Papeles del Psicólogo / Psychologist Papers, 2022 Vol. 43(1), pp. 3-11
https://doi.org/10.23923/pap.psicol.2985
http://www.papelesdelpsicologo.es
http://www.psychologistpapers.com

Impact of ICT on the assessment environment. Innovations through continuous improvement

Paula Elosua
Universidad del País Vasco, España

La tecnología digital ya forma parte de las prácticas relacionadas con el uso de test, y de la evaluación psicológica y educativa; la forma en que se diseñan test, se recogen datos, o los modelos para su análisis han evolucionado. Como consecuencia, nos enfrentemos a retos como la formación del profesional, la colaboración con otras áreas, la expansión del concepto de test, y la deliberación sobre cuestiones éticas y legales. En este entorno dinámico, la experiencia y el conocimiento relacionado con la medición en psicología nos arrogan la capacidad y oportunidad de integrar y expandir los pilares en los que se asienta las buenas prácticas relacionadas con el uso de test: fiabilidad, validez y uso ético. En este trabajo, repasamos el desarrollo de las tecnologías de la información y comunicación, mostrando su impacto en la evaluación psicológica y educativa, con el objetivo de exponer el abanico de innovaciones que marcarán el desarrollo de la investigación y la práctica profesional los próximos años.
Palabras clave: *TIC, Tecnología, Test, Evaluación.*

Digital technology is already part of the practices related to test use as well as psychological and educational assessment; test design, data collection, and psychometric models have all evolved. Consequently, we face a number of challenges such as professional training, collaboration with other areas, expansion of the testing concept, and deliberation on ethical and legal issues. In this dynamic environment, the experience and knowledge related to psychological measurement give us the capacity and opportunity to integrate and expand the pillars on which good test use practices are based: reliability, validity, and ethical use. In this paper, we review the development of the information and communication technologies, showing their impact on psychological and educational assessment, with the aim of exposing some innovations that will mark the development of research and professional practice in the coming years.
Key words: *ICT, Technology, Test, Assessment.*

The impact of information and communication technologies (ICT) on test construction and test use is transforming the assessment environment. The test (as a measurement instrument for the use of the professional, in any of the fields of psychology, education, and health) and the assessment (as a complex activity that integrates knowledge, clinical judgment, reliable information, and psychometric measures (American Psychological Association (APA, 2020)) have opened up to the possibilities offered by ICT. Interactive virtual scenarios, new psychometric models, the use of mobile devices, the collection of process data (log data) or the high volume of data accessible from different sources (big data) have modified the classic static model of measurement with paper/pencil tests. Technology provides the possibility to apply tests in digital formats as well as to score tests in an automated way, but the technology has also been accompanied by new constructs related to competencies in the use of ICT and new ways of measuring soft skills such as leadership, communication, critical thinking, collaborative learning, or teamwork, which are increasingly relevant in the educational, occupational, and social environments (Grundke et al., 2017; OECD, 2016).

The effect of technological development on assessment can be described as a progressive process of assimilation that affects different areas related to the use of tests. From the usage of a computer as a mere support for the administration of a pencil/paper test (Elosua, 2021), to the integration of technology in substantive models that facilitate the measurement of new variables, or the new paradigm of data-driven research, digital technology has impacted all the phases that accompany the construction and use of tests. In the field of educational assessment, for example, three stages have been identified that mark the impact of ICT (Bennet, 2015); first, the infrastructure is developed and technology is used to administer traditional tests with a new support; on this infrastructure and taking advantage of the benefits of computer use, adaptive tests are constructed (Wainer et al., 2000; Weiss, 1982), new item formats appear (multimedia, short constructed responses, static performance tasks such as essays), and the automatic generation of items, or the automation of test correction; in the most advanced stage, technology is integrated in such a way in the assessment process that it forms part of the design of tests based on cognitive models that, making use of simulations and interactive tasks, reproduce "real" environments that facilitate authentic assessment.

In this paper we review the ICT-related innovations that are having the greatest impact in the field of test-related assessment. The objective is to point out the aspects that, from our perspective, are most relevant in the construction and use of tests. With this aim, and in order to offer a general overview, we begin with some brief notes to define and locate the elements of the ICT field that have had the greatest influence on the current situation.

Received: 1 November 2021 - Accepted: 21 December 2021
Correspondence: Paula Elosua. *Universidad del País Vasco. Avda. de Tolosa, 70. 20018 San Sebastián. España.*
E-mail: paula.elosua@ehu.es
....
This work has been partially funded by the Ministry of Science and Innovation (PID2019-103859RB-100).

DEVELOPMENT OF DIGITAL TECHNOLOGY

One of the most salient features of the third industrial revolution, which has given rise to the information society (Rifkin, 2011), has been the development of digital technology. The major milestones are related to the emergence and development of elements such as personal computers or the Internet, which have transformed the ways of human communication.

Personal computers

The idea of the personal computer emerged about 50 years ago; companies such as IBM, HP, and Apple introduced their first machines in the 1980s; the IBM personal computer appeared in 1981, HP marketed its first personal computer in 1980, and The Apple Macintosh was presented in an advertisement during the Super Bowl telecast in 1984. The appearance of the personal computer placed computing power directly in the hands of millions of people, thereby changing the previous paradigm in which individual user access privileges were tightly controlled by system administrators. The subsequent introduction of client-server architecture further led to the linking of personal computers (clients) to computers (servers) that allowed the sharing of large amounts of data (O'Regan, 2021).

