A critical review of antecedents of psychological measurement: Is it necessary to revisit or reorganized the foundations psychometry?

Una revisión crítica de los antecedentes de la medición psicológica: ¿Es necesario revisitar o reorganizar los fundamentos de la Psicometría?

José I. Martínez-Guerrero [a]

Abstract:

The purpose of this article was to revisit and complement the background of psychological measurement, as well as to analyze its foundations from a scientific perspective. A historical-methodological analysis was made of the contributions of great pioneers of measurement at the beginning of the 19th century and even the main contributions of the mid-20th century. Knowing the History of Science in general, and the methodological background of its discipline, both constitute very important ingredients in the scientific training of the researcher. It was found that the roots of Psychometrics come from contributions in Mathematics, Statistics, and Experimental Sciences, which have allowed basing and developing theories, models, techniques, and procedures for the current construction of psychological measurements in social and behavioral sciences. Current psychometrics has quality standards that guide best practices in the development and validation of psychological measurements. The problems and challenges in scientific research have shown that an interdisciplinary approach allows a better understanding of scientific concepts, as well as the possibility of reflection to reorganize the disciplines and expand their foundations with new perspectives.

Keywords:

Psychological measurement; Psychometrics; Test theory; Science History

Resumen:

El propósito del presente artículo fue revisitar y complementar los antecedentes de la medición psicológica, así como analizar sus fundamentos con una perspectiva científica. Se realizó un análisis histórico-metodológico de las aportaciones de grandes pioneros de la medición a principios del siglo XIX y hasta las principales contribuciones de mediados del siglo XX. Conocer la Historia de la Ciencia en general y los antecedentes metodológicos de su disciplina en particular, constituyen ingredientes muy importantes en la formación científica del investigador. Se encontró que las raíces de la Psicometría devienen de las contribuciones en Matemáticas, Estadística y Ciencias Experimentales, lo que ha permitido fundamentar y desarrollar teorías, modelos, técnicas y procedimientos para la actual construcción de mediciones psicológicas en ciencias sociales y del comportamiento. La Psicometría actual cuenta con estándares de calidad que orientan mejores prácticas en el desarrollo y validación de mediciones psicológicas. Los problemas y retos en la investigación científica han mostrado que un enfoque interdisciplinario permite una mejor comprensión de los conceptos científicos, así como la posibilidad de reflexionar para reorganizar las disciplinas y ampliar sus fundamentos con nuevas perspectivas.

Palabras Clave:

Medición psicológica; psicometría; Teoría de los Test; Historia de Ciencia

INTRODUCTION

The purpose of this article is to revisit and complement the background that gave rise to the measurement of the psychological from a scientific perspective. The review begins with the earliest pillars that emerged in the early 19th century and analyzes the main foundations of psychological measurement up to the mid-20th century. The article shares the principle, which has been gaining more recognition in recent decades, about the central importance it has for the scientist's training the fact of knowing and analyzing the History of Science and, in particular, the theoretical and methodological background of his discipline. Science has shown that problems and challenges cross-disciplinary boundaries and its interdisciplinary approach contribute to a better understanding

of scientific methods and concepts. The analysis of the historical background of science can help the reflection and reorganization of disciplines, and thus broaden new perspectives for scientific development and its current technological applications (Ribes, 2005; Barahona, 2009).

Karl Gauss' contributions to scientific measurement

Unlike the traditional history of Psychometrics in classical sources (Boring, 1950; Anastasi, 1966, Anastasi, 1974; Heidbreder, 1960; Nunally, 1970) where the beginning of psychological measurement was located with the anecdotes of European astronomers in their eagerness to identify errors of observation, and from where the so-called "personal equation" of Bessel was formulated; or in other cases, which begin with physiologists, without a greater philosophical context of the

time (Goodwin, 2009); or more specialized beginning with the studies of Francis Galton (Carroll, 1987). In contrast, it is proposed here, in the first place, to recognize as an initial background of Psychometrics the contributions of Karl Friedrich Gauss (1777-1855), with the applications of the normal distribution model, Gaussian bell or probability density function, the error theory and its least-squares method, among other contributions. The History of Science has recorded how Gauss proposed and applied these methods successfully in measuring distances and trajectories of planets and comets, from data of astronomical observations, in order to predict their location within their trajectory (Gauss, 1809); his contributions gave Gauss fame and recognition in Germany and Europe from the first decade of the 19th century (Boyer, 1986).

Although European astronomers continued to apply the "personal equation" to identify errors in their observations until the mid-19th century, they chose to build more accurate devices such as the chronograph and chronoscope to reduce the variability of errors in their observations. Therefore, the personal equation was not used later in Astronomy and much less in the measurement of psychological variables. In any case, the recognition of Bessel should be credited with having been the first to publish in 1838 the measurement of distances to stars applying the stellar parallax technique, even though other astronomers had already used it (Gribbin, 2003).

The contributions of Karl Gauss have constituted fundamental ingredients for the subsequent statistical and psychometric analysis, both in psychological research and from the sociological analysis carried out in 1835 by Quetelet in his work *On man and the development of human faculties*; as well as in the application of methods to predict social phenomena and techniques to estimate measurement errors in various disciplines.

In his book *Disquisitiones Arithmeticae* (1801), Gauss developed and proved central theorems on geometric progressions, including notions such as mathematical congruence that allowed to formalize and unify the Theory of Numbers with a simplified nomenclature; he developed theorems that facilitated the estimation of power residuals, with iterative periods of residuals, and applied analytical methods that formed the foundation of his work (Hawking, 2010). Many of Gauss's contributions have allowed the development of applied mathematics in various fields in natural, social, and behavioral sciences.

The influence of Gauss in the formation of his disciples was of capital importance for later contributions; it is worth mentioning two of his most prominent followers: Bernhard Riemann and Richard Dedekind; who in the second half of the 19th century laid the mathematical foundations for the measurement of continuous variables and the theory of magnitudes, antecedents of the second-order axioms of Peano's measurement in 1887 (Odifreddi, 2006). On the one hand, we have the *Riemann Integral* to measure any interval of the area under the curve of various continuous functions, a basic tool to estimate the probability at each level of the psychometric function. Riemann elaborated in 1854 a theoretical framework for one-dimensional measurements, the extension of it to multidimensional measurements, and its metric relationships for variational sets (Riemann, 1929). Likewise, Riemann postulated the concept of multidimensional space, adopted in various disciplines, and its influence on modern psychometrics in multidimensional IRT models (Reckase, 2009). A treatise would be needed to analyze Riemann's contributions and his

implications in various fields, for example, in the measurement of the curvature of space in Einstein's Theory of Relativity. An interesting fact is that Riemann acknowledged having taken Karl Gauss and Johann F. Herbart as the basis for his analysis in reviewing the geometric foundations of multidimensional spaces and their metric implications (Riemann, 1929). On his part, Dedekind continued the formalization of the precise measurement of continuous variables, the concepts of limit and shearing (cut-off points) to delimit classes of measured values in quantitative data sets.

