



BIROn - Birkbeck Institutional Research Online

Lorch, Marjorie (2022) Defining 'Normal': methodological issues in Aphasia and intelligence research. *Cortex* 153 , pp. 224-234. ISSN 0010-9452.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/48163/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

Defining 'Normal': Methodological Issues in Aphasia and Intelligence Research

Marjorie Perlman Lorch, PhD

Department of Languages, Cultures and Applied Linguistics, School of Arts

Birkbeck, University of London

26 Russell Square, London WC1B 5DQ, United Kingdom

m.lorch@bbk.ac.uk

ORCID: 0000-0001-8505-8815

Abstract

Theodore H. Weisenburg (1876-1934) and a series of colleagues embarked on a research program in 1927 to develop standardized tests to investigate the nature of language and intellectual impairments in aphasic and non-aphasic individuals. This project culminated in two significant contributions to neuropsychological testing (Weisenburg and MacBride, 1935; Weisenburg, Roe and McBride, 1936). After an initial study demonstrated the problematic aspects of Henry Head's aphasia tests (1926), Weisenburg developed a new battery of tests which were given to individuals with aphasia. The significant innovation of this work was the original concept of a matched control group. This included those with other neurological impairments, and a range of non-neurologically impaired individuals with the aim of providing a characterization of what was 'normal'. They identified many crucial participant variables regarding age, education, and socioeconomic status and used population statistics to ensure their control sample was representative. A detailed critical assessment of each of their successive elaborations is examined, focusing on the methodological innovations they represent. The contribution of this work to contemporaries and successive generations of neuropsychologists is examined regarding ongoing issues in clinical testing and research design.

Keywords: aphasia, methodology, normal, assessment, control group

Funding: This research did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

1. Introduction

The term 'normal' is generally avoided today in clinical and behavioural research in favour of alternatives such as healthy, neurotypical or unimpaired (Fenton & Krahn, 2007). In the 19th and 20th century, the word was generally used to designate something or someone that constituted or conformed to a type or standard which represented regular, usual, typical, or ordinary features. In the more specific usage referring to a quality of people, it was used to mean someone who was "physically and mentally sound; free from any disorder; healthy" (Anonymous, 2021). In addition, there was, and still is, a use of the term in statistics to refer to a distribution with specific properties where the mean, median and mode are the same (normal curve). When used in the study of a group of people, this is referred to as a normal sample of individuals. Interest in determining what was normal for a range of physical and psychological characteristics began with the first wave of big data studies in the second half

of the 19th century (Cryle & Stephens, 2017) while the testing of individuals' cognitive abilities began to develop in the first years of the 20th century (White, 2000).

The present study explores the development of the first standardized neuropsychological test battery devised in the 1920s and 1930s to determine the nature of selective language and intellectual impairments due to acquired neurological damage. The focus is on the activities of the Philadelphia based neurologist Theodore H. Weisenburg (1876-1934) (sometimes spelled Weisenberg) (Figure 1) who embarked on a decade long project into patterns of impairment in language and other cognitive abilities in individuals with aphasia. These efforts were original in creating the methodological refinement of a control group to use as a comparator against the performance of the research group of interest.

In a series of studies, Weisenburg and his collaborators tested the performance of a variously constituted groups of aphasic and non-aphasic individuals. They refined the procedure for selecting an appropriate reference group for such clinical research. In the process, they created one of the first standardized set of language and intelligence tests. Weisenburg argued the need to determine the characteristics of normal performance is to inform the interpretation and analysis of the abnormal rather than relying on subjective judgement.

This pioneering interdisciplinary research project significantly contributed to the development of methodological practice in clinical behavioural research. It established the practice of testing the performance of unimpaired individuals and developed a rationale for identifying the relevant participant characteristics for a demographically matched control group in neuropsychology research. A detailed critical consideration of each of the iterations in their research design is presented below with a focus on the methodological innovations they represent. Considerations of current practice in neuropsychology research will be discussed.

1.1 Historical Context

Sir Henry Head's (1861-1940) two volume book *Aphasia and Kindred Disorders of Speech* (1926) was the first major contribution to the subject in the 20th century. The basis of this work was a detailed study of a group of brain damaged individuals' performance on various tests. It was the culminating work of his long career. Head was trained at the National Hospital for the Paralysed and Epileptic, Queen Square, London (now called The National Hospital for Neurology and Neurosurgery) and was influenced by his senior colleague and expert in aphasia, John Hughlings Jackson (1835-1911) (Swash, 2008). Like most English hospitals in the 19th century, the National typically treated patients who were poor, in general poor health, and poorly educated. Those higher on the socio-economic scale were seen in private practice or attended at home and were not typically the subjects of research.

One of the earliest examples of a case series study of the newly described syndrome of aphasia was by Jackson (1864). He recorded the details of 35 hospital patients for etiology (diverse), sex (majority of women), age (ranging from childhood to old age), and marital status. The report also notes the occupations of the men (e.g., plasterer, bargeman) and the presence of chronic conditions (rheumatism, syphilis, alcohol abuse). These patient characteristics are representative of aphasia research reported in the second half of the 19th century in Britain. Head's book represented a significant departure from this approach. Although his primary focus was on the theoretical concept of symbolic behaviour, he also made a novel methodological choice by using 26 military officers who all had suffered head wounds in World War 1 as participants. In contrast to earlier studies, this group was relatively fit, young, well-educated, and privileged. Head argued that the merit of this

homogenous group was there were fewer potentially confounding factors than those in previous studies. An extensive set of tests was also devised to assess a range of language and cognitive performance. His methodological concerns were evident in the detailed description of his procedures and the use of multiple trials to determine the in/consistency of performance.

Head's methodological innovations in behavioural testing were widely lauded on publication (e.g., Bartlett, 1926). However, another review of Head's book, while also overwhelmingly positive, included a challenging criticism:

"It is a question whether the classification that Head has made can stand the test of experience. A fundamental error lies in the fact that his tests were not checked up on normal persons; at least, there is no mention of it. The reviewer questions whether any series of tests can ever adequately evaluate in all patients a standard method of measuring disturbances of speech." (Anonymous, 1927p. 1063)

This challenging statement appeared in the American journal *Archives of Neurology and Psychiatry (Archives)*. Although it has not been possible to determine the identity of the reviewer, Weisenburg was the Editor in Chief of this journal. Moreover, he had been awarded a grant from the Commonwealth Fund to investigate aphasia himself. The precise timing of events is unclear, but it appears that Weisenburg had been planning this work a year before Head's book was published (Weisenburg, 1929b).