The Internet and the WWW

The formal beginning of the Internet is usually placed in the 1960s, when the United States Department of Defense set up the ARPANET (Advanced Research Project Agency) network connecting four American research centers. This first network, together with the invention of "packet switching", which made it possible to increase the speed and effectiveness of teleprocessing, and the appearance of the ASCII code (American Standard Code for Information Interchange), created by a committee of the American Standards Association (ASA) in 1963, made it possible to connect around 30 institutions in the 1970s.

The Internet and the World Wide Web (WWW) are not synonymous. The appearance of the WWW, conceived by Tim Berners-Lee, was established at the *Conseil Européen pour la Recherche Nucléaire* (EONR, European Organization for Nuclear Research) in 1990. The WWW builds on earlier advances related to the development of hypertext, the mouse, and the Internet. Berners-Lee created a system that assigned each web page a standard address, the URL (universal resource locator)—which was accessed via HTTP (hypertext transfer protocol)—and programmed a browser that allowed web pages to be requested, retrieved, and displayed on the local PC between computers connected via the Internet. The invention of the World Wide Web was a revolutionary milestone. It transformed the Internet from a primarily academic use to where it is now an integral part of people's lives. Users can surf the web, hyperlink between millions of computers, and share information quickly and easily. According to Internet World Stats (www.https://internetworldstats.com) it is estimated that Internet penetration today is 65.6%; that is, out of an estimated world population of 7,8751765,587 people, 5,1681780,607 use the internet.

TABLE 1
DIGITAL TECHNOLOGY

Year	Reference
1980	Personal computers
1993	World Wide Web (free)
2003-2006	LinkedIn, Facebook, Twitter
2005	Big data
2013	Internet of things
2020	Big data + Artificial intelligence

Social networks

The ubiquity of smartphones, or simply cell phones, has driven the growth of social networks, which in turn have transformed the forms of human communication. Social networks are applications that allow the creation and exchange of user-generated content; they can be horizontal or generalist and vertical or specialized depending on whether they have defined themes or activities; LinkedIn or Researchgate, for example, would be vertical professional and research networks.

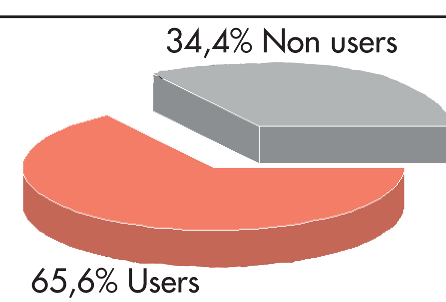
Analyzing the type and quality of information shared in social networks, these could be classified by combining theoretical models from the field of networks with models on social processes (Kaplan & Haenlein, 2010); one axis would be defined by the level of social presence and the richness of the medium, that is, by the amount of information that can be transmitted in a time interval, and the second would be defined by the self-representation or desire to control the impressions made on the other, and the conscious or unconscious disclosure of personal information (self-disclosure).

According to data published by eBizMBA in October 2021, Facebook has 2,200,000,000 different users per month, Youtube has 1,850,000,000, Twitter has 375,000.000, and LinkedIn has 85,000,000 (http://www.ebizmba.com/articles/social-networking-websites).

Artificial intelligence

At the boundary between the third and fourth industrial revolution, defined as the fastest revolution in the field of technology, and not comparable to anything in the history of mankind in terms of impact (Lee, 2018), artificial intelligence (AI) has become an object of study and analysis in scientific, technical, political, and social fields. AI has drawn from disciplines as varied as philosophy, mathematics, economics, neuroscience, psychology, computer engineering, cybernetics, and linguistics. Today we find AI applications in search engines, medical diagnostics, voice recognition, robot control, web search, advertising, and even toys.

FIGURE 1
INTERNET PENETRATION



Note. Data extracted from www.https://internetworldstats.com as of November 2021

Mentioning artificial intelligence inevitably leads to citing the mathematician Alan Turing, whose work in the early 1950s contributed to the debate on thinking machines, consciousness, and intelligence. He created the famous Turing test to assess the consciousness of a machine; in this experiment a judge evaluates a conversation in natural language between two parties, a human and a machine, with the purpose of discerning which is the machine. In this line of work, in 1966, the German Joseph Weizenbaum programmed the ELIZA program at the Massachusetts Institute of Technology (MIT); the ELIZA program is a bot that emulates a Rogerian psychotherapist and interacts with a person sitting in front of a typewriter (similar to an online chat). The software operates by breaking down the user's input into its constituent voice parts and then retyping them in a way that appears to be a flowing dialogue. The author was surprised to find that many users thought the program had real comprehension. This observation led him to reflect on the ethics and implications of the field of artificial intelligence (Weizenbaum, 1976), an issue that permeates the development of AI. The reader can "converse" with a Spanish version of ELIZA at http://deixilabs.com/eliza.html.