Another important antecedent from the early 19th century that must be considered is the contribution of Pierre Simon Laplace (1749-1827) in the measurement of probabilities; and who acknowledged having taken up and applied Gauss's ideas on the measurement of error. Laplace participated in the Committee to determine the new metric system in France and propose the definition of the meter as a fundamental unit of measurement, and he was immersed in the scientific field of his time. In 1812 he published his *Théorie analytique des probabilités*, where he exposes the basic principles of probability calculus to analyze phenomena and their possible applications in various social issues (Hawking, 2010). However, he himself recognized that the history of probability analysis began in the 17th century with Fermat and Pascal, and later with Bernoulli and Bayes in the 18th century.

Herbart's Psychology as a Quantitative Science

Secondly, the contributions of Johann Friedrich Herbart, (1776-1841) should be reconsidered among the early antecedents of Psychometry, who in addition to opposing the dominant psychology of the faculties in previous centuries, did not agree with the classification of the sciences established by Kant, where Psychology was located as a discipline of the spirit and not experimental, so it could not be a science (Heidbreder, 1960). Now it is known from other little-known writings of Kant, that this place of science of human behavior was reserved for Anthropology:

> *"The science of the rules of how a man should conduct himself constitutes practical Philosophy and the science of the rules of effective behavior is Anthropology."* And he adds: *"The same thing that happens with theoretical Physics, so closely linked to experiments, is also possible. do experiments with man ..."* (Kant, 1930, pg. 38).

However, Johann Friedrich Herbart made a clear proposal of Psychology as a quantitative science from the first decades of the 19th century. This fact should remind us of the motto of one of the main current psychometric journals that have been published since 1936: *Psychometrika*:
"*It is a Journal dedicated to the development of Psychology as a quantitative rational science.*" This Journal is the organ of The Psychometric Society, a scientific association founded in 1935 by L. L. Thurstone.

Although Herbart was not a psychologist by origin, he was mainly a philosopher and pedagogue, he occupies an important place in the History of Psychology, particularly in the History of Psychology in Europe (Boring, 1950). Herbart published his work *Lehrbuch zur Psychologie* in 1816 and his *Psychologie als Wissenschaft* in 1824, in which he sets out Psychology as an empirical science to which mathematical methods could be applied to observe and measure basic mental phenomena.

Certain assertions of Herbart and his methodological proposal for Psychology as a quantitative science generated controversies and differences, both with nativist Kantians, as with English empiricists, and even with some later researchers of the physiological tradition. What is clear is that Herbart stated that Psychology should be quantitative and observational, but it was not necessary that it be analytical and experimental, because it should be considered unitary to the mind. It is likely that the latter was what led to some contempt on the part of some experimental psychologists, who later wrote other stories. In the beginning, Herbart was one of the opponents of European idealism from a realistic philosophy perspective (Antisieri & Reale, 1988). In Psychology, he analyzed mental units in terms of representations (Vorstellung) in order to explain basic and complex mental phenomena, as well as their psychological dynamics from said mental units or ideas. He identified that each idea or mental representation has a certain degree of strength or intensity, with special emphasis on the phenomena of inhibition: when it comes to linking a simple idea with other incompatible ones and with more force, it tends to be located below the threshold of consciousness; as well as the opposite process when the ideas are compatible and tend to be assimilated to other ideas found in the "apperceptive mass" consciousness (Heidbreder, 1960). According to Herbart, each idea tends to remain in consciousness and to repel or inhibit others that are incompatible; and therefore he considered that ideas vary in intensity, duration, and degree of inhibition or strength of association in consciousness, variables that could be scientifically measured.

Regardless of the validity of his theoretical proposal, it is paradoxical that it was Herbart who opened the way to Psychophysics and gave elements to Physiological Psychology in the following decades of the 19th century. Due to his desire for precision and measurement of mental phenomena, Herbart considered that the ideas of the mind had two dimensions: duration and strength or intensity. From this, he began his analysis of ideas as different categories and the degree of union between them, as a third magnitude. Herbart's proposal assumed that the conditions for the measurement and mathematical handling of it were met. Later psychologists such as Titchener have recognized that Fechner inherited mental analysis from Herbart (Boring, 1950). In his psychological theory, Herbart postulated the notion of the *threshold* of consciousness as a central concept for explaining mental dynamics in terms of strength and inhibition of conscious ideas. Thus, among Herbart's contributions to Psychophysics, one can see how Fechner took up the notion of measuring the magnitude of data derived from the analysis of ideas and judgments in consciousness. Furthermore, Fechner took up the notion of *threshold* in his psychophysical experiments, just for the measurement of perceptual thresholds.

Importance of physiologists in psychological measurement

Another central ingredient that complements the antecedents of Psychometrics refers to the research work of physiologists in the 19th century, with special emphasis on the contributions of the scientist, physiologist, and psychologist, Herman von Helmholtz (1821-1894). Helmholtz is part of the History of Science, not only as one of the most prominent scientists of his time but because in addition to his contributions in the physiology and visual and auditory sensory processes, he can be placed as a co-founder and link in Psychology experimental, precisely between Fechner and Wundt.

In the investigations on the nerve function in the motor and sensory centers and pathways, during the first half of the 19th century, the relevant works of Johannes Müller (1838) and Ernst Weber (1846) appear as representatives of the physiologists. Müller, with his studies on the theory of specific energies and identification of centers for each sense, had a great influence on Helmholtz, who would continue with research on the physiology of vision and hearing. Weber, with his experiments and sensory measurements of touch, would also be an important precedent for Helmholtz and Fechner. Müller asserted that given the instantaneous speed of nerve impulses it was impossible to estimate their speed; Faced with this challenge, Helmholtz's creativity led him to measure the speed of nerve conduction, with experiments in which he estimated the time it takes for a muscle to contract, with nerves of different lengths, and using a myograph that he himself invented (Boring, 1950).

