Psychometrics
New Insights in the Diagnosis of Mental Disorders

Edited by Sandro Misciagna

Psychometrics - New Insights in the Diagnosis of Mental Disorders

Edited by Sandro Misciagna

Contributors
Guillaume Gronier, Cristian Ramos-Vera, Víthor Rosa Franco, José Vallejos Saldarriaga, Antonio Serpa Barrientos, Kenneth J. Reid, Ek-Uma Imkome, Sandro Misciagna

Notice

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,500+
Open access books available

176,000+
International authors and editors

190M+
Downloads

156
Countries delivered to

Our authors are among the
Top 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Meet the editor

Doctor Sandro Misciagna was born on March 15, 1969 in Italy. He received a degree in medicine in 1995 and in neurology in 1999 at Catholic University in Rome. From 1993 to 1995 he attended a research laboratory involved in cerebellar functions and from 1994 to 2003 he attended the Neuropsychological department involved in human cognitive and behavioural disorders. From 2001 to 2003 he was a teacher of clinical neuropsychology, clinical neurology and cognitive rehabilitation at Catholic University and in 2003 he took Ph.D. in Neuroscience discussing behavioural and cognitive profiles of frontotemporal dementia. As a clinician, he has worked in neurological departments of Italian hospitals, Alzheimer's clinics, and Neuropsychiatric and Neuro-rehabilitative clinics. Since 2016 he works in the Neuroscience department of Belcolle Hospital in Viterbo where he currently deals with the diagnosis and treatments of epileptic patients. Since 2014 he collaborates with IntechOpen as an Academic Book Editor.

Contents

Preface

Psychometrics is a specialized field of psychology devoted to techniques of assessment and measurement and their respective theories. It is concerned with objective measurement of cognitive functions (e.g., intelligence, memory, attention, reaction times, etc.), characteristics of personality (e.g., hypochondria, hysteria, paranoia, social introversion, psychotic traits), emotions, behaviour, and socio-educational qualities. Its constructs are also applied to emotional and mental disorders (especially anxiety and depression).

Psychometric measures are inferred through mathematical models based on observation and statistical analysis conducted on a sample of subjects in comparison with the general population which brings to the planning of mental tests, scales, and open- or close-ended questionnaires.

Psychometrics – New Insights in the Diagnosis of Mental Disorders discusses psychometrics and its uses. It is divided into three sections.

The first section consists of an introductory chapter by Sandro Misciagna that defines psychometrics and discusses its history and main concepts. It presents a brief history of psychometrics from the first experimental studies of Sir Francis Galton, often referred to as the father of psychometrics, to the development of modern mathematical and statistical models applied to the study of human consciousness.

The second section consists of two chapters that discuss the theoretical bases of psychometrics. Chapter 2 by Cristian Ramos-Vera et al. presents the methodological principles of network psychometrics, which is a new approach to the study of latent variables. The authors examine the differences between the traditional approach of latent variable theory and network psychometrics in the context of psychopathology. They discuss the theoretical bases and advantages of this approach, such as its activation of a relationship between symptoms of some disorders or identification and classification of comorbidities. The authors also review analytical models for network psychometrics that are based on probabilistic graphs (non-directional, directional, and chain graphs). They emphasize the potential use of network models with different data sources (genetic, neurological, physiological, and behavioural) in a neuropsychological context. In this context, network psychometrics can measure cognitive variables such as memory, language, attention, executive functions, and processing speed variables. A network approach is proposed by the authors as a model for understanding the diagnosis and treatment of psychological disorders.

Chapter 3 by Guillaume Gronier proposes a methodological framework for psychometric analysis in cross-cultural adaptation of psychological scales. The author discusses the steps necessary for validation scales on the bases of other recommendation methodologies, such as the ones proposed by the International Test Commission, called ITC guidelines. According to these guidelines, the recommendations are divided into six

steps: pre-condition, test development, confirmation, administration, score scales, and documentation. The author describes how to measure internal consistency (Cronbach's alpha, McDonald's omega), statistical bases for validation of the psychological scales (exploratory factor analysis, principal component analysis, confirmatory factor analysis), convergent validity, time constancy, and socio-demographic analyses.

The third and final section of the book consists of three chapters about psychometric instruments and procedures. Chapter 4 by Sandro Misciagna is a review of psychometric methodologies in the assessment of dementia. Most forms of dementia are classified by means of morphological techniques, assays of biomarkers in cerebrospinal fluid, and neuropsychological assessment into degenerative forms, dementia of vascular type, and dementia secondary to other conditions. It is very difficult to make a clear-cut diagnosis of the different types of dementia by means of clinical methods. However, many psychometric tests play a prominent role in the screening and evaluation of patients with cognitive impairment. Some tools can help clinicians in differential diagnosis among the various forms of dementia, such as those that assess clinical aspects, tests that focus on specific cognitive areas, or behavioural inventories. Still, nowadays there is no consensus about the best strategies for screening and assessment of cognitive impairment among elderly subjects. In this chapter, the author reviews the screening tools and psychometric test instruments that healthcare professionals can use for screening and neuropsychological assessment of geriatric individuals with cognitive disorders to help diagnose dementia as well as make differential diagnosis of the most common forms of dementia.

Chapter 5 by Kenneth J. Reid focuses on psychometric analysis of a version of the Student Attitudinal Success Inventory (SASI-I) to assess first-year engineering students. This instrument was a survey designed to collect data on the affective non-cognitive characteristics of incoming students. Data were based on a cohort of undergraduate engineering students enrolled in a Midwestern university over a three-year period from 2004 to 2006. The instrument consisted of 161 items indicating responses on a 5-point Likert scale (1 for strongly agree to 5 for strongly disagree). The nine measures used to assess academic success are academic motivation, metacognition, deep learning, surface learning, academic self-efficacy, leadership, team vs individual motivation, expectancy-value, and major decision. The scale scores of SASI-I demonstrate evidence of reliability and validity even if further studies are needed to evaluate this tool over the years.

Chapter 6 by Prof. Ek-Uma Imkome discusses the assessment of psychiatric and mental disorders regarding symptoms of stressful responses, such as depression, anxiety, and suicidal ideation. The author uses the context of the COVID-19 pandemic since these disorders were particularly evident during this time because of quarantine and social isolation. The author states that the most relevant psychological problems during the COVID-19 pandemic were fear, phobias, anxiety disorders, psychological trauma, and stress. The author reviews the steps for administering and interpreting a development scale to assess mental health. The first step is to determine the possible measurement scale to use, such as the Likert rating scale or questionnaire, and possible methodologies of administration (cross-sectional, self-reporting methodology, web-based surveys). The second step is to analyse the results using statistical methods (exploring factor analysis, confirmatory factor analysis) and then optimize the results. A scale with psychometric properties could help nurses screen and identify individuals with psychiatric disorders.

This book provides a comprehensive overview of theoretical approaches, instruments, and procedures of modern psychometric science with examples of psychometric assessments in human experimental models and possible clinical applications in human disorders.

Dr. Sandro Misciagna
Neurology Department,
Belcolle Hospital,
Viterbo, Italy

Section 1

Definition, History and Theories of Psychometrics

Chapter 1

Introductory Chapter: Psychometrics

Sandro Misciagna

1. Introduction

Psychometrics is a scientific discipline that concerns the theories and construction of models for measurement of psychological data. These models try to establish how psychological latent constructs, such as human intelligence, psychological abilities, or mental disorders, can be measured through use of psychological tests, genetic profiles, or neuroscientific information [1]. This problem is commonly approached by building mathematical measurement models in which latent variables act as a common determinant of a set of observable variables [2]. Latent variables models represent the construct of interest as a latent variable that is the common determinant of a set of scores. Psychometrics involves the formalization of psychological theories and the design of psychological assessment instruments including surveys, scales, and open or closed questionnaires [1]. Psychometricians are usually psychologists with advanced training in psychometrics and measurement theory. However, psychometrics is a highly interdisciplinary field with connections with statistical models, data theory, econometrics, biometrics, measurement theory, and mathematical psychology.

2. Brief history of psychometrics

The birth of psychometrics is generally situated at the end of the nineteenth century when Sir Francis Galton in 1884 created an anthropometric laboratory to determine psychological attributes experimentally [3]. Among the first constructs of interest, he proposed to measure keenness of sight, color sense, and judgment of eyes. Galton, often referred to as the father of psychometrics, attempted to measure such attributes by using a vast variety of tasks, recording performance accuracy as well as reaction times. In his book entitled "Hereditary genius" he described different characteristics that people possess regarding sensory and motor functions such as visual acuity, physical strength, or reaction times. He included among anthropometric measures also mental abilities that could be measured through mental tests.

Francis Galton was probably inspired by Charles Darwin who in 1859 published the book "On the origin of species by means of natural selection" [4]. In this book, Darwin described the role of natural selection in the emergence, over time, of different populations of plants and animals. According to his theory, individuals with more adaptive characteristics were more likely to survive and procreate in certain environments, while individuals with less adaptive characteristics were more likely to go extinct.

Another pioneer in the field of psychometrics was James McKeen Cattel who coined the term "mental tests" and was responsible for the research that led to the development of modern tests [5].

At the same time that Darwin, Galton, and Cattel were making their discoveries, the German educator Johann Friedrich Herbart was interested in discovering the mysteries of human consciousness so he created the first mathematical models of the mind [5]. Inspired by his works, the German physiologist Ernst Heinrich Weber tried to demonstrate the existence of a psychological threshold, arguing that a minimum stimulus was necessary to activate a sensory system. After Weber, the German psychologist Gustav Theodor Fechner devised the law that the strength of a sensation grows as the logarithm of the stimulus intensity.

During the early twentieth century, the interest in measuring human qualities intensified greatly when the US implemented programs to select soldiers using tests that measured a range of abilities relevant to military performance. Such tests produced a great deal of data, which led to questions that inspired the birth of psychometric theory as we currently know regarding in particular the analysis of psychological test data, the properties of psychological tests, and the selection of the best tests suited for a certain purpose. Almost immediately two important properties of the tests were identified [6].

The first property of a test concerns the notion of reliability, which is the question of whether a test produces consistent scores when applied in the same circumstances. One of the first scientists to take interest in this topic was the psychologist and statistician Charles Edward Spearman who wrote in 1904 an article about the theory of measurement reliability [7]. A reliable measure is consistent across time, individuals, and situations that is the question of generalization, from test to test, from examiner to examiner, from situation to situation, or from testing time to testing time [8]. For example, it regards the question of whether an intelligence test produces the same score of intelligence quotient when administered to people with the same level of intelligence. In the article of Spearman, he developed most of the basic statistics about reliability including corrections for attenuation, standard error of measurement, correction of the split half, reliability coefficient for test length, and other statistics that are identified with test reliability [9]. In his classic book "Theory of Mental Tests" written in 1950, Harold Gulliksen extended the simple mathematical models for reliability developed by Spearman and provided an extensive mathematization of reliability based on the concept of parallel tests [10].

The second property of a test concerns the notion of validity, which is the question of whether a test is valid in measuring what it is intended to measure. For example, it regards the question if an intelligence test actually measures intelligence. There are three main types of validity on which the worth of psychological tests is determined, which are predictive validity, content validity, and construct validity. In 1954 well-known experts on test development stated that the predictive validity of a test is its correlation with a criterion [11]. Content validity is a demonstration that the items of a test do an adequate job of covering the domain being measured. For example, many types of tests were developed in the 50s in the civil service, the military, the industry, and schools at all levels of education [11]. Construct validity is related to measures of other constructs as required by the theory. It means that the validity of a test could not be determined by the correlation with a single criterion, but it is necessary to provide numerous relationships with variables with which it logically relates [12]. Other forms of validity are criterion-related validity, which refers to the extent to which a

test predicts a sample of behavior, or concurrent validity when the criterion measure is collected being validated.

The concepts of reliability and validity are still today among the most essential elements in the evaluation of quality of any psychological tests. Reliability is necessary but not sufficient for validity. Furthermore, the definition of reliability is widely accepted while the definition of validity is widely contested [13]. The development of reliability theory culminated in the half of the twentieth century with the work of Lord and Novick in 1968 who presented the currently accepted definition of reliability [14]. According to their definition, reliability is a signal-to-noise ratio or better the ratio of true score variance to observed score variance. This concept was somewhat differently conceptualized in different theoretical frameworks. In fact, according to the latent variable theory of Mellenberg formulated in 1994, it was considered a measurement of precision [15], while according to the generalizability theory of Cronbach formulated in 1972, it was considered as a generalizability measure [16]. However, Lord and Novick's definition typically follows as a special case, which indicates the consistency of the general psychometric framework [14].

Reliability could be estimated in various ways, such as from correlation between two test halves, from the average correlation between test items, and from the correlation between two administrations of the same test at different times [17]. Some discussions are about how to optimally estimate reliability or about the coefficients that should be preferred in a particular contest. For example, consistency over repeated measures of the same test can be assessed with the Pearson correlation coefficient, which is often called test-retest reliability. Internal consistency may be assessed by correlating performances on two halves of a test, which is called split-half reliability. One of the most commonly used indexes of reliability is Cronbach's α, which is equivalent to the mean of all possible split-half coefficients [18]. Coefficient alpha is a correlation between an existing test and a hypothetical test, under the assumption that (1) the average correlation among items in each of the two tests is the same and (2) the average correlation between items on the two tests is the same as the average correlation within items on each of the tests.

Other approaches include the intra-class correlation, which is the ratio of the variance of measurements of a given target to the variance of all targets.

As regards the validity of a test there are no widely accepted methods to determine whether a test is valid or is to estimate the degree of validity of a psychometric test [19]. The question of whether a test measures what it aims to measure raises from the question about the nature of the psychological constructs themselves. One question on validity of a test concerns if psychological constructs could be measured in the same way or if different ways of measurement should be invoked [20]. Another question concerns if psychological constructs give a realistic interpretation or if they are summaries of data [21]. Another question concerns if we could talk about validity of tests at all or if we should talk about the validity of interpretations of the test scores [13]. These examples of questions demonstrate that the validity of a test is one of the most problematic psychometric concepts.

Psychometric theory and its practice have developed largely even if there was not a definitive answer to these questions. With the spread of psychometrics as a psychological science, it got inspiration from mathematic and statistic concepts for the development of measurement models of psychological data, therefore, becoming a largely technical discipline. Generally, such models contain a psychological construct to be measured, such as the expectations of the observed scores. Statistical models

have largely contributed to the development of psychometric theories such as the modern test theory of Rash developed in 1960 [22], the classical test theory of Lord and Novick developed in 1968 [14], the latent class analysis of Lazarsfeld and Henry developed in 1968 [23], the cogeneric model of Jöreskog developed in 1971 [24], and the nonparametric item response model of Mokken developed in 1971 [25]. After the theorization of these models, one of the main topics of the psychometric research became the development of software to fit and estimate them and the development of estimation algorithms [26], software for test analysis [27], and general latent variable modeling [28]. These developments have taken place in the last three decades of the twentieth century.

In 2014, American Psychological Association (APA), the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME) published a revision of the book "Standards for Educational and psychological testing" for development, evaluation, and use of psychological tests [29]. This book covers topics about test validity, reliability, errors of measurements, test designs, use of scales, score linking, how to establish cut-off scores, test administration, testing application, and interpretation of psychometric tests.

3. Main concepts of the psychometric theory

Psychometric models relate to a set of observed variables by mapping positions in the latent structure to distribution of the observed variables. This is usually done by specifying a conditional distribution function of the observables given the latent structure. Thus, the general framework consists of a simultaneous regression of observed variables on a latent variable or a set of latent variables.

Three principal theoretical models are derivatives of this idea that are: (1) the form of the latent structure according to which there is a continuous line or a set of latent classes [30], (2) the form of a regression function that consists in a step function or a logistic function, and (3) the distribution or density appropriate to the observations such as a binomial distribution or a normal density.

According to the linear common factor model when the latent structure is a unidimensional continuum, the regression function is linear and the observables follow a normal density [24]. Instead, according to the two parametric logistic models, the latent structure is a unidimensional continuum and the regression function is logistic, and the observables follow a binomial distribution [31]. Finally, in the latent class model, the latent structure is categorical and the observed variables are binary [23]. The latent structure consists of a representation of the constructs to be measured such as intelligence, while the observed scores are typically the concrete behavioral responses, such as the items to determine the QI in an intelligence test. Consequently, the psychometric model coordinates the correspondence between the observational and the theoretical terms creating a measurement model [32]. This means that the psychometric model is a measurement model that coordinates theory with observations and not in the sense that human behavior can be successfully analyzed in terms of quantitative laws.

The reliability is related to the psychometric model through the concept of measurement precision, which is inversely related to the variance of the observed scores [15]. Therefore, the higher the variance of the conditional distribution of the observables, the lower the measurement precision of the observables. Measurement precision may not be identical for different positions of the latent structure.

In the linear common factor model, measurement precision is identical for all values of the latent variable. In the Rash model, measurement precision is highest for the latent position since the logistic regression of the observable has its inflection point [22]. The reliability of test scores, as theorized in the classical test theory of Lord and Novick, is an unconditional index of measurement precision [14].

A subfield of psychometrics that plays an important role in the analysis of educational tests is the item response theory [33] in which the observed variables are the responses to test items such as the items in a QI test. In the item response theory, the latent function that specifies the regression of the observed variables is known as an item characteristic curve (ICC). Generally, item response theory assumes models with a unidimensional and continuous latent structure. In educational testing, items are typically scored dichotomously (0: incorrect and 1: correct) and the item characteristic curve is bounded from above and below and is often modeled using a nonlinear function. The slope of the item characteristic curve, at a given point on the latent scale, is proportional to the ability of the item to discriminate between positions above and below and determine the amount of the item information at that point. The result of the item information against the latent variable is the item information function (IIF). The item information function can be used in psychometrics to regulate the selection of items.

This idea is on the bases of adaptive testing, which is becoming more and more increasingly important with the advent of computerized test administrations [34]. In adaptive testing, items are administered sequentially and are selected for administration adaptively. In the process of item administration, examinee ability is estimated on the basis of the item responses given so far. The next item to be administered is instead chosen on the basis of the slope of the item information function at the estimated examinee ability. In this way, tests can be shortened without disturbing the reliability.

One of the features often necessary in psychometric tests is measurement invariance [35]. This is especially required when psychometric tests are used to select individuals based on their attitudes such as in student placement or in job selection. Generally, these selection processes operate on heterogeneous populations with respect to background variables such as age, sex, education, and ethnicity. In these cases, the psychometric test should function in the same way across different subpopulations and should not produce bias in the test scores in a specific group. For example, this bias could arise when an intelligence test contains questions that are easier in subpopulations with a specific background regardless of their intelligence level. For example, in a test that contains general knowledge questions, it could be more difficult for ethnic minorities for reasons independent of their level of intelligence.

An alternative to the latent variable model in the psychometric literature concern the multidimensional scaling model [36]. Multidimensional scaling is a method for finding a simple representation of data with a large number of latent dimensions. Multidimensional scaling is a psychometric tool to infer the number of underlying dimension in proximity data. An example is given by the degree of different facial expressions that are judged to be similar. Metric multidimensional scaling is applied when similarity measures are continuous [37] while nonmetric multidimensional scaling is applied when similarity measures are ordinal [38]. In the multidimensional scaling model, individual differences are weighted according to the underlying dimensions differently across subjects [39]. In this way, each subject receives a different weight for each dimension. An important instance of multidimensional scaling with individual differences is the unfolding analysis in which each subject is

assumed to have an ideal point on the dimension underlying the preference data [40]. A stimulus is preferred when it is close to the subject's ideal point.

Other alternative psychometric models that make a connection between theory and data are: (1) taking the construct to be a universe of behaviors, described by Cronbach in 1972 [16]; (2) representing the construct as a common effect of the observed variables, described by Bollen and Lennox in 1991 [41]; and (3) interpreting the construct as a causal system in which the observables influence each other, described by Cramer in 2010 [42].

Recent advances in psychometrics have focused attention on models that deal with more complex situations. Some extensions include the incorporation and development of multilevel and random effects structures in item response theory models [43], factor models [44], or latent class models [45]. In these models, item parameters may become random variables and psychometric analysis of data is common in large-scale assessments.

Another recent innovation is the use of computerized technologies tests that make available response times in addition to ordinary responses. It enables more refined estimation techniques and new models of assessment, especially for multidimensional item response theory models [46], factor analysis of categorical and nonnormal data [47], cluster analysis, discriminatory analysis, cognitive diagnosis models [48], and nonlinear factor models [49]. The factor analysis [50] is a method for determining the underlying dimensions of data that does not require a consensus on appropriate procedures for determining the number of latent factors [51]. Cluster analysis is a psychometric approach to find objects that are like each other's. Multidimensional scaling, factor analysis, and cluster analysis are all multivariate descriptive methods used to distill simpler structures from large amounts of data. As factor analysis, discriminatory analysis is a multivariate descriptive statistic that tries to demonstrate that a multiple discriminant function is a special type of factoring, in which the factors are obtained to optimally discriminate among two or more groups of people on the basis of the scores from a set of tests [52]. For example, in a classic paper by Tiedman and colleagues written in 1952, they demonstrated discriminatory analysis with Airman Classification Battery applied to the problem of assigning air force personnel to eight occupational specialties [53]. However, the models underlying discriminant function are more appropriate with respect to noncognitive variables (such as personality) rather than cognitive variables. From a conceptual point of view, recent psychometric literature is focused on the status of psychometric measurement models, the relation between psychometrics and psychology, and the usefulness of Cronbach's alpha as a measure of reliability [54].

The use of open-source statistical software programs has also enabled psychometricians to develop their own models and share these with other researchers.

4. Psychometric research

4.1 From human sciences to artificial machines

Psychometric experiments studying individual differences are mainly concerned with correlation among variability in response to the same group of subjects to different sources of response elicitation.

As argued by Jum Nunnally, there are three overlapping topics in psychometrics methods [11]. The first psychometrics method is mainly a deductive analysis when

it concerns multidimensional scaling, factor analysis, and item analysis. Most of the deductive models are expressed in mathematical models or have mathematical implications. The deductive models in psychometrics have always been closely wedded to basic research on empirical individual differences with respect to achievements, abilities, personality, and other types of human traits [11].

The second psychometrics method is a mathematical method that concerns basic research on individual differences. Examples are psychometric studies that try to determine the structure of specific human abilities [55].

The third psychometrics method concerns the measurement of individual differences in applied settings such as schools, government, military, industry, and other institutions. The application in this setting depends both on the use of deductive models and basic research on human traits [11].

One of the first psychometric instruments was designed to measure human psychological traits, abilities, and characteristics of personality. The historical approach was developed by Alfred Binet and Theodore Simon who made the Stanford-Binet IQ test to measure human intelligence [56]. Subsequently, these tests were revised and other new important tests were developed such as the Wechsler Intelligence Scale for adults and for children (WAIS). Another focus in psychometrics regards personality testing, even if still nowadays there are not widely accepted way of measuring personality since the theoretical construct of personality is a complex idea. Some of the best-known personality instruments include the Minnesota multiphasic personality inventory (also known as MMPI), the Big Five inventory (or Big 5), the Rorschach Inkblot test, the neurotic personality questionnaire (KON 2006) [57], the Eysenk's personality questionnaire (EPQ-R), the personality and preference inventory, and the Myers-Briggs Type indicator. Numerous test batteries of personality grew out of previous findings from factor analysis; in other cases, factor analysis served to construct subtests of a battery with a small number of factors.

Psychometric approaches have also studied extensively human psychological attitudes, human abilities, and educational evolution. Around 1950s, researchers developed collections of tests with heterogeneous criteria finalized to predict success in a particular job or social activity based on mental attitudes. In fact, they discovered that success could be better predicted by use of a battery composed of tests, each of which was homogeneous with respect to a particular psychological trait. In this area, psychometrics tests are applied in setting for making important decisions about people such as selecting people for pilot training or assigning individuals to different types of treatments. Another example would be in comparing groups of children who have undertaken different types of preschool. In this case, psychometric measures would concern various aspects of achievements in relation to language development. A concise battery of aptitude tests does a good job of predicting school grades and other school performances. Some of the correlations with school grades have ranged up into the 80s and are frequently good indicators for future performances in college. However, the philosophy of education has been influenced by the Skinnerian movement that sustains techniques of behavioral modification. This movement argues that individual initial competencies can be changed after a specific training and then it would be superfluous to apply achievement tests. The answer of the researchers who sustain a psychometric approach is that tests can determine the initial level of competence in order to know where to start in the training program, employ aptitude tests to predict how rapidly the person achieves the competencies, and determine the level of competence that is reached.

With the advent of high-speed computers, researchers developed psychometric hardware that could be useful in helping to solve social problems [58]. This approach could be ideal for distinguishing the natural grouping of people, animals, or material objects on a set of relevant measurements.
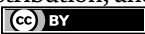
More recently psychometrics is also approaching nonhuman abilities such as learning evolution of machines with particular regard in the area of artificial intelligence so some researchers have proposed an integrated approach under the name of universal psychometrics [59].

Author details

Sandro Misciagna
Neurology Department, Belcolle Hospital, Viterbo, Italy

*Address all correspondence to: sandromisciagna@yahoo.it

IntechOpen

References

[1] Tabachnick BG, Fidell LS. Using Multivariate Statistics. 4th ed. Boston: Allyn and Bacon; 2001

[2] Sijtsma K. Introduction to the measurement of psychological attributes. Measurement. Journal of the International Measurement Confederation. 2011;**44**(7):1209-1219

[3] Galton F. Measurement of character. Fortnightly. Review. 1884;**36**:179-185

[4] Darwin C, Kebler L. On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life. London: J. Murray; 1859

[5] Kaplan RM, Saccuzzo DP. Psychological Testing: Principles, Applications, and Issues. 8th ed. Belmont, CA: Wadsworth, Cengage Learning; 2010

[6] Kelley TL. Interpretation of Educational Measurements. New York: Macmillan; 1927

[7] Spearman C. The proof and measurement of association between two things. The American Journal of Psychology. 1904;**15**:72-101

[8] Nunnally JC, Berstein IH. Psychometric Theory. 3rd ed. New York: McGraw-Hill; 1994

[9] Nunnally JC. Psychometric Theory. New York: McGraw-Hill; 1967

[10] Gulliksen H. Theory of Mental Tests. New York: Wiley; 1950

[11] Nunnally JC. Psychometric theory - 25 years ago and now. Educational Researcher. 1975;**4**(10):7-21

[12] Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychological Bulletin. 1955;**52**:281-303

[13] Newton PE. Clarifying the consensus definition of validity. Measurement: Interdisciplinary Research & Perspective. 2012;**10**(1-2):1-29

[14] Lord FM, Novick MR. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley; 1968

[15] Mellenbergh GJ. A unidimensional latent trait model for continuous item responses. Multivariate Behavioral Research. 1994;**29**(3):223-236

[16] Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles. New York: Wiley; 1972

[17] Kuder GF, Richardson MW. The theory of estimation of test reliability. Psychometrika. 1937;**2**:151-160

[18] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;**16**:297-334

[19] Lissitz RW. The Concept of Validity: Revisions, New Directions, and Applications. Charlotte, NC: Information Age; 2009

[20] Michell J. Measurement in Psychology: A Critical History of a Methodological Concept. Cambridge: Cambridge University Press; 1999

[21] Borsboom D. Measuring the Mind: Conceptual Issues in Contemporary Psychometrics. Cambridge: Cambridge University Press; 2005

[22] Rasch G. Probabilistic Models for some Intelligence and Attainment Tests. Copenhagen: Paedagogiske Institut; 1960

[23] Lazarsfeld PF, Henry NW. Latent Structure Analysis. Mifflin: Houghton; 1968

[24] Jöreskog KG. Statistical analysis of sets of congeneric tests. Psychometrika. 1971;**36**(2):109-133

[25] Mokken RJ. A Theory and Procedure of Scale Analysis. The Hague: Mouton; 1971

[26] Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika. 1981;**46**(4):443-459

[27] Zimowski MF, Muraki E, Mislevy RJ, Bock RD. BILOG-MG: Multiplegroup IRT Analysis and Test Maintenance for Binary Items [Computer Software]. Chicago: Scientific Software International; 1996

[28] Arbuckle JL. IBM SPSS Amos 19 User's Guide. Chicago, IL: SPSS; 2010

[29] American Educational Research Association. Standards for Educational and Psychological Testing. Apa Org. Direct Other Orders to. 1430 K Street, NW, Suite 1200 Washington, DC 20005: AERA Publications Sales; 2014

[30] Lazarsfeld PF. Latent structure analysis. In: Koch S, editor. Psychology. Vol. III. New York: McGraw-Hill; 1959

[31] Birnbaum A. Some latent trait models and their use in inferring an examinee's ability. In: Lord FM, Novick MR, editors. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley; 1968. pp. 397-479

[32] Michell J. Measurement scales and statistics: A clash of paradigms. Psychological Bulletin. 1986;**100**(3):398-407

[33] Embretson SE, Reise SP. Item Response Theory for Psychologists. Mahwah, NJ: Erlbaum; 2000

[34] Van der Linden WJ, Glas CAW. Eds. Computerized Adaptive Testing: Theory and Practice. Kluwer, Norwell, MA. 2000

[35] Meredith W. Measurement invariance, factor analysis, and factorial invariance. Psychometrika. 1993;**58**(4):525-543

[36] Davison ML, Sireci SG. Multidimensional scaling. In: Handbook of applied multivariate statistics and mathematical Modeling copyright 2000 by academic press. Cap. 2000;**12**:323-352

[37] Torgerson WS. Multidimensional scaling: Theory and method. Psychometrika. 1952;**17**:401-409

[38] Shepard RN. Analysis of proximities: Multidimensional scaling with an unknown distance function. Psychometrika. 1962;**27**:125-140

[39] Carroll JD, Chang JJ. Individual differences and multidimensional scaling via an N-way generalization of Eckart-young decomposition. Psychometrika. 1970;**35**:282-319

[40] Coombs C. A Theory of Data. New York: Wiley; 1964

[41] Bollen KA, Lennox R. Conventional wisdom on measurement: A structural equation perspective. Psychological Bulletin. 1991;**110**(2):305-314

[42] Cramer AOJ, Waldorp LJ, van der Maas H, Borsboom D. Comorbidity:

A network perspective. The Behavioral and Brain Sciences. 2010;**33**(2-3):137-193

[43] Fox JP, Glas CAW. Bayesian estimation of a multilevel IRT model using Gibbs sampling. Psychometrika. 2001;**66**(2):271-288

[44] Rabe-Hesketh S, Skrondal A, Pickles A. Generalized multilevel structural equation modelling. Psychometrika. 2004;**69**(2):167-190

[45] Lenk P, DeSarbo W. Bayesian inference for finite mixtures of generalized linear models with random effects. Psychometrika. 2000;**65**(1):93-119

[46] Reckase MD. Multidimensional Item Response Theory. London: Springer; 2009

[47] Molenaar D, Dolan CV, de Boeck P. The heteroscedastic graded response model with a skewed latent trait: Testing statistical and substantive hypotheses related to skewed item category functions. Psychometrika. 2012;**77**:455-478

[48] De la Torre J, Douglas JA. Higher order latent trait models for cognitive diagnosis. Psychometrika. 2004;**69**(3):333-353

[49] Lee SY, Zhu HT. Maximum likelihood estimation of nonlinear structural equation models. Psychometrika. 2002;**67**(2):189-210

[50] Ang RP. Development and Validation of the Teacher-Student Relationship Inventory Using Exploratory and Confirmatory Factor Analysis. The Journal of Experimental Education. 2005;**74**(1):55-74. DOI: 10.3200/JEXE.74.1.55-7

[51] Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. Psychological Bulletin. 1986;**99**(3):432-442

[52] Fisher RA. The statistical utilisation of multiple measurements. Annals of Eugenics. 1928;**8**:376-386

[53] Tiedeman DV, Bryan JG, Rulon PJ. Application of the multiple discriminant function to data from the airman classification battery. In: Research Bulletin, 52-37. San Antonio, Texas: Air Training Command, Lackland AFB; 1952

[54] Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika. 2009;**74**(1):107-120

[55] Thurstone LL. Multiple-Factor Analysis. Chicago: University of Chicago Press; 1947

[56] Stern T, Fava M, Wilens T, Rosenbaum J. Massachusetts General Hospital Comprehensive Clinical Psychiatry. 2nd ed. London, New York: Elsevier Health Sciences. 04 Sep 2015:1060. ISBN: 9780323328999. ISBN: 9780323295079

[57] Aleksandrowicz JW, Klasa K, Sobański JA, Stolarska D. KON-2006 Neurotic Personality Questionnaire. Archives of Psychiatry and Psychotherapy. 2009;**1**:21-22

[58] Edwards AL. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. Journal of Applied Psychology. 1953;**37**:90-93

[59] Locurto C, Scanlon C. Individual differences and spatial learning factor in two strains of mice. The Behavioral and Brain Sciences. 1987;**112**:344-352

Section 2

Theoretical Bases of Psychometrics

Chapter 2

Psychometric Networks and Their Implications for the Treatment and Diagnosis of Psychopathologies

Cristian Ramos-Vera, Víthor Rosa Franco,
José Vallejos Saldarriaga and Antonio Serpa Barrientos

Abstract

In this chapter, we present the main methodological principles of psychological networks as a way of conceptualizing mental disorders. In the network approach, mental disorders are conceptualized as the consequence of direct interactions between symptoms, which may involve biological, psychological, and social mechanisms. If these cause-and-effect relationships are strong enough, symptoms can generate a degree of feedback to sustain them. It is discussed how such an approach contrasts with the traditional psychometric approach, known as the Latent Variable Theory, which assumes that disorders are constructs that exist but are not directly observable. Furthermore, it is also discussed how new neuropsychological hypotheses have been derived in the network approach and how such hypotheses generate direct implications for the understanding of diagnosis and treatment of psychological disorders. Finally, the recentness of the network approach in psychology and how future studies can establish its robustness are discussed.

Keywords: graph theory, network analysis, psychometrics, neuropsychology, clinical measurement

1. Introduction

Network psychometrics is a new approach to the study of latent variables (i.e., psychological constructs) that contrasts with the traditional psychometric approach. In the traditional approach, responses to items on a psychological instrument (e.g., responses to questions such as "Do you sleep poorly?") are analyzed as evidence of an underlying characteristic (or psychopathology) that the researcher or clinician wishes to measure [1]. This idea is formalized in analytic methods, such as Factor Analysis, Item Response Theory, Latent Class Analysis, and Mixture Modeling, among others, which are the main ways to validate psychological and psychiatric instruments [2].

