate measures of working memory. This work led to a renewal of interest in individual differences and correlational approaches within the field of experimental psychology (e.g., Engle et. al., 1999). Much of this work was concerned with the relationship between fluid intelligence and working memory (e.g., Fry & Hale, 1996), a construct originating from cognitive psychology. This trend represents a return to the study of individual differences by experimentalists as practiced by Terman and suggested by Cronbach.

Carroll (1993) re-factored correlation matrices from a very large series of prior studies. Carroll's method involved oblique rotation of initial correlation matrices followed by higher-order factoring of the first-order factor correlation matrices, and on occasion, a third step in this process. From these results Carroll (1993) developed a three-stratum theory consisting of narrow, broad, and general factors. Carroll interpreted the factors at each level, a practice he regarded as theory construction. The highest, or third stratum level, was interpreted as general intelligence (i.e., *g*). Second stratum abilities were interpreted as fluid intelligence, crystalized intelligence, memory, visual perception, auditory perception, retrieval ability and cognitive speediness. Carroll also noted that additional second-order factors might be identified. According to Carroll (1993, page 68):

> It is clear that all the leading figures in psychometrics – Binet, Spearman, Thurstone, and Guilford (to name but a few) – have had an abiding concern for the nature of intelligence; all of them have realized, too, that to construct a theory of intelligence is to construct a theory of cognition.

Carroll's also described these early theorists as having common sense explanation of traits rather than well-developed cognitive theories.

Carroll's views on mental abilities are illuminated by several exchanges he had with his contemporaries. One of these was with Kranzer

& Jensen (1991) over whether *g* was unitary. Kranzer & Jensen (1991) showed that four principal components derived from a battery of "elementary cognitive tasks" independently contributed to the prediction of an estimate of *g*. They reasoned that since these components were orthogonal, *g* must consist of at least four distinct components. Kranzer & Jensen concluded that their results were consistent with the hypothesis of Detterman (1982) that intelligence is the result of a set of orthogonal variables. Carroll (1991) argued that Kranzer & Jensen (1991) only had an estimate of *g*. He produced hypothetical example factor matrices that showed how their results could be produced given a unitary *g* and argued that it was more parsimonious to assume this. Carroll (1994) also had an exchange with Humphries (1994) who presented what he called a behaviorist's view of intelligence, reminiscent of that of Thorndike. Humphries asserted that intelligence is the acquired repertoire of all intellectual skills and knowledge available to an individual at any particular point in time. Humphries also stated that intelligence was not necessarily anything more than a mathematical dimension. Carroll (1994) countered that cognitive abilities such as intelligence are real entities in the individual rather than an acquired repertoire of skills.

The Neuropsychological Approach

Finger (1994) suggests that the field of neuropsychology can trace its roots to Democritus (ca. 460-370 B.C.), who held that rational functions were controlled by the head, and Galen (A.D. 130-200), who associated intellect with the brain. An ancient Egyptian papyrus described the association of head injury with a loss of speech (Sondhaus & Finger, 1988). However, Broca's description of a patient with loss of speech due to a lesion in the left inferior frontal cortex was a particularly critical event in thinking about localization of function that foreshadowed the beginnings of modern neuropsychology (Finger, 1994).

Halstead (1947) distinguished between several forms of intelligence, including psychometric intelligence and biological intelligence. Halstead's distinction between these different forms of intelligence was predominantly in terms of how they were validated. According to Halstead, psychometric intelligence, as measured by test batteries such as the Stanford-Binet, is validated in terms of sociological criteria such as educational attainment. In contrast, Halstead's biological intelligence is validated in terms of sensitivity to brain pathology. Halstead stated that the relationship between these two forms of intelligence was an empirical matter. Halstead (1947) developed a battery from 27 tests that were given to an assortment of individuals both with and without brain pathology. Halstead notes that Thurstone conducted a factor analysis of these data that produced a four-factor solution. Halstead also derived an impairment index that was a composite of these individual factors. Halstead concluded that psychometric intelligence was unrelated to frontal lobe functioning based on his review of studies examining the effects of frontal lobotomies on the Stanford-Binet. In contrast, he reported large effects of frontal lobectomy on his impairment index. From results such as these, Halsted concluded that the frontal lobes were particularly important for biological intelligence. He also concluded that psychometric intelligence, as measured by the Stanford-Binet, was unrelated to biological intelligence. He stated that “one wonders in what sense the term “intelligence” can properly be applied to this test” (Halstead, 1947, page 141).