Weisenburg was Professor of Neurology at the Graduate School of Medicine, University of Pennsylvania and held clinical appointments at the Philadelphia Orthopedic Hospital and Infirmary for Nervous Diseases and the Philadelphia General Hospital (Tilney, 1934). Weisenburg's career path has many similarities to Head's. He also pursued research in sensation and perception. He had an interest in aphasia from the earliest point in his career, while he was the intern to the Philadelphia neurologist Charles K. Mills (1845-1931) (Mills, 1904). Weisenburg served as a Neurosurgeon during World War I and also studied shell shock (Weisenburg, 1919). Like Head, Weisenburg produced his major work on aphasia at the end of his career.

Weisenburg's research into language and cognitive impairments in aphasic individuals was in part an attempt to address what he saw as serious methodological limitations in Head's approach. This led to significant and original refinements in the design of clinical research that influenced practices throughout the 20th century, while concern about the constituency of a control group is an ongoing concern. His efforts to define a suitable reference group for neuropsychological research will be considered in detail below.

2. Weisenburg's consideration of what constitutes normal performance

2.1 The initial attempt

Weisenburg's first steps were directly triggered by Head's work on aphasia. Head had pursued a particular research strategy that Weisenburg found antithetical. Head argued that evidence from his homogeneous group of young aphasic head injury participants had merit over the heterogenous set of individuals employed in previous studies. Head felt that this approach would allow him to discover the theoretical nature of aphasia. Weisenburg took the opposite view. He insisted that the widest possible sample was needed to understand aphasic impairments. He granted that the strength of Head's work was in the meticulous description of his testing procedures but criticized it for not testing individuals without

aphasia. He argued that there was no standard for determining what constituted an impaired performance, and that its identification had rested solely on the examiners' subjective judgement.

Weisenburg's overarching aim was to provide a characterization of impairments in aphasic individuals and improve procedures for clinical diagnosis. However first, he needed to address the difficulties created by the absence of data on unimpaired people that could be used to compare to those with neuropathological conditions. In early 1927, he tasked two of his neuropsychiatry students to administer Head's tests to the six others in their group. They found a range of failures were made by these young, healthy, and well-educated individuals who had been selected to be demographically comparable to Head's aphasic participants. This raised a puzzle. How could such a test battery be used to identify neurological impairment if unimpaired individuals could not successfully perform the tasks? The absence of any clear determination of the range of performances demonstrated by healthy individuals was essential. Weisenburg argued that evaluative judgments on aphasic individuals' responses could not be made without having an adequate knowledge of what the 'normal' responses would be under similar conditions.

Weisenburg presented the results of the study with the six medical trainees along with an additional 23 non-aphasic individuals at a meeting of the American Neurological Association a few months later. Weisenburg subsequently published this with the two students Gerald H. J. Pearson (1893-1969) and Bernard J. Alpers (1900-1981) as co-authors (Pearson et al., 1928). Weisenburg and his colleagues identified the research problem as follows:

“...it must be perfectly clear what constitutes an aphasic response as compared with a normal reaction; or, to put it briefly, there must be standards for comparison, so that the interpretations of results in examining various types of aphasia may be accurate. It is essential to know, therefore, exactly what constitutes a normal response, what may be looked on as a deviation from this normal, and how a normal person without demonstrable speech disorder will respond to the various tests used in the study of aphasia. If this and the errors a normal person will make are known, it is possible to decide accurately what constitutes error in an aphasic response, and to evaluate these responses properly in the study of speech.” (Pearson et al., 1928, p. 281)

In this way, they hoped to identify aspects of performance which were due to non-linguistic factors related to the characteristics of the tasks or participants. Furthermore, they questioned whether poor performance on linguistic tasks could be a general consequence of brain damage. In selecting non-aphasic participants, they considered age, education, socio-economic status and type of illness. The objective was to collect a baseline range of performance that could be used to judge if aphasic individuals' performance was impaired.

The 29 normal participants were classed into several groups. Examination of the rationale for their inclusion and the choice of demographic variables reveal Weisenburg and colleagues' underlying assumptions. This will be discussed with respect to today's standards of rigour to highlight their methodological originality in identifying relevant control factors.

Weisenburg outlined his participant selection criteria as follows:

“The control material, therefore, consisted of two main groups: persons with superior and average intelligence without organic nervous disorder, and persons with organic brain disease of different types. The subjects in both these groups were without speech disorders. In selecting the control material, an attempt was made to obtain young persons without nervous disease, in order to perceive what reactions healthy, non-aphasic persons would give. In addition, however, patients with organic brain

disease but without disturbances of speech were selected in order to approach more closely the condition of the aphasic patient who has brain disease of a highly localized nature." (Pearson et al., 1928, p. 283)

This rationale indicates an appreciation of the utility of two types of control groups: the young and healthy; and the demographically matched neurologically impaired. The non-aphasic participants were divided into six subgroups ranked by general intellectual abilities and socioeconomic status. In the first group, labelled "Superior Intelligence", were the six graduate students in neuropsychiatry who served as a match to Head's young war wounded officers in terms of social strata and intelligence. They were clearly healthy, and what today would be described as 'neurotypical', but with the additional minority characteristic of a very high level of education. Weisenburg points out that they were highly motivated and acknowledged that these students were familiar with these psychometric tests from their training. Weisenburg shows methodological sensitivity in identifying these factors as sources of potential confound.

The second group was described as having "Average Intelligence". There were nine young adult patients from the Philadelphia Orthopedic Hospital (which included nervous diseases). Although designated as "young", their ages are not reported. Their intelligence level was determined using education as a proxy measure. This was described as the upper grades of the primary school, with "several" who had completed the first two years of high school. This would be typical of the general population at a time when secondary school was still limited to the higher socio-economic classes (Cubberley, 1934). However, these participants were not healthy. Three had chronic organic disease of the nervous system (two with chronic encephalitis; one with epilepsy), while the reason the other six were in the hospital is not reported. There was no consideration of how the chronic encephalitis or epilepsy might have affected mental development, and no details of the age of onset of illness are provided as would be expected today.