Helmholtz published in 1850 the results of his experiments on the rate of transmission of nerve impulses. These experiments were the basis for the measurement of reaction times; and that the Dutch physiologist F. C. Donders applied them in 1868, first as a measurement of a simple reaction and later refined the procedures with measurements of compound sensory reactions, discrimination, and choice. As is known, mental chronometry, the first stage of experimental psychology, would be developed from these methods. Helmholtz himself advised Wundt to use Donders' reaction time measurement and Fechner's methods in experiments in the Leipzig laboratory. Ultimately, it would really be Oswald Külpe, in Leipzig and later in Würsburg, who between centuries would refine the technique of measuring reaction times of psychological processes.

Other scientific contributions of Helmholtz in Physics appear in studies of the late 19th century. For example, Ebert recognizes that: "From Faraday's law, Helmholtz was the first to show that in the case of electrolytes it must be considered that every valence is charged with a minimum amount of electricity, valence charge, which at like the electrical elements of the atom, it can no longer be divided (Ebert, 1894, cited in Sánchez-Ron, 2001).

Another area in which Helmholtz contributed to the early development of Psychometrics was his proposal of the theory of measure for the empirical sciences; that is, the geometric analysis of it, the concept of number and its relationship with empirical magnitudes, to formalize a theory of scientific measurement (Helmholtz, 1887). This proposal was intended to be different from the classical theory of measurement because, at least since Galileo and Newton, only physical measurements were conceptualized in the framework of extensive measurements. Helmholtz's approach to formalizing a measure theory was a link towards the formulation of the theory at the beginning of the 20th century first with Hölder's quantity axioms (1901), and later with N. R. Campbell's representational theory of measure and your definition of fundamental and derived measures in science (Campbell, 1920).

Recent analyses have questioned Helmholtz's proposal in the sense that it is very different from current measurement theory in several respects. It is pointed out, for example, that he did not differentiate between numbers and numerals; or that his measurement proposal was more of a nominal assignment of concepts, very different from the current measurement theory (Biagioli, 2016). However, Helmholtz is considered a pioneer

in the approach to measurement theory to establish the basic conditions for the numerical representation of empirical quantities, which had not been updated for more than four centuries, even with the advancement of science.

In his research on perception, Helmholtz considers that Psychology could be an exact science, with experiments, measurement methods, and the use of mathematics, as Herbart had previously proposed (Boring, 1950). Just as Gauss inherited two of his closest disciples, Helmholtz also influenced two of his students and collaborators: Hertz in the measurement of electromagnetic waves in Physics; and Wundt on the measurement of mental processes in the nascent experimental psychology.

Psychophysical and measurement methods of Fechner

The emergence of the scientific measurement of psychological variables has been located in the mid-19th century (Boring, 1950). However, it is necessary to recognize among the antecedents that its foundations appeared from the beginning of the 19th century with the contributions of Karl Friedrich Gauss and others. On the other hand, it is necessary to consider the advances in the physiological experiments of the senses with Johannes Müller and Ernst Weber; Favorable antecedent conditions for the investigations of Fechner, Helmholtz, and Donders, which promoted an environment of scientific observation, measurement, and data recording in Biology and Psychology (Gribbin, 2003; Goodwin, 2009).

Even though, Fechner could be credited with being the first researcher to measure a psychological variable; It should not be forgotten that in the 14th century Nicole de Oresme had collected measurement data on the way in which the virtue of charity varied among her fellow monks. Oresme is recognized in the History of Science for having been the first to establish a fixed coordinate system for the representation of measurement data (Odifreddi, 2006); but in addition to showing the graphical representation of data from observations of objects in motion, he also tried to measure and graph data on human behavior traits such as virtue, which had a very different character from extensive measurements, so he considered them intensive as the pleasure, pain, etc. Oresme tried to make explicit and integrate both extensive and intensive magnitudes into the classical theory of measurement, which had not been revised until the proposals of the late 19th century (Michell, 1990).

Gustav Fechner (1801-1887), physician, physicist, and mathematician, with a clear influence both from Herbart's approaches and from Weber's research, mainly in the application of mathematics in the study of perceptual phenomena, in some concepts and variables that they considered. For example, Herbart proposed the analysis and dynamics of ideas in consciousness and in the elaboration of the relationships between concepts. Starting from the principles of inhibition and association of ideas postulated by Herbart, Fechner set out to apply a mathematical model to express psychophysical relationships. This fact undoubtedly constitutes relevant evidence of how 19th-century psychology in Germany evolved as a quantitative science. Another fundamental component to push Psychology as a natural science would be offered by the physiology of the senses, the use of the experimental method, and the measurement of psychological variables in the laboratory. This formal process began with the publication of Fechner's Elements of Psychophysics. As Edward G. Boring has noted:

"Fechner was to take from Herbart the notion of measuring the magnitude of conscious data, the notion of mathematical analysis of those data and, most important of all, the notion of threshold identification" (Boring, 1950, pg. 279).

At the beginning of his scientific work, Fechner made his debut in 1831 with a research report in Physics on the measurement of direct electric currents; and, as is known later, his image as a scientist would be consolidated with the publication of "Elements of Psychophysics" in 1860. However, it is important to differentiate the scientific contributions from Fechner's general work, which is also framed in a philosophical context, which represented a process of historical transition that goes from the evolution and debate of Kantian ideas and the critical approach to the problems that can be identified as late idealism, which developed in Germany during the first decades of the 19th century; this movement was prolific in ideas and divergent manifestations of human thought (Leehmann, 1964).

Among other important philosophical movements that emerged in the context of Fechner, dualistic metaphysics stands out on the one hand and the development of a neo-Kantian approach on the other. The movements that preceded Fechner's proposals were characterized by a conspicuous rivalry that separated Christian philosophers from their Hegelian rivals. However, in this transition, discussions of some psychological problems such as personality or in particular the problem of the unconscious were raised, which was postulated from Herbart. Once the idealistic philosophers left speculative theism behind, there was a transition from late idealism, philosophical anthropology, and neo-Kantianism, which constituted a bridge for new approaches, both in philosophy and in the sciences of the 19th century.