Halstead’s student, Ralph Reitan, continued this work (Reitan, 1956). In addition, Reitan (1955) added new tests to Halstead’s battery. In contrast to Halstead’s approach, Reitan (1964) did not rely on factor analysis and used profiles of individual test scores to characterize the effects of different forms of brain damage. The goal of this type of analysis was to provide a basis for diagnosis of the characteristics and location of lesions in patients with brain damage. This work provided a basis for the development of the Halstead-Reitan battery that has been used extensively for neuropsychological assessment. However, the ability of these tests to diagnose the nature and location of brain damage proved elusive. For example, in reviewing the effects of brain damage on several of these tests, Reitan and Wolfsen (1994) concluded that the evidence for the popular belief that they were specifically sensitive to frontal lobe damage was tenuous. In contrast, neuroimaging techniques were developed that proved very successful in localizing and characterizing brain damage (Doyle et. al., 1981). With these considerations in mind, Leonberger (1989) suggested that the goals of neuropsychological assessment should change to characterizing the nature of a patient’s cognitive deficit rather than the nature of the patient's brain pathology.

Teuber (1955) described double dissociation as a method to establish localization of function. This method involves showing that two brain regions are functionally dissociated by two behavioral tests, each test being affected by a lesion in one area and not the other. This method was subsequently used in a large number of neuropsychological studies to characterize the nature of deficits in individuals with brain damage. Reviewing this literature, Shallice (1988) suggests that double dissociations indicate that there are functional specializations for sub-processes in perception, memory, speech, and output systems. Double dissociation is a method that is not without its critics however. For example, Plaut (1995) produced a double dissociation by “lesioning” an artificial connectionist network that lacked a modular structure. Plaut asserted that these findings called into question the theoretical implications of reliance on single-case studies.

Scoville and Milner (1957) described cases of profound loss of memory following removal of the hippocampus for the treatment of intractable epilepsy. This memory loss was described both in terms of causal descriptions of the patient’s behavior and also in terms of marked differences between the Wechsler intelligence and memory scales. Subsequent work by Toal (1957) and others (Erickson & Scott, 1977) questioned the adequacy of the Wechsler memory scale for characterizing organic memory problems. Toal described the Weschler memory scale as being based on an ambiguous “common-sense” definition of memory that did not provide a means of determining what was being measured. Scoville and Milner’s report led to a considerable amount of research aimed at characterizing the nature of amnesic deficits and identifying dissociable memory disorders (Butters et. al., 1995). This research resulted in development of specialized memory test batteries (e.g., Delis et. al., 1991) as well as revisions of the Wechsler memory scale (Kent, 2017).

Das et. al. (1975) advocated a process-based approach to understanding mental abilities. They asserted that the brain was the source of cognitive functioning. They based their conceptual model on the information processing approach of Luria and devised a scale to measure these processes (Das & Naglieri, 1997). Interestingly, Golden et. al. (1978) developed an alternative scale based on their understanding of the work of Luria. The scale of Golden et. al. (1978) is an attempt to standardize administration and scoring of Luria’s tests that cover a large assortment of functions. The scale developed by Das and Naglieri (1997) is more focused on Luria’s theoretical account of simultaneous and successive processing. Neither of these test batteries appears to be widely used by neuropsychologists at present (Rabin et,. al., 2016). Tupper (1999) describes these and a number of other approaches to test development as “neo-Lurian methods”.

Chase et. al. (1984) examined the relationship between WAIS scores and cerebral glucose metabolism in patients with Alzheimer’s disease and controls. They report that verbal subtest performance was associated with activity in left parasylvian areas and performance subtest performance was associated with right posterior parietal areas. This study pioneered the use of modern neuroimaging methods for the localization of brain areas associated with individual differences in mental abilities. In addition, relationships between performance on cognitive tasks and neuroimaging has resulted in the generation of novel hypotheses, such as the concept that superior performance is associated with greater mental efficiency (Deary & Caryl, 1997).

Livingstone & Hubel (1987a) presented evidence for separate processing of form, color, movement, and depth in the primate visual system. They showed that human psychophysical data also supported the concept of separate channels of processing for these same visual features (Livingstone & Hubel, 1987b). Their research did not focus on individual differences. However, this theory of separate processing streams proved to be extremely influential in orienting subsequent research in neuroscience towards considering how information is processed in specialized networks. This led to a move away from the strict localizationist thinking that had been the focus of investigators such as Reitan. The concept of "selectively distributed processing" may be seen as a resolution in the debate between localization of function and equipotentiality (Mesulam, 1998).