The term artificial intelligence appears in a Dartmouth summer research project written by computer scientist John McCarthy in 1955. Today, one of the most accepted definitions of AI is the one proposed by Russell and Norvig (2021), according to which AI focuses on the study and construction of agents that do the right thing—the right thing being the goal set for the agent—and defining agent as something that perceives its environment through sensors. In simple statistical terms the right thing to do could be the decision (estimate) that minimizes the loss function. This definition has been accepted by the European Union, which reformulates it as: "Software that is developed using one or more of the techniques and strategies listed in Annex I[1] and that can, for a given set of human-defined objectives, generate output information such as content, predictions, recommendations, or decisions that influence the environments with which it interacts" (European Commission, 2021).

The field of AI is broad, and we find specialized areas in computing, machine learning, natural language processing, computer vision, and robotics. In addition, and given the implications of AI, the area dedicated to the ethical and legal aspects related to its implementation is a constant.

The data

The availability of information from social networks or from devices, sensors, and services that capture data around them (the internet of things) has marked a paradigm shift in the research, the market, and the industry. The origin of the term "big data" dates back to 2005 and is attributed to Roger Mougalas. Today big data refers to information characterized by its Volume, Velocity, and Variety, and which requires specific analytical methods for its treatment (De Mauro et al., 2016; Zicari, 2013). The three Vs characteristics of big data have been expanding with more nouns that attempt to reflect its nature more faithfully; thus, we find determinants such as Value, Veracity, Visualization, Volatility, Validity, and Viability (Maté-Jimenez, 2014). Big data basically operates with machine learning models that search for patterns and relationships for classification and prediction; it is a data-centric, exploratory approach to research that applies analytical techniques in the search for patterns in the data. In contrast, the traditional, deductive, theory-centered approach is associated with the formulation and testing of hypotheses on a data sample. Between the two perspectives, which may be on a continuum, there is a growing position that advocates complementarity for the advancement of knowledge and scientific productivity (Maass et al., 2018).

TABLE 2
CLASSIFICATION OF SOCIAL NETWORKS

			Social dimension		
			Social presence / Richness of the environment		
			Low	Medium	High
Network dimension	Self-representation	High	Blogs *Twitter*	Social networks Facebook LinkedIn	Virtual spaces *Second life*
	Personal information	Low	Collaborative projects Wikipedia BookCrossing Flick *Slideshare*	ContentsYoutube	Games World of Warcraft

Note. Adapted from Kaplan and Haenlein (2010)

[1] Annex I. Machine learning strategies, including supervised, unsupervised, and reinforcement learning, that employ a wide variety of methods, including deep learning. Logic and knowledge-based strategies, especially knowledge representation, inductive programming (logic), knowledge bases, inference and deduction engines, expert and (symbolic) reasoning systems. Statistical strategies, Bayesian estimation, search and optimization methods.

IMPACT OF ICT ON TESTS AS ASSESSMENT TOOLS

The technological innovations described in the previous points and some others that we cannot cover due to space limitations (computational power, robotics, software development, etc.) have impacted the whole process related to the use of tests; item construction, test administration, test scoring, the collection of additional data, the psychometric models to manage them, the use of data as a source of information, etc. are areas of work and research that adapt the traditional test and the practices of the twentieth century to the social demands of the twenty-first century.

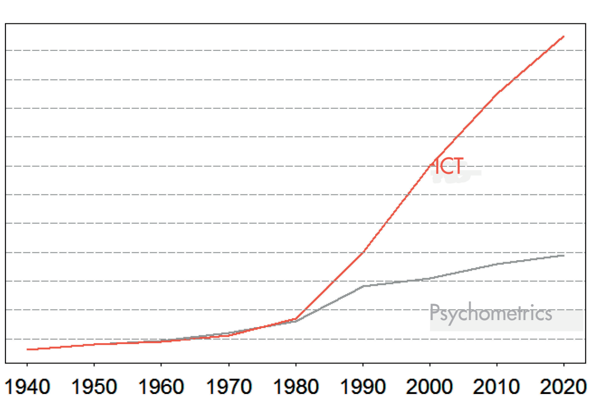
Adaptive tests

The development of computerized adaptive tests (CAT) and the associated psychometric theory built on the basis of item response theory have facilitated the construction of personalized tests and the comparability of scores between people who receive different items (Hambleton, 2006). The adaptive test is a system composed of a bank of items with known psychometric characteristics, and selection algorithms that—depending on the level of ability estimated after each response—choose the item stored in the bank which, due to its properties (parameters), will offer the maximum information about the level of competence of the person being assessed (Olea et al., 2010; van der Linden & Glas, 2000).

Automated item generation

Automated item construction or automated item generation (AIG) is presented as a solution to the growing demand of a market characterized by the implementation of computerized adaptive tests, the application of tests via the Internet, the transparency required of assessment projects, and the increase in educational assessment programs. The application of items, especially in the educational field, has gone from being considered an occasional activity to an ad hoc activity that is accompanied by the requirement to have a large number of items. In view of this need, the handcrafted construction that focuses on a single item that is constructed, reviewed, edited, and calibrated until it reaches the required quality standards is not an efficient process.

FIGURE 2
FIGURATIVE REPRESENTATION OF THE EVOLUTION OF INFORMATION TECHNOLOGIES AND PSYCHOMETRICS



Automatic item generation combines cognitive and psychometric theories that make it possible to construct items with pre-established psychometric properties from a model. In this framework, the unit of analysis is an item model (Gierl et al., 2020) or a cognitive model (Embretson & Yang, 2006). AIG can be represented as a 3-stage process: 1) the content that will serve as the basis for item generation is identified; 2) the item model is constructed; and 3) the algorithms that will generate the items from phases 1 and 2 are programmed. Several item generation models can be consulted in Bejar et. al. (2003), Case and Swanson (2002), or Gierl and Lai (2013).