In theoretical terms, the traditional psychometric approach, known as Latent Variable Theory [3], suppose that the observed behavior (e.g., responses to items on a psychological questionnaire or scale) is the effect of a common cause (in the clinical context, usually assumed to be psychiatric disorders). This approach is used

in different ways in psychology and psychiatry (see the study by Demjaha et al. [4]). Whereas in psychology metric models (i.e., those that assume that psychiatric disorders are quantitative variables) are more commonly used, in psychiatry categorical models (i.e., those that assume that a disorder is or is not present) are more common. These theoretical differences translate into differences in how to diagnose, classify, and even clinically act on psychiatric disorders [4].

Network psychometrics has the main feature in relation to the traditional psychometric approach that it does not necessarily assume that psychological constructs exist [5]. More specifically, network models of psychopathology assume that symptoms form complex cause-and-effect relationships with each other, dynamically reinforcing each other and giving rise to psychiatric disorders [6–8]. However, there are alternative network models that allow different interpretations. Some are even compatible with the Latent Variable Theory. The aim of the present study is to analyze critically the main distinctions between Latent Variables Theory and Network Psychometrics in the context of psychopathologies. As specific objectives, we will critically evaluate Latent Variable Theory in the causal perspective of Pearl [9], present the theoretical foundations of Network Psychometrics, and discuss the theoretical and practical implications for clinical study and action in the context of psychopathology.

2. Latent variables in psychology

Latent Variable Theory, in its various implementations in statistical models, is formally indistinguishable from the so-called common cause model [9]. The models of this theory assume that when the latent variable is tested, the correlations between observable behaviors should disappear. This property is known as "local independence," which is normatively imposed in traditional psychometric models [10]. This implication derives from the fact that correlations between effects with a common cause are suppressed whenever there is no direct causal relationship between these effects and the relationship between the two variables is controlled by the common cause [9].

Thus, the psychometric model and the causal interpretation affirm that the psychological (or psychopathological) construct naturally causes the behaviors. This relationship is certainly not a coincidence: the standard psychometric model is based on the notion that different indicators measure the same thing because they depend on the same property and no other [11]. Another consequence of Latent Variable Theory is that item response can be described in terms of a functional relationship between a single property of individuals and items [1]. Thus, in the case of unidimensional tests (i.e., based primarily on a single construct or disorder), it is assumed that all psychopathology test items are statistically interchangeable [12]. From a pragmatic point of view, Item Response Theory models, such as the Two-Parameter Logistic Model [2], can demonstrate which items are most closely related to the central construct being measured (so-called item discrimination), as well as the sensitivity of items to the magnitude of the construct (so-called item difficulty). However, it cannot be said that there are items that play a more central role in the identification of the construct, and as long as their difficulties and discriminations are adjusted, all items are equivalent. Such implication contrasts with clinical practice, where it is identified that there are more characteristic or more influential symptoms in each psychopathological disorder [4].

Regarding the development of instruments for the measurement, identification, or screening of psychopathological disorders, the common cause model provides psychometrics and psychiatry with a standard approach to test construction and analysis [13]. This approach is implemented with the following steps:

1. create a set of items as a measure of the same construct;

2. collect data and apply a statistical model that formalizes this common causal dependence;

3. eliminate or modify items that do not fit the model, and

4. repeat steps 1 through 3 until the model fits the data adequately.

Following these steps, provided we are changed from the recommended order it is possible to measure virtually any construct [1], although there is criticism as to whether such an approach actually produces a true measure [14, 15]. Such an approach will not be accurate if there is no common factor across items, which some researchers in psychopathology suggest is the case (see the study by Fried et al. [16]).

For example, in one of the most influential works in psychometric history in the clinical and psychiatric context, Krueger [17] defined the two main higher-order factors of his model in terms of two central psychopathological processes: internalizing and externalizing. These latent variables of the measurement model (i.e., the statistical factors) refer to two intrinsically significant psychological mechanisms that, in principle, could be easily observable in the expression of a picture of even heterogeneous behaviors. According to this author, internalization can lead to depression or anxiety, whereas externalization can lead to antisocial or aggressive behaviors. Although the behaviors are very different, these differences would reflect basic processes in the way psychopathology manifests itself.

In Krueger's original approach [17], the underlying causal homogeneity is psychological in nature, but more recent studies propose that the underlying causal homogeneity is neurological or genetic. Overall, there is a growth in studies that seek to reveal the "underlying brain mechanisms" of psychopathology [18]. In essence, however, all of these approaches boil down to the same explanatory model: there is some "deeper" cause of the symptomatology (e.g., a psychological variable, a brain abnormality, a genetic mutation, among others) that explains why people show the observed symptomatology. Certainly, there are many advances in this area (see the study by Rose [19]). However, it is also known that there are a number of socioeconomic influences on the mental health of individuals, which are not considered in the identification, classification, and treatment of disorders (see the study by Silva et al. [20]).

3. The network psychometry approach and psychopathology

The network psychometrics approach assumes that the lack of stronger evidence for the latent origins (whether psychological, neurological, genetic, or otherwise) of psychopathological disorders cannot be a matter of measurement problems or a limited understanding of genetics and the brain. The alternative proposed by the network approach is that this lack of evidence may be the result of an erroneous way of thinking about or assessing the relationship between symptoms and disorders [21].

More specifically, in the network approach to psychopathology, it is assumed that disorders emerge when, over time, specific symptoms become more strongly connected [8]. From a pragmatic point of view, psychopathology is identifiable when the probability of observing a symptom is higher than "normal" (additionally another symptom has been observed).

It is important to specify that many diagnostic systems, such as the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [22], do not make any explicit assumptions about the origins of the symptoms. No explanatory mechanisms of the disorder are presented, but only the main symptoms and clinical criteria. Traditionally in psychopathology, no direct attribution of relationships between symptoms of disorders and the common effect of a latent variable is made directly. The relationships between symptoms in their various contexts are established as criteria (see the study by Ramos Vera; Cramer et al.; Borsboom; and Spitzer et al. [23–26]). For example, a person who often has panic attacks in public places (symptom 1) is likely to be afraid that the attacks will recur (symptom 2) and, consequently, will avoid public places frequently (symptom 3). In another example, a person who cannot sleep (symptom 1) will end up tired and unable to concentrate (symptoms 2 and 3), which may cause him or her to feel guilty about poor performance at school or work (symptom 4). Evidence of this type of relationship between symptoms is common and makes it clear that local independence and equivalence between symptoms are not real for several disorders and their indicators.

It should be noted that, despite not explicitly assuming causal symptom structure, diagnostic systems, such as DSM-5, include such structures at least implicitly [24]. For example, a person who sleeps poorly does not show symptoms of depression if the lack of sleep is attributed to a newborn child, just as a person who frequently washes his or her hands only shows a symptom of obsessive-compulsive disorder, hand washing occurs in response to an excessive obsession with hygiene. From this point of view, it can be argued that diagnostic systems, such as the DSM-5, are not purely empirical or theoretically neutral as is often claimed. It is clear that at least as far as hegemonic diagnostic practice in psychopathology is concerned, common cause models are rejected [25, 26]. Such conceptual positioning may be better elaborated under the network approach, especially in cases where a certain event external to symptoms may activate relationships between symptoms of some disorder for a long time, even in the subsequent absence of such an external event [25, 26].

Another advantage of the network approach is the method by which comorbidities can be identified and classified. Ideally, symptoms should be sufficient and necessary conditions for identifying a disorder. However, in the general clinical context, this is rarely the case (even for some disorders or diseases that are clearly biological in origin). It is more common to state that symptoms nominated as "characteristic" of a psychiatric disorder are simply those more frequent in one group of individuals than in others [21, 25]. The traditional psychometric approach, by favoring symptoms that would be more "characteristic" (i.e., occur together more frequently and thus would be more correlated), would not identify idiosyncrasies derived from an individual's specific symptoms. Consequently, some authors [27] suggest that diagnostic comorbidity could be a consequence of spurious associations and, for this reason, could be reduced by retaining distinctive symptoms, but eliminating nonspecific symptoms, in psychopathological assessments.

In the network approach, symptoms are assessed in relation to their "importance" for the stability of the symptom network as a totality [25, 28]. For example, centrality measures indicate the degree of interconnectedness of a symptom with the other

symptoms in the network. As there are different ways in which one symptom can connect to another, different centrality metrics can demonstrate different degrees of "importance" of the assessed symptom [29]. Thus, for the assessment of comorbidities, using an idiographic network analysis paradigm, it is possible to identify, for each individual and for groups of individuals, which symptoms appear most relevant in the set of networks and which expression of symptom interdependencies allows certain comorbidities to occur in some individuals [24, 30–32].

4. Analytical methods for network psychometry

The models used in network psychometry are derived from the graph theory of mathematics [33]. Graphs (also called networks) are mathematical objects in which nodes represent various elements (such as other mathematical objects, e.g., sets and variables, or even real objects, e.g., individuals and organizations) and edges represent relationships between nodes. In the statistical derivation of graph theory, known as probabilistic graph models [34], nodes are used to represent variables (in the case of psychopathology networks, the variables are usually the possible symptoms) and edges are used to represent the dependency relationships between the nodes. Dependency relationships usually involve correlations or partial correlations, but may also involve nonlinear dependency measures [35]. It is also common to use clustering methods to identify which variables are most strongly connected [36, 37].

Unlike social networks, in which nodes (people) and the relationships between them can be directly observed [38], psychological networks are based on probabilistic graphs [20, 39]. There are three main types of probabilistic graphs, which are given below [34]:

    i. nondirectional graphs (in which the relationships between variables are symmetric);

    ii. directional graphs (in which the relationships between variables are asymmetric); and.

    iii. chain graphs (sometimes also called mixed graphs in which there are both symmetric and asymmetric relationships).

In the study of psychological networks, the use of nondirectional graphs where edges represent partial correlations is the most common [21, 27, 40, 41]. This preference is mainly because nondirectional graphs allow us to derive hypotheses about causal relationships without the need to make explicit assumptions about which variables are cause and which are effect.

Among the nondirectional graph models used in the study of psychological networks, three of them have received special attention in the literature, which are as follows:

    i. correlation networks;

    ii. partial correlation networks; and.

    iii. directional graphs (directed acyclic graphs, DAG).

The first of these is the correlation network [34]. This type of model uses correlations as measures of dependence between variables and is used when one wants to know if there is a direct dependence between variables. These models have two main limitations. First, these models do not allow inferences to be made about causal relationships since, according to the theory of causal calculus [9], these can only be derived from conditional dependencies. The second limitation of this type of model is that, being based on correlations, the dependencies are not affected by the other variables in the network.

The second type of network model, probably the most widely used, is the partial correlations network model (also known as concentration networks) [42]. This type of model uses partial correlations to measure the strength of the linear relationship between variables. There are two main ways to estimate partial correlation network models [41]. The first is to simply calculate the partial correlations of all the variables in the model and remove the edges of the correlations that are not significant. However, this type of practice is sensitive to false positives. For this reason, it has been more common to use regularized partial correlation networks [42]. which minimizes the probability of maintaining spurious relationships. The use of partial correlations is particularly interesting, as such measures can be interpreted as causal relationships between variables [9, 34]. However, care must be taken not to interpret them as mutualistic causal relationships (which is the case in some important references in the literature) [42]. In fact, partial correlation network models can also be referred to as "visual graphs," which are the non-directional representations of DAGs [43]. This means that causal directions, in some cases, can be determined.

The third type of model is known as DAG [9]. Directional graphs of the DAG type imply all the expected causal relationships between the collected variables. DAGs allow one to appreciate the existence of cycles in the network. For example, it is possible that a variable A causes a variable B, which in turn causes a variable C, and that this in turn causes variable A. This condition is used to avoid breaking the basic assumptions of causality, such as localism and realism of natural phenomena, as well as the transitivity of causal relationships. However, working with longitudinal data, it is possible to identify cycles that are valid (i.e., when the transitivity of causal relationships at the same moment in time is respected) [44]. For example, inattention at a time point t = 1 can be the cause of inattention at the same time point t = 1. If this relationship is true, it is only causally valid to say that inattention is also the cause of inattention if inattention at time point t = 1 is the cause of inattention at time point t ≥ 2. DAGs have not been widely used in psychology given that they require explicit assumptions about which relationships are causal or not; however, few causal theories in psychology or psychopathology have the robustness to be used in this way [30, 42].

4.1 The use of network models in the context of psychopathology

It is important to emphasize that the use of network models not only allows us to address the complexity of the relationships between variables but is in fact a different approach to thinking about theories in psychopathology. Network analyses have been fundamental for researchers to work with more diverse data sources (e.g., genetic, neurological, physiological, behavioral, and other data) and to seek more comprehensive ways of theorizing. In this context, network analyses have been complemented by what is known as conjoint modeling [45]. Joint modeling is a statistical approach similar to structural equation modeling, but which allows the use of any alternative model as a measurement model (i.e., "for example, see [46]"). These models are

used, for example, to develop psychological or psychopathological models sensitive to neurophysiological limitations.

The proposal of the mental health-related symptom network model has promoted the application of different types of variables from different levels of psychobiological development to explore new systemic theories that may include cognitive, biological, and social aspects [47–49], as well as risk and protective factors for mental health [50]. This explanation is of great importance in the current context, for example, a network review study reported the first 18 months since the pandemic, symptomatological variables of fear, distress, and stress were used to a greater extent by COVID-19 [28].

These symptoms allow us to understand the etiological mechanisms of the psychological impact of a stressful event, such as the current pandemic. Protective factors, such as resilience or psychological well-being, and psychosocial measures, such as alcohol and drug abuse, were also included [28]. The studies reviewed by Ramos-Vera et al. [28] refer to the use of different clinical variables related to COVID-19, such as preventive behaviors; emergency personnel communication measures, atypical reactions to pandemic stress, anti-mask attitudes; components of COVID-19 dreams and nightmares, insomnia and work fatigue. One of the studies considered variables consequent to the pandemic, such as perceived present and future infection risk, loss of income, and financial worry [51], while another research conducted in Italy by Invito et al. [52] took into account psychological distress and viral contagion beliefs, and added epidemiological characteristics, such as COVID-19 diagnosis, sex status and number of COVID-19 infected and deaths according to the participant's region. Symptom interaction network theory research has spurred several papers seeking to explore the interconnections of the most recurrent physical and psychological symptoms in certain chronic conditions, such as cancer [53], HIV [54], schizophrenia [55], stroke [56], chronic pain [57] chronic bowel disease [58], multiple sclerosis [59], arterial hypertension [7], obesity [60], and COVID-19 [61].

4.2 The use of psychological network models in the context of neuropsychology

Network neuropsychology can be useful in understanding cognitive adaptation and maladaptation in neurological disorders. Since cognitive functions are not isolated from each other, despite being framed in different domains they can be represented as a cognitive network system, additionally, the successful performance of most neuropsychological tasks is based on the interdependence of several cognitive domains [62, 63]. One of the properties of this network variant is the representation of several networks where measurable differences in neuropsychological profiles between distinct groups can be identified. Two previous studies report that differences are identified in the way neuropsychological tasks are associated in the network between those with neurological diagnoses (cognitive impairment and Alzheimer's disease) relative to control groups [64, 65]. Specifically, regroupings of memory, language and semantic variables and executive or attention, working memory and processing speed variables are evidenced in the network system belonging to participants with Alzheimer's disease relative to healthy control models. This feature allows for new explorations of the cognitive network reorganization that may occur throughout the stages of aging, as referred to in the cognitive dedifferentiation hypothesis. It is very likely that aging has an impact on network composition and there is a need to identify topological deviations that may be indicative of age-related neuropathology [66].

Cognitive impairment can be considered as a transdiagnostic dimension of psychopathology [67, 68], therefore, it is possible to consider the study and use of psychopathological symptoms and cognitive performance in network models. An Italian research in patients with a psychiatric diagnosis of schizophrenia included in the network system psychopathological symptoms of disorganization and avolition, positive and negative symptoms related to schizophrenia, in addition to the expressive deficit, akathisia, dystonia, parkinsonism and dyskinesia, and cognitive performance according to six domains: thought processing, attention/vigilance, working memory, verbal learning, visual learning, reasoning and problem solving, and social cognition, [69]. This work found a greater positive relationship of cognitive performance with social cognition and a negative with parkinsonism (this factor was more connected with psychopathological and cognitive measures) and disorganization.

Networks in neuropsychology may also aim to gain insight into changing associative patterns between cognitive constructs following brain damage [70]. For example, research by Iverson et al. [71] estimated the network structure of physical, cognitive, and emotional symptoms associated with attention deficit hyperactivity disorder following concussion. A total of 3074 student athletes were included who reported increased levels of difficulty concentrating and emotional symptoms. Most of the relationships between symptoms were positive, and the most influential symptoms in the network were dizziness and intensity of emotional symptoms. The relationships with the highest magnitude were emotional intensity and psychological distress, as well as forgetfulness and visual problems. There was a structural difference in the network according to sex, with a higher frequency of symptoms in women [71]. These findings demonstrate that similar studies should be encouraged in clinical participants given that from a systems neuroscience perspective, damage to one area of the brain is considered to affect the functioning of other areas adaptively (e.g., compensation, neuronal reserve, degeneration) or maladaptively (e.g., diaschisis, transneuronal degeneration, and dedifferentiation) [72].

Researchers can make supplementary assumptions, such as specifying hierarchical and/or directional relationships between cognitive functions or support other neuropsychological approaches, such as cognitive neuropsychology to create network models. Network theory can also be used to model relationships between tasks, which offers the advantage of conditioning (multivariate control) on all variables in the model, without making any assumptions about the underlying relationships between cognitive functions. In the following, certain studies are detailed with the aim of illustrating findings that would probably not be found using traditional methods of psychometric analysis.

One of the most important contributions to the field of neuropsychology, in the context of network analysis, is the study by Tosi et al. [65]. In this study, differences in networks of neuropsychological variables were evaluated in patients with and without clinical conditions, composed of 165 healthy elderly, 191 patients with Alzheimer's disease (AD), and 129 patients with vascular encephalopathy (VE). These networks included neuropsychological measures in the domains of memory, language, executive functions, attention, and abstract reasoning, in addition to the covariates of age, sex, and years of schooling. Patients with VS obtained better results (greater connection of cognitive abilities) than those with AD even when controlling for covariates, also, two groups of variables focused on memory and frontal-executive functions were identified in these networks.

Another study evaluated the network configuration of neurocognitive measures in adults using four serial assessments approximately one year apart [73].

The sample consisted of two groups of 432 elderly who obtained, at baseline, a cognitive assessment at normal levels. However, after subsequent assessment steps, the first group retained the same cognitive diagnosis, whereas participants in the second group developed mild cognitive impairment or AD dementia. Differences in network structures (connectivity and centrality) were identified between the groups even before AD was diagnosed, with such differences increasing over time.

Ferguson [64] estimated three network structures in adults according to his neuropsychological assessment:

    i. cognitive normality;

    ii. amnestic mild cognitive impairment; and

    iii. AD (Alzheimer's disease).

In these structures, the networks were composed of cognitive variables linked to the domains of attention, working memory, episodic memory, language, fluency, visuospatial ability, and sociodemographic variables (such as age and education). The centrality of episodic memory in the network structure of people with cognitive impairment was higher, whereas processing speed and fluency were more central in the network of people with AD. In addition, two groups of variables were identified in the three networks, the first focused on semantic memory and language, while the second was composed of attention, processing speed, and working memory.

The research by Foret et al. [74] composed of adults with no neurological or psychiatric history aimed to compare two simultaneous networks in men and women that included biomarkers of cognitive impairment risk, components of the metabolic syndrome (obesity, hypertension, dyslipidemia, and hyperglycemia), neuroimaging-based brain age minus chronological age, ratio of white matter hyperintensities to total brain volume, resting-state brain connectivity based on default mode network seed analysis, and ratios of N-acetyl aspartate, glutamate, and myo-inositol to creatine, which were measured by proton magnetic resonance spectroscopy [74]. Differences were found in the connectivity of both networks where women report lower relationships between cardiometabolic risk variables and brain functioning, furthermore, the most influential measures are shown to be apolipoprotein status and waist circumference.

An investigation in Scottish patients with multiple sclerosis evaluated two networks with a difference of a 12-month follow-up period where psychological aspects more prevalent in this clinical condition, such as fatigue, sleep quality, anxiety, and depression, were evaluated [59]. Measures of physical disability, upper extremity dexterity, gait speed, body mass index, and cognitive performance based on the domains of information processing speed, auditory information processing, working memory, and attention span, as well as neuroanatomical variables related to intracranial volume in the natural space were also considered. The results report that fatigue was related to most variables with the exception of brain measures and depression was the most central element in both networks, respectively [59].

The most recent study by Rotstein et al. [75] evaluated psychometric networks of cognitive impairment in more than 1000 American patients with Alzheimer's disease assessed by the cognitive subscale of the Alzheimer's Disease Assessment Scale composed of seven domains: temporal and spatial orientation, attention, learning, memory, abstract thinking, verbal fluency, and naming. Several network systems

were represented between two groups that received treatment with donepezil and placebo at 24 weeks of follow-up, the results showed a statistically significant difference in the global strength of the network integrated by the patients who received medication, evidencing a lower cognitive deterioration in this group.

Also, other network variants that assess dimensionality have been implemented, such as Exploratory Graphical Analysis (EGA; [36]). EGA employs a network algorithm to detect Walktrap communities [76]. Therefore, EGA estimates the dimensionality of multivariate data by combining network analysis with a community detection algorithm, where a community represents a latent variable reported in a factor technique [36, 37]. Consequently, it is a method to detect dimensions in networks, and additionally reports factor loadings of network variables with their respective communities. In addition to using the EBICglasso estimator for regularized partial correlation networks, this variant of the psychometric network can also group the variables in a graphical model composed of a zero-order correlation matrix using the Maximally Filtered Triangulated Graph Method (TMFG; [77]). This method allows regularizing the relationships and selecting the most parsimonious network structure.

The use of the Bootstrap Exploratory Bootstrap Graphical Analysis (bootEGA) module is recommended, to evaluate the structural consistency of an estimated dimensional structure. Structural consistency is understood as the extent to which a dimension is interrelated (internal consistency) and homogeneous in the presence of other related dimensions [78, 79], such a measure provides an alternative but complementary approach to internal consistency measures in the factor analytic framework. In bootEGA estimation, two metrics are required for structural consistency. The first consists of investigating the solidity of the structure of the dimensionality and the second in the robustness of the location of each element within these dimensions. Three steps have been described for this purpose: (1) estimating a network using EGA, (2) then generating new replicate data from a multivariate normal distribution (with the same number of cases as the original distribution), (3) then applying EGA to the replicate data sets, continuing interactively until the desired number of samples (e.g., 500 participants; [80]) is achieved. Therefore, there are two reasons for employing the parametric bootstrap: resampling smaller samples increase the influences that outlier cases may have on the estimated sampling distribution, and (2) its higher accuracy is the detection of the correct dimensionality structure in the simulated populations [80].

Finally, the need for more studies with multilayer networks (network of networks) is highlighted since they allow better statistical accuracy of the joint use of neurophysiological and psychological data [81, 82]. This may be important in the current pandemic context, as COVID-19 can affect the central nervous system and cause neuropsychiatric disorders [83]. Naturally, this clinical condition has a complex etiology, composed of associative networks of inflammatory biomarkers that can be represented in a network system [84], together with other physical and mental health risk phenotypes [84, 85] and neuroanatomical measures [59, 81, 86]. In this sense, network assessment of variables at different psychobiological levels related to COVID-19 can add to findings reported widely in the literature [87–93].

5. Conclusions

The main objective of this research was to critically analyze the main distinctions between Latent Variables Theory and Network Psychometrics in the

context of psychopathologies. To achieve this goal, relevant implications of the common cause model have been presented which, in contrast to the discussion on Network Psychometrics, do not seem to correctly represent some of the empirical evidence. It is important to note that research in using network analysis is still being refined and specific theories are still scarce [94]. However, the observed results have been promising and consolidation of the field will show how important this new line of research can be [24, 41]. On the other hand, although the network approach is not, after all, the most suitable for the study of psychopathology and psychological constructs in general, the exemplified applications, especially those involving variables external to psychological symptoms, are important for the promotion of new hypotheses in the neuropsychological field [95–97], in the face of the inclusion of new network centrality metrics that allow the identification of different structural features following the systemic grouping of transdiagnostic variables in network models [98–100], including longitudinal data to assess how the network is organized over time [101].

In this perspective, network analysis has the potential to change the field of psychopathology, and even neuropsychology, given its tools that allow combining evidence from different contexts and backgrounds in a way that was not previously used, this is essential in the complex assessment of psychosocial and public health risk factors (e.g., addictions and suicidal behavior, see the study by Anderson et al.; Penzel et al.; Hirota et al.; Sanchez-Garcia et al.; and Calati et al. [101–105]). Therefore, future studies that combine data and evidence from different levels of analysis and from different sources may lead to a better understanding of transdiagnostic factors [106–109], cognitive deficits [67], and especially of the integration of neural, behavioral, and symptomatic systems [110–113].

Finally, it is recognized that the understanding and study of psychological variables is a complex task, involving a multitude of variables at multiple levels of analysis (biological, cognitive, and social), which are related to each other in a complex way [114]. However, network analysis may lead to a change in the current epistemological and methodological approach to psychological phenomena so that this complexity can be effectively assessed [115, 116]. Network analysis is unlikely to be one of the best innovations in the field of studying psychological phenomena and problems remain to be solved [28, 96, 117–122], but we believe that the presented discussion highlights positive expectations for the future.

Conflict of interest

The authors have no conflict of interest.

Author details

Cristian Ramos-Vera[1,2]*, Víthor Rosa Franco[3], José Vallejos Saldarriaga[1]
and Antonio Serpa Barrientos[2,4]

1 Universidad Cesar Vallejo, Research Area, Lima, Peru

2 Sociedad Peruana de Psicometría, Lima, Peru

3 Universidade São Francisco, Campinas, Brazil

4 Universidad Nacional Mayor de San Marcos, Lima, Peru

*Address all correspondence to: cristony_777@hotmail.com

IntechOpen

References

[1] Borsboom D, Mellenbergh GJ, Van Heerden J. The theoretical status of latent variables. Psychological Review. 2003;**110**(2):203-2019. DOI: 10.1037/0033-295X.110.2.203

[2] Coulacoglou C, Saklofske DH. Psychometrics and Psychological Assessment: Principles and Applications. London: Academic Press; 2017

[3] McDonald RP. Test Theory: A unified Treatment. New York: Psychology Press; 1999

[4] Demjaha A, Morgan K, Morgan C, Landau S, Dean K, Reichenberg A, et al. Combining dimensional and categorical representation of psychosis: The way forward for DSM-V and ICD-11? Psychological Medicine. 2009;**39**(12):1943-1955. DOI: 10.1017/S0033291709990651

[5] Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial paper. Behavior Research Methods. 2017;**50**(1):195-212. DOI: 10.3758/s13428-017-0862-1

[6] Blanchard MA, Heeren A. Ongoing and future challenges of the network approach to psychopathology: From theoretical conjectures to clinical translations. In: Asmundson G, Noel M, editors. Comprehensive Clinical Psychology. 2nd ed. Amsterdam: Elsevier; 2022. Available from: https://dial.uclouvain.be/pr/boreal/object/boreal%3A237881/datastream/PDF_01/view

[7] Ramos-Vera C, Baños-Chaparro J, Ogundokun R. Network structure of depressive symptoms in Peruvian adults with arterial hypertension. F1000Research. 2022;**10**(19):1-21. DOI: 10.12688/f1000research.27422.3

[8] Robinaugh DJ, Hoekstra RH, Toner ER, Borsboom D. The network approach to psychopathology: A review of the literature 2008-2018 and an agenda for future research. Psychological Medicine. 2020;**50**(3):353-366. DOI: 10.1017/S0033291719003404

[9] Pearl J. Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press; 2009

[10] Mellenbergh GJ. Generalized linear item response theory. Psychological Bulletin. 1994;**115**:300-307. DOI: 10.1037/0033-2909.115.2.300

[11] Mellenbergh GJ. Measurement precision in test score and item response models. Psychological Methods. 1996;**1**(3):293-299. DOI: 10.1037/1082-989X.1.3.293

[12] Diaconis P, Freedman D. Finite exchangeable sequences. The Annals of Probability. 1980;**8**(4):745-764. DOI: 10.1214/aop/1176994663

[13] Furr M. Scale Construction and Psychometrics for Social and Personality Psychology. California: Sage; 2011

[14] Michell J. Is psychometrics pathological science? Measurement. 2008;**6**(1-2):7-24. DOI: 10.1080/15366360802035489

[15] Trendler G. Measurement theory, psychology and the revolution that cannot happen. Theory & Psychology. 2009;**19**(5):579-599. DOI: 10.1177/0959354309341926

[16] Fried EI, van Borkulo CD, Cramer AO, Boschloo L, Schoevers RA, Borsboom D. Mental disorders as networks of problems: A review of

recent insights. Social Psychiatry and Psychiatric Epidemiology. 2017;**52**(1): 1-10. DOI: 10.1007/s00127-016-1319-z

[17] Krueger RF. The structure of common mental disorders. Archives of General Psychiatry. 1999;**56**(10):921-926. DOI: 10.1001/archpsyc.56.10.921

[18] Insel TR, Cuthbert B, N.: Medicine. Brain disorders? Precisely. Science. 2015;**348**(6234):499-500. DOI: 10.1126/ science.aab2358

[19] Rose N. Neuroscience and the future for mental health? Epidemiology and Psychiatric Sciences. 2016;**25**(2):95-100. DOI: 10.1017/S2045796015000621

[20] Silva M, Loureiro A, Cardoso G. Social determinants of mental health: A review of the evidence. The European Journal of Psychiatry. 2016;**30**(4):259-292. Available from: https://scielo.isciii. es/scielo.php?pid=S02136163201600040 0004&script=sci_arttext&tlng=en

[21] Borsboom D, Cramer AOJ. Network analysis: An integrative approach to the structure of psychopathology. Annual Review of Clinical Psychology. 2013;**9**:91-121. DOI: 10.1146/annurev-clinpsy-050212-185608

[22] American psychiatric association diagnostic and statistical manual of mental disorders (DSM–5). American Psychiatric Association; 2014. Available from: https://www.eafit.edu.co/ninos/ reddelaspreguntas/Documents/dsm-v-guia-consulta-manual-diagnostico-estadistico-trastornos-mentales.pdf

[23] Ramos Vera C. Las redes de relación estadística en la investigación de nutrición. Nutrición Hospitalaria. 2021;**38**(3):671-672. DOI: 10.20960/ nh.03522

[24] Cramer AOJ, Waldorp LJ, van der Maas HLJ, Borsboom D. Comorbidity: A network perspective. Behavioral and Brain Sciences. 2010;**33**:137-150. DOI: 10.1017/S0140525X09991567

[25] Borsboom D. A network theory of mental disorders. World Psychiatry. 2017;**16**(1):5-13. DOI: 10.1002/wps.20375

[26] Spitzer RL, First MB, Wakefield JC. Saving PTSD from itself in DSM-V. Journal of Anxiety Disorders. 2007;**21**(2):233-241. DOI: 10.1016/j. janxdis.2006.09.006

[27] Borsboom D, Deserno MK, Rhemtulla M, Epskamp S, Fried EI, McNally RJ, et al. Network analysis of multivariate data in psychological science. Nature Reviews Methods Primers. 2021;**58**(1). DOI: 10.1038/ s43586-021-00055-w1-18

[28] Ramos-Vera C, García-Ampudia L, Serpa-Barrientos A. Una alternativa de análisis de redes en la exploración de los estados de salud mental, condiciones crónicas y COVID-19. Iatreia. In Press 2022:1-22. Available from: https:// revistas.udea.edu.co/index.php/iatreia/ article/view/347261

[29] Castro D, Ferreira F, de Castro I, Rodrigues AR, Correia M, Ribeiro J, et al. The differential role of 00central and bridge symptoms in deactivating psychopathological networks. Frontiers in Psychology. 2019;**10**:e2448. DOI: 10.3389/fpsyg.2019.02448

[30] Bringmann LF, Elmer T, Epskamp S, Krause RW, Schoch D, Wichers M, et al. What do centrality measures measure in psychological networks? Journal of Abnormal Psychology. 2019;**128**(8):892-903. Available from: https://psycnet.apa. org/doi/10.1037/abn0000446

[31] Bringmann LF, Albers C, Bockting C, Borsboom D, Ceulemans E, Cramer A, et al. Psychopathological networks:

Theory, methods and practice. Behaviour Research and Therapy. 2022;**149**:e104011. DOI: 10.1016/j.brat.2021.104011

[32] Fisher AJ, Reeves JW, Lawyer G, Medaglia JD, Rubel JA. Exploring the idiographic dynamics of mood and anxiety via network analysis. Journal of Abnormal Psychology. 2017;**126**(8):1044-1056. DOI: 10.1037/abn0000311

[33] West DB. Introduction to Graph Theory. New Jersey: Prentice-Hall; 2001

[34] Lauritzen SL. Graphical Models. Oxford: Clarendon Press; 1996

[35] Isvoranu AM, Epskamp S, Waldorp L, Borsboom D. Network Psychometrics with R: A Guide for Behavioral and Social Scientists. New York: Routledge; 2022

[36] Golino HF, Epskamp S. Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. PLoS One. 2017;**12**(6):e0174035. DOI: 10.1371/journal.pone.0174035

[37] Golino H, Shi D, Christensen AP, Garrido LE, Nieto MD, Sadana R, et al. Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. Psychological Methods. 2020;**25**(3):292-320. DOI: 10.1037/met0000255

[38] Scott J, Carrington PJ. The SAGE Handbook of Social Network Analysis. Los Angeles: SAGE; 2011

[39] Ramos-Vera C. Las redes de relación estadística en la investigación psiquiátrica: El caso del delirio en el contexto de COVID-19. Revista Colombiana de Psiquiatría (English Ed.). 2021;**50**(3):158-159 DOI:10.1016/j.rcpeng.2021.02.001

[40] Fried EI, Von Stockert S, Haslbeck JMB, Lamers F, Schoevers RA, Penninx BWJH. Using network analysis to examine links between individual depressive symptoms, inflammatory markers, and covariates. Psychological Medicine. 2020;**50**(16):2682-2690. DOI: 10.1017/S0033291719002770

[41] McNally RJ. Network analysis of psychopathology: Controversies and challenges. Annual Review of Clinical Psychology. 2021;**17**:31-53. DOI: 10.1146/annurev-clinpsy-081219-092850

[42] Epskamp S, Fried EI. A tutorial on regularized partial correlation networks. Psychological Methods. 2018;**23**(4):617-634. DOI: 10.1037/met0000167

[43] Andersson SA, Madigan D, Perlman MD. On the Markov equivalence of chain graphs, undirected graphs, and acyclic digraphs. Scandinavian Journal of Statistics. 1997;**24**(1):81-102. DOI: 10.1111/1467-9469.00050

[44] Epskamp S. Psychometric network models from time-series and panel data. Psychometrika. 2020;**85**(1):206-231. DOI: 10.1007/s11336-020-09697-3

[45] Gollini I, Murphy TB. Joint modeling of multiple network views. Journal of Computational and Graphical Statistics. 2016;**25**(1):246-265. DOI: 10.1080/10618600.2014.978006

[46] Turner BM, Forstmann BU, Steyvers M. Joint Models of Neural and Behavioral Data. Switzerland: Springer International Publishing; 2019

[47] Kappelmann N, Czamara D, Rost N, Moser S, Schmoll V, Trastulla L. CHARGE inflammation working group: Polygenic risk for immuno-metabolic markers and specific depressive symptoms: A multi-sample network analysis study. Brain, Behavior, and

Immunity. 2021;**95**:256-268.
DOI: 10.1016/j.bbi.2021.03.024

[48] Moriarity DP, van Borkulo C,
Alloy LB. Inflammatory phenotype
of depression symptom structure: A
network perspective. Brain, Behavior,
and Immunity. 2021;**93**:35-42.
DOI: 10.1016/j.bbi.2020.12.005

[49] Saari T, Smith EE, Ismail Z. Network
analysis of impulse dyscontrol in mild
cognitive impairment and subjective
cognitive decline. International
Psychogeriatrics. 2021;**1-10**.
DOI: 10.1017/s1041610220004123

[50] Lunansky G, Van Borkulo CD,
Haslbeck J, Van der Linden MA,
Garay CJ, Etchevers MJ, et al. The
mental health ecosystem: Extending
symptom networks with risk and
protective factors. Frontiers in
Psychiatry. 2021;**12**:e301. DOI: 10.3389/
fpsyt.2021.640658

[51] Zavlis O, Butter S, Bennett K,
Hartman TK, Hyland P, Mason L,
et al. How does the COVID-19 pandemic
impact on population mental health? A
network analysis of COVID influences on
depression, anxiety and traumatic stress
in the UK population. Psychological
Medicine. 2021;**1-9**. DOI: 10.1017/
S0033291721000635

[52] Invitto S, Romano D, Garbarini F,
Bruno V, Urgesi C, Curcio G, et al. Major
stress-related symptoms during the
lockdown: A study by the Italian Society
of Psychophysiology and Cognitive
Neuroscience. Frontiers in Public
Health. 2021;**9**:e250. DOI: 10.3389/
fpubh.2021.636089

[53] Papachristou N, Barnaghi P,
Cooper B, Kober KM, Maguire R,
Paul SM, et al. Network analysis of the
multidimensional symptom experience
of oncology. Scientific Reports.

