

Article

New Trends in Digital Technology-Based Psychological and Educational Assessment

Paula Elosua¹, David Aguado², Eduardo Fonseca-Pedrero³, Francisco José Abad² and Pablo Santamaría⁴

1 Universidad del País Vasco.

2 Universidad Autónoma de Madrid.

3 Universidad de La Rioja.

4 Hogrefe TEA Ediciones.

ARTICLE INFO

Received: May 18, 2022

Accepted: September 25, 2022

Keywords:

ICT
Technology
Test
Assessment
Measurement

ABSTRACT

Background: The emergence of digital technology in the field of psychological and educational measurement and assessment broadens the traditional concept of pencil and paper tests. New assessment models built on the proliferation of smartphones, social networks and software developments are opening up new horizons in the field. **Method:** This study is divided into four sections, each discussing the benefits and limitations of a specific type of technology-based assessment: ambulatory assessment, social networks, gamification and forced-choice testing. **Results:** The latest developments are clearly relevant in the field of psychological and educational measurement and assessment. Among other benefits, they bring greater ecological validity to the assessment process and eliminate the bias associated with retrospective assessment. **Conclusions:** Some of these new approaches point to a multidisciplinary scenario with a tradition which has yet to be created. Psychometrics must secure a place in this new world by contributing sound expertise in the measurement of psychological variables. The challenges and debates facing the field of psychology as it incorporates these new approaches are also discussed.

Nuevas Tendencias en Evaluación Psicológica y Educativa Apoyada en Tecnologías Digitales

RESUMEN

Antecedentes: La irrupción de la tecnología digital en las áreas de medición y evaluación psicológica y educativa expande el concepto clásico de test de lápiz y papel. Los modelos de evaluación construidos sobre la ubicuidad de los smartphones, las redes sociales o el desarrollo del software abren nuevas posibilidades para la evaluación. **Método:** El estudio se organiza en cuatro partes en cada una de las cuales se discuten las ventajas y limitaciones de una aplicación de la tecnología a la evaluación: la evaluación ambulatoria, las redes sociales, la gamificación y las pruebas de elección forzosa. **Resultados:** Los nuevos desarrollos resultan claramente relevantes en el ámbito de la medición y la evaluación psicológica y educativa. Entre otras ventajas, aportan una mayor validez ecológica al proceso evaluativo y eliminan el sesgo relacionado con la evaluación retrospectiva. **Conclusiones:** Algunas de estas nuevas aproximaciones llevan a un escenario multidisciplinar con una tradición aún por construir. La psicometría está obligada a integrarse en este nuevo espacio aportando una sólida experiencia en la medición de variables psicológicas. Se muestran los temas de debate y retos que ha de abordar el buen quehacer de la psicología en la incorporación de estas nuevas aproximaciones.

Palabras clave:

TIC
Tecnología
Test
Evaluación
Medición

Psychological and educational tests and assessments are being deeply impacted by digital technology. The new scenario brought about by information and communication technologies (ICTs) is broadening the traditional pen and paper testing and moving towards new assessment models and systems that are built in interactive virtual arenas and take advantage of the proliferation of mobile devices (e.g., laptops, tablets, and smartphones), the vast volume of data available from different sources (big data), computational power, and software development (Elosua, 2022). The concept and use of tests as assessment tools is being expanded and adapted to the new world defined by the fourth industrial revolution (Schwab, 2017) where big data, cloud computing, and the internet of things are fundamentally changing the way we live. The boundary between test and data as instruments that analyse information and facilitate decision-making is becoming blurred.

This emergence of data as a new concept somehow questions the traditional notion of psychological and educational test as the basic unit for collecting information on behaviours, attitudes, skills, knowledge or beliefs, among other variables. Furthermore, the internet of things and the widespread use of mobile devices enable the use of new methodologies, introducing an ecological perspective to the field of assessment.

In terms of evolution, three stages have been identified in the impact of ICTs on educational assessment (Bennett, 2015):

1. The use of technology exclusively for carrying out classical assessment.
2. The use of technology to support and improve test construction.
3. The integration of technology as part of the assessment process.

In the field of psychological assessment, these three stages also apply, but a fourth stage could also be added to include the implosion of data as a dynamic, available, ubiquitous, and diverse source of information. Social networks and devices or sensors that capture data have driven a paradigm shift in psychological research by allowing data to be analysed under classical and new paradigms in the study of psychological traits and disorders.