Computerized test application

One of the great advantages of the application of computer-based tests is the possibility to collect process data referring to the interaction of the person being tested with the testing environment. These data, log data, can be used to reconstruct specific behaviors, delve into learning theories, study differences between groups and, in short, extend validation studies. The information on the time spent on each element can be useful, for example, to analyze motivation, fatigue, or speed of execution; in addition, the computerized application allows the collection of multimodal data, multivariate process data such as facial movements, eye movements, audio, magnetic resonance imaging, or computed tomography, which, although they still present difficulties of analysis, will be the subject of research in the coming years (Guidry et al., 2013; Ramalingam & Adams, 2018; Scherer et al., 2015; in this monograph, Suarez et al., 2022).

New item formats

The improvement in the capabilities of personal computers together with the development of the Internet and the availability of mobile devices allows the application of tests/items in formats other than the classic pencil-and-paper format. The new item formats, which are enriched by the use of video or animation, overcome some of the disadvantages associated with items of choice and thus allow the measurement of aspects that are difficult to achieve with the multiple-choice format or its variants.

Suffice it to cite as an example of the expansion and common use of new item formats that the international assessment program PISA (Programme for International Student Assessment), managed by the OECD, introduced computerized administration in 2006; since the 2015 edition, PISA has been designed and implemented digitally (the reader can see examples of PISA items at https://www.oecd.org/pisa/test/).

The animation included in an item can be two-dimensional, simple three-dimensional, photorealistic three-dimensional, or virtual reality (Popp et al., 2016). Virtual reality is a simulation, derived from the gaming industry, that creates the sensation of physical presence (Linowes, 2015; Parisi, 2015). In principle, these items provide for a higher face validity, are accompanied by dynamic and interactive presentations that can generate better acceptance and interest towards the test situation by the individual being tested, and favor the assessment of complex tasks (Bruk-Lee et al., 2013; Ryall et al., 2016). However, multimedia formats could also introduce unwanted factors into the assessment process; for example, the sociodemographic characterization of the avatar is a salient feature that may trigger attitudes or affective experiences that interact in unintended and undesired ways (Sadler et al., 2012). There are several criticisms of this type of item due to its lack of validity in the educational field (Young et al., 2012), although in medicine it has a long tradition and prestige (McGaghie et al., 2010).

Automated essay correction

The appearance in the 1930s of optical readers increased the efficiency of the correction process, and with it the volume of response selection tests; however, the implementation and increasing application of production items (essays, open-ended items) has prompted the search for efficient correction methods for this task format. Automated essay correction began in the 1960s (Page, 1968), and with the sophistication of natural language processing and machine learning, automated essay correction is now accepted as part of educational practice. Several studies have compared it with judge-mediated examination and have shown its effectiveness in the assessment of different correction criteria (Williamson et al., 2012). The most powerful companies related to educational assessment have software designed for this purpose (Pearson's Test of English; ETS Criterion; Accuplacer by College Board, etc.).

The idea behind automated essay correction is that the system is able to convert the student's production (oral or written) into a score (or several scores) or feedback, which is accurate, reliable, and aligned with the constructs to be assessed. The process relies on a large sample of essays representing the range of possible productions and scores. The essays are first corrected by human experts who provide the system with the necessary information to train it to estimate the true score. On that basis, the software learns to associate the distinctive features of each essay with the scores assigned by the experts. Once the model has been built, the system is able to predict the score that the experts would assign to a new trial.

Psychometric modeling

As digital technology permeates the way tests are constructed, administered, and corrected, psychometric models for data analysis and estimation of true scores and measurement errors are adapting to the demands of new environments. The biggest challenges they face could be modeling multidimensionality, and adapting the methodology built for use in controlled and structured environments for use in dynamic environments and with data with a lower level of structuring.

Bifactor models

Bifactor models model the multidimensionality present in many constructs analyzed by psychology and are applied when there is a general factor and specific factors or group factors. Although bifactor models or nested models were proposed in 1937 (Holzinger & Swineford), their application has been extended in the last decade (Rodriguez et al., 2016). Basically, the bifactor model and the second-order factor model could have similar interpretations (Chen et al., 2006), but the former becomes relevant when the interest is focused on the group factors, when the objective is to analyze the relationship between these and the items that comprise them, or when the aim is to study the predictive capacity of the partial scales in depth.

Testlet models

Testlet-based IRT models are formally bifactor models; the difference between the two comes from the tradition of their use, which is linked to factor models and IRT models, or to fields of specialization in psychology or education. The testlet is constructed because of the need for parameter estimation under conditions of violation of local independence associated with the application of context-dependent groups of items (Bradlow et al., 1999; Wainer et al., 2007). In a testlet each item is an indicator of a general dimension, and of a dimension associated with a group of items. The general dimension represents the latent variable of central interest (e.g., reading proficiency) while the rest are incorporated to account for additional dependencies between items belonging to the same subgroup.