2019;**9**(1):1-11.
DOI: 10.1038/s41598-018-36973-1

[54] Zhu Z, Hu Y, Xing W, Guo M,
Zhao R, Han S, et al. Identifying
symptom clusters among people living
with HIV on antiretroviral therapy
in China: A network analysis. Journal
of Pain and Symptom Management.
2019;**57**(3):617-626. DOI: 10.1016/j.
jpainsymman.2018.11.011

[55] Abplanalp SJ, Green MF. Symptom
structure in schizophrenia: Implications
of latent variable modeling vs network
analysis. Schizophrenia Bulletin.
2022;**48**(3):538-543. DOI: 10.1093/
schbul/sbac020

[56] Ashaie SA, Hung J, Funkhouser CJ,
Shankman SA, Cherney LR. Depression
over time in persons with stroke: A
network analysis approach. Journal
of Affective Disorders Reports.
2021;**4**:e100131. DOI: 10.1016/j.
jadr.2021.100131

[57] Gómez Penedo JM,
Rubel JA, Blättler L, Schmidt SJ,
Stewart J, Egloff N. The complex
interplay of pain, depression, and
anxiety symptoms in patients with
chronic pain: A network approach.
The Clinical Journal of Pain.
2020;**36**(4):249-259. DOI: 10.1097/
AJP.0000000000000797

[58] Nemirovsky A, Ilan K, Lerner L,
Cohen-Lavi L, Schwartz D, Goren G,
et al. Brain-immune axis regulation
is responsive to cognitive behavioral
therapy and mindfulness intervention:
Observations from a randomized
controlled trial in patients with Crohn's
disease. Brain, Behavior, & Immunity-
Health. 2022;**19**. DOI: 10.1016/j.
bbih.2021.100407

[59] Chang YT, Kearns PK, Carson A,
Gillespie D, Meijboom R, Kampaite A,

et al. Data-driven analysis shows robust links between fatigue and depression in early multiple sclerosis. medRxiv. 2022. DOI: 10.1101/2022.01.13.22269128

[60] Ramos-Vera C. Serpa A, Vallejos-Saldarriaga J, Saintila J. Network analysis of depressive symptomatology in underweight and obese adults. Journal of Primary Care & Community Health. In Press. 2022. DOI: 10.1177/21501319221096917

[61] Spechbach H, Jacquerioz F, Prendki V, Kaiser L, Smit M, Calmy A, et al. Network analysis of outpatients to identify predictive symptoms and combinations of symptoms associated with positive/negative SARS-CoV-2 nasopharyngeal swabs. Frontiers in Medicine. 2021;**8**. DOI: 10.3389/fmed.2021.685124

[62] Hills TT, Kenett YN. Is the mind a network? Maps, vehicles, and skyhooks in cognitive network science. Topics in Cognitive Science. 2022;**14**(1):189-208. DOI: 10.1111/tops.12570

[63] Mareva S, CALM team, & Holmes, J. Network models of learning and cognition in typical and atypical learners. Journal of Applied Research in Memory and Cognition. Advance online publication; 2021. DOI: 10.1037/h0101870

[64] Ferguson C. A network psychometric approach to neurocognition in early Alzheimers's disease. Cortex. 2021;**137**:61-73. DOI: 10.1016/j.cortex.2021.01.002

[65] Tosi G, Borsani C, Castiglioni S, Daini R, Franceschi M, Romano D. Complexity in neuropsychological assessments of cognitive impairment: A network analysis approach. Cortex. 2020;**124**(3):85-96. DOI: 10.1016/j.cortex.2019.11.004

[66] Koen JD, Srokova S, Rugg MD. Age-related neural dedifferentiation and cognition. Current Opinion in Behavioral Sciences. 2020;**32**:7-14. DOI: 10.1016/j.cobeha.2020.01.006

[67] Abramovitch A, Short T, Schweiger A. The c factor: Cognitive dysfunction as a transdiagnostic dimension in psychopathology. Clinical Psychology Review. 2021;**81**:e102007. DOI: 10.1016/j.cpr.2021.102007

[68] Haywood D, Baughman F, Mullan B, Heslop K. What accounts for the factors of psychopathology? An investigation of the neurocognitive correlates of internalising, externalising, and the P-factor. PsyArXiv. 2022. Available from: https://psyarxiv.com/h97gw/download/?format=pdf

[69] Monteleone P, Cascino G, Monteleone AM, Rocca P, Rossi A, Bertolino A, et al. Prevalence of antipsychotic-induced extrapyramidal symptoms and their association with neurocognition and social cognition in outpatients with schizophrenia in the "real-life". Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2021;**109**:e110250. DOI: 10.1016/j.pnpbp.2021.110250

[70] Iverson GL. Network analysis and precision rehabilitation for the post-concussion syndrome. Frontiers in Neurology. 2019;**10**:e489. DOI: 10.3389/fneur.2019.00489

[71] Iverson GL, Jones PJ, Karr JE, Maxwell B, Zafonte R, Berkner PD, et al. Network structure of physical, cognitive, and emotional symptoms at preseason baseline in student athletes with attention-deficit/hyperactivity disorder. Archives of Clinical Neuropsychology. 2020;**35**(7):1109-1122. DOI: 10.1093/arclin/acaa030

[72] Fornito A, Zalesky A, Breakspear M. The connectomics of brain disorders. Nature Reviews Neuroscience. 2015;**16**(3):159-172. DOI: 10.1038/nrn3901

[73] Baily AR. Network analysis of cognitive symptom domains in alzheimer's disease (AD) [thesis doctoral]. The Vegas: University of Nevada; 2020. Available from: https://digitalscholarship.unlv.edu/thesesdissertations/3986

[74] Foret JT, Dekhtyar M, Cole JH, Gourley DD, Caillaud M, Tanaka H, et al. Network modeling sex differences in brain integrity and metabolic health. Frontiers in Aging Neuroscience. 2021;**13**:e329. DOI: 10.3389/fnagi.2021.691691

[75] Rotstein A, Levine SZ, Samara M, Yoshida K, Goldberg Y, Cipriani A, et al. Cognitive impairment networks in Alzheimer's disease: Analysis of three double-blind randomized, placebo-controlled, clinical trials of donepezil. European Neuropsychopharmacology. 2022;**57**:50-58. DOI: 10.1016/j.euroneuro.2022.01.001

[76] Pons P, Latapy M. Computing communities in large networks using random walks. Journal of Graph Algorithms and Applications. 2006;**10**:191-218. DOI: 10.7155/jgaa.00185

[77] Massara GP, Di Matteo T, Aste T. Network filtering for big data: Triangulated maximally filtered graph. Journal of Complex Networks. 2017;**5**(2):161-178. DOI: 10.1093/comnet/cnw015

[78] Christensen AP, Golino H, Silvia PJ. A psychometric network perspective on the validity and validation of personality trait questionnaires. European Journal of Personality. 2020;**34**(6):1095-1108. DOI: 10.1002/per.2265

[79] Golino H, Moulder R, Shi D, Christensen AP, Garrido LE, Nieto MD, et al. Entropy fit indices: New fit measures for assessing the structure and dimensionality of multiple latent variables. Multivariate Behavioral Research. 2021;**56**(6):874-902. DOI: 10.1080/00273171.2020.1779642

[80] Christensen AP, Golino H. Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A Monte Carlo simulation and tutorial. Psych. 2021;**3**(3):479-500. DOI: 10.3390/psych3030032

[81] Blanken TF, Bathelt J, Deserno MK, Voge L, Borsboom D, Douw L. Connecting brain and behavior in clinical neuroscience: A network approach. Neuroscience & Biobehavioral Reviews. 2021;**130**:81-90. DOI: 10.1016/j.neubiorev.2021.07.027

[82] Simpson-Kent IL, Fried EI, Akarca D, Mareva S, Bullmore ET, Team CALM, et al. Bridging brain and cognition: A multilayer network analysis of brain structural covariance and general intelligence in a developmental sample of struggling learners. Journal of. Intelligence. 2021;**9**(2):e32. DOI: 10.3390/jintelligence9020032

[83] Troyer EA, Kohn JN, Hong S. Are we facing a crashing wave of neuropsychiatric sequelae of COVID-19? Neuropsychiatric symptoms and potential immunologic mechanisms. Brain, Behavior, and Immunity. 2020;**87**:34-39. DOI: 10.1016/j.bbi.2020.04.027

[84] Cathomas F, Klaus F, Guetter K, Chung HK, Beharelle AR, Spiller TR, et al. Increased random exploration in schizophrenia is associated with

inflammation. Schizophrenia. 2021;**7**(1):1-9. DOI: 10.1038/s41537-020-00133-0

[85] Ramos-Vera C. Las redes de correlación en la investigación de la hipertensión arterial y riesgo vascular. Hipertensión y Riesgo Vascular. 2021;**38**(3):156-157. DOI: 10.1016/j.hipert.2021.02.001

[86] Hilland E, Landrø NI, Kraft B, Tamnes CK, Fried EI, Maglanoc LA, et al. Exploring the links between specific depression symptoms and brain structure: A network study. Psychiatry and Clinical Neurosciences. 2019;**74**(3):220-221. DOI: 10.1111/pcn.12969

[87] Chambon M, Dalege J, Elberse JE, van Harreveld F. A: Psychological network approach to attitudes and preventive behaviors during pandemics: A COVID-19 study in the United Kingdom and the Netherlands. Social Psychological and Personality Science. 2021:1-13. DOI: 10.1177/19485506211002420

[88] Gibson-Miller J, Zavlis O, Hartman TK, Bennett KM, Butter S, Levita L, et al. A network approach to understanding social distancing behaviour during the first UK lockdown of the COVID-19 pandemic. Psychology & Health. 2022:1-19. DOI: 10.1080/08870446.2022.2057497

[89] Houston J, Thorson E, Kim E, Mantrala MK. COVID-19 communication ecology: Visualizing communication resource connections during a public health emergency using network analysis. American Behavioral Scientist. 2021:1-21. DOI: 10.1177/0002764221992811

[90] Ramos-Vera C. The dynamic network relationships of obsession

and death from COVID-19 anxiety among Peruvian university students during the second quarantine. Revista Colombiana de Psiquiatria (English Ed.). 2021;**50**(3):160-163. DOI: 10.1016/j.rcpeng.2021.08.002

[91] Ryu S, Park IH, Kim M, Lee YR, Lee J, Kim H, et al. Network study of responses to unusualness and psychological stress during the COVID-19 outbreak in Korea. PLoS One. 2021;**16**(2):e0246894. DOI: 10.1371/journal.pone.0246894

[92] Taylor S, Asmundson GJ. Negative attitudes about facemasks during the COVID-19 pandemic: The dual importance of perceived ineffectiveness and psychological reactance. PLoS One. 2021;**16**(2):e0246317. DOI: 10.1371/journal.pone.0246317

[93] Taylor S, Paluszek MM, Rachor GS, McKay D, Asmundson GJ. Substance use and abuse, COVID-19-related distress, and disregard for social distancing: A network analysis. Addictive Behaviors. 2021;**114**:e106754. DOI: 10.1016/j.addbeh.2020.106754

[94] Van Der Maas HLJ, Dolan CV, Grasman RPPP, Wicherts JM, Huizenga HM, Raijmakers MEJ. A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. Psychological Review. 2006;**113**(4):842-861. DOI: 10.1037/0033-295X.113.4.842

[95] Korem N, Cohen LD, Rubinsten O. The link between math anxiety and performance does not depend on working memory: A network analysis study. Consciousness and Cognition. 2022;**100**:e103298. DOI: 10.1016/j.concog.2022.103298

[96] Ferguson CE. Network neuropsychology: The map and the

territory. Neuroscience & Biobehavioral Reviews. 2022;**132**:638-647. DOI: 10.1016/j.neubiorev.2021.11.024

[97] Burns GL, Preszler J, Ahnach A, Servera M, Becker SP. Multisource network and latent variable models of sluggish cognitive tempo, ADHD-Inattentive, and depressive symptoms with spanish children: Equivalent findings and recommendations. Research on Child and Adolescent Psychopathology. 2022:1-14. DOI: 10.1007/s10802-021-00890-1

[98] Castro D, Ferreira F, Ferreira TB. Modularity of the personality network. European Journal of Psychological Assessment. 2021;**36**(6):998-1008. DOI: 10.1027/1015-5759/a000613

[99] Ferreira F, Castro D, Ferreira TB. The modular structure of posttraumatic stress disorder in adolescents. Current Psychology. 2022:1-13. DOI: 10.1007/s12144-021-02538-1

[100] Jimeno N, Gomez-Pilar J, Poza J, Hornero R, Vogeley K, Meisenzahl E, et al. (Attenuated) hallucinations join basic symptoms in a transdiagnostic network cluster analysis. Schizophrenia Research. 2022;**243**:43-54. DOI: 10.1016/j.schres.2022.02.018

[101] Anderson AR, Kurz AS, Szabo YZ, McGuire AP, Frankfurt SB. Exploring the longitudinal clustering of lifestyle behaviors, social determinants of health, and depression. Journal of Health Psychology. Advance online publication; 2022:e13591053211072685. DOI: 10.1177/ 13591053211072685

[102] Penzel N, Antonucci LA, Betz LT, Sanfelici R, Weiske J, Pogarell O, et al. Association between age of cannabis initiation and gray matter covariance networks in recent onset psychosis. Neuropsychopharmacology.

2021;**46**(8):1484-1493. DOI: 10.1038/s41386-021-00977-9

[103] Hirota T, McElroy E, So R. Network analysis of internet addiction symptoms among a clinical sample of Japanese adolescents with autism spectrum disorder. Journal of Autism and Developmental Disorders. 2021;**51**(8):2764-2772. DOI: 10.1007/s10803-020-04714-x

[104] Sanchez-Garcia M, de la Rosa-Cáceres A, Díaz-Batanero C, Fernández-Calderón F, Lozano OM. Cocaine use disorder criteria in a clinical sample: An analysis using item response theory, factor and network analysis. The American Journal of Drug and Alcohol Abuse. 2022;**1-9**. DOI: 10.1080/00952990.2021.2012185

[105] Calati R, Romano D, Magliocca S, Madeddu F, Zeppegno P, Gramaglia C. The interpersonal-psychological theory of suicide and the role of psychological pain during the COVID-19 pandemic: A network analysis: Suicide and psychological pain. Journal of Affective Disorders. 2022;**302**:435-439. DOI: 10.1016/j.jad.2022.01.078

[106] Smith AR, Hunt RA, Grunewald W, Jeon ME, Stanley IH, Levinson CA, et al. Identifying central symptoms and bridge pathways between autism spectrum disorder traits and suicidality within an active duty sample. Archives of Suicide Research. 2021:1-16. DOI: 10.1080/13811118.2021.1993398

[107] Eadeh HM, Markon KE, Nigg JT, Nikolas MA. Evaluating the viability of neurocognition as a transdiagnostic construct using both latent variable models and network analysis. Research on Child and Adolescent Psychopathology. 2021:1-14. DOI: 10.1007/s10802-021-00770-8

[108] Chattrattrai T, Blanken TF, Lobbezoo F, Su N, Aarab G, Van Someren EJ. A network analysis of self-reported sleep bruxism in the Netherlands Sleep Registry: Its associations with insomnia and several demographic, psychological, and life-style factors. Sleep Medicine. 2022;**93**:63-70. DOI: 10.1016/j.sleep.2022.03.018

[109] Pappa E, Peters E, Bell V. Insight-related beliefs and controllability appraisals contribute little to hallucinated voices: A transdiagnostic network analysis study. European Archives of Psychiatry and Clinical Neuroscience. 2020:1-11 DOI:10.1007/s00406-020-01166-3

[110] Guineau M, Ikani N, Rinck M, Collard R, Van Eijndhoven P, Tendolkar I, et al. Anhedonia as a transdiagnostic symptom across psychological disorders: A network approach. Psychological Medicine. 2022:1-12. DOI: 10.1017/S0033291722000575

[111] Isvoranu AM, Abdin E, Chong SA, Vaingankar J, Borsboom D, Subramaniam M. Extended network analysis: From psychopathology to chronic illness. BMC Psychiatry. 2021;**21**(1):1-9. DOI: 10.1186/s12888-021-03128-y

[112] Letina S, Blanken TF, Deserno MK, Borsboom D. Expanding network analysis tools in psychological networks: Minimal spanning trees, participation coefficients, and motif analysis applied to a network of 26 psychological attributes. Complexity. 2019. DOI: 10.1155/2019/9424605

[113] Kraft B, Bø R, Heeren A, Ulset V, Stiles T, Landrø NI. Depression-related impairment in executive functioning is primarily associated with fatigue and anhedonia. PsyArXiv. 2022. DOI: 10.31234/osf.io/qh47y

[114] Michelini G, Palumbo IM, DeYoung CG, Latzman RD, Kotov R. Linking RDoC and HiTOP: A new interface for advancing psychiatric nosology and neuroscience. Clinical Psychology Review. 2021;**86**:e102025. DOI: 10.1016/j.cpr.2021.102025

[115] Brooks D, Hulst HE, de Bruin L, Glas G, Geurts JJ, Douw L. The multilayer network approach in the study of personality neuroscience. Brain Sciences. 2020;**10**(12):915. DOI: 10.3390/brainsci10120915

[116] Zainal NH, Newman MG. Elevated anxiety relates to future executive dysfunction: A cross-lagged panel network analysis of psychopathology and cognitive functioning components. PsyArXiv. 2021. DOI: 10.31234/osf.io/hrfqa

[117] Hoffart A, Johnson SU. Latent trait, latent-trait state, and a network approach to mental problems and their mechanisms of change. Clinical Psychological Science. 2020;**8**(4):595-613. DOI: 10.1177/2167702620901744

[118] Morvan Y, Fried EI, Chevance A. Network modeling in psychopathology: Hopes and challenges. L'Encéphale. 2020;**46**(1):1-2. DOI: 10.1016/j.encep.2020.01.001

[119] Borsboom D. Possible futures for network psychometrics. Psychometrika. 2022;**87**:253-265. DOI: 10.1007/s11336-022-09851-z

[120] Bringmann LF, Eronen MI. Don't blame the model: Reconsidering the network approach to psychopathology. Psychological Review. 2018;**125**(4):606-615. DOI: 10.1037/rev0000108

[121] Krendl AC, Betzel RF. Social cognitive network neuroscience. Social Cognitive and Affective Neuroscience. 2022:nsac020. DOI: 10.1093/scan/nsac020

[122] Xie S, McDonnell E, Wang Y. Conditional Gaussian graphical model for estimating personalized disease symptom networks. Statistics in Medicine. 2022;**41**(3):543-553. DOI: 10.1002/sim.9274

Chapter 3

Psychometric Analyses in the Transcultural Adaptation of Psychological Scales

Guillaume Gronier

Abstract

Measurement scales play an important role in the methodology of psychological research and practice. They make it possible to obtain scores linked to numerous individual characteristics (feeling of hope, perceived stress, experience, felt well-being, etc.) and thus to draw up a profile of respondents or to compare several situations with each other according to their psychological impact. Most of the research on the construction of these scales is Anglo-Saxon and, therefore, proposes scales in English. However, many non-English speaking countries feel the need to use these scales for their studies, which requires them to be translated into a target language. This proposed chapter describes the steps and psychometric analyses required to adapt an English scale in another language. Based in particular on the recommendations of the International Test Commission and the APA Standards of Practice for Testing, this chapter aims to guide researchers who wish to undertake the translation of a psychological scale. It also includes an analysis of the literature on the translation practices of some one hundred scales, translated and published recently in various scientific journals.

Keywords: translation, questionnaire, scale, psychometric analyses, transcultural adaptation

1. Introduction

Psychology has long mobilised the subjective assessment of individual characteristics using questionnaires or measurement scales. These self-administered scales, i.e., which subjects are invited to respond alone, capture the perception that subjects have of themselves. Without being exhaustive, this may, for example, concern their perceived well-being or ill-being, their perception of certain personality traits, their satisfaction with a product, or their way of apprehending a particular situation. These scales generally have a diagnostic purpose: they provide a score that, once interpreted, gives an evaluation of the subject's perception. While some scales, particularly in the health field, propose thresholds for interpreting their scores, most of them leave the researcher or practitioner free to interpret the meaning of the scores obtained.

The design of these scales is based on a very specific scientific approach, which generally follows the Churchill paradigm [1]. The methodological paradigm for scale construction defined by Churchill aims not only to reduce the common biases in

scale completion (halo bias, social desirability bias, contamination bias and response polarisation bias) but also to verify the internal validity of the scale. The approach is thus based on a succession of stages of item definition, data collection and psychometric analysis, which, as part of an iterative process, ultimately makes it possible to validate the scale that has been designed. Some psychology scales have been validated and used for many years. For example, the Depression Anxiety Stress Scales (DASS) [2] have been used for over 25 years to measure perceived stress and anxiety in clinical psychology.

Therefore, when research requiring the use of certain psychological scales is conducted in languages other than that of the original scale, it seems simpler and more reliable to translate these scales than to create new ones from scratch in the target language. Adapting a scale into new languages thus has the following advantages:

- to be able to compare the scores of the same scale submitted to different audiences of different languages;

- to draw on scales that have often been used in different research studies, thus allowing the advantages and limitations of the scale to be better delineated;

- to continue psychometric analyses of the original scale in new studies; and

- promote research on scale construction.

Like the creation of a new scale, the cross-cultural adaptation of a scale is based on a clearly defined process, of which there are two main steps: the translation of the scale into the target language and the analysis of the psychometric properties of the translation. From a psychometric point of view, the aim is to ensure that the translated version corresponds to the properties of the original version, with particular attention paid to factor correspondence.

This chapter aims to summarise the psychometric analyses necessary for the validation of cross-cultural adaptations of psychology scales. It is thus intended as an aid to researchers and practitioners who wish to adopt a scale into a new language.

2. General methodology for cross-cultural adaptation of psychology scales

Several methodological frameworks describe the steps necessary for cross-cultural adaptation and validation of scales [1–4]. These frameworks are regularly discussed and adapted to provide a more reliable methodology. One of the most common frameworks is the one proposed by the International Test Commission, called the ITC Guidelines for Translating and Adapting Tests (Second Edition) [5]. This guide provides a set of 18 recommendations for conducting and evaluating the adaptation (sometimes also referred to as 'localisation') or simultaneous development of psychological and educational tests for use with different populations.

The 18 recommendations are divided into six main themes: preconditioning, test development, confirmation, empirical analyses, administration, score scales and interpretation and documentation. **Figure 1** summarises the framework described by the ITC.

Among these steps, some require psychometric treatments for the validation of the scales during cross-cultural adaptation, in particular, step 5 'Score scales and

I. Pre-condition
1. Obtain the permission from the holder 2. Evaluate the relevance of the test for the target population 3. Minimize the influence of the cultural and linguistic influences

II. Test development
4. Ensure that the adaptation process takes into account the linguistic, psychological and cultural differences of the target populations. 5. Use appropriate translation designs and procedures to maximize the suitability of the test adaptation. 6. Provide evidence that the test instructions and item content have similar meaning for all intended populations. 7. Provide evidence that the item formats, rating scales, and other procedures are suitable for all intended populations. 8. Collect pilot data on the adapted test.

III. Confirmation
9. Select sample with characteristics that are relevant for the intended use of the test. 10. Provide relevant statistical evidence about the construct equivalence, method equivalence, and item equivalence. 11. Provide evidence supporting the norms, reliability and validity of the adapted version of the test. 12. Use an appropriate equating design and data analysis procedures.

IV. Administration
13. Prepare administration materials and instructions. 14. Specify testing conditions that should be followed closely.

V. Score scales and interpretation
15. Interpret any group score differences. 16. Only compare scores across populations when the level of invariance has been established.

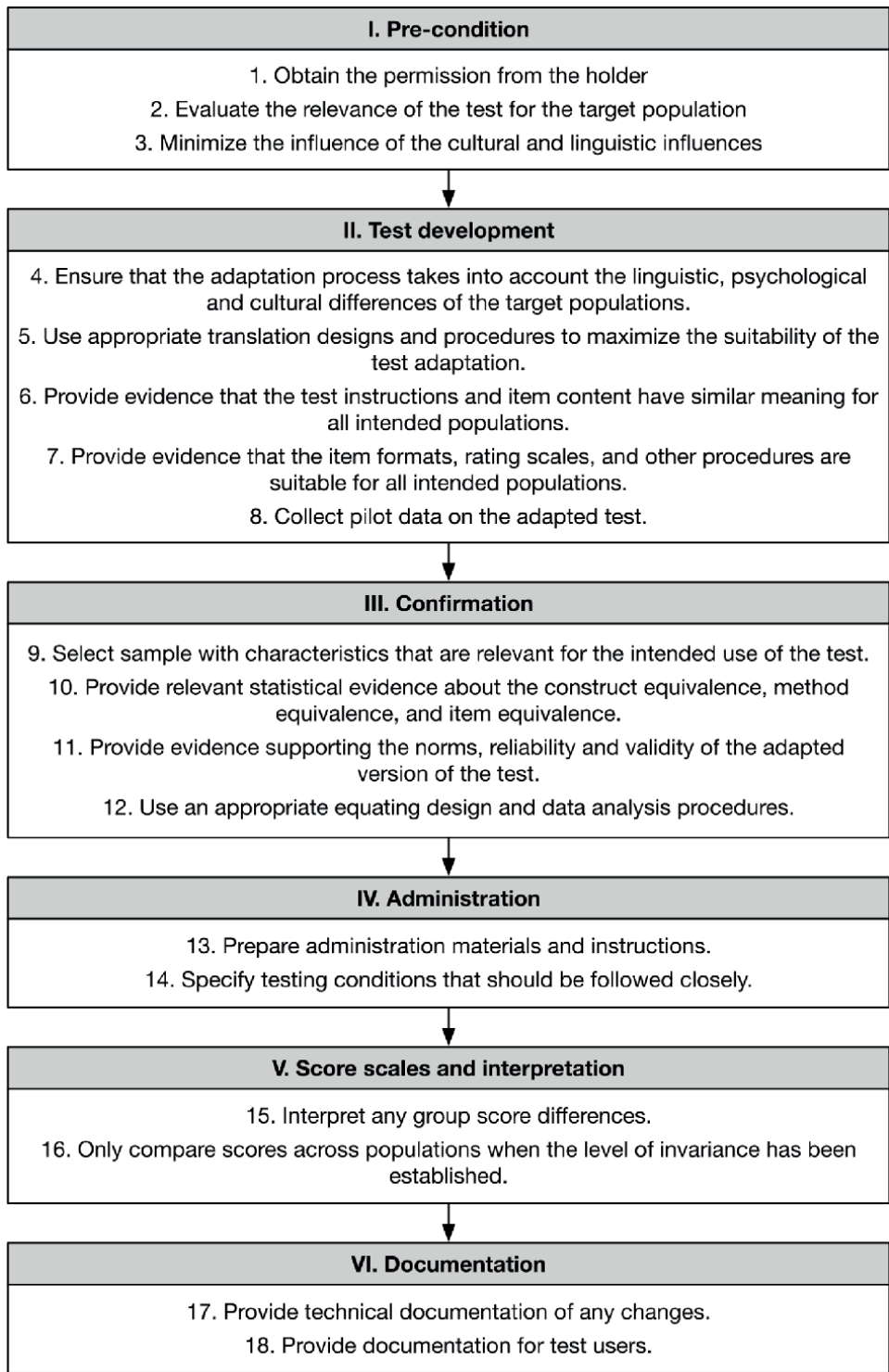| **VI. Documentation** |
| --- |
| 17. Provide technical documentation of any changes.
18. Provide documentation for test users. |

Figure 1.
Synthesis of the International Test Commission guidelines for translating and adapting tests.

interpretation'. Indeed, psychometric analyses are involved in the process of adapting items from the original language to a new language in order to ensure the quality of the translation. Failure to transfer the meaning of the original items can lead to a variation called scale error in the scale scores. As a result of this transfer of meaning, it is possible to create a structure that is different from the original scale structure. Therefore, in an adaptation study, it is necessary to ensure that the translation of the item is done correctly before starting the analysis. A consistent translation process is very important for the elimination of structural differences [6].

The most important and commonly applied psychometric analyses are presented in the following section.

3. Psychometric analyses of scale adaptations

3.1 Measuring internal consistency

According to the models of classical test theory [7], the total score (X) on a test is never fully representative of the true score (V), i.e., the exact quantity that is being measured. There is always an error (ε), so the total score is composed of the true score and the error score. Thus, we note [8]:

$$\text{Score}_{\text{Total}} = \text{Score}_{\text{True}} + \text{Score}_{\text{Error}} \tag{1}$$

The error is assumed to be random with an average of zero, so that it sometimes acts to increase the total score and sometimes to decrease it, but does not bias it in any systematic way. Since any scale has some degree of measurement error, it is never possible to determine the true score, which would be the average of all the scores a person would get if they took the test an infinite number of times [9].

The error is itself divided into two components: the random error, which is normally distributed and has a mean of 0, and the systematic error, which is asymmetrically distributed and has a mean that differs from 0. While the random error does not introduce systematic bias into the measurement, the systematic error, when it differs from 0, will cause the observed score to systematically overestimate or underestimate the true score. Thus, the true score (V) will be composed of the construct of interest (CI) and the systematic error (SE), plus the random error (RE):

$$\text{Score}_{\text{Total}} = \text{Score}_{\text{Construct of Interest}} + \text{Score}_{\text{SystematicError}} + \text{Score}_{\text{RandomError}} \tag{2}$$

Fidelity estimators are used to assess how close the observed score is to the true score.

3.1.1 Cronbach's alpha

One of the most widely used fidelity indices in the humanities and social sciences is most likely Cronbach's alpha [10]. According to Cronbach, internal consistency refers to the homogeneity of the items, i.e. how similar the test items are or, in other words, how well they measure the same dimension of a construct, i.e., its unidimensionality.

Cho and Kim [11] state that the articles by Cortina [12] and Schmitt [13] have done much to inform researchers on the use of alpha, highlighting its advantages and limitations. Other research is more radical and recommends the use of other measures of internal consistency [14, 15]. Indeed, several authors [11–13] have demonstrated that a high alpha value does not necessarily translate into homogeneity or unidimensionality of the items. Rather, alpha indicates how closely the items in a scale are related or correlated to each other.

Yet most studies of cross-cultural adaptation of scales in psychology still rely on the calculation of Cronbach's alpha as a measure of internal consistency or homogeneity; see for example [11–13]. This persistence of alpha in psychometric studies can be explained by the ubiquity of this measure since the 1950s, which allows comparisons to be made between scales. It is indeed common to rely on the alpha of the original scale to ensure the validity of a translation into a target language, by comparing the alpha of the two scales. Moreover, across research, alpha is used as a traditional benchmark for measuring internal consistency, although as we have pointed out this interpretation is biased. Sijtsma [15] finally points out that in practice it is often understood that SPSS statistical software does not offer any calculations other than homogeneity, which is of course wrong. Cho and Kim [11] conclude that alpha has become as popular as some marketing products, which are less effective than others but have a better reputation than others. They, therefore, advise authors, but also editors of scientific journals, to incorporate other indicators of internal consistency, in addition to or instead of alpha.