In this paper we offer a state-of-the-art picture of the new paradigms that are impacting psychological research in four specific areas. We focus on: (a) ambulatory assessment, (b) the use of social networks as a source of information for recruiting personnel, (c) gamification, and (d) psychometric modelling and test construction for controlling response bias. These innovations, among others, have been built on the development of ICT and are all situated between the third and fourth generations of technology-based psychological and educational assessments. If applied correctly, they could help preserve and support the fundamental assessment principals regarding respect for diversity, equity and inclusion.

Ambulatory Assessment

Ambulatory Assessment (AA) can be defined as a systematic and structured procedure for studying people's behaviour (affect, cognition, mental states, etc.) in their natural environment and in real time (or near real time), at multiple moments in time, usually using an electronic device, such as a personal digital assistant (PDA) or smartphone (Fonseca-Pedrero et al., 2022; Myin-Germeys & Kuppens, 2021; Trull & Ebner-Priemer, 2013). AA uses a variety of data sources i.e., multimodal assessment (e.g., psycho-physiological,

biological, self-reported, and behavioural) and represents a conceptual and methodological umbrella that includes experience sampling methodology (ESM), ecological momentary assessment (EMA) and momentary assessment (Stone & Shiffman, 1994).

AA attempts to overcome some of the problems associated with the traditional assessment paradigm such as the lack of ecological validity (less generalizability) and the biases associated with retrospective assessments. At the same time, AA allows researchers to investigate within-person processes and patterns of variation over time, assess context-specific relationships, and deliver feedback in real time (Ebner-Priemer & Trull, 2009; Myin-Germeys & Kuppens, 2021).

AA has grown exponentially in recent years, gaining importance and popularity in different areas of psychology (Heron et al., 2017; Myin-Germeys et al., 2018; van Roekel et al., 2019). Developments and studies have been conducted in the field of clinical psychology, in areas such as anxiety and depression (Hall et al., 2021), psychosis (Bell et al., 2017), and suicidal behaviour (Sedano-Capdevila et al., 2021). Particularly noteworthy are the studies on the mechanisms and dynamics of symptoms, the prediction of the recurrence or onset of symptoms, the monitoring of treatment effects, and the prediction of treatment success (Trull & Ebner-Priemer, 2013).

Due to the novelty and growth in popularity of AA, some general reporting guidelines have been published (Heron et al., 2017; Trull & Ebner-Priemer, 2020). According to Trull and Ebner-Priemer (2020), when carrying out a study under this new paradigm, the following should always be reported on: (a) the sample size and selection, (b) sampling design, (c) selection and reporting of measures, (d) devices and software used, (e) compliance, (f) participant training, monitoring and remuneration, and (g) data management and analysis.

Social Network-Based Assessment in the Organizational Realm

In the context of the war for talent (Frasca & Edwards, 2017), technology has emerged as a key element in the field of human resource management (HRM) (Ryan & Derous, 2019). Social network websites (SNW) have become one of the main technology tools used by professionals in recruitment and selection processes (Nikolaou, 2014; Woods et al., 2020). SNW profiles allow HRM professionals to gather information about applicants' knowledge, skills, abilities and other characteristics and examine the degree to which applicants' qualifications are aligned with the job requirements or fit with the organizational culture (Bangerter et al., 2012). SNWs provide more information than traditional methods (Zide et al., 2014) and at a lower cost (Nikolaou, 2014). A SNW is characterized by allowing the user to: (a) define a profile within a bounded system, (b) articulate a list of contacts with whom to share information, and (c) view and browse their own list and those of other users to identify opportunities for connection and contact (Boyd & Ellison, 2007). For instance, as Collmus et al. (2016) point out, the information contained in a LinkedIn profile includes features of traditional résumés, reference checks, and recommendation letters.

HRM professionals do not only use SNWs such as LinkedIn to identify and contact potential candidates; they also use them to make selection decisions. To do this, they make inferences about the degree to which applicants will fit the job profile and the organization (Kluemper et al., 2012; Roulin & Bangerter,

2013). Thus, from viewing LinkedIn profiles, professionals make inferences about the applicants' personality and competencies, and make predictions about their future performance (Van Iddekinge et al., 2016). A key issue is that this decision-making process can be accompanied by deficits in the reliability of measures, and a lack of validity due to the scarcity of associated evidence. Research has attempted to shed light on how professionals use this information to make inferences and investigate the psychometric properties of LinkedIn as a selection tool.