Kaplan (1988) outlined the Boston Process Approach which was based on the premise that different individuals could arrive at a solution to a problem in different ways. From this perspective it follows that examination of the source of errors is of considerable importance in describing test behavior. Kaplan suggested that most standardized tests were multifactorial. She described cases where errors on tests such as the WAIS-R block design and word-finding difficulties on the Boston Naming Test could result from different strategies used by different individuals. Kaplan advocated examining qualitative differences in behavior that are not apparent when only considering outcomes. Libon et al. (2013) state that Kaplan's approach to identifying the processes used by individual's to solve problems was influenced by her background in Gestalt psychology. They also indicate that Kaplan thought that there was no fundamental difference between assessment and experimentation in neuropsychology. Both assessment and experimentation involve hypothesis testing. Test batteries developed in the tradition of the Boston Process Approach (the Delis-Kaplan Executive Function System and the California Verbal Learning Test) are currently commonly used by clinical neuropsychologists (Rabin et. al., 2016).

Issues and Trends

Interpretations of the history of abilities testing have varied over time. Tyler (1965) noted that early mental testers had widely divergent views as to how to define intelligence. This included notions of judgment, abstract thinking, ability to learn, etc. Tyler (1965, page 62) stated that "The thing that saved psychology from bogging down in a mire of semantic confusion was the predominately practical orientation of mental testers." This view reflected the practical orientation of that time that emphasized prediction of diverse domains of performance. Carroll & Maxwell (1979, page 608) stated:

> A persistent tension has existed between those who believe that human cognitive capacities can be well summarized in a single global concept of intelligence and those who prefer to emphasize the multidimensional character of the concept. The bulk of recent research is predicated on a multifactorial view...

More recently, several reviews of the history of research on human mental abilities have emphasized the progressive nature of theory development in this area (e.g., Flanagan et. al., 2014; Schneider & Flanagan, 2015). This view reflects the current interest in the use of individual differences for theory development. The Cattell-Horn-Carroll (CHC) model is currently popular. Schneider & Flanagan (2015, page 323) state:

> CHC theory is not so much a new theory but an elaboration of very robust findings that were first discovered by Spearman, Thurstone, and many other early researchers.

Geisinger (2000) has described psychological testing as continuing to evolve rapidly. Likewise, Riley et. al. (2017, page 38) state:

> Each new version of the Wechsler Adult Intelligence Scale (WAIS) or Wechsler Intelligence Scale for Children (WISC) involves changes based on findings that have emerged in the basic and applied literatures on the nature of cognitive functions.

In contrast, in his discussion of the history of the WAIS, Boake (2002) describes this most common test of intelligence as remaining almost unchanged through various revisions. He notes that many of the WAIS sub-scales had been around in one form or another for quite some time before the Wechsler-Bellevue was assembled. According to Boake (2002), all of the Wechsler-Bellevue subtests except for Block Design were derived from the Army tests. There have been relatively few changes in content since that time. For example, the only change in content for the 3rd revision was the addition of Symbol Search as a supplemental sub-test (Tulsky et. al.,2003b). Most of the changes were related to updating norms and providing new ways to compute index scores. Boake (2002) states:

> From a historical perspective, the Wechsler- Bellevue Intelligence Scale is a battery of intelligence tests developed between the 1880s and World War I. In their origins, the Wechsler subtests represent the major pre-World War I approaches to cognitive assessment.

Thus, major changes to the WAIS scales have come in interpretation rather than content.

Public controversy over the testing of mental abilities has "waxed and waned" ever since the initial development of these tests (Cronbach, 1975). These controversies include issues of racial, class, and immigrant differences in intelligence test scores. Cronbach (1975, page 11) attributes these to:

> journalists mining scholarly reports for controversial copy, distorting the original to make it more exciting, pointing up disagreements, and sometimes reporting only the iconoclastic side.

However, it is also clear that psychologists have made controversial public pronouncements of their own accord, such as the book by Hernstein & Murray (1994) that discussed, among other things, racial and class differences in intelligence. According to Walsh et. al. (2014, page 245) "Galton, Jackson, Maudsley, and Spearman each invoked science to justify maintaining the societal status quo." According to Jensen (2000), Spearman thought that the measurement of g could be used to determine whether individuals were qualified to vote or have children. While Thurstone believed in the diversity of abilities, he also stated that:

> If the facts support the genetic interpretation, then the accusation of being undemocratic must not be hurled at the biologists. If anyone is undemocratic on this issue it must be Mother Nature" (Thurstone, 1946, page 111).

Neisser et. al. (1996) point out that uncertainty about the nature, origins, and measurement of intelligence make overall generalizations about these issues inappropriate. Neisser et. al. (1996, page 97) conclude:

> In a field where so many issues are unresolved and so many questions unanswered, the confident tone that has characterized most of the debate on these topics is clearly out of place.

According to Cubelli (2005, page 273), the history of neuropsychology:

> could be described either as a sequence of completely new approaches that substituted the preceding ones, or as a pendulum like movement, in which