3.1.2 McDonald's omega

As an alternative to Cronbach's alpha, McDonald's omega [16] is the second indicator of internal consistency that is most often found in cross-cultural adaptation of scales in psychology. It is a fidelity coefficient that takes into account the strength of the association between items and a construct on the one hand, and the link between the items and the measurement error on the other. Thus, according to McDonald, the omega provides a more accurate estimate of the true reliability of the scale.

Several studies justify the use of McDonald's omega as an alternative reliability index to the alpha [14, 17, 18]. Also, some cross-cultural adaptation studies calculate the omega in addition to the alpha [19–21]. However, these studies are far from being the most representative, and none of them completely replace alpha with omega.

3.2 Factor analysis

3.2.1 Exploratory factor analysis

The main purpose of exploratory factor analysis (EFA) is to identify the underlying latent variables or factors of a measure by exploring the relationship among observed variables [22]. Roberson et al. [22] also report that, as an exploratory technique, EFA should not be used as a rigorous verification of the theoretical model; that is, in the case of cross-cultural scale adaptation, as a means of verifying the factorial adequacy of the translated scale with respect to the original scale. Finally, the authors summarise a set of good practices for conducting EFA, in terms of the statistical distribution, sample size, extraction and rotation to be applied and the matrices to be included in the publications. Comrey [23] points out in this respect that too little

information on the application of EFA is given by researchers, which makes it difficult to compare or replicate studies.

In general, EFA is used to extract latent factors from the newly translated scale. The results of this analysis are compared to the structure of the original scale to verify that the same factors are present, with a similar organisation of items within each factor. Many studies of cross-cultural scale adaptation in psychology use this process; see for example [19–21, 24].

3.2.2 Principal component analysis

Principal component analysis (PCA) is one of the most popular multivariate statistical technique in psychometric analysis in psychology. It is also likely to be the oldest multivariate technique, formalised in its current state by Hotelling [25]. According to Abdi and Williams [26], PCA analyses a data table representing observations described by several dependent variables, which are, in general, inter-correlated. Its goal is to extract the important information from the data table and to express this information as a set of new orthogonal variables called principal components. PCA also represents the pattern of similarity of the observations and the variables by displaying them as points on maps. Jolliffe [27] adds that PCA is often used to reduce the dimensionality of a data set, replacing the p variables which have been measured by a much smaller set of m components. In the case of measurement scales in psychology, p represents the items and m the factors, or dimensions, of that scale.

In cross-cultural adaptations of psychological scales, the PCA is applied instead of the EFA. The orthogonal Varimax rotation is the most common one [24], although other rotation methods are also used, but are generally not well documented [28].

Two criteria are frequently used to determine the number of factors to be extracted from the PCA. The first criterion is the widely used eigenvalue. The higher the initial eigenvalue, the more the factor explains a significant portion of the total variance. By convention, any factor with an initial eigenvalue greater than 1 is considered significant. The second criterion is Cattell's kink criterion, and it is a more stringent criterion for determining the number of factors. Here a graph displays all the points that represent the eigenvalues of the components. They are connected by a line. Only those factors that lie before the abrupt change in slope are retained. The points following this change, called the bend break, appear to form a straight horizontal line. A few publications offer eigenvalue graphs [29], but this is not common practice (**Figure 2**).

3.2.3 Confirmatory factor analysis

Confirmatory factor analysis (CFA) is a type of structural equation modelling that assesses the internal validity of an instrument or the relationships between several manifest and latent variables [30]. CFA is used to test the fit between an a priori defined theoretical model and empirically collected data. This means that the researcher must be able to specify how many factors are needed and which variables would load heavily or have near-zero loadings on each factor. Thus, on the basis of various fit indices, it is determined whether the postulated model fits the data well. When the model does not show a good enough fit, the indices exceed a threshold value, thus suggesting the rejection of the model tested.

The CFA technique is particularly well suited to cross-cultural studies. Watkins [31] states that CFA can be used to compare the equivalence of factor structures across
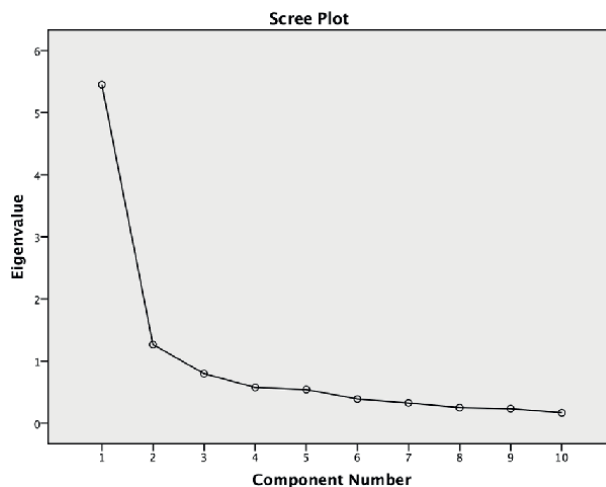
Figure 2.
Illustration of a scree plot [29].

cultures. This can be done either by collecting similar data in each culture or by collecting data in one culture and testing it against the factorial model established in the other culture. DiStefano and Hess [32] note the ubiquity of CFA in construct validation studies in psychology. Indeed, it is observed that most cross-cultural adaptations are validated or invalidated, using CFA; see for example [33–35].

The validation of the adapted scale, in comparison with the theoretical model of the original scale, necessarily relies on the consideration of fit indices, described in the next section.

3.2.4 Fit indices

The validation of the structural model calculated in the CFA is based on a set of fit indices whose thresholds indicate whether the model tested is valid or not. In other words, in the cross-cultural adaptation of psychological scales, the researcher applies the structural model of the original scale to his or her translation using a CFA in the first instance (the items are grouped into the corresponding dimensions), and then observes whether this model can be retained or should be rejected. If it is rejected, one or more other models are then applied until a satisfactory model is found, thus meeting the fit indices. This approach is applied, for example, to the French translation and validation of the Karitane Parenting Confidence Scale [20].

Several fit indices are usually calculated, of which we present here the most used among a larger set of fit indices [36], indicating the thresholds for model acceptance:

- Normed χ^2. A value below 2 suggests good model fit and below 3 acceptable model fit [37].

- Goodness of Fit Index (GFI) [38]. The value should be higher than 0.90 [39].

- Comparative fit index (CFI). The value should be higher than 0.80 [39]. CFI indicates a good model fit for values in the range between 0.95 and 1.00, whereas values in the range of 0.90 to 0.95 signify acceptable fit [37].

- Tucker-Lewis index (TLI). TLI higher than 0.95 is a commonly used cut-off criterion for the goodness of fit [40].

- Root mean square error of approximation (RMSEA). The value should be lower than 0.10 [39]. The values less than 0.05 indicate a good model fit and less than 0.08 an acceptable model fit [37].

- Standardised root means square residual (RMSR). Hu and Bentler [40] suggested the value should be lower than 0.08, while Sivo, Fan, Witta and Willse [41] suggested a cut-off value of 0.05.

- Akaike information criterion (AIC). The lower the value is, the better the fit [39].

- Normed Fit Index (NNFI). A correct model should have a value higher than 0.95 [41].

- Standardised root means square residual (SRMR). The values are expected to stay below 0.10 [37].

3.3 Convergent validity

Calculating convergent validity is an important step in measuring the validity of a scale adapted into another language. This is to ensure that the instrument really measures the construct(s) it is intended to measure and that it provides an adequate measure of the theoretical model on which it is based. A scale with good construct validity should therefore normally have high correlations with other scales measuring the same or similar constructs. Convergent correlations are therefore measured using, most often, Pearson's correlation coefficient.

In the context of cross-cultural adaptation of a scale, the translated scale is compared to one or more scales in the same language, which measure a similar psychological concept. For example, Yang, Zang, Ma et Bai [19] compared the Surgical Fear Questionnaire (SFQ) with the Hospital Anxiety and Depression Scale (HADS). The significance levels (p-value), associated with the correlation coefficients, indicate whether the links between the scales are satisfactory or not.

3.4 Time constancy

The time constancy is measured using the so-called test–retest technique. This technique consists of administering the same scale to the same subjects at two-time intervals. Generally, following the first measurement, the second measurement is carried out after 2 to 4 weeks. The scores at these two-time points are compared using a Pearson correlation coefficient, the Intraclass Correlation Coefficient (ICC) or the Kendall coefficient of concordance [42]. This technique ensures that the scale is stable over time and therefore reliable. A correlation with $0.30 < r < 0.50$ is considered as low, moderate with $0.50 < r < 0.70$ and strong with $r > 0.70$ [39].

In a cross-cultural adaptation of the Implicit Theory of Emotion Scale, Congard et al. [43] interviewed 35 subjects, 21 to 27 days apart. The Pearson correlation coefficient of 0.69 ($p < 0.001$) showed very good reliability of the scale over time.

It should be noted, however, that this technique is not relevant for certain scales in psychology, such as those measuring the perception of a product. Indeed, depending

on the use of the product, the same individual may have very different perceptions of the same product from one week to the next. This is particularly the case for scales measuring usability or user experience [29, 44, 45].

3.5 Socio-demographic analyses

The sensitivity of a cross-cultural adjustment is measured by comparisons between different modalities of the same variable. The difference in scores according to gender is often the first element of comparison. Depending on the variables and the number of modalities of the variables, researchers conventionally apply Student's t or ANOVA when there are more than two modalities.

In the adaptation of the Feelings at School (FAS) scale, Sanchez et al. [46] compared the scores between two different primary school levels (6- and 11-year-olds). The calculation of an ANOVA revealed the presence of a significant effect on the school level.

4. Conclusion

The aim of this chapter was to propose a methodological framework for psychometric analyses in the cross-cultural adaptation of psychological scales. Although the
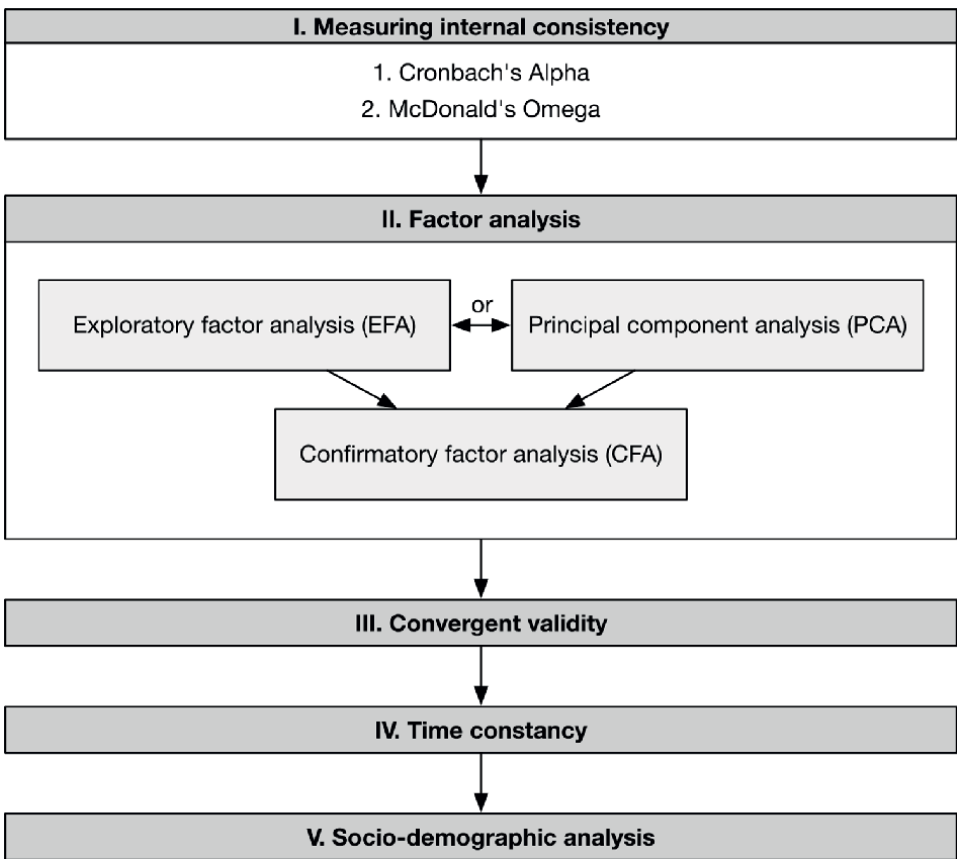
Figure 3.
Methodological framework for psychometric analysis in the transcultural adaptation of psychological scales.

choice of statistical validation tools may change from one study to another, depending on the requirements of the journal for publication or the psychometric skills of the researcher, it is possible to identify a guideline in a succession of steps that can serve as a guide to the cross-cultural adaptation of scales. This methodological line, which takes up the analyses described in this chapter, is described in **Figure 3**. It is imperfect and not exhaustive, but it will be a support that will be suitable for most of the validation of scale translations.

Author details

Guillaume Gronier
Luxembourg Institute of Science and Technology (LIST), Esch-sur-Alzette, Luxembourg

*Address all correspondence to: guillaume.gronier@list.lu

IntechOpen

References

[1] Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. Journal of Evaluation in Clinical Practice. 2011;**17**:268-274. DOI: 10.1111/j.1365-2753.2010.01434.x

[2] Gana K, Broc G, Boudouda NE, Calcagni N, Ben YS. The ITC guidelines for translating and adapting tests (second edition). Pratiques Psychologiques. 2021;**27**:175-200. DOI: 10.1016/j.prps.2020.06.005

[3] Gana K, Boudouda NE, Ben Youssef S, Calcagni N, Broc G. Transcultural adaptation of psychological tests and scales: A practical guide based on the ITC guidelines for translating and adapting tests and the standards for educational and psychological testing. Pratiques Psychologiques. 2021;**27**:223-240. DOI: 10.1016/j.prps.2021.02.001

[4] Sperber AD. Translation and validation of study instruments for cross-cultural research. Gastroenterology. 2004;**126**:124-128. DOI: 10.1053/j.gastro.2003.10.016

[5] International Test Commission. ITC Guidelines for Translating and Adapting Tests. 2nd ed. 2017. Available from: www.InTestCom.org

[6] Orçan F. Exploratory and confirmatory factor analysis: Which one to use first? Journal of Measurement and Evaluation in Education and Psychology. 2018;**9**:414-421. DOI: 10.21031/epod.394323

[7] Steyer R. Classical (psychometric) test theory. International Encyclopedia of the Social & Behavioral Sciences. 2001:1955-1962. DOI: 10.1016/B0-08-043076-7/00721-X

[8] Streiner DL. Starting at the beginning: An introduction to coefficient alpha and internal consistency. Journal of Personality Assessment. 2003;**80**:99-103. DOI: 10.1207/S15327752JPA8001_18

[9] Allen MJ, Yen WM. Introduction to Measurement Theory. Monterey: Brooks Cole; 1979

[10] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;**16**:297-334. DOI: 10.1007/BF02310555

[11] Cho E, Kim S. Cronbach's coefficient alpha: Well known but poorly understood. Organizational Research Methods. 2015;**18**:207-230. DOI: 10.1177/1094428114555994

[12] Cortina JM. What is coefficient alpha? An examination of theory and applications. The Journal of Applied Psychology. 1993;**78**:98-104. DOI: 10.1037/0021-9010.78.1.98

[13] Schmitt N. Uses and abuses of coefficient alpha. Psychological Assessment. 1996;**8**:350-353. DOI: 10.1037/1040-3590.8.4.350

[14] Hayes AF, Coutts JJ. Use omega rather than Cronbach's alpha for estimating reliability. Communication Methods and Measures. 2020;**14**:1-24. DOI: 10.1080/19312458.2020.1718629

[15] Sijtsma K. On the use, the misuse, and the very limited usefulness of cronbach's alpha. Psychometrika. 2009;**74**:107-120. DOI: 10.1007/s11336-008-9101-0

[16] McDonald RP. Generalizability in factorable domains: "Domain validity and generalizability". Educational and Psychological Measurement. 1978;**38**:75-79. DOI: 10.1177/001316447803800111

[17] Revelle W, Zinbarg RE. Coefficients alpha, beta, omega, and the glb: Comments on sijtsma. Psychometrika. 2009;**74**:145-154. DOI: 10.1007/s11336-008-9102-z

[18] Kelley K. Supplemental material for confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. Psychological Methods. 2016;**21**:69-92. DOI: 10.1037/a0040086.supp

[19] Yang G, Zang X, Ma X, Bai P. Translation, cross-cultural adaptation, and psychometric properties of the Chinese version of the surgical fear questionnaire. Journal of PeriAnesthesia Nursing. 2022;**000**:1-7. DOI: 10.1016/j.jopan.2021.08.004

[20] Ribeiro MH, Coulon N, Guerrien A. French translation and validation of the Karitane parenting confidence scale. European Review of Applied Psychology. 2022;**72**:1-8. DOI: 10.1016/j.erap.2022.100759

[21] Villarreal-Zegarra D, Torres-Puente R, Otazú-Alfaro S, Al-kassab-Córdova A, Rey de Castro J, Mezones-Holguín E. Spanish version of Jenkins sleep scale in physicians and nurses: Psychometric properties from a Peruvian nationally representative sample. Journal of Psychosomatic Research. 2022;**157**:110759. DOI: 10.1016/j.jpsychores.2022.110759

[22] Roberson RB, Elliott TR, Chang JE, Hill JN. Exploratory factor analysis in rehabilitation psychology: A content analysis. Rehabilitation Psychology. 2014;**59**:429-438. DOI: 10.1037/a0037899

[23] Comrey AL. Common methodological problems in factor analytic studies. Journal of Consulting and Clinical Psychology. 1978;**46**:648-659. DOI: 10.1037/0022-006X.46.4.648

[24] Bled C, Bouvet L. Validation of the French version of the object spatial imagery and verbal questionnaire. European Review of Applied Psychology. 2021;**71**:100687. DOI: 10.1016/j.erap.2021.100687

[25] Hotelling H. Analysis of a complex of statistical variables into principal components. Journal of Education & Psychology. 1933;**24**:417-441. DOI: 10.1037/h0071325

[26] Abdi H, Williams LJ. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2010;**2**:433-459. DOI: 10.1002/wics.101

[27] Jolliffe IT. Rotation of principal components: Choice of normalization constraints. Journal of Applied Statistics. 1995;**22**:29-35. DOI: 10.1080/757584395

[28] Henry A, Gagnon J. Psychometrics properties of the francophone version of the displaced aggression questionnaire. Pratiques Psychologiques. 2022;**28**:29-42. DOI: 10.1016/j.prps.2021.04.002

[29] Gronier G, Baudet A. Psychometric evaluation of the F-SUS: Creation and validation of the French version of the system usability scale. International Journal of Human Computer Interaction. 2021;**37**:1571-1582. DOI: 10.1080/10447318.2021.1898828

[30] Byrne BM. Structural equation Modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. International Journal of

Testing. 2001;**1**:55-86. DOI: 10.1207/ s15327574ijt0101_4

[31] Watkins D. The role of confirmatory factor analysis in cross-cultural research. International Journal of Psychology. 1989;**24**:685-701. DOI: 10.1080/ 00207598908247839

[32] DiStefano C, Hess B. Using confirmatory factor analysis for construct validation: An empirical review. Journal of Psychoeducational Assessment. 2005;**23**:225-241. DOI: 10.1177/073428290502300303

[33] Fung SF. Validity of the brief resilience scale and brief resilient coping scale in a Chinese sample. International Journal of Environmental Research and Public Health. 2020;**17**:1-9. DOI: 10.3390/ ijerph17041265

[34] Sahlan RN, Todd J, Swami V. Psychometric properties of a Farsi translation of the functionality appreciation scale (FAS) in Iranian adolescents. Body Image. 2022;**41**:163-171. DOI: 10.1016/j.bodyim.2022.02.011

[35] Petot D, Petot JM, Fouques D. Factor structure and psychometric properties of the Achenbach and Rescorla's youth self-report French adaptation. European Review of Applied Psychology-Revue Europeenne de Psychologie Appliquee. 2022;**72**:1-13. DOI: 10.1016/j. erap.2021.100701

[36] Hu L, Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods. 1998;**3**:424-453. DOI: 10.1037// 1082-989x.3.4.424

[37] Schweizer K. Some guidelines concerning the modeling of traits and abilities in test construction. European Journal of Psychological Assessment.

2010;**26**:1-2. DOI: 10.1027/1015-5759/ a000001

[38] Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling. 2002;**9**:233-255. DOI: 10.1207/ S15328007SEM0902_5

[39] Tabachnick BG, Fidell LS. Using Multivariate Statistics. seventh ed. Boston: Pearson; 2019

[40] Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999;**6**:1-55. DOI: 10.1080/10705519909540118

[41] Sivo SA, Xitao FAN, Witta EL, Willse JT. The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. The Journal of Experimental Education. 2006;**74**:267-288. DOI: 10.3200/JEXE.74.3.267-288

[42] Noble S, Scheinost D, Constable RT. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. NeuroImage. 2019;**203**:116157. DOI: 10.1016/j. neuroimage.2019.116157

[43] Congard A, Le Vigouroux S, Antoine P, Andreotti E, Perret P. Psychometric properties of a French version of the implicit theories of emotion scale. European Review of Applied Psychology. 2022;**72**:100728. DOI: 10.1016/j.erap.2021.100728

[44] Erdinç O, Lewis JR. Psychometric evaluation of the T-CSUQ: The Turkish version of the computer system usability questionnaire. International Journal of Human Computer Interaction. 2013;**29**:319-326. DOI: 10.1080/10447318.2012.711702

[45] Lallemand C, Koenig V, Gronier G, Martin R. Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. European Review of Applied Psychology. 2015;**65**:239-252. DOI: 10.1016/j.erap.2015.08.002

[46] Sanchez C, Baussard L, Blanc N. Validation of the French and enriched version of the feelings about school (FAS) with students aged 6 to 11. Psychologie Française. 2022:1-17. DOI: 10.1016/j.psfr.2021.12.001

Section 3

Psychometrics Instruments and Procedures

Chapter 4

Psychometry in Dementia

Sandro Misciagna

Abstract

Grow in aging has led to an increasing number of people presenting with cognitive impairment and dementia. Most forms of dementia are classified by means of morphological techniques, assays of biomarkers in cerebrospinal fluid and neuropsychological assessment, into degenerative forms, dementia of vascular type and dementia secondary to other conditions. It is very difficult to make a clear-cut diagnosis of the different types of dementia by means of clinical methods. However, many psychometric tests play a prominent role in screening and evaluation of patients with cognitive impairment. Some tools can help clinicians in differential diagnosis among the various forms of dementia such as the ones that assess clinical aspects, tests that focus on specific cognitive areas or behavioral inventories. Still nowadays, there is not a consensus about the best strategies for screening and assessment of cognitive impairment among elderly subjects. The purpose of this chapter is to make a review of the screening tools and psychometric test instruments that healthcare professionals can use for screening and neuropsychological assessment of geriatric individuals with cognitive disorders to help diagnosis of dementia and to make differential diagnosis of the most common forms of dementia.

Keywords: psychometric, cognitive impairment, dementia, Alzheimer disease, vascular dementia, neuropsychological assessment, neuropsychology, mild cognitive impairment

1. Introduction

All countries of the world have observed a substantial increase in the number of elderly people. This phenomenon resulted to an increase of chronic health conditions and cognitive impairments. With the consequence of world population senescence, the healthcare professionals need to differentiate expected changes due to aging from pathological conditions due to dementia.

The current state of knowledge allows a detection of neurodegenerative brain changes [1] with neuroimaging techniques such as magnetic resonance imaging [2] or positron emission tomography with amyloid binding tracers [3] and assays of biomarkers such as beta amyloid fragment or phosphorylated tau protein in cerebrospinal fluid [4]. These techniques make possible to identify degenerative forms of dementia (Alzheimer's disease, Lewy bodies dementia and frontotemporal dementia), dementia of vascular type and dementia secondary to other conditions (such as traumatic brain injuries, human immunodeficiency infection, substance-induced dementia, Huntington disease, Parkinson's disease and prion disease).

IntechOpen

However, research studies show that preclinical diagnosis of neurodegenerative conditions is still possible with neuropsychological measurement of cognitive changes [5]. Different cognitive tasks used in combination can provide screening and assessment of cognitive impairment among elderly people, since if used together they can bring more information. Psychometric tests are also useful to differentiate pseudo-dementias [6] or other form of primary dementias that may mimic dementia of Alzheimer type such as frontotemporal dementia and Jacob–Creutzfeldt disease, predict increased or reduced risk of dementia and describe disease evolution in affected individuals.

There is no consensus regarding the best strategies for screening of cognitive impairment among elderly patients even if several brief instruments are recommended [7].

Most individuals and their caregivers would want to know a diagnosis of dementia as soon as possible to allow them to make decisions regarding future plans when they are still able to do it [8]. Furthermore, studies conducted to prevent cognitive decline and disability have demonstrated that pharmacological treatments and early interventions on healthy life-style factors such as social interactions, leisure activities, cognitive stimulation, Mediterranean diet and regular physical activity should be encouraged in patients with mid cognitive impairments as possible protectors against neurodegenerative disease of aging and progression of cognitive deficits [9].

2. Screening batteries for detection of cognitive impairment

Clinicians have developed many neuropsychological instruments that are best suited for middle-aged and older people. Brief screening tools are useful for identifying individuals with cognitive disorders, staging their severity, tracking progression over time and response to treatments. Most of them have a general applicability. The sensitivity of the screening test is defined as the number of positives correctively identified by the test as a percentage of the total number of the positives in the population studied (percentage of demented subjects). Conversely the specificity of the screening test is the number of negatives correctly identified by the test as a percentage of all the true negatives (percentage of not demented subjects). Screening tests are summarized in **Table 1**.

Some researchers consider the *clock drawing test* (CDT) as a possible instrument for the screening of dementia [10]. In this test subjects are asked to follow a two-step instruction: The first instruction consists in drawing a clock with all the numbers on it, while the second instruction consists in putting the hands of the clock on a specific time [11]. CDT is a very brief cognitive test (administration time < 5 min), easy to apply, but it is vulnerable to different interpretations of the final result, given by the different ways to analyze the clock that was drawn. The most common score in screening patients with Alzheimer disease is a total score of 10 points and a cut-off score of 7 [11]. A score of 10 corresponds to the best representation of the clock, while a score of 1 corresponds to the worst representation of the clock. Other authors propose a hierarchical classification system of errors from 1 (mild) to 5 (severe) with a total score of 5, determined by the highest level of error based on a classification system [12]. According to these authors, the errors could be classified as mild visuospatial errors, errors in denoting the time, perseverations, severely disorganization of the space and total inability to make any reasonable attempt at a clock [12].

The CDT has 67 to 97.9% sensitivity and 69 to 94.2% specificity in screening cognitive impairment [13]. It cannot be used among people with visual or motor

Battery	Total score	Sensitivity %	Specificity %	Cut-off scores	Administration time	Components of the battery
CDT	10	67–97.9	69–94.2	<7	4-5 min	A, VS, E, PR
Dem-Tect	18	100	92	≤12	8–10 min	M, L, E
MIS	8	43–86	93–97	≤4	4 min	M
Mini-Cog	5	76–99	85.3–96	≤3	2 min	A, M, L. VS, E
MSQ	10	92.3–100	86.5–100	≥3	2 min	O
SPMSQ	10	92.3–100	86.5–100	≥3	2 min	O, A, MT
BDRS	28	43	94	≥4	4–5 min	ADL
VFT	NA	37–89.5	43–97	12/13 or 14/15	4–5 min	M, L, E
SLUMS	30	92–100	76–100	23.5 or 25.5	7 min	O, M, L, A, C, VS, E
RAVLT-IR	75	97.9	17.9	≤30	10–15 min	A, M, L
RAVLT-DR	15	95.7	28.6	≤3	30 min	M, L
RAVLT-RT	15	91.5	46.4	<10	30–60 min	M, L
RAVLT-FC	15	92.6	67.9	<13	30–60 min	M, L
ACE-R	100	84–94	89–100	82 or 88	20 min	O, M, L, A, VS, E
CAMCOG	106	92	96	<80	30 min	O, L, M, A, PR, CA, AT, PE, VS
MMSE	30	88.3	86.2	23/24 or 24/25	6–10 min	O, M, L, A, C, VS, PR
MoCA	30	90–100	87	26	6–10 min	O, M, L, A, E, C, VS, PR

Abbreviations of the battery: CDT = Clock Drawing Test; Dem-Tect = Dementia Detection Test; MIS = Memory Impairment Screening Test; Mini-Cog = Mini Cognition Test; MSQ = Mental Status Questionnaire; SPMSQ = Short Portable Mental Status Questionnaire; BDRS=Blessed Dementia Rating Scale; VFT = Verbal Fluency Test; CAMCOG = Cambridge Cognitive Examination; MMSE = Mini Mental State Examination; MoCA = Montreal Cognitive Assessment; RAVLT-IR = Rey auditory verbal learning test-Immediate recall; RAVLT-DR = Rey auditory verbal learning test-Delayed recall; RAVLT-RT = Rey auditory verbal learning test-Recognition trial; RAVLT-FC = Rey auditory verbal learning test-Forced choice recognition.
Abbreviations of the components: A = attention; ADL = activities of daily living; AT = abstract thinking; C = calculation; E = executive functions; L = language; M = memory; MT = mental tracking; O = orientation; PE = perception; PR = Praxis; VS = visuo-spatial functions.

Table 1.
Screening tools for detection of cognitive impairment.

difficulties that prevent them from using properly paper and pen. There is not a consensus on whether the CDT can distinguish mild cognitive impairment (MCI) from dementia even if this test can assess motor and executive functions, memory and verbal comprehension and has been used to differentiate dementia from cognition in different studies [14]. Finally, there are also multiple scoring methods for interpreting CDT (with different degree of complexity), and there is no consensus on the best method [13].

The *DemTect* is an instrument used by Kalbe to detect dementia [15]. It takes from 8 to 10 minutes to complete and includes different cognitive tests: immediate memory of a word list, late evocation of the same list, a numerical transcoding task, and a digit span test and a semantic verbal fluency test. Total score is independent from age and education. Maximum score is 18. A score of 13–18 points is adequate for age, a score of 9–12 points identifies MCI with sensitivity of 80% and specificity of 92%, a score of 8 points or below identifies dementia with sensitivity of 100% and specificity of 92% [15].

The *memory impairment screening test* (MIS) consists in a brief 4 items battery of neuropsychological tests (delayed and cued recall of words) that has been recommended as a preliminary test in screening for dementia, in conjunction with other screening tools [16]. Total score is 8, and the cut-off score is 4. Sensitivity of MIS in screening dementia ranges from 43 to 86%, and specificity ranges from 93 to 97% [13]. MIS can be applied within four minutes and does not require the ability to write [17]. However, it is sensitive to reading abilities that means educational level of the subjects tested influences the results.

The *Mini-Cog* is a brief screening tool that includes CDT and a test of immediate and late evocation of three words [18] with a total score of 5. Administration time requires about two minutes. The Mini-Cog has a moderate-high sensitivity (76–99%) and a moderate-great specificity (85.3–96%) [19]. In a study conducted by Fowler et al., where the Mini-Cog was applied together with MIS, authors concluded that Mini-Cog was suitable for routine screening of dementia within primary care [20]. The Mini-Cog classifies dementia from cognitive normal and is not influenced from language or education [18].

The *mental status questionnaire* (MSQ) is a brief and simple screening test that consists in ten questions [21]. *Orientation questions* deal with place and time, and *general information questions* concern personal information (age, month and year of birth, name of the current and the past president). Each incorrect response receives 1 point; the maximum score is 10. According to the authors, scores from 0 to 2 indicate absence or just moderate brain dysfunction, scores from 3 to 8 a moderate brain dysfunction and scores from 9 to 10 reflect a severe brain dysfunction. MSQ scores correlate significantly with measures of brain metabolism in cortical and subcortical areas of the brain [22]. Sensitivity of MSQ in screening dementia ranges from 92.3 to 100%, while specificity ranges from 86.5 to 100% [13]. This instrument is more accurate in identifying moderate to severely impaired patients and normal subjects, but produces a high rate of false negatives among mildly impaired patients [23]. Fillebaum suggests that just two items (date of birth and name of the previous president) are sufficient as a brief screening technique [24].

The *short portable mental status questionnaire* (SPMSQ) is another brief and simple ten-point test for screening of cognitive impairment [25]. Seven questions involve spatial, temporal and personal orientation (e.g. current date and place of the examination); two questions consist in asking the names of current and previous presidents, while the last tasks consist in tests of concentration and mental tracking. Sensitivity of MSQ in screening dementia ranges from 92.3 to 100%, while specificity ranges from 86.5 to 100% [13]. The SPMSQ is able to discriminate between cognitively intact patients and patients with three level of cognitive impairment severity. On the bases of a regression analysis study, the best items to identify cognitive impairment are the date of birth, the name of the previous president and the name of the current day of the week [24]. However, like most of the brief screening instruments, the SPMSQ does not identify mildly impaired subjects from early-demented patients [24].

The *blessed dementia rating scale* (BDRS) is a brief rating scale that registers functional behavior changes reported by informants [26]. The subjects are rated on a list of 22 items, covering changes in the previous six months in abilities in personal care/eating, dressing, sphincter control, performances in daily activities (moneys, finding one's way), in personality, interests and drive. The BDRS contains items that score 1 and items weighted according the severity. Total possible score of this scale is 28 for the most deteriorated patients. Persons receiving a score less than 4 are considered unimpaired; persons with a score between 4 and 9 are considered with mild impairment; patients with a score of 10 or higher are considered with a moderate to severe impairment [27]. In a study conducted to recognize dementia in general practice, BDRS showed a sensitivity of 43% and a specificity of 94% [28]. Blessed and colleagues find a highly significant correlation between mean plaque counts and scores of this scale in a group of patients examined before they died and later came to autopsy [26].

The *verbal fluency test* (VFT) [29] is a very brief test for evaluation of language, memory and executive functions. The phonological VFT version consists in asking to list words that begin with a particular letter of the alphabet in a time interval of a minute. Instead, the semantic VFT version consists in asking to list categories of colors, animals, fruits or cities in one minute. The VFT is a simple test that can be easily applied and is very effective in evaluating language abilities and executive functions since the test requires the ability to self-regulate working memory through the ability to search and retrieve information stored in long-term memory. The VFT has a range of sensitivity from 37 to 89.5% and a specificity from 43 to 97% [13]. The VFT is quite accurate for screening early stages of cognitive impairment or dementia. It is also able to distinguish between individuals with or without normal cognition. This test is simple and does not require any materials other than a device to keep track of the time and number of words produced. Performances in the test are influenced by subject's level of age and education, and consequently raw scores must be corrected for these variables.