Modeling of forced-choice items

As opposed to the commonly known Likert item in which a person responds to a question on an ordered response scale, forced-choice items or ipsative items force a choice between two or more statements that can be ordered according to the preferences shown (in this monograph, Abad et al., 2022). This is intended to control response bias (acquiescence, social desirability, central tendency, severity, etc.) and to improve the assessment process (Brown & Maydeu-Olivares, 2011, 2018; Chan, 2003).

Network analysis

Network analysis in psychology is an alternative to the traditional psychometric view that relates a construct to several indicators of which it is the cause. In network modeling, indicators are perceived as proxies of variables that interact with each other; for example, from the traditional perspective, symptoms such as lack of energy, sleep problems, or low self-esteem are caused by depression, whereas in the network model these symptoms constitute a network of mutual interaction. This involves a different approach to the modeling and study of psychological phenomena (Epskamp et al., 2018; Fonseca-Pedrero, 2018).

Big data. Social networks, wearables, and mobile devices

The social science research tradition built on Cattell's (1966) representation of the data cube as an ordered set consisting of three axes (cases, variables, and temporal moments) has given way to massive, unstructured, high-dimensional data. This irruption, in a way, questions the traditional concept of the test as the basic unit for collecting information on behaviors, attitudes, or beliefs. Social networks such as Facebook, LinkedIn, or Twitter are continuous sources of data that are giving rise to a new line of research in psychology (in this monograph, Andrés et al., 2022); depression, suicidal ideation, personality, and personnel selection, for example, are being analyzed through the analysis of information taken from social networks (Conway & O'Connor, 2016; Dwyer et al., 2018; Skaik & Inkpen, 2020; Woo et al., 2020). But in addition, the internet of things or the use of mobile devices opens the possibility of using methodologies based on experience sampling methodology (ESM; Myin-Germeys et al., 2018; Stieger et al., 2018) or ambulatory assessment (in this monograph, Fonseca-Pedrero et al., 2022), which bring an ecological perspective to psychological assessment.

Ethical aspects

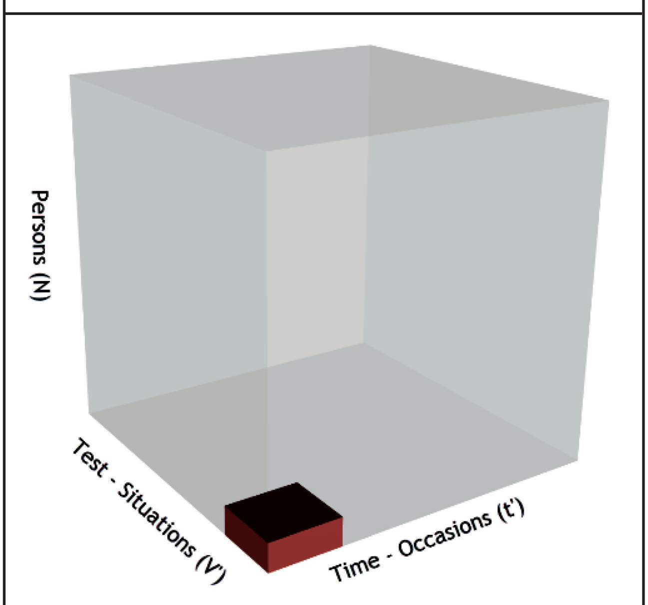
Parallel to the increase in AI applications, a debate is developing on the ethical, legal, and social implications. Many national and international organizations have created ad hoc committees of experts to prepare documents and guidelines on AI. In Spain, in July 2020, the Advisory Council on artificial intelligence was constituted, the European Community published in 2018 the coordinated plan on artificial intelligence in which the watchdog role is given to *AI Watch*, and the EC is currently working on its regulation. The constitution of such committees is a clear indicator of the technological, economic, political, and social impact of new technologies.

In a review of guidelines and standards in which 84 documents are analyzed, it is concluded that the problems most discussed in the reports that study the ethical issues involved with AI are related to transparency, fairness, equity, and bias (Jobin, Ienca, & Vayena, 2019). These are concepts that—on the other hand—have been, and are, the subject of standards on test construction and use, with which we are in permanent contact (in this monograph, Hernández et al., 2022).

DISCUSSION

The foundations of current theoretical psychometrics were built in the second half of the 20th century; the classical test theory model, the formulation of the factorial model, and the first proposals in item response theory also known as "new psychometrics" correspond to that time (Lord & Novick, 1968; van der Linden & Hambleton, 1997). Since then, the impulse of digital technology and the accessibility and power of software and hardware have enabled the widespread use of these psychometric models, thus shortening the gap between theoretical and applied psychometrics (Elosua, 2012).