Group three comprised seven left hemiplegic patients from the Philadelphia General Hospital. These were described as elderly men with social and intellectual levels much inferior to those in the second group. However, Weisenburg specifically selected them as comparable matches for the aphasic individuals, as they had similar neuropathology and were drawn from the same demographic population. This group represents what would become the typical control for studies of aphasic individuals in clinical research.

They also included seven individuals classified into three separate groups. Two patients with "generalized cerebral arterio-sclerosis" were considered distinct from the left hemiplegic patients because their intellectual and social status was judged to be higher. Four elderly patients with "stationary tabes" without signs of cerebral involvement were classed as having a "restricted intellectual background". Finally, one patient with "general paralysis" was placed in a separate group and only a partial examination was completed. It is interesting that these seven individuals were not classed together as having chronic and diffuse neuropathology, in contrast to those with focal left hemisphere involvement. Instead, Weisenburg and colleagues saw them as having distinct characteristics that needed to be considered separately.

There is no mention of the participants' sex, however the male pronoun is used throughout. More surprising is the fact that the participants' handedness was not documented, as this variable had been considered relevant in accounts of aphasia and the organization of brain-behaviour relations since the 1860s (Harris, 1993). The assembly of these groups of individuals has some interesting implications. First, there is a group of healthy individuals of high intelligence intended to provide a sample that matched Head's. However, in this initial study no other 'normal' participants were included. Use of this type of control group is still in

widespread use today where drawing on university students is common. The other participants were hospitalized and suffering the effects of illness primarily affecting different parts of the nervous system but without aphasic symptoms. For these there was an attempt to include individuals with a wide range of ages and levels of education.

The investigators chose to report their results in a way that emphasized the patterns of performance by the different groups on Head's tests. It is notable that the errors were analysed in detail, but only presented as frequencies of errors made by a given group rather than by individual. For example, they observe that for the coin tests, errors were made by three out of seven participants with left hemiplegia, while for the tasks involving common objects, all successfully completed it except for three participants from this group. It is unclear whether these are the same or different from those who had difficulties with the coin tasks as participants are not individually identified. This manner of creating composite scores by groups of unequal numbers of individuals makes it difficult for any closer inspection or reanalysis of their findings. This first attempt to report performance by groups was a somewhat idiosyncratic approach and would not be considered viable by today's analytical practice. However, they do make some important observations which have remained of interest in neuropsychology.

For example, with respect to body schema and executing actions to visual copy, and spoken or written command, they report that 'touch your nose with the right hand' was the only item successfully carried out by all participants, while errors were made by all their participants in performing simple commands with the left and right hands and sides of the body. They document that half of the graduate students made mistakes on copying the movement of touching the left eye with their right hand, and eight out of nine from the "average intelligence" group could not correctly complete these visual imitation tasks. Errors were made by the majority in all the groups of participants to oral command and visual copy. The written commands generated perseverative errors by many participants. The fact that all the items were presented together on one card was felt to make it difficult for participants to keep their place. This they identified as a procedural shortcoming to be rectified in future studies. Generally, they conclude that these tasks were not useful for identifying language impairments but should be used to investigate apraxia. This became a common component of neuropsychological testing in the second half of the 20th century (Poeck & Orgass, 1971; Semenza & Goodglass, 1985).

Weisenburg and colleagues also observed many errors on other tasks, some of which are still in widespread use today. Errors were made on the letters of the alphabet which were attributed to their participants' low levels of education and literacy rather than difficulties with language per se. This task was also felt to require a degree of sustained attention. The authors conclude that it is not a test to be used to identify aphasia. No groups' performance was found to be error free on Head's tests of reading and writing sentences from pictures. The analysis of the pattern of errors on other tasks indicated the effects of age, education, and artefacts due to the stimulus cards. While all but one individual could recognize and name colours, none successfully performed the clock task. In the development of neuropsychological testing, the value of clock drawing and clock reading tasks was recognized for revealing a range of different neuropsychological deficits through a qualitative investigation of erroneous performance (Freedman et al., 1994).

The overall findings of Pearson, Alpers and Weisenburg's study were considered in a contemporaneous notice (Fleming, 1928). It highlighted the frequent mistakes made by the control participants which were no different from those made by aphasic patients. Fleming concluded that many of Head's tasks were "almost worthless" and appear to simply reflect

levels of intelligence. However, he still suggests that Head's tests are "as good as, if not better than, any others devised for the testing of aphasia" (Fleming, 1928, p. 994).

Weisenburg would work to improve on both the selection of control participants and the standardization of test scoring over the next six years.

2.2 Weisenberg argues the value of normal

Weisenburg presented further developments of his project at the New York Academy of Medicine on April 9, 1929, and subsequently published (Weisenburg, 1929b). He asserted the value of determining the errors made by normal people, arguing that this would lead to important refinements in clinical testing. His preliminary work had led him to conclude that when testing speech functions, other psychological processes must also be investigated. In addition to the careful selection of tests and use of normal control data to judge abnormal performance, Weisenburg suggested that the psychological state of both the participant and examiner must also be considered. If the patient is tired or "in no mood" to be tested, or the examiner is weary, impatient or "not in accord with his patient" the results will be "misleading" (Weisenburg, 1929b, p. 380). These are discerning considerations of bias in clinical testing.

The proceedings also included a transcription of the discussion that followed Weisenburg's presentation. The carefully qualified but primarily negative recorded comments capture the views of some elite clinicians of the day and reveal a range of reactions to his efforts to study the normal. It appears that Weisenburg had hit upon an issue that others were actively reflecting upon but unwilling to acknowledge as a useful endeavour. Frederick Tilney (1875-1938), a close colleague of Weisenburg's, remarked that he had been questioning what can be considered a normal performance in the context of aphasia with respect to sensation, perception, movement, and general intelligence. He had observed that experience and education will affect these functions in normal people. At the same time, Tilney sounded a note of warning against an attempt to establish "the norm". He referred to derogatory comments by the British physician Leyden in the *Lancet* about the American drive for medical standards and the use such indicators. Tilney cited Leyden's point that the Babinski sign can be found in healthy runners as an instance of overinterpretation of certain observations. (The author has been unable to identify this publication). Tilney's argument was seconded by Walter Timme (1874-1956) who suggested that if a symptom can be found in both healthy and pathological individuals, then the dividing line between them and the determination of normalcy was not possible.