Of course, this transition made it possible to emancipate the natural sciences and promote incipient scientific developments in Physiology and Psychology (Reale and Antiseri, 1988). Fechner focuses, from Medicine and Physics, on the reconciliation of the thought of natural science with a new scientific idealism; where everything material and individual has a correlate of consciousness that, in a hierarchical way, covers from the basic and individual to the collective, and even the entire world. For this reason, Fechner has been considered a pantheistic philosopher with a scientistic approach, for having conceived the world as a natural hierarchy with different levels or units of consciousness. It is important to emphasize that Fechner's scientific contributions: the foundation and development of his methods of observation and measurement for Psychophysics are independent of his metaphysical approaches and his attempt to propose a general vision of the world. With this contextual background, Gustav Theodor Fechner's research work was developed.

Fechner's scientific contribution is limited to the implementation of experimental methods of psychophysical research and procedures for measuring perceptual thresholds. His research led him to establish a model that postulates the mathematical relationship between the intensity of the physical stimulus and the magnitude of the sensory experience in a logarithmic relationship: $S = k \log (I)$. In the next century, Stevens would try it with an exponential psychophysical model: $S = a(I)^n$ (Stevens, 1956). Since then he began in Psychology, in a systematic way, the mathematical modeling in the measurement of psychological variables.

Theory of Evolution and individual differences

On the other hand, the evolutionary study on the diversity of biological species in the 19th century was developed from the beginning of that century with treatises on Botany and Zoology by Erasmus Darwin and Lamarck, which undoubtedly promoted the practices of systematic observation, registration data, and later the practices of measurement of biological, psychological and social phenomena and processes. Late in the century, Darwin published the theoretical proposal that evolutionary processes and the resulting variability of species could be explained by the mechanism of natural selection; that is, as a result of the opportunities for reproduction and survival of individuals better adapted to the demands in different environments (Darwin, 1859). After publishing his famous book, The Origin of Species, Charles Darwin himself publishes well-known works by psychologists *The descent of man (1871)* and *The expression of de Emotions in Man and Animals (1872)*. In this context, as a background, the interest arises to study the differences between animals and man.

It is worth remembering that after his travels and careful observations of the enormous variability of plants and animals, Darwin had no doubt that the process of evolution was a fact; but he still did not have the fundamental mechanism to explain the observed facts. The central ideas emerged after Darwin read Malthus's published essay on the population principle, which explained how populations could grow in geometric progression, doubling at certain intervals of time; and that these phenomena could naturally maintain a balance, whether the populations were controlled by plague, predators and especially by the limited amount of food; also, in the case of human beings with wars. Therefore, a good part of the offspring will die without having the opportunity to reproduce, if nature takes its course (Gribbin, 2003). Of course, it is not necessarily the only mechanism that explains the evolution of all the biological variability in the different species. It is now known that Darwin was a long way from having knowledge of the mechanisms of genetic transmission, of the characteristics of DNA and RNA, and of genetic codes.

With the explanatory elements of demographic pressure, the struggle for survival and reproduction, as the main promoters of evolution, Darwin included that of the individuals best adapted to environmental demands; that is, those that are in better adaptive conditions. Here it is convenient to clarify the meaning of this last concept, very unfortunate and ambiguous for the scientific precision that was required. The word *Aptus*, according to the Classic Latin Dictionary, means willing, adjusted, accommodated. That is, adapted in the sense of a fitting puzzle piece; but not in the sense of sports competition or of evaluating which individual is "the best" as in the Olympic games in ancient Greece; but neither is it the sense of "greatness or nobility, the latter term implying skill, grace or knowledge" (Ribes, 2005).

It is important to be very careful with such interpretations of the term "aptus"; because besides being questionable at present, it can confuse biological adaptation with intellectual capacity; and it could be used for ideological purposes or to politically justify the situation of social classes. These implications carry unjustified consequences, in contrast to the authentic and objective interest in studying and measuring individual differences in human behavior.

In this context the figure Sir Francis Galton (1833-1911) arises, by the way; Darwin's cousin, who was first interested in studying the inherited characteristics of talented people in his famous work Hereditary Genius (1869). Galton analyzed data from families, samples of brilliant college students, and prominent individuals in Britain, attempting to analyze relationships between inherited determinants and the proficient ability of grandparents and their descendants. In his theory, Galton defined competent ability, which included three basic components: intelligence, enthusiasm, and a commitment to work hard to achieve high levels of competence. Of course, Galton did not measure these characteristics of people and, in any case, his tests were not valid indicators of those attributes. Galton started his Anthropometric Laboratory in 1884 where he recorded a large amount of data on measures of reaction times to different stimuli, the strength of motor responses, sensory discrimination, among other anthropometric data. (Galton, 1889).

In addition to constructing the first tests of basic psychomotor and sensory abilities, Galton designed questionnaires to obtain data from large groups of people; but above all, he applied the statistical analysis of data with procedures derived from Gauss's contributions, in order to identify differential performances in the test results and the relationships between the measured variables. Thus, Galton's pioneering work sits in this other tradition of psychological measurement. As is known, the main followers of him, Pearson, and Spearman would continue a systematic work of measuring intellectual abilities and analyzing them statistically to look for relationships between the data obtained.

For his part, James Mckeen Catell (1860-1944), after doing his doctoral studies with Wundt in Leipzig and before returning to America, had a brief stay in London where he visited the Galton Laboratory to learn about its measurement methods and its data analysis techniques. On his return to the United States, Catell was the first American psychologist to found a psychometrics laboratory at the University of Pennsylvania in 1889 and later another at Columbia University.

In Catell's classic article published in *Mind* in 1890, the term *Mental Test* appeared for the first time. With his training in Psychophysics and Experimental Psychology, and with Galton's conviction of facing challenges to more accurately measure simple intellectual performance skills, Catell collected a large amount of data from the application of psychological skills tests, with the purpose of evaluating the basic capabilities of college students. However, disappointment would soon come when, in their attempt to evaluate the predictive capacity of their test results, very low correlations were obtained with independent indicators of physical and intellectual performance, teacher evaluations, grade achievement, and students' school progress (Wissler, 1901). Even with the above results, James M. Catell continued to make efforts to apply a greater number of tests, obtain and analyze more data on mental abilities in young university students, but apparently with less than encouraging results; which sparked criticism from his colleagues and controversies for the low correlations between the tests; therefore, they were considered by his colleagues as questionable indicators to evaluate and predict intelligence or school success.