The *Saint Louis University Mental Status examination* (SLUMS) is a 30-point screening questionnaire [30]. It explores orientation, attention, memory, calculation, executive functions, language and visuo-spatial functions. It takes about 7 minutes to complete. The optimal cut-off scores are 23.5 for subjects with less school education and 25.5 for subjects with high school education [30]. The cut-off scores for dementia are 19.5 for subjects with less school education and 21.5 for subjects with high school education [30]. The sensitivity is 92–100%, while the specificity is 76–100% [30].

The *Rey auditory verbal learning test* (RAVLT) is an easy screening test based on assessment of verbal memory [31]. It consists of five presentations with recall of a 15-word list, one presentation of a second 15-word list and a sixth recall trial which altogether take from 10 to 15 minutes. Retention is generally examined after 30 minutes. The immediate free recall trial (RAVLT-IR) gives information on immediate word span recall and provides a learning curve that reveals learning strategies. The delayed recall trial (RAVLT-DR) gives information on how well the patient recalls what was once learned. The score for each trial is the number of words correctly recalled. Total score of the immediate free recall is the sum of the five trials; the maximum score is 75. Total score of the delayed recall trial is the number of words recalled with a maximum score of 15.

Many clinicians and researchers also include a recognitions trial (RAVLT-RT) that is assessed after a time of 30–60 minutes first developed by Lezak [32]. In the recognition task, the examiner asks the patient to identify as many words as possible from

the first list when shown a list of 50 words containing words that are semantically associated or phonemically similar of the 15-words target list.

Porech et al. presented in 2016 an auditory verbal learning forced-choice recognition task (RAVLT-FC) and proposed its use as part of a routine neuropsychological assessment [33]. In this procedure, the examiner reads a pair of words and asks the subject to choose the word from each pair that had been presented in the original list. The RAVLT-FC is then composed of 15 FC items, each consisting of the target list of 15 words paired with 15 distractors. RAVLT-FC total score consists in the number of words correctly identified; maximum score is 15. Authors have considered for each trial cut off scores that minimize risk of false positive and that have high specificity values. According to a recent study on the validity of the Rey auditory verbal learning test, RAVL-IR (total 1–5 trial) at a level of cut-off less than 30 has a specificity of 97.9% and sensitivity of 17.9%; RAVL-DR (long delay recall) at a level of cut-off less than 3 has a specificity of 95.7% and sensitivity of 28.6%; RAVL-RT (recognition task) at a level of cut-off less than 10 has a specificity of 91.5% and sensitivity of 46.4%; RAVL-FC (forced-choice trial total score) at a level of cut-off less than 13 has a strong specificity (92.6%) and a sensitivity of 67.9% [34]. As in all learning tests, age effects are prominent and tend to affect all the relevant measures [35].

The Addenbrooke's cognitive examination-revised (ACE-R) is a brief neuropsychological battery with an approximately administration time of 20 minutes. It includes five subdomain scores that are orientation/attention, memory, verbal fluency, language and visual–spatial abilities. A cut-off as 88 is indicative of presence of dementia with a sensitivity of 94% and a specificity of 89%. However, the cut-offs are defined as 82, with a sensitivity of 84% and a specificity of 100%. The ACE-R is recommended both in general hospital settings than in memory clinics [36].

The *Cambridge cognitive examination* (CAMCOG) is a brief neuropsychological battery since it takes about 30 minutes to complete. The CAMCOG consists in eight major subscales that are orientation, attention, memory, calculation, language, abstract thinking, perception and praxis. The CAMCOG has been designed to detect dementia and mild degrees of cognitive impairment [37]. The total score is 106, and the cut-off score for diagnosis of cognitive impairment is less than 80. The sensitivity is 92%, specificity is 96% [38], and there is not a ceiling effect [37].

The *mini–mental state examination* (MMSE) is one on the most widely used brief screening instrument for recognition of cognitive impairment [39]. It takes approximately 8 minutes to complete [40] and is generally considered a gold standard in clinical practice for dementia [41]. MMSE has been validated for application both in the community and in primary care in many countries [42]. It is a screening tool for cognitive impairment and identification of individuals for a more complex evaluation. MMSE is a rapid tool influenced by levels of age and formal education. The MMSE explores orientation (in time and space), attention and concentration (spell of words backwards and subtractions), memory (immediate and delayed recall of 3 words), language (denomination of 3 objects, repetition of words, execution of a three- stage command, reading and obey to an order and writing a sentence) and visuospatial construction (ability to copy a geometrical figure). Each correct response receives 1 point of score; the maximum score is 30 [43]. A suggested cut-score of 23 versus 24 (23/24) was recommended for detecting dementia in a primary care setting in persons with at least 8 years of education [39]. Scores of 21 to 24, 10 to 20 and 9 or less indicate, respectively, mild, moderate and severe cognitive impairment [44]. Sensitivity and specificity of MMSE are 88.3% and 86.2%, respectively [13].

However, MMSE is not suitable for the screening of the initial phases of dementia and is not useful to evaluate executive functions. It is effective in discriminating patients with moderate or greater cognitive deficits from control subjects [45]. It is sensitive for the follow up of progressive deterioration in dementing patients [46]. Item analyses indicate that the three-word recall is the most sensible item to dementia while the second most failures are orientation for date [47]. Given to its susceptibility to ceiling effect [48] and sociodemographic factors [49], the MMSE should not be used in isolation to definitively diagnose or rule out of dementia [50].

The *montreal cognitive assessment* (MoCA) is a brief battery of 30 point tests frequently used to screen mild forms of cognitive impairment (MCI) [51]. The MoCA has better sensitivity and specificity than MMSE in detecting MCI and dementia [52]. The MoCA covers all cognitive domains and has more tests of executive functions. Using a cut-off of 26 the MoCA assesses different cognitive domains such as orientation, memory, attention, executive functions, naming, language, abstraction and visuospatial abilities. According to some authors, it is the test with the highest predictive value for differentiating MCI and Alzheimer's dementia (AD) from normal individual cases, with a sensitivity of 90% and 100%, respectively [53]. The MoCA also has a high specificity in identifying 87% of healthy controls [54]. However, MoCA has significant correlations with age and level of formal education as other cognitive batteries such as mini–mental state examination [39], clock drawing test [55], Cambridge cognitive examination or verbal fluency test [29]. The MoCA is a tool that provides a superior overall assessment in the early stages of cognitive decline [56], but has the disadvantage of taking longer than MMSE, and it presents limitations with regard the capabilities of illiterate individuals to perform proposed tasks. The MoCA is also useful in identifying non-anamnestic forms of MCI, behavioral variant of frontotemporal dementia [57] and mild cognitive impairment in patients with Parkinson's disease [58].

3. Functional assessment of subjects with cognitive deficits

As suggested by Jiang et al. [59], changes in instrumental activities daily living for domestic works are common in patients with mild cognitive impairment. Therefore, it is recommendable the use of functional scales during the screening of subjects with cognitive deficits (**Table 2**).

The *informant questionnaire on cognitive decline in the elderly* (IQCODE) is an example of functional scale [60] with an administration time less than 20 minutes. It consists in a 26 items questionnaire applied to individuals that accompanying the patient. The examiner asks various questions about different performances of the patient in different activities of daily living (ADLs). The IQCODE has a sensitivity of 75% to 83 and a specificity from 65–90% for a cut-point of approximately 3.3 [13]. According to some researchers, this scale has better precision of results than MMSE in cases of MCI [61].

The *Pfeifer functional assessment questionnaire* (*PFAQ*) consists in simple questions about performance of elderly people regarding their functional ability in ten activities of daily living such as paying bills, making out business papers, shopping, playing a game of skill, making a cup of coffee, preparing a meal, keeping track of current events, paying attention to TV programs, remembering appointments and travel [62]. For every activity the examiner assigns four level of function from 0 (normal) to 3 (it requires assistance). Total score is 30. This

Functional scale	Total score	Sensitivity %	Specificity%	Cut-off	Administration time	Cognitive and functional domains of daily living
IQCODE	26–130	75–83	65–90	3.3	10–20 min	M, O, J, PS, CA, HH, PC
PFAQ	30	88.3	76.5	>10	15–20 min	M, O, J, PS, CA, HH, PC
CDR	18	95	94	>6	20–30 min	M, O, J, PS, CA, HH, PC

Abbreviations of the functional scale: IQCODE = Informant Questionnaire on Cognitive Decline in the Elderly; PFQAQ = Pfeiffer Functional Activity Questionnaire; CDR = Clinical Dementia Rating scale.
Abbreviations of the functional domains: CA = community affairs; J = judgment; HH = home and hobbies; M = memory; O = orientation; PS = problem solving.

Table 2.
Functional evaluation scales.

questionnaire can be completed in 15–20 minutes. In a clinical study, the use of PFAQ combined with VFT showed a sensitivity of 88.3% and a specificity of 76.5%, suggesting that these tests could be useful for screening of cognitive impairment among elderly subjects [29].

The *clinical dementia rating scale* (CDR) is another functional tool to assess behavior and cognition among elderly and to establish the degree of dementia [63]. The administration time is about 20 minutes. This instrument is a 0–3-point numeric scale that explores cognitive and behavioral functions, in order to assess the influence of cognitive impairment on functional capacity to perform activity daily living. The domains explored are six: memory, orientation, judgment, problem solving, community affairs, home and hobbies and personal care. For each domain, a score of 0 means an absence of impairment, while a score of 3 means a severe impairment; total score is 18.

In a study where CDR was used for screening of dementia, the authors found a sensitivity of 95% and a specificity of 94% [64].

The CDR is used in clinical practice and research studies to stage dementia severity and monitor disease progression over times. Since its first version, researchers have developed a modified version of CDR that includes domains of language, behavior and personality disorders to capture a range of symptoms beyond memory impairment associated with less common dementia types [65].

4. Assessment of cognitive functions in demented patients

Comprehensive test batteries provide a baseline of an individual with dementia and monitor symptoms progression over time. Most of the psychometric test planned for the examination of dementia consist in batteries that incorporates pre-existing neuropsychological tests that their creators brought together [66]. Each battery generally contains published tests or tasks specifically developed for the battery. The neuropsychological evaluation typically includes a clinical interview and assessment of different cognitive domains [67].

Orientation assessment consists in asking information about spatial, temporal and personal components.

Complex attention has different domains that include sustained attention, divided attention, selective attention and processing speed. Sustained attention is the ability to maintain attention over time and is tested in tasks where the subjects have to execute specific tasks in a determined time. Divided attention consists in executing two or more tasks in the same period of time. Selective attention consists in maintenance of attention despite distracting stimuli. Examples of attention tasks consist in mental calculations tasks, backwords spans and barrage tasks. Processing speed measures times of reaction in the attention tasks.

Battery	Sub-tests	Administration time	Main cognitive functions explored
IBMD	TO; BVRT; COWAT	10–20 minutes	O, VS, L, M, E, PR, VS
DAB	FTT; FDS; NT; ViMT; VeMT; TT; WFT; SDST; CD; NCT	45 min	A, L, M, E, AT, VS
CERAD	VFT; NT; MMSE; WLMT; CP; WLRecall; WLRecog	20–30 min	O, L, E, M, A, VS, PR, C
MFI	MMSE; RCPM; SDST	15–25 min	O, M, L, A, VS, AT
NSB	GENERAL NEUROPSYCHOLOGY TESTS	30–45 min	O, A, M, L, VS, PR
MDB	VFT; PC; WLMT; RCPM; ViMT; CD	40–75 min	VS, L, M, AT, VS, PR
CCT	PinfT; CAT; PIntT; MT; PRS; MF; VR; RLDO	45 min	O, AT, VS, M
DRS	FDS; FTT; CD; VeMT	30–45 min	A, E, PR, AT, M
ABCD	MSQ; SR; WLMT; NT; VC; CD	45–90 min	O, L, M, PR
CSD	VT; VR; VSR; VeMT; OM	120 min	L, E, VS, M
ADAS	MSQ; BSQ	45 min	O, L, M, PR

Abbreviations of the batteries: IBMD=Iowa Battery for Mental Decline; DAB=Dementia Assessment Battery; CERAD=Consortium to Establish a Registry of Alzheimer Disease; MFI = Mental Function Index; NSB=Neuropsychological Screening Battery; MDB = Mental Deterioration Battery; CCT = Cognitive Competency Test; DRS=Dementia Rating Scale; ABCD = Arizona Battery for Communicate Disorders of Dementia; CSD=Cognitive Scales for Dementia; ADAS = Alzheimer's Disease Assessment Scale.

Abbreviations of the sub-tests: BSQ = behavioral state questionnaires; BVRT = Benton Visual retention test; CAT = card arrangement test; CD = copying designs; CP=Construction praxis; COWAT = Controlled oral association test; FDS=Forward digit span; FTT = finger tapping test; NCT = number cancelation test; MF = management of finances; MMSE = Mini Mental State examination; MSQ = Mental status questionnaires; MT = memory test; NT = naming test; OM = object memory; PC = phrase construction; PInfT = personal information test; PintT = picture interpretation test; PRS=Practical reading skills; RCPM = Raven colored progressive matrices; RLDO = route learning and directional orientation; SDST = symbol digit substitution test; SR = story recall; TO = temporal orientation; TT = token test; VC = verbal comprehension; VeMT = verbal memory test; ViMT = Visual memory test; VFT = Verbal Fluency test; VR = verbal reasoning; VSR = visuo spatial reasoning; VT = vocabulary test; WFT = Word fluency test; WLMT = word list memory test; WLRecall = word list recall; WLRecog = word list recognition.

Abbreviations of the functions explored: O = orientation; A = Attention; M = memory, VS = visuo-spatial function; E = executive functions; L = Language; PR = Praxis; C=Calculation, AT = abstract thinking, PE = perception.

Table 3.
Neuropsychological batteries for assessment of cognitive functions in dementia.

Memory comprises recent memory (ability to encode new information), semantic memory (memory of facts), long-term autobiographical memory (memory of personal events) and implicit memory (procedural learning). These sub-domains can be studied using verbal and non-verbal materials. For example, recent and long-term verbal memory are studied using tasks of free recall of words, delayed recall of words, cued recall and recognition. Spatial long-term memory can be studied using tasks of delayed reproduction of geometrical figures.

Executive and processing functions include different sub-components such as working memory, planning, decision-making, feedback error corrections and mental flexibility. Clinicians have many techniques to assess executive functions that include trail-making tests, planning tasks, problem-solving tasks, inhibition tasks based on Stroop interference effect [68], backwords spans, tasks of abstraction, matrix reasoning, verbal judgments and category fluencies.

Language has subdomains that include expressive abilities (naming, fluency, vocabulary and word finding), grammar, syntax and receptive language. Expressive language abilities are generally studied by asking the patient to name objects visually presented. Grammar and syntax errors can be observed during naming and fluency tests. During language assessment examiners generally study also reading, writing, trans-codification abilities and verbal comprehension.

Perceptual-motor functions include visual perception, visual-constructional reasoning, motor coordination, praxis and gnosis. The usual method to detect these deficits is to ask the patients to draw simple and complex geometrical figures such as the Rey–Osterrieth complex figure [69] or block design. Examiners can test visual perception function using line bisection tasks and visual perception tasks. Perception tasks can include facial and colors recognition tasks.

The entire neuropsychological exam can take from several minutes to several hours on the bases of the battery of tests used. The most common batteries used in assessment of cognitive functions are summarized in **Table 3**.

5. Neuropsychological batteries

The *Iowa screening battery for mental decline* is one of the shortest battery with an administration time of less than 20 minutes [70]. It consists of just three tests: *temporal orientation*, *Benton visual retention test* and *controlled oral word association test*. This battery is able to discriminate patients with dementia due to many aetiologies (degenerative, vascular, mixed or other aetiologies) from normal subjects. The authors use this battery as a screening test and submit to further evaluations subjects with possible dementia.

Another battery of tests generally used for assessment of dementia is the *dementia assessment battery* (DAB) [71]. This neuropsychological battery consists in administration of ten tests. The *finger tapping test* consists a 15 seconds trial. *Forward digit span* consists in two-digit sets. *The naming test* assesses the abilities to denominate four sets of items of the 60 Boston naming test. *Visual memory test* consists in four sets of three items of geometric designs similar to those of the Benton visual retention test. *Verbal memory test* consists in four nine-items lists to be repeated three times in the recall trial and recognized among other words in the recognition trial. *Token test* is an oral understanding task that comes from the multilingual Aphasia examination battery. *Word fluency test* is a task of generation of lists of words that comes from the

multilingual Aphasia examination battery. *Symbol digit substitution test* is a five-symbol form of digit symbol that comes from Wechsler intelligence scale. *Copying designs* is a constructive apraxia task that uses for models Benton visual retention test figures. *Number cancelation test* is a task specifically developed for this battery. The dementia assessment battery takes about 45 minutes to administer.

Probably the best known of dementia batteries is that developed by the *Consortium to Establish a Registry of Alzheimer's Disease* (CERAD) [46]. This battery consists of seven tests. *Verbal fluency* for semantic category is a test in which the subject has to tell a list of words in the category of animals [72]. The *naming test* consists in the presentation of 15 items of the Boston naming test with five words each of low, medium and high frequency of occurrence. *Mini–mental state* is the battery for examination of mental status. *Word list memory test* is a task consisting in three learning trials of list of ten words. *Constructional praxis* consists in copying four geometric figures. *Word list recall* is the delayed recall of the list of the ten words of the word list memory test. *Word list recognition* consists in the identification of the ten target words among ten distractors. This neuropsychological battery is generally completed within 20 to 30 minutes. The CERAD is both a diagnostic and a follow-up tool.

The *mental function index* (MFI) is a screening battery [73] that incorporates three tests: the *mini–mental state examination* (MMSE), The R*aven's colored progressive matrices* (RCPM) and the *symbol digit modalities test* (SDMT). The three tests required 15–20 minutes for normal and mildly demented individuals; very obsessive, very depressed, or confused persons required up to 25 minutes. The scores of the three tests enter into a discriminant function equation to arrive to a mental function index. A score equal or superior to zero is typical in demented patients, while a negative score characterizes no demented patients. This index has a high level of agreement with diagnosis made by neurologists and is useful to discriminate among normal subjects, depressed patients and demented patients. It can be used as a follow-up tool to put in evidence increasing deterioration over a period of several years.

The *neuropsychological screening battery* [45] consists in 18 tests that cover the major areas of cognitive functioning. The battery is used in an original or an abbreviated form and requires 30–45 minutes. Cut-off scores have been developed for middle age subjects. The battery is enough effective to discriminate between Alzheimer's patients and normal elderly control subjects by comparing their performance in the different tasks.

The *mental deterioration battery* (MDB) is a neuropsychology battery generally used in Italy [74]. It is composed by three tests that explore verbal functions and three tests that explore visuo-constructive functions. The verbal tasks are *the word fluency*, *phrase construction* and *Rey's 15 words memory*, while non-verbal tasks are *The Raven's colored progressive matrices*, *immediate visual memory test* and *copying drawings*. Phrase construction test consists in asking to the subject to compose sentences from two or three words. Immediate visual memory test uses some of the items of the colored matrices that the patient has to recognize for three seconds among four alternative response choices. In copying drawings, the subject copies a star, a cube and a house on a blank paper, then copies the same figures on a paper containing "landmarks". The administration time of this neuropsychological battery is about 40–75 minutes. Studies conducted on this battery have demonstrated that non-aphasic subjects with predominant dysfunction in left hemisphere have prevalent deficits in verbal tasks while subjects with predominant dysfunction in right hemisphere have prevalent deficits in visuospatial tasks [75]. Word fluency demonstrated particular sensitivity to

dysfunction in anterior cerebral areas, while copying drawings demonstrated particular sensitivity to posterior cerebral areas.

Some psychometric test has been specifically designed to measure competency of elderly patients with dementia [66].

The *cognitive competency test* [76] is a battery validated on a sample of subjects with an age range from 50 to 93 years. The approximate administration time is 45 minutes. The battery consists of eight tests for each of which have been developed cutting scores. *Personal information test* requires the subject to write his information on an application form such as the ones used to open a bank account. *Card arrangement test* requires the subject to give the correct sequence of practical activities such as doing the laundry or making a phone call. *Picture interpretation test* consists in asking to the subject to explain what is happening in a set of five pictures. *The memory test* consists in a task of immediate and delayed recall of everyday activities such as time and place of an appointment or a short list of words. *Practical reading skills* consist in a presentation of pictures of daily situations and asking the subject to choose the proper response. In *management of finances,* the subject receives an envelope containing ten money-related items such as bills, a credit card application or a blank check and the instructions to sort these items for a bank deposit, pay a bill or doing other financial operations. *Verbal reasoning* consists in solving practical questions about time management or personal care. Finally, *route learning and directional orientation* consist in tasks about the correct use of a map of towns and routes, ability to discovery a route or trace a simple path.

The *dementia rating scale* (DRS) consists in five scales that examine the areas that characterize dementia of Alzheimer's type [77]. *Attention* is examined using digit forwards and backward tasks. Initiation and *perseveration* is examined studying the capabilities to repeat a series of one-syllable rhymes, perform a double alternating hand movements and copying tasks. *Construction abilities* are studied in a set of tasks of copy a set of figures and lines. *Conceptual functions* are explored by similarities items. *Memory functions* consist in tests of delayed recall of lists of words, sentences and designs. In the research used with this battery authors reported age effects, while educational levels did not affect performances [78]. Mattis reports that the examination of demented patients can take 30–45 minutes. The total score of this battery of tests discriminates Alzheimer's patients from normal control subjects [23] and mildly impaired patients from control subjects [79]. Subscales appear sensitive to different neuropathological conditions. For example, mildly and moderated Alzheimer's patients were more impaired in the attention and concept formation subscales [78], patients with frontal damage resulted more impaired only in initiation and perseveration subscales, and Korsakoff patients had worse scores in memory subscales [80].

The *Arizona battery for communication disorders of dementia* (ABCD) is a 14 sub-tests battery constructed for examining speech, language, verbal memory and communication deficits of demented patients [81]. The administration time is from 45 to 90 minutes. The battery consists in mental status, story recall, word learning, description and naming tests, verbal comprehension, but also drawing and copying tasks. Most of these tests have been subjected to reliability and validity evaluation [82]. The subtests of this battery are able to discriminate Alzheimer's patients from both normal subjects and aphasic stroke patients [82].

Some *cognitive scales for dementia* [83] are useful to distinguish levels of dysfunction in patients with dementia in particular of Alzheimer's typology. This cognitive battery consists in a set of six scales developed to assess patients at mild to moderate levels of deterioration to normal elderly subjects. The scales contain from 48 to 122

items. The scales are vocabulary, verbal reasoning, visual–spatial reasoning, verbal memory and object memory. A series of mazes are used to measure executive functions. Testing may take as long as two hours. This battery lowers the floor level and allows for gradations in patient's performances.

The *Alzheimer disease assessment scale* (ADAS) is a 21-item scale which combines a mental states examination (items from 1 to 11) and a behavioral examination (items from 12 to 21) [84]. The mental status questions concern orientation, language (speech and comprehension), memory (recall and recognition of a list of words), constructional praxis (ability to copy geometric figures) and ideational praxis (ability to prepare and send a letter). Most items are scored on a scale from one to five. Higher scores indicate severity of dysfunction. Interrater reliability is high for both the cognitive and the behavior sections of the scale. Items analysis shows data differences in relationship to the progression of the severity of the dementia. The ADAS is administered in approximately 45 minutes.

6. Psychometric tools for differential diagnosis of cognitive impairment

Psychometric assessment with specific tools can help clinicians in the process of differential diagnosis of cognitive impairments and dementia (see **Table 4**).

Mild cognitive impairment (MCI) is a clinical condition generally involving memory (in amnestic forms) but not functioning so that it do not meets clinical criteria for dementia [85]. Other major criteria included by the International Working Group consist in suboptimal performance on cognitive tests without evidence of functional limitations [86]. The assessment and follow-up of these subjects is necessary due to their increased risk for developing dementia [87]. The most commonly test used in clinical practice for diagnosis of MCI are *MMSE*, *MoCA*, *ACE-R, DemTect, SLUMS, IQCODE and CAMCOG.* The sensitivity of MMSE in the diagnosis of MCI varies from 18.1% to 85.7%, while specificity varies from 48–100% [88]. The MoCA for screening of MCI has a sensitivity of 90% and a specificity of 87% [54] and is probably the best alternative to the MMSE in detecting MCI among patients older than 60 [89]. Also, the ACE-R has high sensitivity (84%) and specificity (100%) in the screening of MCI [90]; however, both MoCA and ACE-R have failed to discriminate patients with MCI minimally educated and healthy controls matched for age and education [91]. DemTect has a sensitivity of 80% in screening of MCI [15]. Another psychometric instrument for the assessment of MCI is the *Saint Louis University Mental Status Examination* (SLUMS) with a sensitivity of 92–95% and a specificity of 76–81%. SLUMS is probably better than MMSE in detecting MCI [30].

Alzheimer's disease (AD) is the most common form of degenerative dementia. The typical presentation consists in memory impairment and executive dysfunction interfering with daily life activities [92]. Many tests have been validated for screening of patients with moderate Alzheimer's disease such as *MMSE*, *MoCA*, *MIS*, *Mini-Cog*, *CDT, CAMCOG and RAVLD.* Study of delayed memory impairment conducted using *RAVL-DR* (Rey auditory verbal learning test—delayed recall of a list of 15 words) have demonstrated that it can predict AD with an accuracy of 75.9% [93]. Some authors suggest that a supraspan learning task in the immediate recall task can give information about short-term retention and learning capabilities [94]. Patients with defective learning abilities (such as patients with Alzheimer's dementia) show a better recall of the words at the end of the list than those at the beginning (known as "*recency effect*"). On the other hand, normal subjects have generally better recall of

Type of Cognitive impairment	Psychometric tools for differential diagnosis
Mild cognitive impairment	MMSE, MoCA, ACE-R, DemTect, IQCODE, CAMCOG, SLUMS
Alzheimer's disease	MMSE, MoCA, MIS, Mini-Cog, CDT, RAVLT, CAMCOG, NPI, Behave-AD
Frontotemporal dementia	FAB, Exit-25
Vascular cognitive impairment	HIS, TMT, DS, MMSE, CAMCOG, ACE-R, MoCA
Parkinson's disease dementia	MoCA, SCOPA-COG, ACE-R, PD-CRS

Abbreviations of the psychometric tools: ACE-R = Addenbrooke cognitive examination revised, Behave-AD = Behavioral pathology in Alzheimer's disease interview, CAMCOG = Cambridge Cognitive Examination, CDT = Clock drawing test, DS=Digit span, Exit-25 = Executive interview, FAB = frontal assessment battery, HIS=Hatchinski ischemic score, IQCODE = Informant Questionnaire on Cognitive Decline in the Elderly, MIS = memory impaired screening test, MMSE = Mini mental state examination, NPI = neuropsychiatric inventory, PD-CRS=Parkinson disease cognitive rating scale, RAVLT = Rey auditory verbal learning test, SCOPA-COG = Scales for outcomes of Parkinson disease-cognition, SLUMS=Saint Louis University Mental Status examination, TMT = trail making test.

Table 4.
Psychometric tools for differential diagnosis of the most common forms of cognitive impairment.

the words at the beginning of the list then most of the other words (known as "*primacy effect*"). Many subjects with good learning capability repeat the list in almost the same order as it is given [66]. Patients with early Alzheimer's type dementia have a very low recall for the first presentation of the 15-words list and performances characterized by many more intrusions than patients with other diagnostic groups [95].

Some psychometric tools such as *neuropsychiatric inventory* [96] or behavioral pathology in Alzheimer's disease focus the attention on different aspects of behavior disorders that are common in this form of dementia. The *behavioral pathology in Alzheimer's disease* (Behave-AD) [97] reviews seven categories of behavior symptoms. These are paranoid and delusional ideation (e.g. hallucinations or activity disturbances), aggressivity, diurnal rhythm disturbances, affective disorders, anxieties and phobias. The behavioral symptoms often create problems for caregivers of demented patients but could be pharmacologically treated. Each of the behavioral symptoms is rated on a four-point scale in which a score of zero indicates that the symptom is not present, while a score of 3 indicates that the symptom is present and not tolerated by the caregiver. A follow up of behavioral symptoms thought the course of the disease is possible with this instrument.

Frontotemporal dementia is a degenerative form of dementia less common than Alzheimer's disease. Neuropsychological instruments focused on detection of impairments in executive functions are important for the diagnosis of this form of dementia [98]. An example is the *frontal assessment battery* (FAB), a short battery that takes about 10 minutes to complete [99]. The score of FAB ranges from 0 to 18. Subjects with a lower score have a more severe impairment [100]. The FAB is effective to differentiate patients with frontal lobe impairment from healthy subjects [100]. The *executive interview* (EXIT-25) is another screening toll exploring executive functions that takes about 10 minutes to complete [101]. FAB and EXIT-25 have a similar diagnostic power for distinguishing frontotemporal dementia from Alzheimer's disease [98].

Cerebrovascular diseases are a common cause of cognitive impairments known as *vascular cognitive impairment* and vascular dementia. Clinical tools such as the *Hatchinski ischemic score* [102] sometimes modified [103] help clinicians to distinguish vascular dementias from other primary dementia types such as Alzheimer's disease. Some items have a score of 2 points such as abrupt onset of dementia, fluctuating

course, history of strokes, presence of focal neurological symptoms and presence of focal neurological signs. Other items have a score of 1 point such as stepwise deterioration, presence of nocturnal confusion, relative preservation of personality, presence of depression, somatic complaints, emotional continence, history of hypertension and evidence of associated atherosclerosis. The total score of the Hatchinski Ischemic score is 18. Studies conducted by Hatchinski and his colleagues [102] demonstrated that the ischemic score clearly differentiated patients suffering from Alzheimer's dementia from patients with multi-infarct dementia. In fact, the higher the score the more likely is that the patient is suffering from vascular dementia. Prominent cognitive deficits in vascular cognitive impairment are executive and attentional functions while episodic memory is relatively intact. The best neuropsychological tools for assessment of patients with small vessels disease are the *trail making test* and *digit spans* [104]. The *MoCA* and the *informant-based cognitive screening test* (IBCST) are screening tools for diagnosis of multidomain cognitive impairment in stroke and post-stroke dementia [105]. Other screening tools that can be used for differential diagnosis of post-stroke dementia include the *MMSE*, the *Rotterdam-CAMCOG* [37] and the *Addenbrooke's cognitive examination* [36].

Cognitive impairment in Parkinson disease generally includes fronto-subcortical dysfunctions such as attention, frontal and executive functions deficits. Less common are cortical dysfunctions such as memory and visuospatial deficits. The *MoCA* is a tool that can be suited for screening of cognitive impairment in Parkinson disease [106]. Another valid instrument sensitive to cognitive deficits in Parkinson's disease is the *scale for outcomes of Parkinson's disease-cognition* (SCOPA-COG). SCOPA-COG is a 10-item scale with a maximum score of 43 [107]. The *ACE-R* with a cut-off less than 89 has a sensitivity of 69% and a specificity of 84% in detecting mild cognitive impairment in Parkinson disease. Finally, the *Parkinson's disease cognitive rating scale* (PD-CRS) is a tool that identifies both fronto-subcortical than cortical deficits associated with Parkinson's disease [108]. The PD-CRS has a sensitivity of 94% and specificity of 94% in diagnosis of dementia in Parkinson's disease [108].

7. Conclusions

There are many neuropsychological tools for screening and assessment of cognitive functions among elderly people. Scales, inventories and other tests designed for the screening of dementia contain items and tasks that are sensitive to the most common dementing processes especially recent and remote memory and some aspects of attention. The *mini–mental state examination* is still the most commonly used in screening of dementia. Other commonly used screening tools include the *montreal cognitive assessment* and the *clinical dementia rating scale*.

For a detailed cognitive profile of the dementia and differential diagnosis of the different forms of dementia, examiners must explore many cognitive areas that include memory, attention, executive functions, language and visuo-spatial functions. It involves the use of neuropsychological batteries of tests, each one measuring a distinct cognitive ability with greater sensitivity and specificity than a screening toll. Diagnostic accuracy may be enhanced by combining data from several of the instruments described in this chapter.

Neuropsychological measurements play an important role in identification conditions of normal aging, dementia assessment, prediction of development of cognitive impairment, measurement of residual functional abilities and identification of possible targets of intervention [109].

However, still nowadays it is necessary to create new cognitive instruments used by general practitioners within primary care services as routine procedures in order to reduce negative effects of dementia on elderly people.