These sentiments were echoed by Smith Ely Jelliffe (1866-1945) who remarked that while it might be considered desirable to "seek units and definite criteria", this effort was futile. No matter how many people were tested, this would not capture the variability present in the multitude of people in the world. He acknowledged the desire to characterise absolutes but insisted the effects of culture and training prevented this. He belittled the value of Weisenburg's efforts: "We have learned this evening something about "normal" persons for coin discrimination in New York, as contrasted with Philadelphia; and other kinds of normal persons." (Weisenburg, 1929a, p. 1299)

On a different tack, Jelliffe raised a concern about the limited and subjective basis of people's notion of normal. He argued from a psychoanalytic perspective that a person's idea of normal was largely created by their perception of themselves and didn't include those that were different from them. He stressed the distinction between averages and ideals. Rather than framing the normal, this should be done in terms of variables and variations. He reminisced about his folly of youth. Under the influence of the Italian criminologist Cesare Lombroso's (1835-1909) ideas, he tried to ascertain what a normal ear was as compared to

ears of insane and degenerate people. Jelliffe recounted how he began to see pathology everywhere. He then vowed to focus on variation rather than normal, as the latter always has the implication of pathological.

In his reply to the points raised by Jelliffe, Weisenburg suggested that because individuals based their notion of what is normal on limited subjective experience this was even more reason to collect objective evidence of the range from the population. His final comment underscores his commitment to the necessity of determining the normal for clinical practice:

“So far as I am concerned, this work has made me much more careful in my clinical examinations and I find that with knowledge of the type of errors that normal persons make, my clinical interpretation is better and I am likely to make fewer mistakes. Moreover, I have learned to take into consideration the patient’s psychologic aspect as weak as my own.” (Weisenburg, 1929a, p. 1301)

This discussion reveals a range of views from the 1920s which blur the difference between determining a normal range and the average.

2.3 Normal range and average

A widespread interest in employing large sets of measurements to determine the average for physical and mental properties began in the mid-19th century. However, their significance held different meanings in different contexts and applications. For example, Adolphe Quetelet’s (1796-1874) book *Sur l’homme et le développement de ses facultés, ou Essai de physique sociale* considers the measurement of the length of the average soldier’s stride, the determination of which was felt to have practical utility (Quetelet, 1842). He suggested that that any deviation from these averages would represent deformity and disease. This type of approach permeated the second half of the 19th century and had a great influence on the nascent field of psychology. Wilhelm Wundt asserted: “It can be stated without exaggeration that more psychology can be learned from statistical averages than from all philosophers, except Aristotle.” (Translated from Wundt, 1862, p. xxv)

The interest in averages continued to grow and the normal distribution was increasingly deployed in psychology research (and regrettably, also to eugenics) by psychometrician Francis Galton (1822-1911) and his disciple Karl Pearson (1857-1937). In the early 20th century, Galton blended the statistical concept of the normal average with the medical concept of health in the measurement of intelligence (Morrison, 2002). At this time, there was renewed interest in how the normal distribution related to psychological and social aspects of human behaviour. What was sought by many was a determination of the average or most typical performance. In contrast, Weisenburg’s interest was in determining the difference between the pathological and unimpaired people who were not free from error. This implies the need to identify the lower limit of the range of normal performance.

3. Methodological innovations on a larger scale

Weisenburg had the growing appreciation that “there is such a variety in what is termed ‘normal’ that, unless adequate data are carefully assembled, dependence can not be placed upon them.” (Weisenburg, 1929b, p. 377). He continued to develop his methods over succeeding years. While Weisenburg had initially enlisted his junior neuropsychiatry colleagues to assist him in this project, his next step was to recruit cross-disciplinary collaborators from educational psychology, something that was relatively unusual at the

time. He chose a new collaborator from outside of clinical neurology to bring in expertise in psychological test development and standardization. On the recommendation of Agnes Rogers (1884-1943) who was professor of educational psychology at nearby Bryn Mawr College, Weisenburg recruited a new graduate assistant for his aphasia assessment project (Risser, 2018). Katharine E. McBride (1904-1976) (Figure 2) had demonstrated her potential in the paper she co-authored with Rogers on the training of intelligence-test examiners (Rogers et al., 1928). McBride began working with Weisenburg in 1929 and used this work for her PhD dissertation at Bryn Mar (McBride, 1932). McBride represents one of the first women to make a significant contribution to neuropsychology through her substantial role in this project.

In 1931, Weisenburg presented a paper on his new approach to the study of aphasia at the First International Neurological Congress in Berne, Switzerland (Sachs, 1932). The meeting was attended by 881 neurologists representing 42 countries and provided a significant forum for disseminating his work in progress. He reviewed his findings that the same type of errors are made by normal and pathological subjects on all the existing tests of aphasia and argued for the need to develop a new approach to clinical assessment to address this problem (Watson, 1932). Weisenburg emphasized the need to identify psychological functions that could be selectively and partially impaired due to focal cortical damage.

Two years later, the main findings of Weisenburg and McBride's project were presented in his Charles K. Mills Memorial Lecture to the American Neurological Association (Weisenburg, 1934). He described the initial screening of 314 individuals with suspected aphasia, the more detailed examination of 234, and final selection of 60 patients for extensive testing along with 85 non-aphasic controls. This sample now included a wider range of aphasic individuals who Weisenburg categorised as: male and female, right and left-handed, "White" and "Colored", having a range of ages, educational attainments and occupational status. Some of the tests they employed had been used by Kurt Goldstein (1878-1965), but others were newly devised. However, Weisenburg stressed that the originality of the study was in the use of a group of matched healthy and neurologically impaired non-aphasic patient controls which allowed the performance of aphasic individuals to be analysed as by the degree of deviation from the median normal adult. This was the culmination of his sustained efforts to compile a new and improved battery of standardized test for aphasia with established norms. The manuscript of the resulting book, co-authored with McBride, was submitted to the Commonwealth Fund publishers a few weeks before Weisenburg's death on August 3, 1934.