Another possible meeting point between the two traditions of scientific psychology, experimental measurement, and the measurement of individual differences, could be located in the figure of the German psychologist Herman Ebbinghaus (1850-

1909). Among the first systematic efforts to measure learning and memory, the pioneering work of Ebbinghaus appeared published in 1885, where he used the experimental method, to show quantitative results of memory, its graphic representation, and objective measures of learning and forgetting in the function of time. With his contributions, Ebbinghaus had managed to establish a paradigm to measure performance in learning tasks, the memory process, and the speed of forgetting in humans (Anderson, 1995). With this methodological basis, lines of research on the learning of verbal stimuli in Experimental Psychology were promoted during the first decades of the 20th century. In addition to his work as the author of psychology books and editor of a scientific journal, Ebbinghaus developed an original group test with completion questions on study habits and time management of schoolchildren, which he was entrusted with in Breslau in 1897 (Heidebreder, 1960). This prototype and completion items proved to be very useful for the development of further tests and in the psychometric practice of the following decades.

Influences and limitations of Geometry in measurement

Before closing the analysis of the scientific heritage of the 19th century, it is necessary to point out that among the basic concepts that have been handled in psychometrics, for example, to understand the properties of current psychometric models; in particular the concepts of Invariance, Additivity, and Monotony, as properties of functions, were derived from Euclidean Geometry. This Geometry consolidated its long period in the 19th century, with Hilbert's Foundations of Geometry of 1899. However, the mathematical and scientific development in that same century saw the emergence of new approaches and analysis of other Non-Euclidean Geometries with Gauss, Bolyai, Lobachewsky, and Riemannun, a very interesting and productive century that promoted modern science (Odifedi, 2006).

After the Bible and other classic books, Euclid's Elements is one of the most consulted and read for more than two millennia. The most important contribution of the Elements is the systematic presentation that Euclid made of the basic postulates of Geometry, a synthesis of the mathematical knowledge of Classical Greece. Another central contribution was the deductive logic that he used to formalize and demonstrate the geometric postulates and their metric implications. (Euclides, Book I).

The influence and impact of Geometry on the evolution of human knowledge and the development of civilization has been more than evident. However, in addition to a certain vagueness in Euclid's definitions, several of the fundamental postulates of plane Geometry are not fulfilled in other Non-Euclidean Geometries: Hyperbolic or Elliptical (Ramírez & Sienra, 2009). For example, in Non-Euclidean Geometries the axiom of parallels is false. It is evident that the concept of parallels was taken over by the Psychometry of Geometry. It should be noted in passing that in the Classical Test Theory (CTT) in many cases the assumption of parallelism is not fulfilled, because in practice the "parallel tests" obtain different reliability coefficients, due to the estimation of different correlations (Traub, 1994).

Returning to Geometry, it is interesting to note that Euclid did not formalize or make explicit a precise definition of the area or its measurements. The only thing he refers to in the second book of the Elements is *"In every area of a parallelogram, call*

"gnomon" any one of the parallelograms located around its diagonal, together with two complements" (Euclides, Book II). In the Elements, the only reference is made to the characteristics and similarities in the sides of a parallelogram, the opposite angles to each other, and how, for example, the diagonal of a rectangle divides the area into two equal parts. However, Euclid exposed certain common notions, from which some basic geometric properties can be deduced: 1) In a symmetric figure, if it undergoes some type of transformation (translation, rotation, reflection, inclination, etc.) it can be observed that in Equal surfaces remain the same areas: *Invariance* of areas. 2) A surface is obtained by adding together a finite number of surfaces and has an area equal to the sum of the areas of its *n* surfaces: *Finite additivity*. 3) In a flat figure a surface contained in another has an area that is always smaller, which can grow but can only be less or equal: *Monotony*.

The concept of invariance has transcended in the development of Geometry to such a degree that, before the emergence of new Non-Euclidean Geometries, and algebraic advances, Klein proposed a general definition of Geometry as the study of invariants under a group of transformations (Ramírez-Galarza & Sienra-Loera, 2009).

In current psychometrics the concept of Invariance in the measurement has special importance; not only as a relevant goal to achieve stability in measurements but as a necessary condition of psychometric models and practices to ensure that the calibration of the measurement instruments is independent of the sample used and that the measurement of the subjects is independent of the instrument. used and thus achieve invariant measurements (Engelhard, 2018). In the Classical Test Theory, this goal had not been achieved; and it is not until the development of the Rasch Model and the Item Response Theory that this possibility is available, but the measure must be justified, have well-calibrated instruments, and show that the empirical data of the measurements fit the psychometric model used and thus allow stable replications in the measurement of psychological variables.

Regarding the finite additivity and monotonic growth of the functions, which represent the relationship between the probability of response and the level of the attribute or trait measured in the psychometric models, they have been very useful concepts in the analysis and psychometric estimates in the models of current psychological measurements.

Psychological measurements in the first half of the 20th century

Analyzing the background and the subsequent development of Psychometry, considering the first contributions in psychological measurement in the 20th century, represents a great challenge. Therefore, by virtue of the fact that it constitutes a very vast field, an exhaustive analysis of each of the contributions is not intended, because it is such a large number of characters that it would be the subject of a book or, better still, of several volumes. On the other hand, only those pillars that allow contributing to certain critical reflections of the measurement of psychological attributes from a scientific perspective will be pointed out here.

The two central figures who undoubtedly drove the research and development of one of the traditions in psychometry were Alfred Binet (1857-1911) and Charles Spearman (1863-1945). The pioneering articles of this century were On general intelligence and its objective measurement (Spearman, 1904) and The new methods of diagnosing the intellectual level (Binet

& Simon, 1905). The Binet tests were intended to assess children's intellectual abilities (attention, memory, comprehension, imagination, reasoning) in order to identify and predict their ability to learn at school and thus be able to classify them into different degrees of intellectual development, depending on your age. The French Ministry of Education entrusted Binet with an objective diagnosis with scientific instruments for the planning and grouping of children in normal public schools or to direct them to special education centers if they had very low levels of ability. In addition to having a useful and justified tool to assess the intellectual capacity of children, he proposed the concept of mental age that completed the quantifiable indicator that had been proposed by William Stern (1900) to assess the rate of mental development of children: The Intellectual Quotient (IQ). As can be seen, it is a developmental indicator based on the age of the children.

Based on his studies on psychopathology and degrees of mental impairment within the clinical tradition, Binet sought the development of scales in mental tests, designing various simple tasks of increasing complexity and calibrating them with many children of different ages to predict intellectual performance of schoolchildren. Thus, she would achieve a psychometric function of the intellectual capacity of French children, from different degrees of mental deficiency to high levels of intellectual capacity, based on the difficulty of each task and depending on the chronological age of the infants. This psychometric approach to assessing children in France was purported to be a measure of the average rate of cognitive development.