Author details

Sandro Misciagna
Neurology Department, Belcolle Hospital, Viterbo, Italy

*Address all correspondence to: sandromisciagna@yahoo.it

IntechOpen

References

[1] Tomlinson BE. The structural and quantitative aspects of the dementias. In: Roberts PJ, editor. Biochemistry of Dementia. Chichester: Wiley; 1980. pp. 15-52

[2] Schoonenboom NS, van der Flier WM, Blankenstein MA, et al. CSF and MRI markers independently contribute to the diagnosis of Alzheimer's disease. Neurobiology of Aging. 2008;**29**:669-675

[3] Klunk WE, Engler H, Nordberg A, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh compound-B. Annals of Neurology. 2004;**55**:306-319

[4] Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Annals of Neurology. 2009;**65**:403-413

[5] Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement. 2011;**7**:280-292

[6] Wells CE. Pseudodementia. Am. J. Psychiat. 1979;**136**:895-900

[7] Yassuda MS, da Silva HS, Lima-Silva TB, et al. Normative data for the brief cognitive screening battery stratified by age and education. Dement Neuropsychol. 2017;**11**(1):48-53

[8] Prorok JC, Horgan S, Seitz DP. Health care experiences of people with dementia and their caregivers: A meta-ethnographic analysis of qualitative studies. CMAJ. 2013;**185**(14):E669-E680

[9] Kivipelto M, Solomon A, Ahtiluoto S, et al. The Finnish geriatric intervention study to prevent cognitive impairment and disability (FINGER): Study design and progress. Alzheimer's Dement. 2013;**9**(6):657-665

[10] Reiner K, Eichler T, Hertel J, Hoffmann W, Thyrian JR. The clock drawing test: A reasonable instrument to assess probable dementia in primary care? Current Alzheimer Research. 2018;**15**(1):38-43

[11] Sunderland T, Hill JL, Mellow AM, et al. Clock drawing in Alzheimer's disease. Journal of the American Geriatrics Society. 1989;**37**(8):725-729

[12] Shulman KI, Shedletsky R, Silver IL. The challenge of time: Clock-drawing and cognitive function in the elderly. International Journal of Geriatric Psychiatry. 1986;**1**(2):135-140

[13] Lin JS, O'Connor E, Rossom RC, et al. Screening for cognitive impairment in older adults: A systematic review for the U.S. preventive services task force. Annals of Internal Medicine. 2013;**159**(9):601-612

[14] Lourenço RA, Ribeiro-Filho ST, Moreira Ide F, Paradela EM, Miranda AS. The clock drawing test: Performance among elderly with low educational level. Braz J Psychiatry. 2008;**30**(4):309-315

[15] Kalbe E, Kessler J, Calabrese P, et al. DemTect: A new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. International Journal of Geriatric Psychiatry. 2004;**19**(2):136-143

[16] Buschke H, Kuslansky G, Katz M, et al. Screening for dementia with the

memory impairment screen. Neurology. 1999;**52**(2):231-238

[17] Cordell CB, Borson S, Boustani M, et al. Alzheimer's association recommendations for operationalizing the detection of cognitive impairment during the Medicare annual wellness visit in a primary care setting. Alzheimer's Dement. 2013;**9**(2):141-150

[18] Borson S, Scanlan JM, Chen P, Ganguli M. The mini-cog as a screen for dementia: Validation in a population-based sample. Journal of the American Geriatrics Society. 2003;**51**(10):1451-1454

[19] Borson S, Scanlan J, Brush M, et al. The mini-cog: A cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. International Journal of Geriatric Psychiatry. 2000;**15**(11):1021-1027

[20] Fowler NR, Perkins AJ, Gao S, Sachs GA, Boustani MA. Risks and benefits of screening for dementia in primary care: The Indiana University cognitive health outcomes investigation of the comparative effectiveness of dementia screening (IU CHOICE) trial. Journal of the American Geriatrics Society. 2020;**68**(3):535-543

[21] Kahn RL, Miller NE. Assessment of altered brain function in the aged. In: Storandt M, Siegler I, Ellis M, editors. The Clinical Psychology of Aging. New York: Plenum Press; 1978

[22] de Leon MJ, George AE, Ferris SH. Computed tomography and positron emission tomography correlates of cognitive decline in aging and senile dementia. In: Poon LW, editor. Handbook for Clinical Memory Assessment of Older Adults. Washington, DC: American Psychological association; 1986

[23] Kasniak AW. The neuropsychology of dementia. In: Grant I, Adams KM (Eds).

Neuropsychological Assessment of Neuropsychiatric Disorders. New York: Oxford University Press. 1986

[24] Fillebaum GG. Comparison of two brief tests of organic brain impairment, the MSQ and the short portable MSQ. Journal of the American Geriatrics Society. 1980;**28**:381-384

[25] Pfeiffer E. SPMSQ: Short Portable Mental Status Questionnaire. Journal of the American Geriatric Society. 1975;**23**:433-441

[26] Blessed G, Tomlison BE, Roth M. The association between quantitative measures of dementia and of senile changes in the cerebral grey matter of elderly subjects. British Journal of Psychiatry. 1968;**114**:797-811

[27] Eastwood MR, Lautenschlaegar E, Corbin S. A comparison of clinical methods for assessing dementia. Journal of the American Geriatrics Society. 1983;**31**(6):342-347

[28] Mant A, Eyland EA, Pond DC, Saunders A, Chancellor AHB. Recognition of dementia in general practice: Comparison of general practitioner's opinion with assessment using the mini mental state examination and the Blessed dementia rating scale. Family Practice. 1988;**6**(3):184-188

[29] Isaacs B, Akhtar AJ. The set test: A rapid test of mental function in old people. Age and Ageing. 1972;**1**(4):222-226

[30] Tariq SH, Tumosa N, Chibnall JT, et al. Comparison of the Saint Louis university mental status examination and the mini-mental state examination for detecting dementia and mild neurocognitive disorder—A pilot study. The American

Journal of Geriatric Psychiatry. 2006;**14**(11):900-910

[31] Rey A. L'examen Clinique en psychologie. Paris: Presses Universitaries de France; 1964

[32] Lezak MH. Neuropsychological Assessment. 2nd ed. New York, NY: Oxford University Press; 1983

[33] Poreh A, Bezdicek O, Korobkova I, Levin JB, Dines P. The Rey auditory verbal learning test forced-choice recognition task: Base-rate data and norms. Applied Neuropsychology: Adult. 2016;**23**:155-161

[34] Ashendorf L, Sugarman MA. Evaluation of performance validity using a Rey auditory verbal learning test forced-choice trial. The Clinical Neuropsychologist. 2016;**30**(4):599-609

[35] IvniK RJ, Malec JF, Smith GE, et al. Mayo's older Americans normative studies: Updated AVLT norms for ages 56-97. The Clinical Neuropsychologist. 1992;**6**:83-104

[36] Larner AJ, Mitchell AJ. A meta-analysis of the accuracy of the Addenbrooke's cognitive examination (ACE) and the Addenbrooke's cognitive examination-revised (ACE-R) in the detection of dementia. International Psychogeriatrics. 2014;**26**(4):555-563

[37] Huppert FA, Brayne C, Gill C, et al. CAMCOG—A concise neuropsychological test to assist dementia diagnosis: Socio-demographic determinants in an elderly population sample. The British Journal of Clinical Psychology. 1995;**34**(4):529-541

[38] Roth M, Tym E, Mountjoy C, et al. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference

to the early detection of dementia. The British Journal of Psychiatry. 1986;**149**(6):698-709

[39] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research. 1975;**12**(3):189-198

[40] Swain DG, O'Brien AG, Nightingale PG. Cognitive assessment in elderly patients admitted to hospital: The relationship between the abbreviated mental test and the mini-mental state examination. Clinical Rehabilitation. 1999;**13**(6):503-508

[41] Arevalo-Rodriguez I, Pedraza OL, Rodriguez A, et al. Alzheimer's disease dementia guidelines for diagnostic testing: A systematic review. Am J Alzheimer's Dis Other Demen. 2013;**28**:111-119

[42] Rait G, Fletcher A, Smeeth L, et al. Prevalence of cognitive impairment: Results from the MRC trial of assessment and management of older people in the community. Age and Ageing. 2005;**34**(3):242-248

[43] Cockrell JR, Folstein MF. Mini-mental state examination. In: Abou-Saleh MT, Katona CLE, Anand KA, editors. Principles and Practice of Geriatric Psychiatry. 2002 ed. Chichester, West Sussex, UK: John Wiley & Sons, Ltd.; 2002. pp. 140-141

[44] Young J, Meagher D, Maclullich A. Cognitive assessment of older people. BMJ. 2011;**343**:d5042

[45] Filley CM, Davis KA, Schmitz SP, et al. Neuropsychological performance and magnetic resonance imaging in Alzheimer's disease and normal aging. Neuropsychiatry, Neuropsychology

and Behavioural Neurology. 1989;**2**:81-91

[46] Morris JC, Heyman A, Mohs RC, et al. The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. clinical and neuropsychological assessment of Alzheimer's disease. Neurology. 1989;**39**:1159-1165

[47] Galasko D, Klauber MR, Hofstetter CR, et al. The mini mental state examination in the early diagnosis of alzheimer's disease. Archives of Neurology. 1990;**47**:49-52

[48] Tombaugh TN, McIntyre NJ. The mini-mental state examination: A comprehensive review. Journal of the American Geriatrics Society. 1992;**40**:922-935

[49] Jones RN, Gallo JJ. Education bias in the mini-mental state examination. International Psychogeriatrics. 2001;**13**:299-310

[50] Pangman VC, Sloan J, Guse L. An examination of psychometric properties of the mini-mental state examination and the standardized minimental state examination: Implications for clinical practice. Applied Nursing Research. 2000;**13**(4):209-213

[51] Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society. 2019;**67**(9):1991

[52] Trzepacz PT, Hochstetler H, Wang S, et al. Relationship between the Montreal cognitive assessment and mini-mental state examination for assessment of mild cognitive impairment in older adults. B M C Geriatr. 2015;**15**:107

[53] Cecato JF, Montiel JM, Bartholomeu D, Martinelli JE. Poder preditivo do MoCa na avaliação neuropsicológica de pacientes com diagnóstico de demência. Rev Bras Geriatr Gerontol. 2015;**17**(4):707-719

[54] Nasreddine ZS, Phillips NA, Be'dirian V, et al. The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society. 2005;**53**(4):695-699

[55] Huntzinger JA, Rosse RB, Schwartz BL, Ross LA, Deutsch SI. Clock drawing in the screening assessment of cognitive impairment in an ambulatory care setting: A preliminary report. General Hospital Psychiatry. 1992;**14**(2):142-144

[56] Sobreira E, Pena-Pereira MA, Eckeli AL, et al. Screening of cognitive impairment in patients with Parkinson's disease: Diagnostic validity of the Brazilian versions of the Montreal cognitive assessment and the Addenbrooke's cognitive examination-revised. Arquivos de Neuro-Psiquiatria. 2015;**73**(11):929-933

[57] Coleman KK, Coleman BL, MacKinley JD, et al. Detection and differentiation of frontotemporal dementia and related disorders from Alzheimer disease using the Montreal cognitive assessment. Alzheimer Disease and Associated Disorders. 2016;**30**:258-263

[58] Marras C, Armstrong MJ, Meaney CA, et al. Measuring mild cognitive impairment in patients with Parkinson's disease. Movement Disorders. 2013;**28**:626-633

[59] Jiang C, Xu Y. The association between mild cognitive impairment and doing housework. Aging & Mental Health. 2014;**18**(2):212-216

[60] Jorm AF, Broe GA, Creasey H, et al. Further data on the validity of the informant questionnaire on cognitive decline in the elderly (IQCODE). International Journal of Geriatric Psychiatry. 1996;**11**(2):131-139

[61] Morales JM, Bermejo F, Romero M, Del-Ser T. Screening of dementia in community-dwelling elderly through informant report. International Journal of Geriatric Psychiatry. 1997;**12**(8):808-816

[62] Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. Journal of Gerontology. 1982;**37**(3):323-329

[63] Morris JC. The clinical dementia rating (CDR): Current version and scoring rules. Neurology. 1993;**43**(11):2412-2414

[64] Juva K, Sulkava R, Erkinjutti T, Ylikiski R, Valvanne J, Tilvis R. Usefulness of the clinical dementia rating scale in screening for dementia. International psychogetiatrics. 1995;7:17-24

[65] Knopman DS, Weintraub S, Pankratz VS. Language and behavior domains enhance the value of the clinical dementia rating scale. Alzheimer's & Dementia. 2011;7:293-299

[66] Lezak MH, Howieson BD, Loring DW. The Neuropsychological Examination: Procedures. Neuropsychological Assessment. Fourth ed. New York: Oxford University Press; 2004

[67] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-V). Arlington (VA): American Psychiatric Association; 2013

[68] Stroop JR. Studies of interference in serial verbal reactions. Journal of Experimental Psychology. 1935;**18**(6):643

[69] Rey A. L'examen psychologique dans les cas d'ence'phalopathie traumatique. Archives de Psychologie. 1941;**28**:286-340

[70] Eslinger PJ, Damasio AR, Benton AL, Van Allen M. Neuropsychologic detection of abnormal mental decline in older persons. Journal of the American Medical Association. 1984;**253**:670-674

[71] Teng EL, Wimer C, Roberts E, et al. Alzheimer's dementia: Performance on parallel forms of the dementia assessment battery. Journal of Clinical and experimental neuropsychology. 1989;**11**:899-912

[72] Monsch AU, Bondi MW, Butters N, et al. Comparison of verbal fluency tasks in the detection of dementia of the Alzheimer's type. Archives of Neurology. 1992;**49**:1253-1258

[73] Pfeiffer RI, Kurosaki TT, Chance JM, et al. Use of the mental function index in older adults: Reliability, validity and measurement of change over time. American journal of epidemiology. 1984;**120**:922-935

[74] Miceli G, Caltagirone C, Gainotti G. Gangliosides in the treatment of mental deterioration. A doubleblind comparison with placebo. Acta psychiatrica Scandinavica. 1977;**55**:102-110

[75] Miceli G, Caltagirone C, Gainotti G, et al. Neuropsychological correlates of localized cerebral lesions in non-aphasic brain-damaged patients. Journal of Clinical Neuropsychology. 1981;**3**:53-63

[76] Wang PL, Ennis KE. The Cognitive Competency Test. Toronto: Mt Sinai

Hospital, Neuropsychology Laboratory; 1986

[77] Mattis S. Dementia Rating Scale (DRS). Odessa, FL: Psychological Assessment Resources; 1988

[78] Vitaliano PP, Breen AR, Russo J, et al. The clinical utility of the dementia rating scale for assessing Alzheimer's patients. Journal of Chronic Disorders. 1984;**37**:743-753

[79] Prinz PN, Vitaliano PP, Vitiello MV, et al. Sleep, EEG and mental changes in senile dementia of the Alzheimer's type. Neurobiology of Aging. 1982;**3**:361-370

[80] Janowsky JS, Shimamura AP, Kritchevsky M, Squire LR. Cognitive impairment following frontal damage and its relevance to human amnesia. Behavioural Neuroscience. 1989;**193**:548-560

[81] Bayles KA, Tomoeda CK. Arizona Battery for Communication Disorders of Dementia. Gaylord MI: National Rehabilitation Services; 1989

[82] Bayles KA, Boone DR, Tomoeda CK, et al. Differentiating Alzheimer's patients from the normal elderly and stroke patients with aphasia. Journal of Speech and Hearing Disorders. 1989;**54**:74-87

[83] Christensen KJ, Multhaup KS, Nordstrom S, Voss K. A cognitive battery for dementia: Development and measurement characteristic. Psychological Assessment. 1991;**3**:168-174

[84] Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. American Journal of Psychiatry. 1984;**141**:1356-1364

[85] Petersen RC, Smith GE, Waring SC, et al. Mild cognitive impairment: Clinical characterization and outcome. Archives of Neurology. 1999;**56**(3):303-308

[86] Winblad B, Palmer K, Kivipelto M, et al. Mild cognitive impairment-beyond controversies, towards a consensus: Report of the international working group on mild cognitive impairment. Journal of Internal Medicine. 2004;**256**:240-246

[87] Petersen RC, Stevens JC, Ganguli M, et al. Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review) report of the quality standards Subcommittee of the American Academy of neurology. Neurology. 2001;**56**(9):1133-1142

[88] Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. Journal of Psychiatric Research. 2009;**43**(4):411-431

[89] Ciesielska N, Sokolowski R, Mazur E, et al. Is the Montreal cognitive assessment (MoCA) test better suited than the mini-mental state examination (MMSE) in mild cognitive impairment (MCI) detection among people aged over 60? Meta-analysis. Psychiatr Pol. 2016;**50**(5):1039-1052

[90] Mioshi E, Dawson K, Mitchell J, et al. The Addenbrooke's cognitive examination revised (ACE-R): A brief cognitive test battery for dementia screening. International Journal of Geriatric Psychiatry. 2006;**21**(11):1078-1085

[91] Lonie JA, Tierney KM, Ebmeier KP. Screening for mild cognitive impairment: A systematic review. International Journal of Geriatric Psychiatry. 2009;**24**(9):902-915

[92] Scheltens P, Blennow K, Breteler MM, et al. Alzheimer's disease. Lancet. 2016;**388**(10043):505-517

[93] Callahan BL, Ramirez J, Berezuk C, et al. Predicting Alzheimer's disease development: A comparison of cognitive criteria and associated neuroimaging biomarkers. Alzheimer's Res Ther. 2015;**7**(1):68

[94] Craik FIM. Age differences in human memory. In: Birren JE, Schaie KW, editors. Handbook of the Psychology of Aging. New York: Van Nostrand Reinhold; 1977

[95] Bigler ED, Rosa L, Schultz F, et al. Rey auditory verbal learning and Rey-Osterrieth complex figure design performance in Alzheimer's disease and closed head injury. Journal of Clinical Psychology. 1989;**45**:277-280

[96] Saari T, Koivisto A, Hintsa T, et al. Psycometrico properties of the neuropsychiatric inventory: A review. Journal of Alzheimer's disease. 2022;**86**(4):1485-1499

[97] Reisberg B, Borenstein J, Franssen E, et al. Behave-AD: A clinical rating scale for the assessment of pharmacologically remediable behavioral symtomatology in Alzheimer's disease. In: Altman HJ, editor. Alzhemer's Disease. New York: Plenum; 1987

[98] Bentvelzen A, Aerts L, Seeher K, et al. A comprehensive review of the quality and feasibility of dementia assessment measures: The dementia outcomes measurement suite. Journal of the American Medical Directors Association. 2017;**18**(10):826-837

[99] Dubois B, Slachevsky A, Litvan I, et al. The FAB: A frontal assessment battery at bedside. Neurology. 2000;**55**(11):1621-1626

[100] Wildgruber D, Kischka U, Faßbender K, et al. The frontal lobe score: Part II: Evaluation of its clinical validity. Clinical Rehabilitation. 2000;**14**(3):272-278

[101] Royall DR, Mahurin RK, Gray KF. Bedside assessment of executive cognitive impairment: The executive interview. Journal of the American Geriatrics Society. 1992;**40**(12):1221-1226

[102] Hachinski VC, Iliff LD, Zilkha E, et al. Cerebral blood flow in dementia. Archives of Neurology. 1975;**32**:632-637

[103] Loeb C. Clinical diagnosis of multi-infarct dementia. In: Amaducci L, Davison AN, Antuono P, editors. Aging of the Brain and Dementia, Aging. Vol. 13. New York: Raven Press; 1980. pp. 251-260

[104] O'sullivan M, Morris R, Markus H. Brief cognitive assessment for patients with cerebral small vessel disease. Journal of Neurology, Neurosurgery, and Psychiatry. 2005;**76**(8):1140-1145

[105] Lees R, Selvarajah J, Fenton C, et al. Test accuracy of cognitive screening tests for diagnosis of dementia and multidomain cognitive impairment in stroke. Stroke. 2014;**45**(10):3008-3018

[106] Dalrymple-Alford J, MacAskill M, Nakas C, et al. The MoCA well-suited screen for cognitive impairment in Parkinson disease. Neurology. 2010;**75**(19):1717-1725

[107] Marinus J, Visser M, Verwey N, et al. Assessment of cognition in Parkinson's disease. Neurology. 2003;**61**(9):1222-1228

[108] Pagonabarraga J, Kulisevsky J, Llebaria G, et al. Parkinson's disease-cognitive rating scale: A new cognitive scale specific for Parkinson's

disease. Movement Disorders. 2008;**23**(7):998-1005

[109] Pasternak E, Smith G. Cognitive and neuropsychological examination of the elderly In: DeKosky ST, Asthana S, editors. Handbook of Clinical Neurology, Vol. 167 (3rd series) Geriatric Neurology. University of Florida, Gainesville, FL, USA: Elsevier B.V.; 2019. DOI: 10.1016/B978-0-12-804766-8.00006-6

Psychometric Analysis of an Instrument to Study Retention in Engineering

Kenneth J. Reid

Abstract

Although engineering programs admit highly qualified students with strong academic credentials, retention in engineering remains lower than most other programs of study. Addressing retention by modeling student success shows promise. Instruments incorporating noncognitive attributes have proven to be more accurate than those using only cognitive variables in predicting student success. The Student Attitudinal Success Instrument (SASI-I), a survey assessing nine specific noncognitive constructs, was developed based largely on existing, validated instruments. It was designed to collect data on affective (noncognitive) characteristics for incoming engineering students (a) that can be collected prior to the first year and (b) for which higher education institutions may have an influence during students' first year of study. This chapter will focus on the psychometric analysis of this instrument. Three years of data from incoming first-year engineering students were collected and analyzed. This work was conducted toward investigating the following research questions: Do the scale scores of the instrument demonstrate evidence of reliability and validity, and what is the normative taxonomy of the scale scores of first-year engineering students across multiple years? Further, to what extent did the overall affective characteristics change over the first year of study?

Keywords: affective, cluster analysis, engineering, noncognitive, normative taxonomy, retention, SASI-I

1. Introduction

Engineering programs tend to admit students who are academically talented, defined by strong grade point averages and standardized exam scores. Unfortunately, many of these students leave engineering, often at the end of their first year of study. Further, the structure of engineering plans of study makes it difficult for students to transfer into engineering, meaning that engineering programs show a significantly lower percentage of retained students when compared to other disciplines.

A plethora of publications regarding undergraduate engineering student retention have been written. Efforts to reform undergraduate engineering vary from the introduction to first-year engineering programs, to pedagogical improvements, to curricula focused on design, mentorship programs, etc. Studies have attempted to

develop predictive models of student success and retention. Studies have demonstrated evidence of strong predictive power of noncognitive attitudes over purely cognitive measures of students in retention and future academic performance [1–9].

This paper focuses on the psychometric analysis of the initial version of the Student Attitudinal Success Inventory (SASI-I), which was shown to be valid and reliable, and further discusses the use of normative taxonomy of first-year engineering students across multiple years to assess engineering students' multifaceted noncognitive attributes. Further, studies of shifts in noncognitive attributes over the first year of study have repeatedly shown trends in an unfavorable direction [8, 10, 11]. The scale was used at the end of the year to examine trends in student normative taxonomy (as operationalized by cluster membership) over the course of an academic year. This chapter will introduce this analysis technique to the study of noncognitive characteristics of first year engineering students.

2. Driving research questions

Specific research questions which led to this analysis include:

1. What is the evidence of reliability and validity for the SASI-I instrument consisting of a number of affective / attitudinal factors related to student success, based on incoming first-year engineering student responses?

2. What evidence supports the use of normative taxonomies (clusters) to establish SASI-I factor stability over different cohorts of incoming first-year engineering students?

3. How do the normative taxonomies of affective student characteristics change over the first year? To what extent do students' group memberships change over time?

3. Instrumentation

Data were based on separate cohorts of undergraduate engineering students enrolling in a large Midwestern university over a three-year period of 2004 (cohort 1; N = 1,605), 2005 (cohort 2; N = 1,777), and 2006 (cohort 3; N = 1,779). **Table 1**

	Preyear			Postyear		
	Number	**Male**	**Female**	**Number**	**Male**	**Female**
Cohort 1	1,605	1,297 (80.8%)	308 (19.2%)	722	563 (78.0%)	159 (22.0%)
Cohort 2	1,777	1,502 (84.2%)	275 (15.5%)	627	516 (82.3%)	111 (17.7%)
Cohort 3	1,779	1,482 (83.3%)	297 (16.7%)			

Table 1.
Demographics of student cohort groups, pre- and postsurveys.

shows cohort demographics. Each cohort consisted of all entering first-year students who were admitted to the college of engineering but had not yet started their first semester, a small number of whom did not subsequently enroll or attend classes at the institution.

The SASI-I was used to collect data from entering engineering students and was used to assess their affective / attitudinal characteristics. Data were used to identify the normative taxonomy of cohorts of incoming students [12].

Students completed the 161 item SASI-I instrument online as part of a required set of activities at orientation, indicating their responses on a Likert scale (from 1 = *Strongly Agree* to 5 = *Strongly Disagree*). The SASI-I was administered along with other tests (e.g., math placement, chemistry) prior to the first semester and again at the end of the academic year. Students who did not complete each of the three parts of the SASI-I assessment were excluded from analysis.

The self-report measures sought to assess students' affective / attitudinal beliefs across the following constructs, each theorized within the literature to be critical to academic success [1, 12].

- *Academic Motivation*: consisting of 25 items in four subfactors: Control, Challenge, Curiosity and Career.

- *Metacognition*: comprised of 20 items in four subfactors: Planning, Self-monitoring/Self-Checking, Cognitive Strategy and Awareness.

- *Deep Learning*: consisting of 10 items in two subfactors, Motive and Strategy.

- *Surface Learning*: consisting of 10 items, originally with identical subfactors to Deep Learning. Factor analysis showed two different subfactors, Memorization and Studying.

- *Academic Self-Efficacy*: consisting of ten individual items that do not form specific subfactors.

- *Leadership*: consisting of 20 items with four subfactors, Motivation, Planning, Self-Assessment and Teammates.

- *Team vs. Individual Orientation*: consisting of 10 items in two subfactors, Individual and Team Dynamic.

- *Expectancy-Value*: consisting of 32 items in five subfactors: Academic Resources, Community Involvement, Employment Opportunities, Persistence and Social Engagement.

- *Major Decision*: consisting of 21 items in four subfactors: Certainty of Decision, Difficulty in Decision, Personal Issues, and Urgency. One question was shown not to load to any of the subfactors and is assessed on its own (Independence).

The multilevel structure, where each item loads to a superordinate construct or general factor (for example, Major Decision) and one subfactor or subordinate factor within the domain of the construct (for example, the Certainty of Decision subfactor under Major Decision) supports analysis at multiple levels.

Construct	Origin	N_s	N_t	Subconstruct	N_i
Intrinsic motivation	Academic Intrinsic Motivation Scale (AIMS) [13]	4	25	Career	5
				Challenge	6
				Control	7
				Curiosity	7
Academic self-efficacy	Based on [14]	1	10	Self-efficacy	10
Expectancy-value	Based on [15]	5	32	Expected use, acad resources	5
					4
				Community involvement	8
				Employment opportunities	7
					8
				Persistence	
				Social engagement	
Deep learning approach	Revised two-factor Study Process Questionnaire (R-SPQ-2F) [16]	2	10	Motive	5
				Strategy	5
Surface learning approach		2	10	Memorization	3
				Studying	7
Metacognition	Based on State Metacognitive Inventory [17]	4	20	Awareness	5
				Cognitive strategy	5
				Planning	5
				Self-checking	5
Leadership	Based on [18]	4	23	Motivation	5
				Planning	4
				Self-assessment	7
				Teammates	7
Team vs. individual orientation	Based on [19]	2	10	Individual dynamic	5
				Team dynamic	5
Major decision	Based on [20]	5	21	Certainty of decision	3
				Difficulty of decision	10
				Personal issues	4
				Urgency	3
				Independence	1
Total	*9 constructs*	29	161		161

Table 2.
Constructs, subconstructs, and number of items in each construct.

Table 2 shows a summary of constructs, subconstructs, and number of items in each construct.

4. Psychometric analysis

4.1 Internal consistency of scales and subscales

Internal consistency of scale scores was investigated using Cronbach's coefficient alpha for each scale and subscale, for each cohort. Values for Cronbach's coefficient alpha exceeding 0.80 are desired [21, 22]. As Cronbach's coefficient alpha is sensitive to the number of items within a construct, the Spearman-Brown formula [23, 24] was used to estimate Cronbach's alpha in any subscale containing less than 10 items.

4.2 Construct and subscale structure

Factor analytic procedures were used to test the scales' multidimensional structures. Each item loads to a construct or general factor (for example, Surface Learning) and one subscale or factor within the domain of the construct (for example, the Studying subscale under Surface Learning) since the scales were based on multidimensional constructs.

Subscale definitions were examined through confirmatory factor analysis (CFA) for those constructs with an *a priori* structure. For those constructs developed specifically for this instrument, Exploratory Factor Analysis (EFA) was used to establish the subscale structure.

SAS (version 9.1.3) *proc factor* with a promax rotation was used for EFA. Promax rotation allows for the rotation of the axes to a position allowing optimal loadings for a set of items. The ideal number of factors was determined using both the Kaiser criterion, in which the number of factors is indicated by factors whose eigenvalues are greater than 1, and examination of the scree plot of eigenvalues.

Confirmatory Factor Analysis was performed to test the factor structure of each construct. Fit was assessed based on the 2004 cohort of students using LISREL™. Each construct (with the exception of *self-efficacy*) was specified using a path diagram showing each latent variable loading to the overall construct and one individual subscale. The null hypothesis (H_0) is that the data will adequately fit the proposed structure for each construct at the item level. In cases where EFA was used to specify the subscale structure (constructs developed for this instrument), a randomly selected subset of the data (n = 500) was used in the EFA procedure; a mutually exclusive subset of the data (n = 1000) was then used to verify the structure using CFA.

Confirmatory Factor Analysis fit was assessed using a number of criteria, including the chi-square statistic, the Comparative Fit Index (CFI), the Goodness of Fit Index (GFI) and the Root Mean Square Error of Approximation (RMSEA). The chi-square statistic is reported, but its sensitivity to sample size means that it is rarely used as the sole criterion to judge model fit [25]; with a sample size in excess of 1500 students, rejection of the null hypothesis is expected. Instead, Hu and Bentler [25] suggest that acceptable model fit (no Type I or Type II errors) is indicated when CFI > 0.95 and RMSEA <0.08, with an excellent fit indicated when RMSEA <0.05. Tanguma [26] demonstrated that the CFI and GFI were relatively unaffected by sample size for sufficiently large samples (n > 500), with acceptable model fit indicated by values of GFI > 0.90.

4.3 Cohort group normative taxonomy

To show factor stability over time, McDermott's [27] three stage cluster analysis was used on each cohort to determine normative taxonomies of students, clustering students with similar response patterns. Each cohort was compared to each other cohort to measure the similarity from year to year. Further, results from each postsurvey sample (cohorts 1 and 2, 2004 and 2005) are established and compared to each other and to the pre-first year normative taxonomy.

The first step of the analysis was converting raw scores to normalized z-scores to equally weight each of the nine affective/attitudinal constructs with respect to each other. Subsequently, mutually exclusive groups of approximately equal size partitions (B = number of blocks) of the data using Ward's minimum variance method [28].

For example, data for the 2005 cohort consists of nine blocks: seven with 197 and two with 199 students, with a normalized z-score for each affective / attitudinal construct for each student. For each random block, criteria for determining the optimal number of clusters (K) were: R^2 statistic (indicating the proportion of the variance accounted for by the clusters), the pseudo-F statistic over the pseudo-t^2 statistic [29] and Mojina's first stopping rule [30].

The second stage involved formation of a ($B \cdot K$) x ($B \cdot K$) similarity matrix reflecting the consequence of merging any two clusters. Each resulting cluster was considered as input to the cluster analysis procedure, resulting in the final number of clusters indicated for the complete data set. The resulting homogeneity coefficients indicate, in this case, the consistency from year to year for each cohort group.

Finally, the third stage applied k-means iterative (nonhierarchical) partitioning to relocate potentially misassigned individual cases to improve homogeneity coefficients. The profile of each individual is examined to ensure membership in the ideal cluster: misassigned profiles are reassigned to a profile more closely matching the individual. All analysis was done in SAS with modifications to code developed by Paul McDermott [27].

Final clusters were expected to satisfy an average within-cluster homogeneity coefficient $\bar{H} > 0.6$ [31]. Cattell's Cluster Similarity Coefficient, r_p, [32] was calculated to demonstrate cluster similarity between clusters within and between cohort groups. Higher coefficient values demonstrate better congruence: excellent similarity is shown with values greater than 0.95 while values between -0.7 and $+0.7$ show poor factor similarity [33].

4.4 Differences in taxonomy over the course of an academic year

Analysis of the response data showed that an assumption of normality was not valid; therefore, nonparametric tests were used to analyze the data for differences over the course of the academic year. Comparisons for statistically significant differences were done using Mann-Whitney nonparametric tests of comparison. Mann-Whitney nonparametric test results were found using SAS for Windows (version 9) proc. npar1way with the *Wilcoxon* and Monte Carlo (*MC*) options. The *MC* option produces Monte Carlo estimates of exact p values and is used specifically for large data sets. Specifying the Monte Carlo estimate results in an estimated value of p as well as upper and lower bounds of the confidence interval of the actual p value (using alpha = 0.01) with a significant savings in computational time. In addition to nonparametric tests, standard two tailed t-tests were also computed and results compared with those of nonparametric tests.

4.4.1 Effect size: Cohen's d

Statistical significance of differences is influenced by large sample sizes, and a statistically significant difference does not *necessarily* imply a meaningful or important difference – only that a true difference of means most likely exists. As the size of the population increases, even very small differences tend to become significant. The effect size, or Cohen's d, is a measure of the magnitude of the effect or the importance of the difference [33–35]. Cohen's d is found by:

$$d = \frac{M_1 - M_2}{\sigma_{pooled}} \tag{1}$$

where M_1 and M_2 are the means of the male and female population. The pooled standard deviation, σ_{pooled}, is the root-mean-square of the standard deviations of the two populations [33]. That is:

$$\sigma_{pooled} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}} \qquad (2)$$

When the two standard deviations are similar (as is typically the case), the root mean square differs very little from the simple average of the two variances.

Hyde [36, 37] defined ranges for effect sizes as part of the Gender Similarity Hypothesis as: near-zero, $d \leq 0.10$; small, $0.11 < d \leq 0.35$; moderate, $0.36 < d \leq 0.65$; large, $0.66 < d \leq 1.0$; and very large, $d > 1.0$.

5. Results

5.1 Structure of the overall instrument

For those constructs with a predefined structure (Motivation, Metacognition, Deep Learning and Surface Learning), Exploratory Factor Analysis (EFA) was performed to verify that the items loaded to the constructs as specified in the literature. EFA results agreed in all but one case, Surface Learning. In the constructs without a predefined structure (Academic Self-efficacy, Leadership, Team vs. Individual Orientation, Expectancy-Value and Major Decision), EFA was used to define the multidimensional structure. Confirmatory Factor Analysis was used to assess the fit of the multidimensional structure of the constructs. Fit indices including chi-square, CFI, GFI and RMSEA for each construct are shown in **Table 3**.