3.1 Weisenburg and McBride's Aphasia Book

The authors' stated approach was a departure from the work of Head and Goldstein: to be empirical and pragmatic, without commitment to any theory or model. They state their justification for this:

"On the one hand, tests are bound to differ as long as theoretical differences exist; on the other hand, theoretical differences must exist, and must even appear greater, as long as the tests lack uniformity. Whatever the way out of the dilemma, it will undoubtedly be found through the study of large numbers of patients by comparable methods." (Weisenburg & McBride, 1935, p. 92)

This view was felt to be a strength by some clinicians, but a weakness for those interested in the newly devised symbolic theories of impairment such as Head's, or the loss of abstract attitude described by Kurt Goldstein (Geroulanos & Meyers, 2016). In contrast, Weisenburg and McBride's objective was to devise a means to identify language disorders and other

psychological changes in aphasic individuals for the purpose of diagnosis and treatment. They included tests of spoken language, reading, writing, spelling and arithmetic as well as tests of "language intelligence", "memory", and "non-language intelligence". (For a detailed critical assessment of the tests see Risser, 2018).

For the final cohort, Weisenberg and McBride recruited participants who represented diverse medical, economic status, educational and occupational histories to represent a good sample of the general population. Their selection criteria for inclusion were as follows: under 60 years of age, adequate sight and hearing, English-speaking by birth--although speaking an additional language was admitted. Anyone with severe mental disorders was excluded. Of 314 persons initially screened, 234 persons met the criteria for inclusion. More detailed testing was done on selected group of 60 individuals with aphasia, 38 patients with unilateral lesions without aphasia, and a "normal" group of 85 individuals. The aphasic individuals had various types of neuropathology: 37 vascular, 15 tumours (14 tested post-op), 8 trauma (7 post-op). This was a deliberate methodological decision to select for diversity. It also demonstrates an increasing refinement in Weisenburg's appreciation for the selection criteria of normal controls from his initial attempt presented above. The final sample was explicitly selected to reflect the characteristics of the population of aphasic patients.

Weisenburg and McBride's objective was to gather evidence of test performance to identify patterns of spared and impaired abilities across modalities with different types of stimuli and requiring different types of responses. This was to ensure a comprehensive picture of the complexity that existed in the clinical population. Weisenburg argued for the interrelatedness of mental functions; impairments in isolation were relatively rare. For the majority, many areas of performance would be affected to some but differing degrees. Only by testing the full range of language and other cognitive tasks could the particular pattern of deficits and intact abilities in each aphasic individual be identified. This could serve as the basis for rehabilitation, something Weisenburg had been interested in from the beginning of his career (Mills, 1904). The book includes one of the earliest systematic attempts to use individual patterns of performance to devise personalized treatment plans (Thompson & Worrall, 2008).

An anonymous review summarized the magnitude of this contribution:

"This book embodies the results of the most extensive study of aphasia ever undertaken and will long remain the standard work of reference on the subject in the English language, if not in any language...This extensive and accurate employment of 'controls' is one of the main features of the book which gives it great value to the clinician as well as to the psychologist.... They represent a good sample of the average population." (B-H, 1936, p. 168)

3.2 The testing of normal intelligence

A second monograph was published a year later to complete the set of publications arising from the Commonwealth fund project. It focused specifically on adult intelligence and analysed performance on a range of tasks by non-impaired individuals of different types. This monograph was completed by McBride after Weisenburg died (Weisenburg et al., 1936). It was co-authored by research assistant Anne Roe (1904-1991) who became the first tenured woman Professor of Education at Harvard University. The work represented the culmination of Weisenburg's project initiated in 1926. His initial intention had been purely clinical-- to determine the range of performances in unimpaired individuals for use in evaluating aphasic impairments. McBride added a wider remit to the project. She was

interested in determining the range of intellectual abilities in the general population. For her, the project's utility extended beyond a medical application to educational psychology.

The selection of participants in the final study represents a significant refinement from those selected for the initial investigation nine years earlier. This reflected their growing appreciation of potentially confounding variables and a more sophisticated view of the function and composition of a control group. Their participants were patients on surgical and orthopedic wards who had no evidence of neurological or mental illness. Selection was further restricted to those patients with bone fractures or dislocations and those recuperating from surgical interventions but in otherwise "good working condition".

Their exclusion criteria for this sample also demonstrates growing awareness of how chronic illness might impact on psychological functions. An interesting footnote presents the rationale for excluding patients with glandular conditions such as tuberculosis and syphilis (which had been included in the 1929 study), citing medical evidence that these chronic diseases were associated with "mental changes" (Weisenburg et al., 1936, p. 34, fn. 2). They also were persuaded by evidence that infantile paralysis was also associated with mental consequences (see Hellal & Lorch, 2010) and excluded these individuals as well.

Several steps were taken to insure a good sample of age, education, occupation, and social status. All eligible patients in the three Philadelphia hospitals where Weisenburg had access were invited to participate, but only 79% of those who were invited volunteered. Weisenburg and colleagues found that those who did not accept were primarily in their 40s and 50s. They noted that this introduced a small sampling biased towards the younger end of the age scale but rejected the use of inducements as that would introduce other biases. Interestingly, their reasoned choice to use voluntary sampling, thereby reducing the number of middle-aged participants, was later criticized as a limiting factor in the utility of their data (Corsini & Fassett, 1953).

In McBride's preface, she identifies the limitations to collecting a truly random sample for psychological testing purposes. She argued that the use of surgical and orthopedic patients represents a fair sample of the "middle levels" of the population. The virtue of using long stay in-patients was recognized for their availability as a captive audience and desire to offset boredom. Their participants expressed a willingness to do something that might benefit others. Another motivation was identified as a reluctance to refuse a request from a doctor. It was acknowledged that participants attached prestige to their involvement in the study. In gaining consent, the participants were informed that the purpose was to study reading, writing and arithmetic, rather than its actual focus on intelligence. All of these refinements indicate appreciation for biases that existed in the initial Weisenburg, Pearson and Alpers study. They continued to administer the tests individually, even for those tasks which could have been administered to a group. They argued that this provided the opportunity to observe patients' methods of working, but also to ensure cooperation by building rapport between the participant and examiner. This same set of circumstances and procedural choices was used on a large scale and to great effect in the neuropsychological testing of individuals in the US Veterans Administration Hospitals from the 1960s onwards (Freed, 2009).