Building on the Realistic Accuracy Model (RAM; Funder, 2012) and the Signaling Theory (Donath, 2008; Spence, 1973), Roulin and Levashina (2019) found that the reliability and validity of the inferences about candidates' skills, personality and cognitive ability are different depending on the characteristics assessed. Inferences regarding skills and personality have lower inter-rater reliability than either cognitive ability inferences ($ICC = .60$) or hiring recommendations ($ICC = .67$). Temporal reliability is similar for all the characteristics (correlations between .41 and .66). Further differences exist depending on the ability or inferred personality trait (skills and traits such as conflict management and conscientiousness are less reliable while estimations of communication skills and agreeableness are more reliable). Regarding the convergent validity of these inferences based on self-report measures, the findings seem to make it clear that personality traits are not inferred accurately, and that the only skills that are inferred more accurately are leadership, planning and communication skills (correlations between .23 and .27). Cognitive ability also seems to be inferred accurately. Finally, regarding predictive validity, the global analysis of the profiles seems to correlate moderately (between .20 and .25) with the development of a successful professional career.

Reliability and validity seem to improve through two strategies: the use of itemized signals for profile information (Fernández et al., 2021; Roulin & Levashina, 2019), and the use of clusters of LinkedIn information. By using the latter strategy, Aguado et al. (2019) established that the information on LinkedIn profiles can be grouped into four large dimensions (LinkedIn Big Four): (a) breadth of professional experience, (b) social capital, (c) interest in updating knowledge, and (d) breadth of non-professional information. These four dimensions have proven useful in predicting employee outcomes such as business and management potential, absenteeism, and productivity. Also, they seem to constitute valid signals of the generic competencies of users (Andrés et al., 2022). In addition, when standardized tools such as rubrics are used to assess these dimensions, the inter-rater and temporal reliability increases substantially (Andrés et al., 2022).

Finally, research seems to suggest that the use of LinkedIn does not generate an adverse impact in terms of gender, and only a small to moderate adverse impact when ethnicity (white vs. non-white) is examined (Roulin & Levashina, 2019).

Gamification

Technology has also facilitated interactive assessment through computerized games, a field which has made extraordinary advances in recent years. In gamification, the subject being evaluated interacts with the computer game with a view to achieving a series of specific results. The subject's behaviour throughout this interaction is automatically assessed and an estimation of the

psychological characteristic being measured is produced. It is not exactly new, it is what Cattell called T-data (Cattell & Warburton, 1967). In fact, within the area of Objective Personality Tests (e.g. Ortner & Proyer, 2015) different tools have been developed to assess risk tendency (Lejuez et al., 2002; Aguado et al., 2011), impulsivity (Elliot et al., 1996), heuristic thinking (Jasper & Ortner, 2014), achievement motivation (Ziegler et al., 2010), and Attention-Deficit Hyperactivity Disorder (Delgado-Gómez et al., 2020). Despite the advantages of these tools in the observation of behaviours, compared to self-assessment, and in terms of resistance to faking (Ortner & Proyer, 2015), the convergent validity results point to the need to delve further into their relationship with the gold standard for measuring personality, i.e. self-reports.

Adopting a different approach, some authors have developed the idea of using serious games for assessment purposes (e.g., Kato & Klerk, 2017). Serious games are games that are used for a purpose other than entertainment, usually for skill training and assessment (e.g., Caballero-Hernández et al., 2017). Serious games are characterized by using game design frameworks (e.g., rewards, objectives, simulated scenarios) together with advanced technological support (e.g., online interactive software, immersive 3D environments, virtual reality). From a psychometric perspective, the use of serious games is based on the hypothesis that they have the potential to offer more valid measures than traditional approaches. This is possible because: (a) they offer evaluative scenarios with greater ecological validity, (b) serious games make it possible to carry out behaviour-based assessment instead of relying on self-reports, and (c) they allow the use of technological advances in data collection and analysis, as well as in the presentation of information (Lumsden et al., 2016).

However, despite its promising potential in assessment, the field of gamification has received little attention in the literature on psychometric and psychological assessment (Ryan & Ployhart, 2014). As a result, gamification strategies have often not been subjected to the quality and rigorous standards of psychometric measurement (Landers, 2015). Therefore, reliable theories need to be developed on the impact of different elements on game assessment results (Armstrong et al., 2016).

Forced-Choice Testing and New Psychometric Models

Most of the self-report scales for the measurement of non-cognitive traits use a Likert scale format, in which a respondent indicates his or her level of agreement with a statement. However, even though meta-analysis studies support the predictive role of self-report scales in applied contexts (Judge et al., 2013), it is well known that this format is sensitive to response biases such as social desirability or acquiescence bias (Kreitchmann et al., 2019). In selection contexts, social desirability and faking increases applicant mean scores in the perceived desirable direction while reducing the reliability and variability of scores (Salgado, 2016).