For his part, Spearman made theoretical and methodological contributions, in addition to providing empirical data for psychometric research, in the measurement of human intelligence and capabilities. Spearman is also located as the creator of the Factor Analysis technique to identify latent variables of individual differences in the intellectual performance of people. As a theorist, he proposed the "G" factor of general intelligence and its specific factors and was the initiator of the Traditional Theory of Tests. These contributions opened paths and developments that achieved very relevant methodological advances in Psychometrics.

After the massive applications of group tests to measure intellectual and adaptive abilities during the First World War, for the purposes of selection of military personnel with the Army Alpha and Army Beta tests and their application on a large scale; and other batteries in education and industry in the US. The fundamental pillar for the advancement of psychometric methods and models was undoubtedly Louis Leon Thurstone (1887-1955). In the first stage, with his proposals for the design and analysis of scales to measure attitudes in the 1920s (1925 and 1928). Later with his theory of multiple factors of intellectual capacities or primary aptitudes (1938); He also contributed in an outstanding way with systematic methods to develop psychological measurement scales and educational evaluation. It is important to note that Thurstone when seeking to standardize his scales with statistical parameters, considered the normal distribution of the reference group and managed to liberate the psychometric relationship from the anchoring of chronological age that the Binet scale had. Although Thurstone's objective was not to use the accumulated normal function as a mathematical model, only for standardization purposes, this step constituted a very significant advance in the use of a statistical model for the development of scales with standardized units in the measure of psychological

variables independent of the particular instruments (Thurstone, 1928).

As noted at the beginning of the article, we also owe Thurstone the founding of the first Psychometric Association in 1935 and the creation of the *Psychometrika* in 1936, which has been a very relevant organ for the dissemination of original articles on theories, model, methods. and psychometric techniques, as well as their mathematical and statistical foundations, for more than eighty years. It should be noted that just the same year Guilford's book Psychometric Methods appeared (1936).

After a period of crisis in the first decades due to questions that Psychology did not have scientific measures that fulfilled Hölder's axioms of additivity, and did not conform to Campbell's definition of fundamental measures, another figure emerges that promoted the measurement Psychology near the Middle of the Century: Stanley S. Stevens (1906-1973). It is true that Stevens partly took up the representational approach to measurement, but it also had an influence from Bridgman's (1927) operationalism, which is why he considered it important to specify the set of operations that would allow the variables of interest to be recorded or measured. In addition to the limitations and problems that the abuse of operational definitions imply, for the representational theory it is necessary to demonstrate that the number system in the measure is reproduced in the quantitative structure of empirical quantities; that is, an isomorphism between numerical measures and the structure of the measured attribute must be ensured. However, for Stevens complying with Hölder's axioms and Campbell's definition of fundamental measures was only one type of measurement. So what he did was broaden and generalize the concept of measure to other ways of assigning numbers, based on certain rules: nominal, ordinal, interval, and ratio; but limited by allowed statistics, which must be respected to maintain the same level of measurement.

Starting with the famous article by Stevens (1946) published in *Science*, Psychology could return to the path of "scientific measurements" by including its indicators and measurements in some of the measurement levels of its expanded scheme but respecting the statistics admissible to keep each scale-invariant. With a too open definition of measurement and his classification of measurement levels, Stevens promoted the measurement work in different areas of Psychology. However, it should be remembered that the origin of his proposal had been the disapproval of his psychophysical scale to measure auditory sensations in the British Association for the Advancement of Science, precisely because it did not meet the additivity criteria of scientific measurement. From a distance, it is interesting to reflect on the limitations of Stevens' scheme, among others, because the nominal assignment of numerical labels does not constitute a measurement in any way, but a simple numerical coding; secondly, it must be recognized that if a large amount of psychological data is only adjusted at the ordinal level, it is necessary to review its pertinent statistical management and recognize its limited scope, because data from the variables measured with different instruments cannot be compared, different populations and diverse cultural contexts.

On the other hand, the ordinal scores of the psychological tests had limitations inherent to the Classical Test Theory (CTT). The CTT with its ambiguous concept of the true score, its assumptions, and a score expressly linked to the instrument with which it is sought to measure, did not constitute an independent measure of the attribute or characteristic of the individuals, but an auxiliary definition referred to the interaction between the

subject and the overall score of the instrument. By measuring only the particular reactions of individuals linked to a specific context of measurement, the results become dependent on the instrument and the comparison sample in each application with which a person is evaluated; In other words, invariant measurements are not obtained from the evaluated individuals, because the score depends on the difficulty of the test applied to measure the same attribute and on the group of subjects with whom it is compared.

Nowadays, it has been common practice to make hypotheses about the measurement of internal constructs, cognitive attributes, traits, dispositions, etc., but it turns out that technically there are underlying problems in the theory of true scoring, if compliance with it is seriously considered. the important basic assumptions of CTT. In addition, the main indicators and elements of the model of the true score are not part of the model and are not on the same scale; The most obvious thing is to recognize that the scale of measurement of the test score does not depend on the individual being measured, but is dependent on the instrument used to measure, which is an inconvenience or at least a limitation. However, it is clear that the *ad hoc* CTT methods and techniques have been very useful in psychometric procedures and have made it possible, in a pragmatic way, to develop and apply a large number of instruments and obtain data that have supported a good part of studies in Psychology and Social Sciences; in particular, in various fields applied throughout the twentieth century, and which continues to be of great use in practice.

In the 1940s and almost half a century, it is worth mentioning that relevant contributions emerged on statistical methods and techniques, as well as applications in large-scale psychological and educational measurement. In addition to the theoretical and methodological contributions, in 1947 the Educational Testing Service was created in the USA, where great personalities of very relevant Psychometrics congregate. Important applications and publications on the development of tests such as the ACE, CEEB, and SAT, among others, have been promoted in this institution. Many of these researchers have made methodological contributions.