The authors described their 70 normal participants as "White", with 47 males and 23 females, with 65/70 born in the USA and the remainder in Scotland or Ireland. Two participants were noted to be Jewish and 2 came from homes where an additional language (Italian) was spoken regularly. Weisenburg and his colleagues validated the representative nature of the sample using two objective data sets. The distribution of occupations in the sample was matched against the 1930 Philadelphia census data and the level of education

was matched against the contemporaneous Army draft data. These efforts to ensure the representative nature of their selected sample indicate a high degree of rigour was achieved by using objective measures derived from large data sets.

Individual tests were used to measure aspects “language intelligence”, “educational achievement” and “non-language” components and the intercorrelation between tasks was analysed. The psychometric findings for individual case studies were reported in units of probable error from the population norm. Interestingly, they report that an additional “Negro group” of 15 people were also tested but their data was not included. A footnote explains that they were generally less well educated and found to have lower occupational scale ratings. The authors acknowledged that this was likely due to inequality of opportunity. The “Negro” group mean scores for all but one test were lower than the “White” group who had higher socio-economic status. While they report that the “Negro” group scored higher on Porteus Mazes, this result did not reach significance. They suggested that finding group differences in performance on this task is an interesting and novel observation, but no further interpretation is offered. (Weisenburg et al., 1936, p. 40, fn 12). This footnote reflects the extreme care in their reporting. They chose to include this data and identified confounding factors, but also resisting speculation on non-significant findings.

A review of the book in the *Psychological Bulletin* judged the normal sample to be of the highest psychometric interest (Wells, 1936). It was hailed it as “the first instance of a major attack on a psychopathological problem by the combined resources of medicine and psychology.... Psychologically the work is quite the most thorough that has been attempted in the field and gives evidence of much care and skill.” (Wells, 1936, p. 291).

4 Weisenburg and McBride’s contribution to neuropsychology

Immediately following the publication of these two books, they came to the attention of Donald O. Hebb (1904-1985), considered to be the founder of neuropsychology. He cited their work in several early studies on aphasia and intelligence. For example, Hebb and Penfield (1940) used Weisenburg and McBride’s norms from both books in the testing of a patient who had a frontal lobe resection. In Hebb’s studies of the effects of brain injury on mental abilities, he judged their contribution to be valuable for the use of a good normal control group and their analysis of performance in brain damaged individuals is based on independent measurement of many tasks. Hebb was particularly interested in Weisenburg and McBride’s findings that aphasic individuals tend to produce inferior results on nonverbal as well as verbal tests. He replicated this in his study of the effect of brain injury on test scores (Hebb, 1942a). Hebb also revalidated selected tests from Weisenburg and McBride’s battery (Hebb, 1942b). Most significantly, he cited their work at several points in his groundbreaking and influential book *The Organization of Behavior* (Hebb, 1949). He identified the work of Weisenburg and colleagues as the first study of neuropsychological impairments that employed adequate controls and acknowledged their massive and highly valuable undertaking in collecting such data from this large number of healthy individuals.

Although the *Aphasia* book quickly went out of print, it was still being used as a text in speech therapy courses in the 1950s (Risser, 2018). The book was reprinted in 1964 at time of increased activity in the development of clinical tests. The authors of many of these new assessments acknowledged the significance of Weisenburg and McBride’s innovative use of normal controls and the importance of its methodological procedures for their own aphasia test designs (e.g., Schuell et al., 1964). *Adult Intelligence* was published in a condensed version with commentary in a collection of readings on key research on individual differences in intelligence (Weisenburg et al., 1961). The editors highlight the study’s methodological

rigour and use of a control group selected from the “mentally normal” hospital population. Their findings were also viewed as a valuable contribution to research on the differential effects of age on specific mental tasks. Both the monographs were subsequently republished in dozens of editions and continue to be held in hundreds of libraries around the world (WorldCat.org).

Weisenburg and McBride’s approach to aphasia testing, and neuropsychological assessment more widely, has been perpetuated in successive developments of diagnostic tests (Kertesz, 2018). Many of their tasks are recognizable in current batteries such as the current Boston Diagnostic Aphasia Exam BDAE-3 (Goodglass, Kaplan and Barresi, 2000). The aphasia battery they developed was the definite turning point in standardized psychometrics. Risser (2018) judged their writing about clinical testing methods to have maintained its value for today’s students of neuropsychology.

5 Discussion and Conclusion

Weisenburg’s project has been detailed here to highlight how he addressed the need for methodological innovation in clinical testing. The steps he and his collaborators took to ensure the selection of normal samples which matched the clinical group on relevant participant variables have been analysed through the consideration of a series of publications from 1927 to 1936.

In their first attempt, Weisenburg, Pearson and Alpers matched the homogenous characteristics of Head’s (1926) aphasic young army officers on age, education, and socioeconomic status. Upon subsequently rejecting Head’s tests as invalid because non-aphasic individuals could not perform them, they also rejected his rarefied approach to participant selection. Over successive iterations they increased the number of control participants and the variables to be matched. In their final studies, they selected participants that were a representative sample of the general population for education and socioeconomic status using objective demographic information. Today, one might question whether the sample size of Weisenburg and McBride’s study was sufficiently large, however the standardization of recent aphasia tests employ similar numbers of participants (Freed, 2009).

What constitutes a matched control group continues to be a point of active discussion in neuropsychology. Strauss and colleagues (2006) discuss the difficulties inherent in addressing the numerous competing issues when selecting control participant characteristics, many of which Weisenburg identified. In a recent letter to the editor of *Cortex*, Bonato and colleagues (2012) highlight the fallacy apparent in many group studies in neuropsychology over the past 30 years. They suggest that the composition of the control group typically differs from the clinical group of interest not only by the absence of specific pathology, but also in other characteristics that may be associated with it. They argue that this approach, which rules out controls with “collateral impairments”, arises from underlying assumptions of modularity. Interestingly, Weisenburg and his colleagues held the view that language and other cognitive processes were intertwined and could not be isolated, or demonstrated as being so clinically, in those with neurological impairments. For them, it was a matter of relative levels of performance and that these would necessarily vary from individual to individual due to the nature of their pathology.