Motivation, Metacognition and Deep Learning constructs: Subscales that were unchanged from those originally presented in their literature included those under

	χ^2	χ^2 df	$pr > \chi^2$	GFI	RMSEA	CFI
Motivation	902.6	240	<.0001	0.99	0.053	1.00
Metacognition	767.53	141	<.0001	0.99	0.067	1.00
Leadership	522.16	197	<.0001	0.99	0.041	1.00
Team vs. Individual orientation	63.45	22	<.0001	1.00	0.012	1.00
Expectancy value	1410.1	417	<.0001	0.98	0.049	1.00
Major decision	805.16	153	<.0001	0.99	0.065	1.00
Deep learning	49.73	22	.00064	1.00	0.036	1.00
Surface learning (original subscales)	152.5	25	<.0001	0.97	0.072	0.942
Surface learning (revised)	31.34	24	0.14	0.99	0.013	1.00
Academic self-efficacy			*n/a*			

Surface Learning original and revised are listed to show improvement with revision of subscale structure.
χ^2 = Chi-squared; χ^2 df = Chi-square degrees of freedom; GFI = Goodness of Fit Index; RMSEA = Root Mean Square Error of Approximation estimate; CFI = Bentley's Comparative Fit Index

Table 3.
Confirmatory Factor Analysis results for each construct.

motivation, metacognition and deep learning. Factor analysis supported the subscales as originally specified. CFA results show an acceptable fit for each of these constructs with values for GFI > 0.9, CFI > 0.9 and RMSEA <0.08.

Surface Learning: The subscales of surface learning were originally defined the same as those of deep learning: *Motive* and *Strategy*. However, EFA results indicated that the subscale *Strategy* itself loaded into two separate factors, which is typically not indicative of a homogeneous construct. EFA results on the entire Surface Learning construct showed individual items clearly loading into one of two factors, which were redefined as *Memorization* and *Studying* based on context of the questions. CFA results indicate a significant improvement in fit. The redefined structure with modified subscales resulted in a value of chi-square that was not significant, meaning that the data did indeed fit the theoretical structure, even with a very large sample size.

Academic Self-Efficacy: EFA performed on the academic self-efficacy construct indicated no subscales.

Leadership, Team vs. Individual Orientation, and Expectancy-Value: Subscales for each of these constructs were defined based on exploratory factor analysis: results were validated using CFA on a mutually exclusive subset of the data. **Table 3** shows the results of the CFA for each construct, with values for GFI > 0.9, CFI > 0.9 and RMSEA <0.05, verifying the validity of the structure for each construct.

Major Decision: Results of the EFA showed the items in the Major Decision construct loaded to five subscales and CFA results showed an acceptable fit for this structure (GFI = 0.99, CFI > 0.99 and RMSEA = 0.065). The Major Decision scale contained one question which was shown not to load on any particular subscale and is presented independent of the subscales (*Independence*). Three items were shown to negatively correlate to the remainder of the scale, and were reverse scored during the analysis.

5.2 Internal consistency of scales and subscales

Reliability of the instrument is demonstrated with acceptable values of Cronbach's alpha for each construct and subscale for each cohort [21, 22, 38]. Complete results are shown in **Table 4**. Values of Cronbach's alpha are shown after reverse-scoring two of the items in the Major Decision scale.

Cronbach's coefficient alpha values for all scales exceed 0.8 with two exceptions: Surface learning (α = 0.79) and Team vs. Individual Orientation ($\alpha \geq 0.75$), demonstrating the homogeneous nature of each construct [22, 39]. The lack of variation in values of Cronbach's coefficient alpha for different student cohort groups is one indication of stability and repeatability of the scales over time. Using the Spearman-Brown formula [24, 40] results in an estimate for Cronbach's coefficient alpha for a construct when interpolated to a specified number of items. In cases where there were fewer than 10 items in a subscale, the Spearman-Brown formula was used to assess values of Cronbach's coefficient alpha for a consistent number of items. Some subscales with very few items were outside of this range; the small number of items within each subscale certainly contributed to low values of alpha. In each case, values of alpha were very consistent from cohort to cohort, further demonstrating the internal consistency of the constructs and subscales.

5.3 Cluster analysis: Pre-year survey

McDermott's three stage cluster analysis was used to derive the core profiles for each of the three years of cohort data from cohort 1 to cohort 3. The primary goal of

Number of items		Cohort 1	Cohort 2	Cohort 3
Motivation	25	0.93	0.93	0.93
Career	5	0.82	0.81	0.83
Challenge	6	0.80	0.79	0.80
Control	7	0.79	0.80	0.79
Curiosity	7	0.81	0.81	0.82
Metacognition	20	0.92	0.92	0.93
Awareness	5	0.77	0.79	0.79
Cognitive strategy	5	0.78	0.79	0.79
Planning	5	0.76	0.78	0.79
Self-checking	5	0.72	0.74	0.76
Deep learning	10	0.81	0.83	0.82
Motive	5	0.69	0.70	0.69
Strategy	5	0.67	0.71	0.71
Surface learning	10	0.79	0.79	0.79
Memorization	3	0.60	0.59*	0.62
Studying	7	0.80	0.79	0.80
Leadership	23	0.89	0.89	0.89
Motivation	5	0.69	0.69	0.69
Planning	4	0.70	0.70	0.67
Self-assessment	7	0.78	0.77	0.77
Teammates	7	0.68	0.70	0.68
Team vs. Individual orientation	10	0.75	0.79	0.76
Individual	5	0.61	0.63	0.59*
Team dynamic	5	0.69	0.74	0.73
Expectancy value	32	0.90	0.91	0.91
Academic resources	5	0.75	0.76	0.76
Community involvement	4	0.71	0.72	0.72
Employment opportunities	8	0.81	0.82	0.82
Persistence	7	0.80	0.81	0.81
Social engagement	8	0.81	0.80	0.80
Major decision	21	0.85	0.85	0.85
Certainty of decision	3	0.49*	0.49*	0.53*
Difficulty in decision	10	0.92	0.92	0.92
Personal issues	4	0.60	0.61	0.56*
Urgency	3	0.57*	0.58*	0.59*
Independence	1	—	—	—
Academic Self-efficacy	10	0.90	0.90	0.90

indicates alpha values less than 0.6

Table 4.
Values of Cronbach's coefficient alpha for each construct and subscale.

this analysis was to identify normative taxonomies of individuals who exhibit similar profiles within a given cluster, yet dissimilar profiles across clusters. Similar taxonomies are indicated by a consistent number of core profiles, and profiles of similar magnitude and shape as determined by the cluster homogeneity coefficient $\bar{H} > 0.6$ [31] and Cattell's similarity coefficient [32] (similar: $r_p > 0.95$, dissimilar: $|r_p| < 0.7$) [41].

Cluster analysis resulted in three core profiles for each cohort. The shape and pattern of each profile was consistent from cohort to cohort, showing strong repeatability and stability.

Figure 1 shows an overlay of plots of the means of each construct for each cluster of students. Specific constructs are shown on the x-axis, with center values (normalized z-scores) for each construct for each cluster on the y-axis. There is no significance to the order of the constructs on the x-axis. Center means for each core cluster within each cohort are shown in **Table 5**. The identification of three distinct clusters and the shape of each cluster of students are significant. Students who tended to rate themselves at least one standard deviation stronger than other students tended to do so across the board, except for their propensity toward surface learning: the sharp spike seen in the plots indicates these students view their learning style as deep (developing an understanding and appreciation of material) as opposed to surface (memorization). As might be expected, those students rating themselves below the affective / attitudinal norms indicated a propensity toward surface learning. Students in cluster 1 responded approximately one standard deviation below the norm (with the exception of Surface Learning) while students in cluster 3 responded one standard deviation above the average (again, with the exception of Surface Learning). Students in cluster 2 clustered about the norms for each construct. There is no significance to the cluster numbers; they are used only to distinguish between groups.

Table 6 shows the number of students within each profile. Final clusters satisfied an average within cluster homogeneity coefficient $\bar{H} > 0.99$ (**Table 6**), demonstrating that the clusters were indeed homogeneous. The plots in **Figure 2** show year to year consistency in shape with minimal variation in values for each construct.
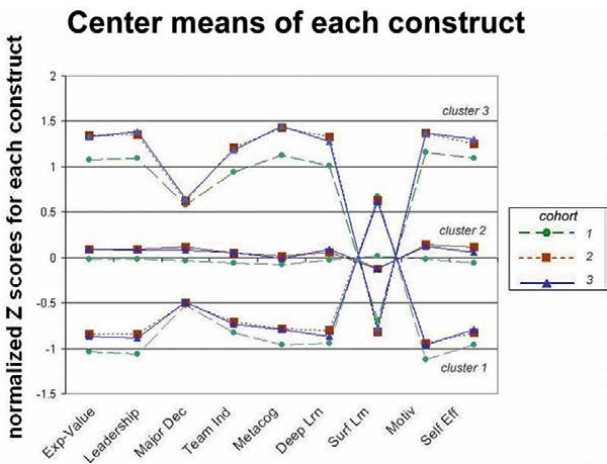

Figure 1.
Overlay of plots of normalized center means for each construct, shown for each cohort of students. Data points are shown as connected to illustrate similarly / dissimilarity of each cluster, and are not meant to imply a relationship between constructs.

Cluster	Cohort	$Z_{\text{ex-val}}$	Z_{lead}	Z_{majdec}	Z_{team}	Z_{meta}	Z_{deep}	Z_{surf}	Z_{motiv}	Z_{self}	n
1 (lower)	2004	−1.03	−1.06	−0.52	−0.83	−0.96	−0.94	0.67	−1.12	−0.96	400
	2005	−0.85	−0.85	−0.50	−0.71	−0.78	−0.80	0.63	−0.94	−0.83	611
	2006	−0.87	−0.89	−0.49	−0.73	−0.79	−0.87	0.61	−0.96	−0.79	570
2 (middle)	2004	−0.02	−0.02	−0.03	−0.06	−0.08	−0.03	0.02	−0.02	−0.06	804
	2005	0.09	0.09	0.12	0.05	0.01	0.06	−0.13	0.14	0.11	840
	2006	0.09	0.08	0.08	0.06	−0.01	0.09	−0.12	0.13	0.05	892
3 (upper)	2004	1.07	1.09	0.58	0.94	1.12	1.01	−0.69	1.16	1.09	401
	2005	1.35	1.35	0.62	1.21	1.43	1.33	−0.82	1.37	1.25	326
	2006	1.33	1.39	0.65	1.19	1.45	1.28	−0.78	1.37	1.30	317

Table 5.
Center means for each construct for core cluster.

	n (2004)	$\bar{H}$ (2004)	n (2005)	$\bar{H}$ (2005)	n (2006)	$\bar{H}$ (2006)
Cluster 1 (lower)	400 (25%)	0.997	.11 (35%)	0.997	570 (32%)	0.997
Cluster 2 (middle)	804 (50%)	0.998	840 (47%)	0.998	892 (50%)	0.998
Cluster 3 (upper)	401 (30%)	0.997	326 (18%)	0.997	317 (18%)	0.997

Table 6.
Number of students (n) and average within cluster homogeneity coefficient ($\bar{H}$) in each cluster for cohorts 1–3 (2004–2006).
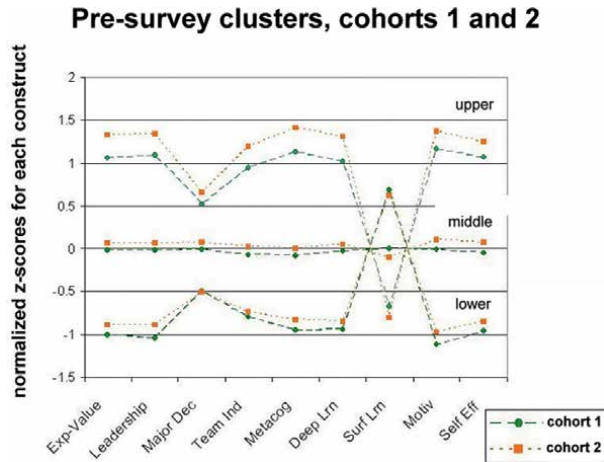
Figure 2.
Overlaid plots of center means for each subscale for clusters 1 and 2 for presurvey data.

Stability is demonstrated with minimal variability in the center values for each cohort over multiple years.

Cattell's Cluster Similarity Coefficients were calculated to objectively determine the similarity between each cluster for each year. Excellent cluster similarity is demonstrated with values of Cattell's coefficient $r_p > 0.95$; values of Cattell's coefficient comparing clusters expected to be highly similar are consistently $0.94 < r_p < 1.00$ (see **Table 7**), demonstrating these clusters to be similar and stable over this time period. Clusters expected to be dissimilar show comparison values well below values indicating similarity, demonstrating these clusters to be dissimilar to each other. Strong coefficients of similarity and similar percentages of students within each cluster show that student responses tend to remain stable from year to year.

Interestingly, the smallest spread among cohorts was found in Major Decision, indicating that students in each profile appear to differ *least* in their initial intent to pursue an engineering degree. The widest disparities between clusters appear to be among motivation and metacognition; this may be expected given the strong academic backgrounds of incoming engineering students.

5.4 Clusters in the postsurvey sample population

Cluster analysis on the postsurvey data shows that the sample population surveyed at the end of the first year clustered into four distinct groups; a significant finding as it differs from the three groups found in the presurvey. The sample population was divided into mutually exclusive groups as input to McDermott's three-stage cluster analysis; during this process, the ideal number of clusters is assessed for each subgroup, then carried through the analysis to arrive at a final answer. The criteria include multiple measures to establish the ideal number of clusters taken in concert with each other. In most cases, four clusters were indicated by Mojina's first stopping rule while the Cubic Clustering Criteria (CCC) and the pseudo-F statistic over the pseudo-t^2 statistic [29] indicated between 4 and 6 clusters. However, Milligan and Cooper [42] found that the CCC often indicates too many clusters. In each subgroup, a four cluster solution was indicated, leading to a final 4-cluster solution for the sample population in cohort 1 and 2 postsurvey data. **Figure 2** shows an overlay plot of the clusters in the postsurvey data for 2004 and 2005.

cluster 2004	2005			2006		
	1	**2**	**3**	**1**	**2**	**3**
1	**0.98**	0.54	0.08	**0.99**	0.53	0.08
2	0.29	**0.98**	–.42	0.30	**0.98**	–.42
3	0.39	–.25	**0.94**	0.37	–.26	**0.94**

cluster 2005	2006		
	1	**2**	**3**
1	**1.00**	0.41	0.18
2	0.44	**1.00**	–.36
3	0.17	–.37	**1.00**

Values show the coefficient for the cluster in the year indicated for each row / column. Bold values show comparison of clusters theoretically similar.

Table 7.
Cattell's Cluster Similarity Coefficient for each cluster 2004–2006.

An examination of the four clusters shows a clearly "upper" cluster and "lower" cluster, where students tended to respond to the SASI either higher or lower (respectively) than their peers (except for Surface Learning as expected). Two clusters emerge near the middle of the responses. As seen in **Figure 2**, these clusters are similar with the exception of Deep and Surface Learning and Major Decision. As a group, students who responded near the average of their peers and tended toward surface learning tended to be significantly lower in the decision to continue in engineering – this group is designated "middle (low)". Conversely, students who tended away from surface learning tended to indicate decisiveness toward their major ["middle (high)"]. Further examination of these two groups showed that, although the sample population was heavily skewed toward those remaining in engineering, retention in the "lower" and "middle (low)" groups was lower than retention in the "upper" and "middle (high)" group, 95% vs. 98% within the sample population. However, the low numbers of students who did not continue in engineering in the program under study do not allow a definitive conclusion to be drawn from these percentages.

Table 8 shows values of Cattell's similarity coefficients, indicating similarity between clusters which are presumed to be similar, and dissimilarity between all other clusters. Notably, the "upper" profiles did not demonstrate excellent similarity (with $r_p = 0.75$), although they are acceptably similar (with $r_p > 0.7$). The "middle (high)" and "middle (low)" clusters do show a degree of similarity, as expected, but are dissimilar enough to justify two distinct clusters of students by Cattell's similarity coefficient ($r_p < 0.6$) (**Figure 3**).

5.5 Cluster analysis of the aggregate postsurvey population

Because of the similarity of the cluster solution between cohorts 1 and 2, the data will be taken in aggregate for much of the analysis. McDermott's cluster analysis was repeated on the aggregate population by combining the data, sorting the students randomly and forming mutually exclusive datasets as previously described. Eight blocks of 184 students were used as input and, as expected, a four cluster solution was indicated once again. Cattell's similarity coefficient showed the resultant cluster solution was similar to the solution based on cohorts 1 and 2 where expected, with values of $r_p > 0.84$ (**Table 9**).

5.6 Differences in constructs, pre- to postsurvey

Ideally, students should improve not only in their cognitive abilities through the first year, but also in their desirable noncognitive characteristics. Of the nine constructs in the SASI-I, a desirable outcome would be an increase in the students' self-perception in eight of the constructs (Motivation, Metacognition, Deep Learning, Academic Self-Efficacy, Leadership, Team vs. Individual Orientation, Expectancy-Value, and Major Decision) with a lower propensity toward Surface Learning. An examination of the means of student responses from the presurvey to the postsurvey shows us that postsurvey responses went down significantly (except for Surface Learning, which increased) over the first year of study. **Table 10** shows the mean values and effect sizes for all differences. While there was a statistically significant movement (p < 0.001) in the nondesired direction for each construct, only Surface Learning and Expectancy-Value also showed a large effect size, or a large 'importance' of shift in mean values.

	$\bar{H}$ (2004)	$\bar{H}$ (2005)
Cluster 1	0.997	0.997
Cluster 2	0.998	0.998
Cluster 3	0.997	0.997

		2005	
cluster 2004	1	2	3
1	**0.98**	0.54	0.08
2	0.29	**0.98**	−0.42
3	0.39	−0.25	**0.94**

Cluster 1 = "lower"; cluster 2 = "middle"; cluster 3 = "upper".

Table 8.
Average within cluster homogeneity coefficient ($\bar{H}$) and Cattell's similarity coefficient for cohorts 1 & 2, presurvey data.
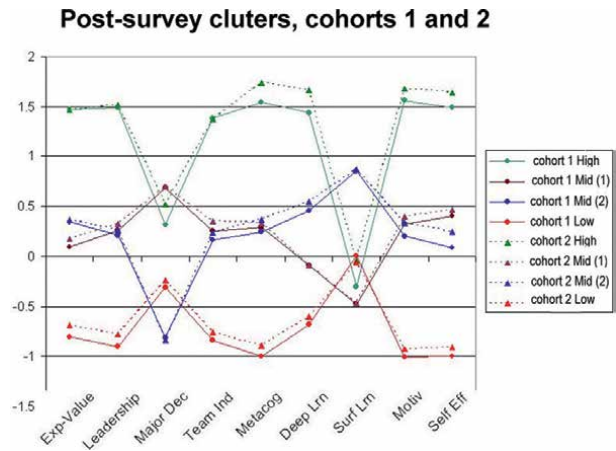
Figure 3.
Overlay plot of 2004 and 2005 postsurvey clusters. Mid (1) indicates a cluster about the average with a higher value of Major Decision. Mid (2) indicates a cluster about the average with a lower value of Major Decision.

population	cluster	Cohort 2			
		Post high	**Post mid (high)**	**Post mid (low)**	**Post low**
Cohort 1	Post High	**0.75**	0.54	0.27	−0.24
	Post Mid (high)	−0.01	**0.94**	0.59	0.48
	Post Mid (low)	0.07	0.48	**0.99**	0.22
	Post Low	−0.49	0.08	0.10	**0.94**

Bold text indicates clusters which are presumed similar. Mid (high) indicates a cluster about the average with a higher value of Major Decision. Mid (low) indicates a cluster about the average with a lower value of Major Decision.

Table 9.
Cattell's similarity coefficients, cohorts 1 and 2, postsurvey clusters.

Comparing the presurvey cluster analysis results with the postsurvey cluster analysis results gives an indication of the similarity of the profiles prior to the first year and at the end of the first year. The fact that the postsurvey data results in a four cluster solution indicates that shifts have certainly occurred with the emergence of an additional cluster, or the further division of student responses.

Figure 4 shows an overlay plot of presurvey clusters from 2004 and postsurvey aggregate results from 2004 to 2005. There is a clear similarity in appearance in the "upper" and "lower" clusters from the presurvey and postsurvey. The middle clusters show a clear deviation for Surface Learning and Major Decision. **Table 11** shows values of Cattell's similarity coefficients, indicating that the "upper" and "lower" clusters are indeed similar (r_p = 0.84 and r_p = 0.93, respectively). The presurvey "middle" group shows acceptable similarity to the two middle clusters of the postsurvey data (r_p = 0.84 and r_p = 0.79). Clusters presumed to be dissimilar are indeed shown to be dissimilar ($|r_p|$ < 0.53).

While these results are important, an investigation based on students tending to shift from one cluster to another through the course of the year should shed additional light on these trends.

population	cluster	Cohort 1 and 2 aggregate postsurvey			
		Post High	Post Mid (high)	Post Mid (low)	Post Low
Cohort 1	Post High	**0.84**	0.49	0.22	−0.29
	Post Mid (high)	0.06	**0.97**	0.62	0.40
	Post Mid (low)	0.13	0.48	**0.98**	0.16
	Post Low	−0.46	0.12	0.17	**0.98**
Cohort 2	Post High	**0.97**	0.11	0.05	−0.45
	Post Mid (high)	0.23	**1.00**	0.57	0.19
	Post Mid (low)	0.17	0.56	**0.99**	0.18
	Post Low	−0.37	0.32	0.31	**0.99**
Aggregate	Post High	**1.00**	0.19	0.11	−0.41
	Post Mid (high)	0.19	**1.00**	0.57	0.24
	Post Mid (low)	0.11	0.57	**1.00**	0.25
	Post Low	−0.41	0.24	0.25	**1.00**

Bold text indicates clusters which are presumed similar. Mid (high) indicates a cluster about the average with a higher value of Major Decision. Mid (low) indicates a cluster about the average with a lower value of Major Decision.

Table 10.
Cattell's similarity coefficients, comparing aggregate population postsurvey data to each of cohorts 1 and 2 and aggregate postsurvey data.
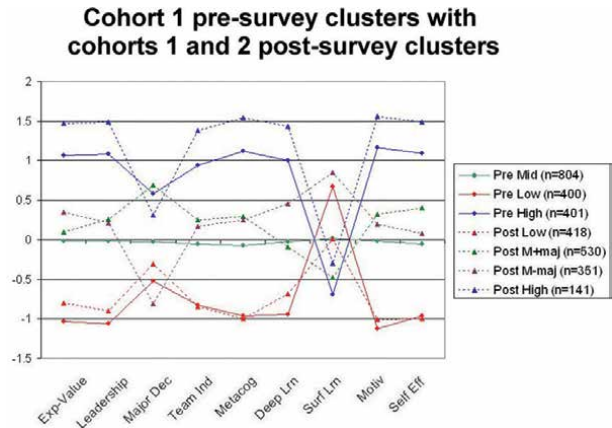

Figure 4.
Presurvey clusters (cohort 1 shown) and postsurvey clusters (cohorts 1 and 2 aggregate).

5.7 Cluster drift: changes in cluster membership over the first year

While the discovery of a three-cluster presurvey and four-cluster postsurvey solution is significant, the movement of students from a precluster to a postcluster is also of interest.

Table 12 is a frequency table showing the number of students moving from each pre-survey cluster to each postsurvey cluster, including the number of male and female students within each group. One indication of cluster stability from the

Construct	Presurvey		Postsurvey				
	Mean	σ, pre	Mean	σ, post	Post-pre	Cohen's *d*	Effect size
Surface Learning	2.487	0.493	3.028	0.628	0.542	0.959	Large
Expectancy-Value	3.915	0.359	3.582	0.500	−0.333	−0.765	Large
Motivation	4.111	0.407	3.816	0.526	−0.295	−0.628	Moderate
Major Decision	3.515	0.441	3.222	0.521	−0.293	−0.608	Moderate
Self-Efficacy	4.232	0.471	3.950	0.589	−0.281	−0.528	Moderate
Deep Learning	3.663	0.477	3.422	0.606	−0.241	−0.442	Moderate
Team vs. Individual *	3.940	0.389	3.793	0.510	−0.147	−0.323	Small
Metacognition *	3.923	0.408	3.803	0.523	−0.120	−0.255	Small
Leadership *	3.926	0.372	3.818	0.486	−0.108	−0.249	Small

* = Statistically significant difference, small to near-zero effect size.

Table 11.
Differences in mean student responses, cohort 1, pre- and postsurvey, Effect Size shown.

Population	Cluster	Cohorts 1 and 2 aggregate data, postsurvey			
		Post High	Post Mid (high)	Post Mid (low)	Post Low
Cohort 1	Pre High	**0.84**	0.52	0.29	−0.26
	Pre Middle	0.01	*0.84*	*0.79*	0.53
	Pre Low	−0.45	0.12	0.21	**0.93**

Bold text indicates clusters which are presumed similar.
Italics indicate clusters from the middle regions which are expected to be somewhat similar.
Mid (high) indicates a cluster about the average with a higher value of Major Decision. Mid (low) indicates a cluster about the average with a lower value of Major Decision.

Table 12.
Cattell's similarity coefficients, 2004 presurvey and 2004–2005 aggregate postsurvey clusters.

presurvey to the postsurvey is that 55% of students remained within their original cluster: this assumes students from the presurvey "middle" cluster remain in one of the two middle clusters in the postsurvey group.

Some other observations can be made immediately:

- Most students who were in the presurvey "upper" cluster moved to either the postsurvey "upper" cluster (n = 81) or the middle cluster with high Major Decision ["middle (high)"] (n = 88).

- Most students who were in the presurvey "lower" cluster moved to the postsurvey "lower" cluster (n = 189), although a large percentage went to the middle cluster with high Major Decision ["middle (high)"] (n = 132 or 31% of students from the presurvey "low" cluster), a favorable shift in student noncognitive characteristics.

- Most students in the postsurvey "upper" cluster came from the presurvey "upper" cluster (n = 81).

- Most students in the postsurvey "lower" cluster came from the presurvey "lower" cluster (n = 189), although a large percentage came from the presurvey "middle" cluster (n = 155, or 41% of the postsurvey low cluster).

- The fewest number of students (n = 6) took the transition from the "lower" presurvey cluster to the "upper" postsurvey cluster. Second to this was the least favorable transition, from presurvey "upper" to postsurvey "lower" (n = 36).

- The largest number of students transitioning was from the presurvey "middle" cluster to the middle cluster with a high Major Decision score ["middle (high)"]: n = 276, or nearly 21% of the population.

If an unfavorable shift is defined as one where a student downgrades their original cluster membership, for example, from presurvey "middle" to postsurvey "lower", or remains in the postsurvey "lower" cluster, 515 students, or 39%, shifted unfavorably in their noncognitive characteristics. Defining a favorable shift as one where students' trajectories are from a lower presurvey cluster to a higher postsurvey cluster or remain in the postsurvey "upper" cluster results in 354 students (or 27%) shifting in a favorable direction. It should be noted that, because the means of each cluster decreased from the presurvey to the postsurvey, students classified as having a positive shift may in fact have lower scores in their self-perception of their noncognitive attributes; if this indeed constitutes a positive shift remains to be explored.

A visual representation of the shift from presurvey cluster membership to postsurvey cluster membership is shown in **Figure 5**. Line weight represents the number of students transitioning from a presurvey to a postsurvey cluster. Postsurvey clusters have been labeled to illustrate their separation from one another.

5.8 Cluster membership and indicators of student success

Table 13 shows progress toward degree (operationalized by credits at the end of semester 4). Neither membership in presurvey cluster nor membership in postsurvey cluster was indicative of more successful progress toward degree. No significant difference was seen and no trends seen from one cluster to another, from either presurvey cluster membership or postcluster membership.

Table 14 shows retention to the end of the second year. In this case, no significant difference in indicated by presurvey cluster membership; however, students in the "upper" or "middle (high)" cluster were significantly more likely to remain in engineering with 93.6% retention vs. students in the "middle (low)" or "lower" postsurvey cluster, with 87.6% retention (z = 3.67). This trend is visible regardless of student presurvey cluster. While this data set is biased in that overall first year retention was very high compared to the overall student population in engineering, the emergence of a significant difference in second year retention based on postsurvey cluster membership indicates the potential for postsurvey cluster membership as an indicator of retention.

Grade point averages at the end of the first and second year are shown in **Tables 15** and **16** respectively. Neither GPA shows any significant difference based on presurvey cluster membership, when taken in aggregate or taken within individual postsurvey cluster memberships. In other words, there is no evidence that presurvey cluster memberships in indicative of improved GPA after one or two years.
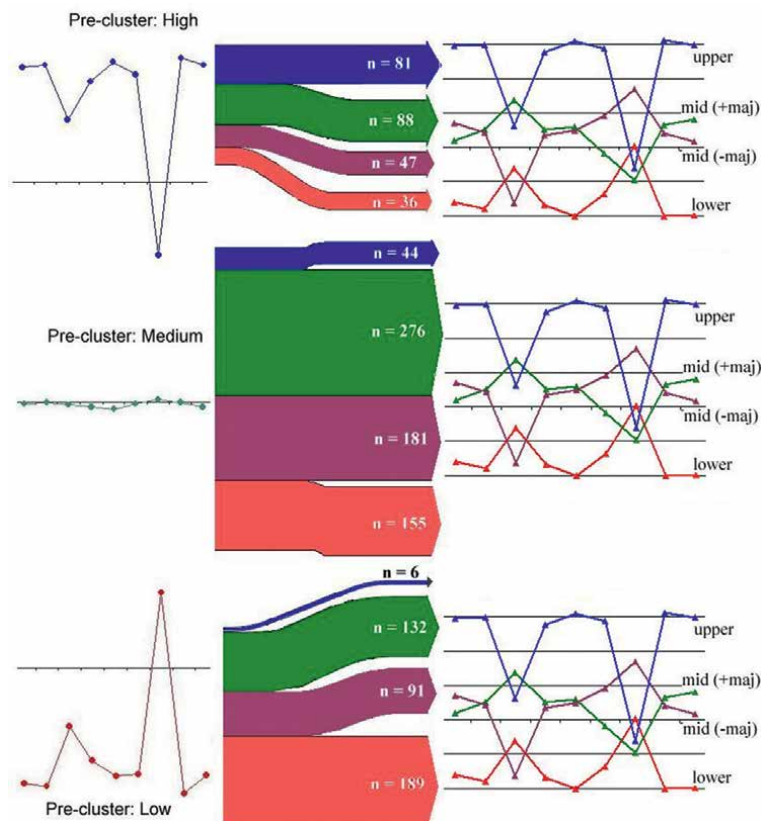
Figure 5.
Visual representation of student trajectories from presurvey cluster membership to postsurvey cluster membership. Line weight represents number of students transitioning cluster membership.

Postsurvey cluster membership does appear to be indicative of GPA to an extent. Improvement in GPA is seen as students progress from the "lower" to the "upper" postsurvey cluster membership when data are taken in aggregate, and within individual presurvey clusters, for both one year and two year GPAs. Using postsurvey cluster membership, a significant difference and small (but near moderate) effect size was found when students in the "upper" and "middle (high)" cluster were combined and compared to students in the "middle (low)" and "lower" clusters: one year GPAs were 3.09 and 2.87 respectively (p < 0.001, d = 0.334); two year GPAs were 3.00 and 2.79 respectively (p < 0.001, d = 0.286). Therefore, it appears that postsurvey cluster membership is indicative of improved student success while presurvey cluster membership is not indicative of improved student success as operationalized by GPA.

Examination of the postsurvey clusters shows that one construct, Major Decision, distinguishes the "upper" and "middle (high)" clusters from the "middle (low)" and "lower" clusters, which is where significant differences emerge. Further examination of differences in indicators based on this construct show that a statistically significant difference is found in two-year retention rate, end of first year GPA and end of second year GPA based on Z score of Major Decision as a single construct (**Table 17**). The effect size of the difference is small for GPAs from year one (d = 0.267) and year two (d = 0.190). No difference was found in student progress toward degree (**Table 18**).

Cohorts 1 & 2 aggregate, presurvey	cluster		Cohorts 1 & 2 aggregate, postsurvey			
			Post High	Post Mid (high)	Post Mid (low)	Post Low
	Pre High	»	81 (61 M, 20 F)	88 (66 M, 22 F)	47 (42 M, 5 F)	36 (32 M, 4 F)
	Pre Middle	»	44 (41 M, 3 F)	276 (187 M, 89 F)	181 (148 M, 33 F)	155 (127 M, 28 F)
	Pre Low	»	6 (3 M, 3 F)	132 (103 M, 29 F)	91 (77 M, 14 F)	189 (169 M, 20 F)

Number of male and female students indicated in parentheses.
Mid (high) indicates a cluster about the average with a higher value of Major Decision. Mid (low) indicates a cluster about the average with a lower value of Major Decision.

Table 13.
Number of students shifting from each presurvey cluster to each postsurvey cluster (cohorts 1 and 2 aggregate data).

Population	Cluster	Cohort 1 & 2 aggregate data, postsurvey				
		Post High	Post Mid (high)	Post Mid (low)	Post Low	
Cohort 1	Pre High	61.1	60.2	60.5	58.2	60.3
	Pre Middle	57.8	60.7	60.1	59.3	60.0
	Pre Low	63.4*	62.3	57.5	60.6	60.5
		60.1	61.0	59.4	59.9	

Mid (high) indicates a cluster about the average with a higher value of Major Decision. Mid (low) indicates a cluster about the average with a lower value of Major Decision.
* indicates very small sample size, n = 6.

Table 14.
Progress toward degree as measured by credits at the end of year 2.

Population	Cluster	Cohort 1 & 2 aggregate data, postsurvey				
		Post High	Post Mid (high)	Post Mid (low)	Post Low	
Cohort 1	Pre High	97.5%	94.2%	87.2%	85.3%	90.4%
	Pre Middle	88.6%	91.2%	87.7%	89.0%	89.5%
	Pre Low	66.7%*	98.5%	85.6%	87.8%	90.4%
		93.1%	93.7%	87.0%	88.0%	
		93.6%		87.6%		

Mid (high) indicates a cluster about the average with a higher value of Major Decision. Mid (low) indicates a cluster about the average with a lower value of Major Decision.
* indicates very small sample size, n = 6.

Table 15.
Retention: registration for fourth semester.

		Cohort 1 & 2 aggregate data, postsurvey				
Population	**Cluster**	**Post high**	**Post mid (high)**	**Post mid (low)**	**Post low**	
Cohort 1	Pre High	3.07	3.11	2.80	2.69	2.98
	Pre Middle	3.14	3.03	2.90	2.90	2.97
	Pre Low	*3.15**	3.18	2.91	2.87	2.98
		3.10	3.08	2.89	2.86	
		3.09			2.87	

Mid (high) indicates a cluster about the average with a higher value of Major Decision. Mid (low) indicates a cluster about the average with a lower value of Major Decision.
* indicates very small sample size, n = 6.