As an alternative, a forced-choice format requires the test taker to partially or totally rank a set of assertions into blocks of two or more statements, based on how well they feel described by them. Response biases such as acquiescence do not affect forced-choice tests, and, if the blocks are built with items that are well-matched in terms of social desirability, they should be less susceptible to the effects of faking. The following are typical formats (Hontangas, 2016): (a) choose the item that best describes you from two

statements, (b) choose the item that best describes you from more than two statements, (c) choose the item that best describes you and the item that least describes you, and (d) rank the alternatives according to the degree to which they describe you.

Forced-choice tests can be fully ipsative, quasi-ipsative or normative, depending on the degree of ipsativity of the obtained scores (Hicks, 1970). Ipsativity means that a person's trait scores depend on his or her other trait scores and has been suggested to be the main drawback of forced-choice tests; it makes inter-individual comparisons difficult and biases typical psychometric analysis results (Hicks, 1970). In the worst-case scenario, in fully ipsative measures, all test taker scores add up to a constant, distorting the internal structure of the test and the predictive validity of the scores. According to certain meta-analyses, quasi-ipsative assessments offer stronger predictive validity (Salgado et al., 2015) and are less susceptible to faking (Martínez & Salgado, 2021).

Forced-choice tests are designed to address the issue of social desirability in non-cognitive measures in selection tests. However, to generate reliable, faking-resistant, non-ipsative scores, optimum forced-choice test design and scoring are critical. First, the items of each block should be well-matched in terms of social desirability, based on expert consensus (Pavlov et al., 2021). Second, the items should be paired optimally (e.g., Kreitchmann et al., 2021). Finally, the scoring could be performed using novel item response theory (IRT) models such as multi-unidimensional pairwise preference (MUPP; Stark et al., 2005) or Thurstone IRT (TIRT; Brown & Maydeu-Olivares, 2011). These models were developed to shape the probability of agreement with an item in a forced-choice block.

Recent advances include the use of genetic algorithms for the optimal assembly of forced-choice blocks (Kreitchmann et al., 2021), the determination of the optimal selection rule in forced-choice Computerized Adaptive Tests (CATs), and the construction of CATs on the fly, in which blocks are assembled optimally at the time of administering the CAT (Kreitchmann et al., 2022).

Furthermore, the use of novel multidimensional models is also showing promise. A bi-factor multidimensional structure is to be expected in many psychological domains, in which items measure both a general factor (e.g., the domain of extraversion) and a group factor (e.g., facet of gregariousness). Nieto et al. (2018) studied the performance of bi-factor CATs, demonstrating that the use of multidimensional models increases efficiency. Recent developments in the algorithms for the estimation of bi-factor structures (García-Garzón et al., 2020) should make it simple to create forced-choice tests based on more realistic bi-factor structures.

In conclusion, while forced-choice tests have been around for a long time, it is only through methodological and technological advances that adequate modelling and scoring of forced-choice data, as well as the optimal design of forced-choice blocks and computerized adaptive applications, have become possible. Although it is challenging to construct an effective forced-choice test, well-designed IRT-based adaptive forced-choice tests will undoubtedly improve measurement in recruiting and selection processes.

Discussion

The paper has described advances in the field of digital technology-based psychological and educational measurement and assessment. Some of these developments address classical measurement problems such as response bias and modelling

dimensionality, and can be seen as a natural progression in the history of psychometric research. In addition, the implementation of new technologies which bring know-how from the fields of engineering or computer science opens up a new horizon of possibilities. Social networks, the internet of things, the process data or the use of mobile devices are making it possible to use methodologies based on experience sampling, gaming and virtual environments. These could help to improve the ecological validity of assessment and better understand people's thoughts, feelings, and behaviours (Bogarín, 2018; Fonseca-Pedrero et al., 2021; López-Mora, 2021; Parsons, 2012). It could be said that the social science research tradition built on Cattell's (1966) representation of the data cube is moving towards huge amounts of unstructured, high-dimensional data which need new data analytical approaches.

All of these new approaches can be seen as part of the broader field of Artificial Intelligence (AI), which is permeating science and society and includes specialized areas in computing, machine learning, natural language processing, computer vision, and robotics. There are many definitions of AI, but one of the most accepted is that proposed by Russell and Norvig (2021). According to these authors, 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 agent is defined 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 (Elosua, 2022). This definition has been accepted by the European Union, which reformulates it as: Software that is developed using one or more of the following techniques and strategies 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 (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). (European Commission, 2021).