Among the relevant contributions in Educational Psychometry and Evaluation, in the middle of the 20th century, the work of Lee Joseph Cronbach (1916-2001) is located. The classical psychometric theory and methods were not sufficient for Cronbach, both in methodological and statistical aspects, as in the understanding and validity of the measured constructs that were reflected in the performance results in the tests (Martínez Guerrero, 2006). Cronbach contributed to systematize techniques to estimate the internal consistency of psychological measurement instruments (Cronbach, 1951); he designed methods to analyze construct validity, as the center of psychometric properties (Cronbach, 1955); he contributed to psychometrics a broader conceptual framework and a methodology for the validation of tests (Cronbach, 1971); developed a proposal to identify various sources of error in the measurement, in order to improve its reliability, and proposed the analysis of generalizability as evidence of validity, which culminated in the Theory of Generalizability (Cronbach, 1972). Before closing these reflections on the main antecedents of measurement in Psychology, Education, and Social Sciences in the first half of the 20th century, I will summarize and briefly point out two very promising measurement foundations with great potential: the axiomatic approach of conjoint measurement (Luce and Tukey, 1964; Fraser, 1980); and

Rasch's model (Rasch, 1960) in Item Response Theory (Birnbaum, 1968; Hambleton, Swaminathan & Rogers, 1991). Among the challenges that researchers and psychometrists have faced in measuring psychological variables is the problem that, in the numerical statements that represent the data, they must have a clear empirical interpretation within the relational system being measured (Adams, Fagot and Robinson, 1965). In addition, for other colleagues, the question is to identify which are the valid inferences that can be made from the empirical relations measured (Michell, 1990).

Faced with these challenges, the Conjoint Measurement Theory considers important, firstly, the need to demonstrate empirically that the variable to be measured as a quantitative structure. The approach conceives that certain variables have relationships whose properties are quantitative; but the quantitative properties of any variable cannot be assumed a priori, they must be subjected to empirical testing. Second, the theory includes methods to identify the quantitative structure of the variables; and, in addition, it proposes to identify the quantitative structure of the constructs through the analysis of joint ordinal relationships between the values of the variables to be measured.

In Psychology and Education, large amounts of data can be obtained from ordinal variables, with which it is possible to identify their quantitative structure and thus manage to order the values in their relationships with other relevant variables in a joint system. It is clear that certain assumptions and metric conditions must be meet so that the combinations of the values of the variables considered are ordinal or interval, integrate a set of ordered values and thus meet jointly with the Hölder additivity criterion in scientific measurement. The reader is encouraged to delve deeper into this approach (Fraser, 1980; Michell, 1990).

Congruent with the Conjoint Measurement Theory, it is interesting to note that Rasch (1961) proposed a joint relationship measurement model between two quantitative variables: the ability of individuals and the difficulty of the items in a test. In the Rasch Model, a construct to be measured is conceptualized as a continuum, in which the location of the examinees can be identified, reflecting their variability in the measured capacity; and the location of test items, which reflect different levels of difficulty. Rasch's model postulates that the probability of an examinee of identifying a correct answer to an item is governed only by the distance between the examinee's ability and the difficulty of the item (Masters, 2018):

$$\beta n - \delta i = \ln (Pni_1 / Pni_0)$$

The Rasch Model then postulates that there is a close relationship between βn, the capacity of individual n and δi, the difficulty of item i; and this difference is a function of the logarithm of the proportion between Pni_1 the probability of correct answer and Pni_0 the probability of incorrect answer. Based on this reasoning, Rasch developed a logistic model with multiple implications for psychological and educational measurement. In addition, he himself laid the foundations for developing extensions to a family of models, which has resulted in enormous potential in its application to a wide range of variables and contexts (Wright & Mock, 2004).

Therefore, the basic Rasch logistic model is:

$$P (Xni=1|\theta n) = \exp (\theta n - \delta i) / 1 + \exp (\theta n - \delta i)$$

The Rasch Model was undoubtedly a turning point in the measurement of psychological variables; and from this approach, a new stage began that was complemented by Birnbaum's proposal with its latent trait models (Lord & Novick, 1968); which allowed the development of the current Item Response Theory, with various families of dichotomous, polytomous, multidimensional, parametric and non-parametric models, etc. and its current applications in psychological, educational and social measurement (van der Linden & Hambleton, 1997; Bond & Fox, 2015; Wilson & Fisher, 2017).

FINAL CONSIDERATIONS

As can be seen, the roots and foundations of the measurement of psychological variables come from Geometry, Mathematics, and Experimental Sciences, in particular from sensory physiological research, as well as the ingeniousness, creativity, and persistence of pioneers who developed procedures, instruments, and techniques to have theories and measurement methods that justify and support them.

If we leave aside for a moment theoretical preferences, school or psychological tradition, even methodological approach, it is evident to identify that the seeds of quantitative psychology were sown and cultivated by mathematical pioneers, philosophers, physiologists, and experimental psychologists during the nineteenth century. Knowing and reflecting on original sources allows revisiting with new perspectives the main lines that began their development: the experimental tradition with Fechner, Helmholtz, Wundt, Ebbinghauss and Thorndike, and others, as well as the psychometric one with Galton, Pearson, Catell, Binet, Spearman, and others. It is also important to relocate original influences and interrelations between researchers of these main lines of measurement of psychological variables (Jones & Thissen, 2007).

Likewise, from the works of Spearman, Thurstone, Stevens, Cronbach, and others, and the results of multiple pioneering efforts and followers during the 20th century, an active and dynamic current Psychometry has been developed, which includes theories, models, methods, techniques, strategies, and procedures to measure psychological and social variables, in all areas of research in Psychology and in social sciences, as well as in various interdisciplinary fields in Education, Health, Economics, Organizations, Institutions, Communities, etc. It is interesting to emphasize the multidisciplinary character of Psychometry from its foundations.

Based on the Rasch Model and other 20th century advances, modern psychometry has more and better theoretical developments of the constructs to be measured, with better foundations and multivariate statistical methods, which allow gathering and analyzing better evidence of the validity of the measures. Thus, current psychometrics constitutes an interdisciplinary field that has notably expanded the scope and precision in measurements and analysis of variables in social and behavioral sciences.

It would be very difficult in an article to comment and reflect on the theoretical and methodological contributions of Psychometry in the second half of the twentieth century, or on the current diversity of models and methods that have been projected in the last two decades; although that was not the objective of the present; but it would be an interesting challenge for Mexican psychometricians and the reason for other articles on these topics.

I conclude with a quote from Benoit Mandelbrot (1924-2010), recognized as the father of the Fractal Geometry of nature, who postulated the analysis of invariant geometric patterns at different scales and their applications in various fields: *"… interest in the History of Science is good for the soul of the scientist"* (Mandelbrot, 1997, pg. 40).

REFERENCES

Adams, E. W., Fagot, R. F. & Robinson, R. E. (1965). A Theory statistics. *Psychometrika,* 30, 99-127.