FIGURE 3
CATTELL'S DATA BOX AND BIG DATA



But innovation is not only focused on the socialization of the confirmatory model, which is also important for strengthening validation studies. Along with it, the concept and use of tests has been expanded and enriched; assessment environments are now diverse, complex, and dynamic. Several authors claim that we are in the midst of the fourth industrial revolution (Schwab, 2017); this is a revolution characterized by big data, cloud computing, and the internet of things. If the third industrial revolution was associated with the scientific explosion and information technology, the fourth is a development of the previous one. The penetration of digital technology in assessment has brought us closer to a new territory in which contributions from areas of knowledge such as engineering, computational linguistics, computer science, and artificial intelligence are opening new fields of exploration, and questioning the traditional concept of testing; in addition, "data" has become part of our lives. This new reference coming from applications, mobile devices, and social networks is making it possible to analyze behaviors and make classifications and predictions. It is true that the procedures for obtaining data are different, but the two approaches have similarities in their objectives. The boundary between the test and the data as instruments that, based on the analysis of information, facilitate decision making is diluted. The question, or questions, in view of this situation, are several: are psychologists prepared to undertake this task? Do the more technological profiles (engineers, computer scientists, etc.) have sufficient substantive background to tackle it? Coordination between areas, psychologist training in languages such as R or Python (Elosua, 2009, 2011), and teamwork (Adjerid & Kelley, 2018; König et al., 2020; Oswald, 2020) may be useful means that allow us to face the diversity and dynamism of this time. The experience gained by psychometrics in the field of measurement, the understanding of the psychological, and the clear awareness about problems related to validity and bias are values that psychology brings, and from which big data can clearly benefit.

But not everything is big data; we have presented a range of current lines of work to which the professional or academic may have several reactions; one of them is to feel that the innovations presented are alien to their usual work. The paper/pencil test and Likert scales are still omnipresent in Spanish psychology; test construction (Muñiz & Fonseca-Pedrero, 2019) and adaptation (Muñiz et al., 2013) are known and fertile territories for academia, and continue to contribute knowledge to psychology. Digital technology, however, demands something more. Many of the innovations are studied, analyzed, and implemented in companies that are dedicated to the market related to the construction and use of tests, especially in the educational and organizational fields. Studies on test use have been pointing out the difference between sectors with respect to attitudes and use of technology in assessment (Muñiz et al., 2020). This difference becomes more acute when we deal with digital technologies (in this monograph, Santamaría & Sánchez-Sánchez, 2022).

The traditional paper/pencil test is still alive, but today it coexists with technological developments that have created an environment in which the virtual and digital are increasing in importance and attraction. The test emerges as an instrument of social support, and therein lies its survival. From the point of view of assessment, the digital era creates an environment shared by psychology, education, engineering, and data science, in which we are obliged to participate proactively.

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

Abad, F. J., Schames, R., Sorrel, M., Nájera, P., García-Garzón, E., Garrido, L. E., & Jiménez, M. (2022). Construyendo tests adaptativos de elección forzosa “on the fly” para la medición de la personalidad [Building Adaptive Forced-Choice Tests “On The Fly” For Personality Measurement]. *Papeles del Psicólogo, 43*(1), 29-35. https://doi.org/10.23923/pap.psicol.2982

Adjerid, I. & Kelley, K. (2018). Big data in psychology: A framework for research advancement. *The American Psychologist, 73*(7), 899-917. https://doi.org/10.1037/amp0000190

American Psychological Association. (2020). *APA guidelines for psychological assessment and evaluation*. Downloaded from https://www.apa.org/about/policy/guidelines-psychological-assessment-evaluation.pdf

Andrés, J. C., Aguado, D., & de Miguel, J. (2022). ¿Qué hay detras de LinkedIn? Midiendo a través de rúbricas las LinkedIn Big Four Dimensions [What’s behind LinkedIn? Measuring the LinkedIn Big Four Dimensions through Rubrics]. *Papeles del Psicólogo, 43*(1), 12-20. https://doi.org/10.23923/pap.psicol.2979

Bejar, I. I., Lawless, R., Morley, M. E., Wagner, M. E., Bennet, R. E., & Revuelta, J. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *Journal of Technology, Learning and Assessment, 2*(3), 1-29.

Bennett, R. E. (2015). The changing nature of educational assessment. *Review of Research in Education, 39*(1), 370-407. https://doi.org/10.3102/0091732X14554179

Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*(2), 153-168. https://doi.org/10.1007/bf02294533

Brown, A. & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*, 460-502. https://doi.org/10.1177/0013164410375112

Brown, A. & Maydeu-Olivares, A. (2018). Modelling Forced-Choice Response Formats. In P. Irwing, T. Booth, & D. J. Hughes. (Eds.), *The Wiley Handbook of Psychometric Testing* (pp. 523-569). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118489772.ch18

Bruk-Lee, V., Drew, E. N., & Hawkes, B. (2013). Candidate reactions to simulations and media-rich assessments in personnel selection. In M. Fetzer & K. Tuzinski (Eds.), *Simulations for personnel selection* (pp. 43-60). Springer New York. https://doi.org/10.1007/978-1-4614-7681-8_3

Case, S. M. & Swanson, D. B. (2002). *Constructing written test questions for the basic and clinical sciences*. National Board of Medical Examiners.

Cattell, R. B. (1966). Data box: Its ordering of total resources in terms of possible relational systems. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 67-128). Rand-McNally.

Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika, 30*(1), 99-121. https://doi.org/10.2333/bhmk.30.99

Chen, F. F., West, S., & Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189-225. https://doi.org/10.1207/s15327906mbr4102_5

Comisión Europea [European Commission]. (2021). *Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial [Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules in the field of artificial intelligence]*.

Conway, M. & O’Connor, D. (2016). *Social media, big data, and mental health: Current advances and ethical implications*. https://doi.org/10.1016/j.copsyc.2016.01.004

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review, 65*, 122-135.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology, 14*(1), 91-118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Elosua, P. (2009). ¿Existe vida más allá del SPSS? Descubre R [Is there life beyond SPSS? Discover R]. *Psicothema, 21*, 652-655.