Table 16.
End of first year GPAs (4.0 scale).

		Cohort 1 & 2 aggregate data, postsurvey				
Population	**Cluster**	**Post high**	**Post mid (high)**	**Post mid (low)**	**Post low**	
Cohort 1	Pre High	3.07	2.93	2.73	2.52	2.88
	Pre Middle	3.04	2.92	2.80	2.80	2.87
	Pre Low	*3.14**	3.15	2.84	2.82	2.93
		3.06	2.99	2.81	2.78	
		3.00			2.79	

Mid (high) indicates a cluster about the average with a higher value of Major Decision. Mid (low) indicates a cluster about the average with a lower value of Major Decision.
* indicates very small sample size, n = 6.

Table 17.
End of second year GPA (4.0 scale).

	Progress toward degree	**2-year retention***	**GPA: 1 year***	**GPA: 2 year***
$Z_{maj} > 0$	60.7 credits	93.1%	3.06	2.96
$Z_{maj} < 0$	59.7 credits	86.0%	2.89	2.82
significance, effect size	p = 0.097 d = 0.09	z = 4.23	p < 0.001 d = 0.267	p = 0.097 d = 0.19

* indicates a statistically significant difference.

Table 18.
Student success indicators based on postsurvey Major Decision construct.

6. Conclusion

The Student Attitudinal Success Instrument I, an instrument to assess nine affective / attitudinal characteristics (and their associated subscales) of incoming students prior to the beginning of their program of study, was evaluated using data collected from large cohorts of incoming engineering students.

The SASI-I is shown to be a psychometrically sound instrument for the population of first-year engineering students at a large institution in the Midwest (United States). Internal consistency of scale scores was investigated. Factor analysis was used to establish and verify the structure of the factors and subfactors. Cronbach's coefficient alpha values for all scales exceed 0.8 (with two exceptions, 0.75 and 0.79); confirmatory factor analysis results verify the theoretical factor structure of each. McDermott's three-stage cluster analysis was used to define the normative taxonomies of three years of student data. Cluster analysis results in a stable, repeatable 3-cluster solution over multiple years: clusters expected to be highly similar were shown to have values of Cattell's coefficient > 0.94, providing evidence of stability.

This instrument is a tool to collect data prior to beginning classes in the first year of engineering, thus, it can provide valuable input to any model predicting retention past the first year (when most attrition in engineering occurs). In addition, this tool provides information on characteristics for which intervention methods may be developed, thus increasing likelihood of student success. Unlike other assessment instruments which rely on data collected during or after the first year or data for which a school may not have an influence, this tool provides necessary input for the creation and adoption of first-year programs at the earliest possible time.

Inputs to model(s) to be developed include these affective / attitudinal constructs in addition to cognitive data; such models will allow for guidance for individual students or small groups who may particularly benefit from specific intervention programs [6, 43–45]. Cluster membership offers a potential model input that has previously not been presented in the literature within engineering. Additionally, students who may not experience a benefit from these interventions may be able to opt out of some first-year programs, thus increasing the value of the course / program content they will experience during their first year.

Finally, while this instrument is effective, additional affective / attitudinal characteristics which could prove to be predictive of retention have been incorporated into the SASI-2 [4]. Ideally, the size of the existing instrument could be reduced to allow for inclusion of additional constructs without increasing the number of items. Additional affective / attitudinal constructs to be investigated and eventually proposed for inclusion should include only those constructs for which first-year intervention programs can have an effect.

Acknowledgements

Author details

Kenneth J. Reid
University of Indianapolis, Indianapolis, Indiana, USA

*Address all correspondence to: reidk@uindy.edu

IntechOpen

References

[1] Reid K, Imbrie PK, Lin JJ, Reed T, Immekus J. Psychometric properties and stability of the Student Attitudinal Success Instrument: The SASI-I. International Journal of Engineering Education. 2016;**32**(6):2470-2486

[2] Bernold L, Spurlin J, Anson C. Understanding our students: A longitudinal study of success and failure in engineering with implications for increased retention. Journal of Engineering Education. 2007;**96**:263-274

[3] Bothaina A, Al-Sheeb A, Abdella G. Modeling of student academic achievement in engineering education using cognitive and non-cognitive factors. Journal of Applied Research in Higher Education. 2019;**11**(2):178-198

[4] Yoon S, Imbrie PK, Lin JJ, Reid K. Validation of the Student Attitudinal Success Inventory II for engineering students. American Society for Engineering Education Annual Conference. 2014. Indianapolis, IN

[5] Rosen J, Glennie E, Dalton B, Lennon J, Bozick R. Noncognitive Skills in the Classroom: New Perspectives on Educational Research. Research Triangle Park, NC: RTI International; 2010

[6] Imbrie P, Lin JJ, Reid K. Comparison of four methodologies for modeling student retention in engineering. American Society for Engineering Education Annual Conference. 2010. Louisville, KY

[7] Ting S. Predicting academic success of first-year engineering students from standardized test scores and psychosocial variables. International Journal of Engineering Education. 2001;**17**(1):75-80

[8] Besterfield-Sacre M, Atman C, Shuman L. Characteristics of freshman engineering students: Models for determining student attrition in engineering. Journal of Engineering Education. 1997;**86**(2):139-149

[9] Ohland M, Sheppard S, Lichtenstein G, Eris O, Chachra D, Layton R. Persistence, engagement, and migration in engineering programs. Journal of Engineering Education. 2008;**97**(3):259-278

[10] Virguez L, Murzi H, Reid K. A quantitative analysis of first-year engineering students' engineering-related motivational beliefs. International Journal of Engineering Education. 2021;**37**:6

[11] Besterfield-Sacre M, Moreno M, Shuman L, Atman C. Gender and ethnicity differences in freshman engineering student attitudes: A cross-institutional study. Journal of Engineering Education. 2021;**90**(4):477-490

[12] Reid K. Development of the Student Attitudinal Success Instrument: Assessment of first year engineering students including differences by gender. [dissertation]. West Lafayette, IN. Purdue University. 2009

[13] French B, Oakes B. Measuring academic intrinsic motivation in the first year of college: Reliability and validity evidence for a new instrument. Journal of the First-Year Experience. 2003;**15**(1):83-102

[14] Bandura A. Perceived self-efficacy in cognitive development and functioning. Educational Psychologist. 1993;**28**(2):117-148

[15] Wigfield A, Eccles J. Expectancy–value theory of achievement motivation. Contemporary Educational Psychology. 2000;**25**:68-81

[16] Biggs J, Kember LD. The Revised Two-Factor Study Process Questionnaire: R-SPQ-2F. British Journal of Educational Psychology. 2001;**71**(1):133-149

[17] O'Neil H, Abedi J. Reliability and validity of a state metacognitive inventory: Potential for alternative assessment. The Journal of Educational Research. 1996;**89**(4):234-245

[18] Hayden D, Holloway E. A longitudinal study of attrition among engineering students. Engineering Education. 1985;**75**(7):664-668

[19] Mcmaster J. Desired attributes of an engineering graduate. AIAA Advanced Measurement and Ground Testing Technology Conference. 1996. New Orleans, LA

[20] Osipow S. Assessing career indecision. Journal of Vocational Behavior. 1999;**55**(1):147-154

[21] Henson R. Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. Measurement and Evaluation in Counseling and Development. 2001;**34**(3):177-189

[22] Nunnally J, Bernstein I. Psychometric Theory. 3rd ed. New York: McGraw-Hill; 1994

[23] Alsawalmeh A, Feldt L. Testing the equality of two independent coefficients adjusted by the Spearman-Brown formula. Applied Psychological Measurement. 1999;**23**(4):363-370

[24] Bodnar G. Statistical analysis of multiple-choice exams. Journal of Chemical Education. 1998;**57**(3):188-190

[25] Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999;**6**(1):1-55

[26] Tanguma J. Effects of sample size on the distribution of selected fit indices: A graphical approach. Educational and Psychological Measurements. 2001;**61**(5):759-776

[27] McDermott P. Megacluster analytic strategy for multistage hierarchical grouping with relocations and replications. Educational and Psychological Measurement. 1998;**58**:677-686

[28] Ward J. Hierarchical grouping to optimize an objective function. American Statistical Journal. 1993;**58**:236-244

[29] Duda R, Hart P. Pattern Classification and Scene Analysis. New York: John Wiley; 1973

[30] Mojena R. Hierarchical grouping methods and stopping rules: An evaluation. The Computer Journal. 1977;**20**(4):359-363

[31] Tyron B, Bailey D. Cluster Analysis. New York: McGraw Hill; 1970

[32] Cattell R. The Scientific Use of Factor Analysis in Behavioral and Life Sciences. New York: Plenum Press; 1978

[33] Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988

[34] Cohen J. A power primer. Psychological Bulletin. 1992;**112**(1):155-159

[35] Rosnow R, Rosenthal R. Computing contrasts, effect sizes, and counternulls on other people's published data:

General procedures for research consumers. Psychological Methods. 1996;**1**(4):331-340

[36] Hyde J, Linn M. Gender similarities in mathematics and science. Science. 2006;**314**:599-600

[37] Hyde J. The gender similarity hypothesis. American Psychologist. 2005;**60**(6):581-592

[38] Streiner D. Starting at the beginning: An introduction to coefficient alpha and internal consistency. Journal of Personality Assessment. 2003;**80**(1):99-103

[39] Tuckman B. Conducting Educational Research. New York: Wadsworth Group; 1999

[40] Alsawalmeh A, Feldt L. Testing the equality of two independent α coefficients adjusted by the Spearman-Brown formula. Applied Psychological Measurement. 1999;**23**(4):363-370

[41] Freedman R, Stumpf S. What can one learn from the learning style inventory? Academy of Management Journal. 1978;**21**(2):275-282

[42] Milligan G, Cooper M. An examination of procedures for determining the number of clusters in a data set. Psychometrika. 1985;**50**:159-179

[43] Imbrie PK, Lin JJ, Reid K, Malyscheff A. Using hybrid data to model student success in engineering with artificial neural networks. Proceedings of the Research in Engineering Education Symposium. Davos, Switzerland, 2008

[44] Imbrie PK, Lin JJ, Oladunni T, Reid K. Use of a neural network model and noncognitive measures to predict student matriculation in engineering,

American Society for Engineering Education National Conference, Honolulu, HI, 2007

[45] Geisinger B, Raj RD. Why they leave: Understanding student attrition from engineering majors. International Journal of Engineering Education. 2013;**29**:914-925

Development and Assessment of Scales in the Area of Psychiatry and Mental Health during the COVID-19 Pandemic

Ek-Uma Imkome

Abstract

Nowadays, mental health problems and psychiatric disorders have a high prevalence and are caused by co-factors. They can relapse and be exacerbated by internal and external factors such as stressful life events, poor coping skills, and COVID-19. The early detection of specific signs and symptoms is complicated. Frontliner clinical nurses must assess patient signs and symptoms as soon as possible. For this process, they require a quick and early detection measurement tool that precedes the interview, physical examination, and laboratory tests. A scale with good psychometric properties will help nurses screen and identify individuals as high-risk or non-high-risk, the severity of their symptoms (mild, moderate, or severe), and provide efficient nursing care.

Keywords: measurement, mental health, psychiatry, psychometric properties, Covid-19 pandemic

1. Introduction

Measurement is essential for healthcare providers to understand the population's health status and trends over time and to measure the effectiveness of interventions to improve it. COVID-19 has spread worldwide at an unprecedented rate and scale. People are experiencing its various psychological effects, ranging from severe symptoms to stressful responses, such as depression, anxiety, suicidal ideation, concern about infection, uncertainty and helplessness from the prolonged pandemic, and loneliness from quarantine and social isolation [1–3]. These psychological problems persist without being identified or treated and can lead to more serious psychological diseases [4–6]. The current pandemic has caused people to become exhausted in their daily lives. Therefore, it is essential to understand the psychological problems experienced by the general population and cope with them during the pandemic to protect them from psychological illness.

A scale in the area of psychiatry and mental health has been developed using valid and reliable measures. Various indicators can measure aspects of health; that is, instruments that summarize the data related to a unique phenomenon. Excellent

quality instruments should measure what is hypothetical, provide the same result if measured by different people in similar circumstances, and measure and reflect changes only in the situation concerned. Sound mental health and psychiatric scales should reflect an aspect of a chosen target, staging of the problem, and cost–benefit. It measures the state of mental health and related needs. It should inform its users whether the set targets are being achieved.

To choose the right scale, knowledge of the process of scale development and assessment of the scale in psychometric properties is essential. This chapter aims to provide suggestions for developing and choosing the scale.

2. Assessment of mental health and psychiatric needs and needs index models

The need for care, perceived need for care, demand for care, and use of care are four concepts. Mental and psychiatric problems are linked to various physical, psychological, social, and economic needs. Three groups of these needs have been assigned by WHO, which are associated with impairment, disabilities, and handicaps. The needs for mental health can be determined at either the individual or population level, and needs are estimated at the population level using four methods: (a) the survey method, (b) analysis of utilization data, (c) analysis of socioeconomic factors, and (d) a combination of techniques. The need for intervention due to mental health problems is not satisfactory. To bridge this gap, there is a need for a new scale to assess mental health needs and psychiatric problems during the COVID-19 pandemic.

3. Scale in the field of mental health and psychiatry during the COVID-19 pandemic

There are some new measures to assess psychological problems during the COVID-19 outbreak in the database during 2020–2022 as follows:

a. Stress

- COVID-19 student stress scale [7]

b. Fear

- The fear of COVID19 scale [8]

- The fear of COVID-19 scale [9]

c. Phobia

- The COVID-19 phobia scale [10]

d. Anxiety [11]

- The coronavirus anxiety scale [12]

e. Depressive

- COVID-19 depression scale for healthcare workers [13]

f. Posttraumatic stress disorders

- Posttraumatic stress disorder questionnaire [14]

Scale	Sample	Administration	Psychometric properties
Stress			
COVID-19 Student Stress Scale [7]	514 Italian university students	The Likert scale consists of seven items. The scale consists of three factors (relationships and academic life, isolation, and fear of contagion).	The scale of fit values were found to be good (v2 /df = 0.56; CFI = 0.95; TLI = 0.95; RMSEA = 0.06), and cronbach's alpha coefficient = .71.
Fear			
The fear of COVID19 scale [8]	717 Iranian participants	This 7-item scale uses a five-point Likert-type Scoring ranges from "strongly disagree (1) to "strongly agree (5)." A high score shows a high level of fear of the COVID-19 outbreak. The scores on this scale range between 7 and 35. This scale has no reverse-scored items.	The corrected item-total correlation range from 0.47 to 0.56, and factor loadings range from 0.66 to 0.74. The properties evaluated using classical test theory and the Rasch model were satisfactory on the seven-item scale. Internal consistency (α = .82) and test–retest reliability (ICC = .72). Concurrent validity was measured by the Hospital Anxiety and Depression Scale (with depression, r = 0.425, anxiety, r = 0.511) and the Perceived Vulnerability to Disease Scale with perceived infectability (r = 0.483) germ aversion (r = 0.459).
The fear of COVID-19 scale [9]	1304 participants from 75 cities in Turkey. Of all the participants, 917 (70.3%) were females, and 387 (29.7%) were males whose ages ranged from 18 to 64 years (M = 29.47, SD = 10.54)	A unidimensional seven-item, five-point Likert scale, ranging from 1 to 5 (1-strongly disagree, 5-strongly agree). There are no reverse-scored items. Scores range between 7 and 35. A high score indicates a high level of fear of the COVID pandemic.	CFA analyses showed that fit indices were all within the acceptable limit [$\chi 2$ (13, N = 1304) = 299.47, p < .05; SRMR = .061; GFI = .936; NFI = .912; IFI = .915; CFI = .915]. The factor loadings of the Fear of COVID-19 Scale ranged from .484 to .723.
Phobia			
The COVID-19 phobia scale [10]	1250 participants (765 women, 61.2%) with a mean age of 37.53 years (SD = 16.94, range = 17–89 years).	A self-report scale to measure levels of coronavirus (COVID-19) phobia. This scale was rated on a five-point scale from 1 to 5 (strongly disagree to strongly agree. The scores on the scale can range between 20 and 100, and a higher score indicates a greater phobia in the respective subscales and total scale.	Kaiser measure of sampling adequacy for EFA was 0.926 Bartlett test of sphericity was significant $\chi 2$ (df = 190) = 14,396.195, p < .001. $\chi 2$ (df = 125) = 446.93, $\chi 2$/df = 3.57, p < .001, GFI = 0.97, AGFI = 96, NFI = 0.98, IFI = 0.98, TLI = 0.98, CFI = 0.98, and RMSEA = 0.03 [90% confidence interval = 0.03 and 0.03].

Scale	Sample	Administration	Psychometric properties
Anxiety			
The coronavirus anxiety scale [11]	Coronavirus Anxiety Scale: A brief mental health screener for COVID-19	A pool of 20 candidate items was created based on the psychology of fear and anxiety literature [15–18]. Each item was written to capture a a unique manifestation of this particular form of anxiety in 4 dimensions: a. cognitive (i.e., repetitive thinking, worry, processing biases, dreaming, planning) b. behavioral (i.e., dysfunctional activities, avoidance, compulsive behaviors) c. emotional (i.e., fear; anxiety; anger) d. physiological (i.e., sleep disturbances, somatic distress, tonic immobility) dimensions of coronavirus anxiety. All item was rated on a five-point scale to assess the symptom frequency, ranging from "not at all (0)" to "nearly every day(4)" over the preceding 2 weeks. This measurement arrangement is based on the DSM-5's cross-cutting symptom quantity, the adult self-rated version (APA, 2013, pp. 734), consistent with the American Psychiatric Association's system of determining psychiatric symptoms over time and response to treatment.	Confirmatory factor analyses demonstrated that the CAS measures a reliable (α = 0.92), unidimensional construct with a structure that was shown to be invariant across gender, race, and age. Construct validity was demonstrated with correlations between CAS scores and demographics, coronavirus diagnosis, history of anxiety, coronavirus fear, functional impairment, alcohol/drug coping, religious coping, hopelessness, suicidal ideation, and social attitudes. The CAS also demonstrated solid discrimination ability for functional impairment (AUC =0.88), while the original cut score of ≥ 9 (76% sensitivity and 90% specificity) showed the most robust diagnostic effectiveness among scores.
Depress			
COVID-19 Depression Scale for healthcare workers [13]	320 Health-care workers (HCWs), including physicians of various medical specialties, dental specialists, and nurses.	Self-report questionnaires of the 10-item COVID-19 depression scale are composed of a two-component structure identified as 1. work-related anxiety 2. psychological distress All the items were formulated as a 5-point Likert-type scale for response options which ranged from 1 = I strongly agree, to 5 = I strongly disagree	The mean CDS-HW score of the study participants was observed to be 23.67 ± 2.82, and the scale demonstrated good internal consistency reliability (Cronbach's alpha: 0.741)
Post-traumatic stress disorders			

Scale	Sample	Administration	Psychometric properties
Post-Traumatic Stress Disorder Questionnaire [14]	Of 2286 respondents, there were 1706 women (74% of the sample). The mean age of the participants was 29.61 (SD = 11.42), and the age ranged between 18 and 74 years.	A self-report questionnaire (COVID-19-PTSD), consisting of 19 items, was developed starting from the PTSD Check List for DSM-5 (PCL-5) questionnaire and was administered to analyze its psychometric properties. The item is a five-point Likert scale ranging from "not at all (0)" to extremely (4)." The questionnaire asks the respondents to "Referring to the COVID-19 pandemic and the social distancing events implemented to contain it, specify how you feel for each of the following dimensions."	The confirmatory factor analysis showed that a seven-factor model (Intrusion, Avoidance, Negative effect, Anhedonia, Dysphoric arousal, Anxious arousal, and Externalizing behavior) fits the data. Significant correlations were found among COVID-19-PTSD scores, general distress, and sleep disturbance.

4. Implication

The developer should focus on the development process to properly develop the new scale, as given in **Table 1**.

For this section, the core concerns in choosing a good measurement are listed in **Table 2**.

Process	Implication
Step 1: Identification of the dimensions of the scale	1. This step includes exclusive deductive methods using a literature review and inductive methods by conducting interviews. Additionally, combined deductive and inductive methods are used for a new scale [19].
	2. Qualitative research: A deductive and inductive approach to scale development, full literature review, subjective analysis, assessment of content validity, and recruitment of a more significant number of interviewers should be a concern. Theoretical analysis: This step involves underpinning the scale theoretically some Ref. to be here to refine experts' items or used opinions. Some developers use target population opinions, and only one expert or population judge in the area of deep specialization [19–23]. The fundamental theories that support the scale development process are classical test theory (CTT)—known as classical psychometry, and item response theory (IRT)—known as modern psychometry [24].
	3. Content Validation: The panel of experts comprising 3–5 people with experience in the concerned field are asked to sort behaviors into groupings and indicate which items are applicable. This item sort is distributed in Microsoft Word format, editable by those not co-located with the researcher. This helps solidify the classifications made by the researcher and provides content validity for the survey.
Step 2: Item pool generation	• Creating questionnaire draft • Gathering of initial items
Step 3: Determination of the measurement scale	• Likert rating scale • Yes or no

Process	Implication
Step 4: Expert review of the initial item pool	• Validity of the questionnaire • Content validity: At least three professionals
Step 5: Revision and inclusion of items	• Pilot study • Missing data: Includes numbers that have been grouped, aggregated, rounded, censored, or truncated, resulting in partial loss of information.

Step 6: Administration of the items to a development sample
 	
• Survey for validity and reliability
	1. Methodological:
	
	  • The cross-sectional methodology is a limitation of scale development. The longitudinal approach in scale development facilitates a better acceptance of the analyzed variables and assesses the predictive validity.
	
	  • Self-reporting methodology is a limitation of new scale development. The self-reporting nature of quantitative studies raises the possibility of participant bias, social desirability, demand characteristics, and response sets which affect the validity of the findings. Incorporating objective or independent measures to supplement the subjective evaluation of the variables studied in developing the new scale and improving the interpretation of findings should be of concern.
	
	  • Web-based surveys are limitations of a new measurement created. The coverage bias is bias due to sampled individuals not having—or choosing not to access—the Internet) and nonresponsive bias due to the informants of a study differing from those who did not respond in terms of demographic or attitudinal variables) (Kim et al., 2011). In-person surveys or survey interviews are suggested as these methods reduce problems related to concerns about confidentiality and the potential for coverage and nonresponsive bias [25].
	
	2. Sample characteristics:
	
	  • Homogeneous and convenient samples are the limitations of generalization.
	
	  • Small sample size: The new measurement should have a larger sample size (minimum ratio of 10:1) to increase the credibility of the results and thus obtain a more exact outcome in the psychometric analysis. Using separate samples is recommended for EFA and CFA [26]. However, some researchers use the same sample due to the rarity of samples, vulnerability groups, illegal/criminals (i.e., methamphetamine use/drug use/violence), the social stigma of patient and family caregivers, and high stress and anxiety to get involved in the research from their history of drug use and criminal [27].
	
	3. Difficulty controlling all variables: Knowing the target construct in detail during the item generation and allowing all possible and vital variables to be investigated and forbidden. Hypothesizing and testing potential variables that could be controlled during the scale development process is the way to help control variables.

Step 7: Evaluation of the items	Psychometric analysis: The robust demonstration of the construct validity and reliability. Inadequate choice of the instruments or variables to be correlated with the study variable.
	
	1. Validity
	
	  • Exploratory factor analysis (EFA): Analyzing a certain amount of subjectivity by identifying and labeling factors. The sample statistical results for EFA-- KMO, Bartlett test of sphericity.
	
	  • Confirmatory factor analysis (CFA): Analyzing and assigning items to factors, testing the hypothesized structure of the data, and statistically comparing alternative models. Examples of results from CFA are CFI, GFI, and RMSEA.
	
	  • Most studies opt to combine EFA and CFA to analyze the new instruments.
	
	  • Convergent validity is used after EFA and CFA
	
	  • Discriminant validity, predictive/nomological validity, and criterion validity are also used to test the construct validity
	
	2. Reliability: The most common technique is internal consistency, and test–retest reliability is the second technique. Item-total correlation/inter-item reliability, Split-half reliability, Inter-judge reliability

Process	Implication
Step 8: Optimization of the tool length Final questionnaire selection	1. The brevity of the scale: The short scales can critically conciliate the instrument's reliability [28]. Many scale items tend to be more reliable, with higher alpha values [19]. Deciding to remove items from the scale may decrease Cronbach's alpha. An alpha value between 0.8 and 0.9 is considered ideal.
	2. Item: Items are easy to answer, and reverse-scored items require good scale development.

Table 1.
Process of scale development.

Core concern	Implication
Content representativeness and relevance,	The scale should use a representative sample from a more significant performance domain and support the intended use of the assessment.
Thinking skills and processes (Substantive evidence)	The thinking processes and skills used to complete the instrument and it is significant to building knowledge in research and innovation that should be checked successfully before deciding to use the scale [12].
Internal structure evidence evaluates the relationship among the assessment tasks or parts of the instrument.	The internal structure evidence that evaluates the scale relationship among the assessment tasks or parts of the instrument should be checked.
External structure	The external structure should be composed of:
	a. How well do the assessment results correlate with other variables or criteria.
	b. Within this category are predictive validity evidence (the extent to which the individual's future can be predicted by prior performance on an assessment instrument) and concurrent validity evidence (the extent to which an individual's current status on a criterion can be estimated from current performance on an assessment instrument).
Reliability over time, assessors, and content domain (Reliability evidence)	The psychometric properties should be sound [19, 22].
Cost, efficiency, practicality, and instructional features (practicality evidence)	The impediments to the proper use of the assessment, such as complexity, training requirements, and cost, should be examined.
Manual instructions	The instructions should contain the details as below:
	a. Manualized instructions for raters
	b. Instructions about determining the application methods of the new scale.
	c. It defined the development of operational strategies that will enable the application of the measurement and the format in which it will be obtainable, determining how the participant's response will be given for each item and how the respondent should respond to each item [24].
	d. They define how the scale scores would be analyzed.
	e. The manual should be short and without confusion to the subjects, contain one or more examples of how the items should be answered, and ensure that the subject is not tense while answering the question.
Stressful life events	Recently, a new scale developed to evaluate psychological troubles based on stress responses considers the COVID-19 outbreak as an external stressor. Stress response refers to a set of affective, cognitive, somatic, and behavioral manifestations within the range of functional integrity. Stressors impact the increase or decrease in one's adaptive capacity. The composition may vary depending on which emotional, cognitive, physical, and behavioral factors of stress responses are focused on. Therefore, choosing and measuring stress responses to COVID-19 should be a scope of the definition of stress.

Table 2.
Core concerns in choosing the measurement.

5. Conclusion

This chapter presents the type of measurement in psychiatry and mental health, the steps of scale development, assessment of mental health and psychiatric needs, needs index models, points of concern, and implications. Various concepts and methodological strategies have been identified and discussed, along with suggestions for choosing appropriate scales and scale development. We believe this chapter makes essential contributions to the literature, mainly because it provides a comprehensive set of recommendations to increase the quality of future practices in the scale development process and selection.

Author details

Ek-Uma Imkome
Faculty of Nursing, Department of Mental Health and Psychiatric Nursing,
Thammasat University, Thailand

*Address all correspondence to: ek-uma@nurse.tu.ac.th

IntechOpen

References

[1] Abbas J, Wang D, Su Z, Ziapour A. The role of social media in the advent of COVID-19 pandemic: Crisis management, mental health challenges and implications. Risk Manag Healthc Policy. 2021;**14**:1917-1932. DOI: 10.2147/rmhp.S284313

[2] O'Connor RC, Wetherall K, Cleare S, McClelland H, Melson AJ, Niedzwiedz CL, et al. Mental health and well-being during the COVID-19 pandemic: Longitudinal analyses of adults in the UK COVID-19 Mental Health & Wellbeing study. The British Journal of Psychiatry. 2021;**218**(6):326-333. DOI: 10.1192/bjp.2020.212

[3] Su Z, McDonnell D, Wen J, Kozak M, Abbas J, Šegalo S, et al. Mental health consequences of COVID -19 media coverage: The need for efective crisis communica- tion practices. Globalization and Health. 2021;**17**(1):1-8. DOI: 10.1186/s12992-020-00654 4

[4] Arslan G, Yıldırım M, Tanhan A, Buluş M, Allen KA. Coronavirus stress, optimism-pessimism, psychological infexibility, and psychological health: Psychometric properties of the Coronavirus Stress Measure. International Journal of Mental Health and Addiction. 2020;**19**(6):2423-2439. DOI: 10.1007/s11469-020-00337-6

[5] Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. The Lancet. 2020;**395**(10227):912-920. DOI: 10.1016/S0140-6736(20)30460-8

[6] Yıldırım M, Solmaz F. COVID-19 burnout, COVID-19 stress and resilience: Initial psychometric properties of COVID-19 Burnout Scale. Death Studies. 2022;**46**(3):524-532. DOI: 10.1080/07481187.2 020.1818885

[7] Zurlo MC, Cattaneo Della Volta MF, Vallone F. COVID-19 student stress questionnaire: Development and validation of a questionnaire to evaluate students' stressors related to the coronavirus pandemic lockdown. Frontiers in Psychology. 2020;**11**:2892. DOI: 10.3389/fpsyg.2020.576758

[8] Ahorsu DK, Lin CY, Imani V, Safari M, Grifths MD, Pakpour AH. The fear of COVID-19 scale: Development and initial validation. International Journal of Mental Health and Addiction. 2020. DOI: 10.1007/s11469-020-00270-8

[9] Satici B, Gocet-Tekin E, Deniz ME, Satici SA. Adaptation of the Fear of COVID-19 Scale: Its Association with Psychological Distress and Life Satisfaction in Turkey. International Journal of Mental Health and Addiction. 2021;**19**(6):1980-1988.DOI: 10.1007/s11469-020-00294-0

[10] Arpaci I, Karataş K, Baloğlu M. The development and initial tests for the psychometric properties of the COVID-19 Phobia Scale (C19P-S). Personality and Individual Differences. 2020;**164**:110108. DOI: 10.1016/j.paid.2020.110108

[11] Lee SA. Coronavirus Anxiety Scale: A brief mental health screener for COVID-19 related anxiety. Death Studies. 2020;**44**(7):393-401. DOI: 10.1080/07481187.2020.1748481. Epub 2020 Apr 16. PMID: 32299304

[12] Morgado FF, Meireles JF, Neves CM, Amaral AC, Ferreira ME. Scale development: Ten main limitations

and recommendations to improve future research practices. Psicologia: Reflexão e Crítica. 2018;**30**(1):1-20. DOI: 10.1186/s41155-016- 0057-1

[13] Divvi A, Kengadaran S, Katuri LS, et al. Development and validation of English version of COVID-19 Depression Scale for health-care workers. Journal of Education and Health Promotion 2021;**10**(1). DOI: 10.4103/jehp.jehp_1610_20

[14] Forte G, Favieri F, Tambelli R, Casagrande M. COVID-19 pandemic in the Italian population: Validation of a post-traumatic stress disorder questionnaire and prevalence of PTSD symptomatology. International Journal of Environmental Research and Public Health. 2020;**17**(11):4151. MDPI AG. Retrieved from. DOI: 10.3390/ijerph17114151

[15] American Educational Research Association, et al., editors. Standards for Educational and Psychological Testing. American Educational Research Association. 2014

[16] Barlow DH. The nature of anxiety: Anxiety, depression, and emotional disorders. In: Rapee RM, Barlow DH, editors. Chronic Anxiety: Generalized Anxiety Disorder and Mixed Anxiety-Depression. Guilford Press; 1991

[17] Ekman P. Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. Times Books/Henry Holt and Co.; 2003

[18] Öhman A, Flykt A, Lundqvist D. Unconscious emotion: Evolutionary perspectives, psychophysiological data and neuropsychological mechanisms. In: Lane RD, Nadel L, editors. Cognitive Neuroscience of Emotion. Oxford University Press; 2000

[19] De Vellis RF. Scale Development: Theory and Applications. 2nd ed. Vol. 26. Thousand Oaks, CA: Sage Publications; 2003

[20] Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. Psychological Assessment. 1995;**7**(3):309-319. DOI: 10.1037/1040-3590.7.3.309

[21] Clarke A, Friede T, Putz R, Ashdown J, Martin S, Blake A, et al. Warwick-edinburgh mental well-being scale (WEMWBS): Mixed methods assessment of validity and reliability in teenage school students in England and Scotland. BMC Health and Quality of Life Outcomes. 2011;**11**:487

[22] DeVellis RF. Scale Development: Theory and Applications. 4th ed. Thousand Oaks, CA: Sage. SAGE; 2017

[23] Nunnally JC. Psychometric Theory. New York: McGraw Hill; 1967

[24] Pasquali L. Instrumentação Psicológica: Fundamentos e Práticas. Porto Alegre: Artmed; 2010

[25] Reed LL, Vidaver-Cohen D, Colwell SR. A new scale to measure executive servant leadership: Development, analysis, and implications for research. Journal of Business Ethics. 2011;**101**:415-434. DOI: 10.1007/s10551-010-0729-1

[26] Zheng J, You L, Lou T, Chen N, Lai D, Liang Y, et al. Development and psychometric evaluation of the dialysis patient-perceived exercise benefits and barriers scale. International Journal of Nursing Studies. 2010;**47**:166-180. DOI: 10.1016/j.ijnurstu.2009.05.023

[27] Imkome E. Develop and assess the psychometric property test on burdened care caregiver scale-Thai version for Schizophrenia and co-occurring

methamphetamine use. F1000Research. 2022;**10**:484. DOI: 10.12688/f1000research.52288.2

[28] Raykov T. Alpha if item deleted: A note on loss of criterion validity in scale development if maximizing coefficient alpha. British Journal of Mathematical and Statistical Psychology. 2008;**61**:275-285. DOI: 10.1348/000711007X188520

Edited by Sandro Misciagna

Psychometrics is the science of assessing and measuring mental capacities and processes, which are constructs not easily observed directly. It is concerned with objective measurement of cognitive functions, characteristics of personality, emotions, behaviour, socio-educational qualities, and mental disorders. This book provides a comprehensive overview of psychometrics and its use in assessing psychological disorders. Chapters discuss the history of psychometrics and its theoretical bases and psychometric methodologies for assessing dementia, mental health, and more.

IntechOpen