Anastasi, A. (1966). *Psicología diferencial*. Aguilar

Anastasi, A. (1974). *Psychological Testing.* Macmillan.

Anderson, J. (1995). *Learning and Memory: An integrated approach*. Wiley.

Antiseri, D. & Reale, G. (1988). *Historia del pensamiento filosófico y científico*. Tomo III. Herder

Barahona, A., Suárez, E., & Martínez, S. (2009). *Filosofía e Historia de la Biología.* UNAM.

Biagioli, F. (2016). Empirical and formal conditions in Helmholtz's theory of measurement. https://www.academia.edu/4125126/Empirical_and_Formal_Conditions_in_Helmholtzs_Theory_of_Measurement

Boring, E. G. (1950). *A History of Experimental Psychology.* CenturyCrofts

Boyer, C. B. (1986). *Historia de la Matemática*. Alianza Editorial.

Bridgman, P. W. (1927). *The logic of modern physics*. Macmillan.

Campbell, N. R. (1920). *Physics: the elements*. Cambridge University Press

Carroll, J. B. (1987). Medición de la Inteligencia. En Robert J. Sternberg, Inteligencia humana: La naturaleza de la inteligencia y su medición. Paidós.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, vol 16, no. 3, 297-334.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in Psychological Tests. *Psychological Bulletin*, 52, 281-302

Cronbach, L. J. (1971). Test Validation. In R. L. Thorndike (Ed), *Educational Measurement*. American Council on Education.

Cronbach, L. J. (1972). The Dependability of behavioral measurements: theory of generalizability for scores and profiles, Wiley.

Engelhard, G. (2018). Cogitations on Invariant Measurement. In Wilson, M. & Fisher *Psychological and Social Measurement*. Springer International Publishing.

Fraser, C. (1980) Measurement in Psychology. *British Journal of Psychology.* Vol 71-1

Galton, F. (1889). Natural Inheritance. Macmillan.

Gauss, K. F. (1801). *Disquisitiones Arithmeticae.* Leipzig. Traducción inglesa de A. Clark, 1965. Yale University Press.

Gauss, K. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis Solem ambientium.* FA Perthes.

Goodwin, C. J. (2009). *Historia de la Psicología Moderna*. Limusa Wiley.

Gribbin, J. (2003). Historia de la Ciencia: 1543-2001. Crítica, S. L.

Guttman, L. (1950). The basis for scalogram analysis. En S.A. Stouffer, L. Guttman, E. Suchman, P. F. Lazarfeld, S. A. Star, & J. A: Clausen (Eds). *Measurement and Prediction*. Princeton University Press.

Hambleton, R., Swaminathan, H. & Rogers, H. (1991). Fundamentals of item response theory. Sage.

Hawking, S. (2010). *Dios creó los números enteros*. Crítica.

Heidbreder, E. (1960). *Psicologías del siglo XX*. Editorial Paidós.

Helmholtz, H. (1887) Zählen und Messen, erkenntnistheoretisch betrachtet. In Hertz, P & Schlick, M. (1921). Herman von Helmholtz: Epistemological Writings.

Hölder, O. (1901). Die Axiome de Quantitat die Lehre von Mass. Berichte ueber die Verhandlugen der Koniglich Sachsischen Gessellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Class, S3, 1-64.

Jones, L. V. & Thissen, D. (2007). A History and Overview of Psychometrics. In C. Rao & S. Sinharay (Eds). *Handbook of Statistics*: *Psychometrics, vol 26.* Elsevier.

Kant, I. (1930) *Lectures on Ethics*. London: Methuen, Ltd.

Lord, F. M. & Novick, M.R. (1968). *Statistical Theories of Mental Tests Scores*. Addison-Wesley Reading Mass.

Luce, R. D. & Tukey, J. W. (1964). Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1-27.

Mandelbrot, B. (1997) *La Geometría Fractal de la Naturaleza*. tusQuets.

Martínez Guerrero, J. (2006). *La medida de estrategias de aprendizaje en estudiantes universitarios.* Madrid: UCM. Universidad Complutense de Madrid.

Michell, J. (1990). *An introduction to the logic of psychological measurement.* Lawrence Erlbaum.

Nunnally, J. C. (1978). *Psychometric theory*. (Second edition). McGraw Hill

Odifreddi, P. (2006). *La Matemática del Siglo XX. De los conjuntos a la complejidad*. Katz Editores.

Ramírez-Galarza, A. & Sienra-Loera, G. (2009). Invitación a las Geometrías no Euclidianas. UNAM.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen and Lydiche.

Rasch, G. (1961). On general laws and the meaning of measurement in Psychology. *Proceedings of the Berkeley Simposium of Mathematical, Statistics, and Probability*, 4, 321-333. University of California Press.

Rasch, G. (1966). An ítem analysis which takes individual differences into account. *The British Journal of Mathematical and Statistical Psychology,* 19, 49-57.

Ribes, E. (2005). ¿Qué es lo que se debe medir en Psicología? La cuestión de las diferencias individuales. *Acta Comportamentalia*, Vol 13, No. 1.

Riemann, B. (1929). *On the Hypotheses Which Lie at the Foundations of Geometry*. In D. E. Smith (trans). A Source Book in Mathematics.

Reckase, M. D. (2009) Multidimensional Item Response Theory. Springer.

Sánchez Ron, J. M. (2001). *Historia de la Física Cuántica*. Crítica Drakontos.

Stern, W. (1900) *Über Psichologie der individuellen Differenzen.* Leipzig: Barth.

Stevens, S. S. (1946) On the theory of scales of measurement. *Science,* 103, 667-680.

Stevens, S. S. (1956) The direct estimation of sensory magnitudes-loudness. *American Journal of Psychology*, 69, 1-25.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529-554

Thurstone, L. (1937). Psychology as a Quantitative Rational Science. *Science*, 85, 27-32

Traub, R. E. (1994). *Reliability for the Social Sciences*. Sage.

Van de Schoot, R., Schmidt, P. y Beuckelaer, A. Eds. (2015). *Measurement Invariance*. Lausanne: Frontiers Research Topics.

Van der Linden W. J. & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. Springer-Verlag.

Wissler, C. (1901). The correlation of mental and physical traits. *Psychological monographs*. No 3, 1-62.

Wright, B. D & Mock, M. (2004). An overview of the family of Rasch Measurement Models. In Benjamin Wright, and others *Introduction to Rasch Measurement: Theory, Models and Application.* JAM Press.