Elosua, P. (2011). *Introducción al entorno R [Introduction to the R environment]*. Universidad del País Vasco. https://www.researchgate.net/publication/264697893_Introduccion_al_entorno_R

Elosua, P. (2012). Tests publicados en España: usos, costumbres y asignaturas pendientes [Tests published in Spain: Uses, customs, and unresolved matters]. *Papeles del Psicólogo, 33*, 12-21.

Elosua, P. (2021). Aplicación remota de test: riesgos y recomendaciones [The remote application of tests: risks and recommendations]. *Papeles del Psicólogo, 41*(2), 33-39. https://doi.org/10.23923/pap.psicol2021.2952

Embretson, S. & Yang, X. (2006). 23 Automatic item generation and cognitive psychology. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. 26, pp. 747-768). Elsevier. https://doi.org/10.1016/S0169-7161(06)26023-1

Epskamp, S., Maris, G., Waldorp, J., & Borsboom, D. (2018). Network psychometrics. In P. Irwing, T. Booth, & D. J. Hughes. (Eds.), *The Wiley Handbook of Psychometric Testing* (pp. 953-986). John Wiley & Sons. https://doi.org/10.1002/9781118489772.ch30

Fonseca-Pedrero, E. (2018). Análisis de redes en psicología [Network analysis in psychology]. *Papeles del Psicólogo, 39*, 1-12. https://doi.org/10.23923/pap.psicol2018.2852

Fonseca-Pedrero, E., Ródenas, Gabriel, Pérez-Albéniz, A., Al-Halabí, S., Pérez, M., & Muñiz, J. (2022). La hora de la evaluación ambulatoria [The Time of Ambulatory Assessment]. *Papeles del Psicólogo, 43*(1), 21-28, https://doi.org/10.23923/pap.psicol.2983

Gierl, M. J. & Lai, H. (2013). Instructional topics in educational measurement (ITEMS) module: using automated processes to generate test items. *Educational Measurement: Issues and Practice, 32*(3), 36-50. https://doi.org/10.1111/emip.12018

Gierl, M. J., Lai, H., & Matovinovic, D. (2020). Augmented intelligence and the future of item development. In Hong Jiao & Robert W. Lissitz (Eds.), *Application of artificial intelligence to assessment* (pp. 1-25). Information Age Publishing-IAP.

Grundke, R., Squicciarini, M., Jamet, S., & Kalamova, M. (2017). *Having the right mix: The role of skill bundles for comparative*