Bonato and colleagues suggest that one of the crucial ways to mitigate the fallacies of interpretation arising from using control groups is to match them to the experimental groups on all those variables that may affect performance. What has been recounted here is the pioneering work by Weisenburg and colleagues almost a century earlier to address this very

problem. Their appreciation for control variables of significance was demonstrated in several increasingly larger and more refined studies in the 1920s and 1930s. They were original in arguing that clinical testing must use control participants to identify impaired test performance. Participants included in the control group should be described on the range of variables considered to affect such performance and the norms could be used to identify the degree of deviation in performance. In this way, they made a significant and lasting contribution to the methodological rigour in neuropsychological research.

Weisenburg and McBride (1935) contributed many key observations in aphasia such as the finding that high frequency names are less vulnerable to word finding difficulties than low frequency names. The data they collected from their control group of individuals with left hemiplegia was foundational in understanding of the cognitive impairments arising from damage to the non-dominant hemisphere (Meyer & Jones, 1957). They also documented the incidence of crossed aphasia in a representative sample of the population, thus addressing the problem of interpreting outliers arising from previous published single cases. Rapp (2011) identifies this as an issue of continuing “peril” in cognitive neuropsychology.

Age-related cognitive changes was another issue that Weisenburg and colleagues identified. They included participants only up to the age of 60, arguing that general cognitive changes would be more likely in anyone older would add a confound. Weisenburg, Roe and McBride (1936) found no evidence of age effects in vocabulary, reading, spelling, and arithmetic. They suggested that their findings are consistent with the hypothesis that continued use prevents decline but point out an alternative interpretation that these represent abilities which mature late and/or decline slowly. The issue about age-related neuropsychological changes continued to attract discussion in more recent work (e.g., Lezak, 1987).

The evolution of their ideas about the value of determining the normal range of performance in assessment reflects increasing sophistication and refinement regarding crucial demographic variables. They were also original in recognizing sources of bias when assessing performance in clinical populations. Their attempt to balance measures that reflect privileged or deprived life experiences is another area of significant innovation in their work. Their refusal to interpret the lower performance in the small data sample from those described as “Negro” participants is an indication of their cautious approach.

The issue of how to choose a control group for clinical research into neuropsychological impairments remains an active area of consideration today. Fellows (2012) discusses the need to select a reference group for studies of neuropsychological impairments as a fundamental aspect of study design. The most common current approach is to employ a healthy control group matched on demographic characteristics known to be of relevance to address potential confounds. An alternative is to select participants who have other physical or mental impairments that do not involve the neuropsychological function of interest to control for the generalized effects of brain damage and the confounds present in those who are unwell and/or hospitalized. The practice of using both such participants in control groups was pioneered by Weisenburg and McBride.

Current reviews continue to reveal the lack of valid and reliable normative data for neuropsychological tests (e.g., Howieson, 2019). Moreover, the various issues that Weisenburg and colleagues identified as sources of confound in interpreting test performance are still problematic areas in 21st century neuropsychological research (Franzen, 2020). The substantial contribution of Weisenburg’s and McBride’s work in devising a means for collecting matched samples from unimpaired individuals and to the methodological rigour of neuropsychological research more generally deserves recognition. Their views on the meaning of ‘normal’, and how it should be deployed in interpreting clinical findings, still provide insight and utility today.

Acknowledgement

I would like to thank the archivists at The Special Collections Department, Bryn Mawr College Libraries for providing materials relating to the work of Katharine E. McBride.

References

- Anonymous. (1927). Book Review: Aphasia by Henry Head. . *Archives of Neurology and Psychiatry*, 18(6), 1062-1063.
- Anonymous. (2021). 'normal'; definition B. In *OED Online*. Retrieved Accessed 10 October 2021, from www.oed.com/view/Entry/128269
- B-H, O. (1936). Book review: Aphasia by Weisenberg and McBride, 1935. *Indian Medical Gazette*, 168-169.
- Bartlett, F. (1926). Critical Notice of Head's Aphasia and Kindred Disorders of Speech. *British Journal of Psychology*, 17(2), 154.
- Bonato, M., Sella, F., Berteletti, I., & Umiltà, C. (2012). Neuropsychology is nothing without control: A potential fallacy hidden in clinical studies. *Cortex*, 48(3), 353-355.
- Corsini, R. J., & Fassett, K. K. (1953). Intelligence and aging. *The Pedagogical Seminary and Journal of Genetic Psychology*, 83(2), 249-264.
- Cryle, P., & Stephens, E. (2017). *Normality: A critical genealogy*. University of Chicago Press.
- Cubberley, E.P. (1934) Public Education in the United States. (Boston: Houghton Mifflin.
- Fellows, L. K. (2012). Group studies in experimental neuropsychology. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 647–659). American Psychological Association.
- Fenton, A., & Krahn, T. (2007). Autism, neurodiversity, and equality beyond the "normal". *Journal of Ethics in Mental Health*, 2(2).
- Fleming, G. W. T. H. (1928). Aphasia : A Study of Normal Control Cases. (Arch. of Neur. and Psychiat., February, 1928.) Pearson, G. H. J., Alpers, B. J., and Weisenburg, T. H. *Journal of Mental Science*, 74(307), 793-794.
- Franzen, S., Van Den Berg, E., Goudsmit, M., Jurgens, C.K., Van De Wiel, L., Kalkisim, Y., Uysal-Bozkir, Ö., Ayhan, Y., Nielsen, T.R., Papma, J.M., 2020. A Systematic Review of Neuropsychological Tests for the Assessment of Dementia in Non-Western, Low-Educated or Illiterate Populations. *Journal of the International Neuropsychological Society* 26, 331–351.. doi:10.1017/s1355617719000894
- Freed, D. (2009). A short history of the Veterans Administration's influence on aphasia assessment tools. *Aphasiology*, 23(9), 1146-1157.
- Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K., & Delis, D. (1994). *Clock drawing: A neuropsychological analysis*. Oxford University Press.
- Geroulanos, S., & Meyers, T. (2016). Integrations, vigilance, catastrophe: the neuropsychiatry of aphasia in Henry Head and Kurt Goldstein. In *Plasticity*