advantage and industry performance in GVCs. OECD Science, Technology and Industry Working Papers, OECD Publishing.
Guidry, B. W., Rupp, D. E., & Lanik, M. (2013). Tracing cognition with assessment center simulations; Using technology to see in the dark. In M. Fettzer & Tuzinski (Eds.), *Simulations for personal selection* (pp. 231-257). Springer Science+Business Media.
Hambleton, R. K. (2006). Psychometric models, test designs and item types for the next generation of educational and psychological tests. In D. Bartram & R. K. Hambleton (Eds.), *Computer-based testing and the internet: Issues and advances* (pp. 77-90). John Wiley & Sons Ltd.
Hernández, A., Elosua, P., Fernández-Hermida, J. R., & Muñiz, J. (2022). Comisión de Test: Veinticinco años velando por la calidad de los test [Test Commission: Twenty-five years working on test quality]. Papeles del Psicólogo, 43(1), 55-62. https://doi.org/10.23923/pap.psicol.2978
Holzinger, K. J. & Swineford, S. (1937). The bi-factor method. *Psychometrika, 47*,41-54.
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389-399. https://doi.org/10.1038/s42256-019-0088-2
Kaplan, A. M. & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons, 53*(1), 59-68. https://doi.org/10.1016/j.bushor.2009.09.003
König, C., Demetriou, A., Glock, P., Hiemstra, A., Iliescu, D., Ionescu, C., Langer, M., Liem, C., Linnenbürger, A., Siegel, R., & Vartholomaios, I. (2020). Some advice for psychologists who want to work with computer scientists on big data. *Personnel Assessment and Decisions*, 17-23. https://doi.org/10.25035/pad.2020.01.002
Lee, K. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin.
Linowes, J. (2015). *Unity virtual reality: Explore the world of virtual reality by building immersive and fun VR projects using unity 3D*. Packt Publishing.
Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley
Maass, W., Parsons, J., Purao, S., Storey, V. C., & Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: Opportunities and challenges for information systems research. *Journal of the Association for Information Systems, 19*(12), https://doi.org/ 10.17705/1jais.00526
Maté-Jimenez, C. (2014). Big data. Un nuevo paradigma de análisis de datos [Big data. A new paradigm for data analysis]. *Anales de mecánica y electricidad, 5*. https://revista-anales.icai.es/web/n_29/seccion_1.html
McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education, 44*, 50-63.
Muñiz, J., Elosua, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: Segunda edición [International Test Commission Guidelines for test translation and adaptation: Second edition]. *Psicothema, 25.2*, 151-157. https://doi.org/10.7334/psicothema2013.24
Muñiz, J. & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test [Ten steps for test development]. *Psicothema, 31*, 7-16. https://doi.org/10.7334/psicothema2018.291
Muñiz, J., Hernández, A., & Fernández-Hermida, J. R. (2020). Utilización de los test en España: el punto de vista de los psicólogos [Test use in Spain: the psychologists' viewpoint]. *Papeles del Psicólogo, 41*(1), 1-15. https://doi.org/10.23923/pap.psicol2020.2921
Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry, 17*(2), 123-132. https://doi.org/10.1002/wps.20513
OECD. (2016). *Skills for a digital world*. OECD/Directorate for Science, Technology.
Olea, J., Abad, F. J., & Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests [Computerized tests and other new types of test]. *Papeles del Psicólogo, 31*(1), 94-107.
O'Regan, G. (2021). *A brief history of computing* (3rd edition). Springer.
Oswald, F. L. (2020). Future research directions for big data in psychology. In S. E. Woo, L. Tay, & R. W. Proctor (Eds.), *Big data in psychological research*. (pp. 427-441). American Psychological Association. https://doi.org/10.1037/0000193-020
Page, E. B. (1968). The use of the computer in analyzing student essays. *International Review of Education 14*, 210–225. https://doi.org/10.1007/BF01419938
Parisi, T. (2015). *Learning virtual reality: Developing immersive experiences and applications for desktop, web, and mobile*. O'Reilly Media.
Popp, E. C., Tuzinski, K., & Fetzer, M. (2016). Actor or avatar? Considerations in selecting appropriate formats for assessment content. In E. F. Drasgow (Ed.), *Improving educational and psychological measurement* (pp. 79-103). Routlege.
Ramalingam, D. & Adams, R. J. (2018). How can the use of data from computer-delivered assessments improve the measurement of twenty-first century skills? In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and teaching of 21st Century Skills: Research and Applications* (pp. 225-238). Springer International Publishing. https://doi.org/10.1007/978-3-319-65368-6_13
Rifkin, J. (2011). *The third industrial revolution: how lateral power is transforming energy, the economy, and the world*. Palgrave Macmillan.
Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological methods, 21*(2), 137-150. https://doi.org/10.1037/met0000045
Russell, S. & Norvig, P. (2021). Artificial intelligence: a modern approach. In *Series in Artificial Intelligence* (4th edition). Prentice-Hall.
Ryall, T., Judd, B. K., & Gordon, C. J. (2016). Simulation-based assessments in health professional education: A systematic review. *Journal of Multidisciplinary Healthcare, 9*, 69-82. https://doi.org/10.2147/JMDH.S92695
Sadler, M. S., Correll, J., Park, B., & Judd, C. M. (2012). The world is not black and white: Racial bias in the decision to shoot in a multiethnic context. *Journal of Social Issues, 68*, 286-313. https://doi.org/10.1111/j.1540-4560.2012.01749.x
Santamaría, P. & Sánchez-Sánchez, F. (2022). Cuestiones abiertas en el uso de las nuevas tecnologías en la evaluación psicológica [Open questions in the use of new technologies in psychological

assessment]. *Papeles del Psicólogo, 43*(1), 48-54. https://doi.org/10.23923/pap.psicol.2984

Scherer, R., Greiff, S., & Hautamäki, J. (2015). Exploring the relation between time on task and ability in complex problem solving. *Intelligence, 48*, 37-50. https://doi.org/10.1016/j.intell.2014.10.003

Schwab, K. (2017). *The Fourth Industrial Revolution*. Crown Publishing Group.

Skaik, R. & Inkpen, D. (2020). Using social media for mental health surveillance: a review. *ACM Computing Surveys, 53*(6), 129:1-129:31. https://doi.org/10.1145/3422824

Stieger, S., Lewetz, D., & Reips, U. (2018). Can smartphones be used to bring computer-based tasks from the lab to the field? A mobile experience-sampling method study about the pace of life. *Behavior Research Methods, 50*(6), 2267-2275. https://doi.org/10.3758/s13428-017-0991-6

Suárez-Álvarez, J., Fernández-Alonso, R., García-Crespo, F. J., & Muñiz, J. (2022). El uso de las nuevas tecnologías en las evaluaciones educativas: La Lectura en un mundo digital [The use of New Technologies in Educational Assessments: Reading in a digital world]. *Papeles del Psicólogo, 43*(1), 36-47. https://doi.org/10.23923/pap.psicol.2986

van der Linden, W. J. & Glas, G. A. W. (Eds.). (2000). *Computerized Adaptive Testing: Theory and Practice*. Kluwer Academic Publishers.

van der Linden, W. J. & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. Springer.

Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge University Press.

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., & Mislevy, R. J. (Eds.). (2000). *Computerized Adaptive Testing: A Primer* (2.ª ed.). Routledge. https://doi.org/10.4324/9781410605931

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement, 6*, 473-492.

Weizenbaum, J. (1976). *Power and human reason: from judgments to calculation*. W.H. Freeman & Co Ltd.

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice, 31*(1), 2-13. https://doi.org/10.1111/j.1745-3992.2011.00223.x

Woo, S. E., Tay, L., & Proctor, R. W. (Eds.). (2020). *Big data in psychological research*. American Psychological Association.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our princess is in another castle: a review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61-89. https://doi.org/10.3102/0034654312436980

Zicari, R. V. (2013). *Big data: Challenges and Opportunities*. Champan and Hall/CRC.