- and Pathology: On the Formation of the Neural Subject* (pp. 112-158). Fordham University Press.
- Harris, L. J. (1993). Broca on cerebral control for speech in right-handers and left-handers: a note on translation and some further comments. *Brain and Language*, 45(1), 108-120.
- Head, H. (1926). *Aphasia and Kindred disorders of speech*. Cambridge University Press.
- Hebb, D. O. (1942a). The effect of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society*, 275-292.
- Hebb, D. O. (1942b). Verbal test material independent of special vocabulary difficulty. *Journal of Educational Psychology*, 33(9), 691.
- Hebb, D. O. (1949). *The organization of behavior : a neuropsychological theory*. John Wiley & Sons.
- Hebb, D. O., & Penfield, W. (1940). Human behavior after extensive bilateral removal from the frontal lobes. *Archives of Neurology & Psychiatry*, 44(2), 421-438.
- Hellal, P., & Lorch, M. P. (2010). The Emergence of the Age Variable in Nineteenth Century Neurology: Considerations of Recovery Patterns in Acquired Childhood Aphasia. In S. B. Finger, F. and Tyler, K.L (Ed.), *Handbook of Clinical Neurology* (Vol. 95, 3rd series, pp. 845-852). Elsevier.
- Howieson, D. (2019). Current limitations of neuropsychological tests and assessment procedures, *The Clinical Neuropsychologist*, 33(2), 200-208.
- Kertesz, A. (2018). The History of Aphasiology. In W. B. Barr & L. A. Bieliauskas (Eds.), *The Oxford Handbook of History of Clinical Neuropsychology*. Oxford University Press.
- Lezak, M. D. (1987). *Neuropsychological assessment*. Oxford University Press.
- McBride, K. E. (1932). *A psychological study of aphasia*. Unpublished PhD thesis. Bryn Mawr College]. Pennsylvania.
- Meyer, V., & Jones, H. G. (1957). Patterns of cognitive test performance as functions of the lateral localization of cerebral abnormalities in the temporal lobe. *Journal of Mental Science*, 103(433), 758-772.
- Mills, C. K. (1904). Treatment of aphasia by training. *Journal of the American Medical Association*, 43(26), 1940-1949.
- Morrison, M. (2002). Modelling populations: Pearson and Fisher on Mendelism and biometry. *The British journal for the philosophy of science*, 53(1), 39-68.
- Pearson, G., Alpers, B. J., & Weiseburg, T. H. (1928). Aphasia: a study of normal control cases. *Archives of Neurology & Psychiatry*, 19(2), 281-295.
- Poeck, K., & Orgass, B. (1971). The concept of the body schema: A critical review and some experimental results. *Cortex*, 7(3), 254-277.
- Quetelet, L. A. (1842). *A treatise on man and the development of his faculties*. Chapman. (English translation of Sur l'homme et le développement de ses facultés, ou Essai de physique sociale. 1835)
- Rapp, B., 2011. Case series in cognitive neuropsychology: Promise, perils, and proper perspective. *Cognitive Neuropsychology* 28, 435–444.
- Risser, A. H. (2018). Katharine McBride, 1935, and “Aphasia”. In *The Oxford Handbook of History of Clinical Neuropsychology*. Oxford University Press.
- Rogers, A. L., Durling, D., & McBride, K. (1928). The constancy of the IQ and the training of examiners. *Journal of Educational Psychology*, 19(4), 257.

- Sachs, B. (1932). First International Neurological Congress. *Archives of Neurology & Psychiatry*, 27(1), 212-219.
- Schuell, H., Jenkins, J. J., & Jimenez-Pabon, E. (1964). *Aphasia in Adults*. Harper.
- Semenza, C., & Goodglass, H. (1985). Localization of body parts in brain injured subjects. *Neuropsychologia*, 23(2), 161-175.
- Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press.
- Swash, M. (2008). Henry Head and the development of clinical neuroscience. *Brain*, 131(12), 3453-3456.
- Thompson, C. K., & Worrall, L. (2008). Approaches to aphasia treatment. *Aphasia rehabilitation. The impairment and its consequences*, 3-24.
- Tilney, F. (1934). Obituary: Theodore H. Weisenburg *Archives of Neurology and Psychiatry*, 32(4), 824-849.
- Watson, G. (1932). The First International Neurological Congress. *The American Journal of Psychology*, 44(2), 384-389.
- Weisenburg, T. H. (1919). Traumatic neuroses in war and in peace. *Journal of the American Medical Association*, 73(8), 596-599.
- Weisenburg, T. H. (1929a). A study of normal. *Archives of Neurology and Psychiatry*, 22(6), 1295-1301.
- Weisenburg, T. H. (1929b). The study of normals. *Journal of the American Medical Association*, 93(5), 377-380.
- Weisenburg, T. H. (1934). A study of aphasia: the Charles K. Mills memorial lecture. *Archives of Neurology & Psychiatry*, 31(1), 1-33.
- Weisenburg, T. H., & McBride, K. E. (1935). *Aphasia: a clinical and psychological study*. The Commonwealth Fund.
- Weisenburg, T. H., Roe, A., & McBride, K. E. (1936). *Adult Intelligence : a psychological study of test performances*. The Commonwealth Fund.
- Weisenburg, T. H., Roe, A., & McBride, K. E. (1961). Changes in Adult Intelligence. In J. J. Jenkins & D. G. Paterson (Eds.), *Studies in individual differences: The search for intelligence* (pp. 414–424). Appleton-Century-Crofts.
- Wells, F. (1936). Review of Aphasia: A Clinical and Psychological Study. *Psychological Bulletin*, 291-294.
- White, S. H. (2000). Conceptual foundations of IQ testing. *Psychology, Public Policy, and Law*, 6(1), 33.
- Wundt, W. M. (1862). *Beiträge zur theorie der sinneswahrnehmung*. CF Winter.

Figure 1: Theodore H. Weisenburg (1876-1934)



[source: Wikimedia commons]

Figure 2: Katharine Elizabeth McBride (1904-1976)



[source: Bryn Mawr College Libraries, Special Collections]

Acknowledgement

I would like to thank the archivists at The Special Collections Department, Bryn Mawr College Libraries for providing materials relating to the work of Katharine McBride which were used in this research.