The use of natural language processing in automated personality assessment already offers high quality standards (Tausczik & Pennebaker 2010; Schwartz et al., 2013; Youyou et al., 2015), automatic analysis of movement (Delgado-Gómez et al., 2016). Analysis of voice characteristics has been successfully used in personality assessment (Mairesse et al., 2007), in the prediction of sleep apnoea (Espinoza-Cuadros et al., 2016), and for the detection of the early stages of Alzheimer's disease (Weiner et al., 2016). The application of virtual scenarios and metaverse has generated assessment tools for the measurement of affective disorders, cognitive processes, spatial abilities, or attention and memory processes (e.g., Knight & Titov, 2009; Law et al., 2006; Powers & Emmelkamp, 2008).

However, applying this new technology in the fields of psychological and educational assessment brings with it a series of issues that need to be addressed (Elosua, 2022): (a) a comprehensive analysis of the contribution of each innovation in terms of validity of the measurement needs to be carried out, (b) a debate on the ethical, legal, and social implications of the new methods of collecting/analysing data must take place, (c) the different fields of knowledge need to be coordinated, and (d) psychologists need to

be trained in the new approaches to data analysis. In this regard, we understand that it is not easy to include sophisticated techniques during a psychometric program for undergraduate students, but we should make our students aware of the need for openness in order to adapt and take advantage of an environment in which analytical techniques in the field of artificial intelligence are progressing rapidly, and where interdisciplinary is a necessary condition to keep high scientific, and ethics standards.

Studies dealing with the associated ethical, scientific and social implications systematically report concerns about transparency, fairness, equity, and bias (Jobin et al., 2019). In the field of AA, for instance, privacy and confidentiality issues have been reported (Fonseca-Pedrero et al., 2022; Palmier-Claus et al., 2011). Given the right to privacy of job applicants, the use of recreational SNWs (such as Facebook or Instagram) in recruiting and selection processes may affect the legal liabilities of companies (Roulin & Levashina, 2019). In addition, negative reactions by applicants to the use of recreational SNWs have already been reported (Aguado et al., 2019; Stoughton et al., 2015; Stoughton, 2016). The lack of tradition and perceived intrusion in this particular field of psychological assessment is also linked to a lack of rigor and psychometric quality standards (Landers, 2015) as well as to the need for sound theories on the impact of the different game elements on the measurements obtained (Armstrong et al., 2016). There is also an open debate around the application of machine learning and/or other AI models to the field of assessment which alerts us of the consequences of a blind focus on maximizing predictive accuracy (Fokkema et al., 2022). In this vein, some interesting attempts are being carried out to integrate analytic techniques derived from AI with theoretical psychometric research applied to learning and assessment; to date, computational psychometrics (CP) seems one of the most promising. CP basically explores analytical models for new types of data and studies how to integrate them to define components of teaching, learning, and assessment (Langenfeld, et al., 2022; von Davier et al., 2021).

Furthermore, despite the technological advances of recent years, an analysis of the professional practice indicates approximately 90% of tests are pen and paper (Santamaría & Sánchez-Sánchez, 2022). These results show two clear trends: a traditional trend which is still reliant on the classical test model and a data-driven model built around the availability of data. The former is a deductive and theory-centered approach and the second is data-centric and exploratory; it applies machine-learning models that search for patterns and relationships which are used for classification and prediction purposes. With a view to advancing our knowledge and increasing our scientific productivity, it would be of great use to find a convergent area between the two models (Maass et al., 2018). To do so, teamwork among different disciplines and training (Adjerid & Kelley, 2018; König et al., 2020; Oswald, 2020) would be key to addressing the diversity and dynamism of this new age. What psychometrics have taught us about psychological measurement and what we already know about the science of psychology offer a very promising combination with which to face future challenges.

Funding

This work has been partially funded by the Ministry of Science and Innovation (PID2019-103859RB-100).

References

- 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>
- Aguado, D., Andrés, J. C., García Izquierdo, A. L., & Rodríguez, J. (2019). LinkedIn® Big Four®: Job performance validation in the ICT sector. *Journal of Work and Organizational Psychology*, 35, 53-64. <https://doi.org/10.5093/jwop2019a7>
- Aguado, D., Rubio, V. J., & Lucía, B. (2011). The Risk Propensity Task (PTR): A proposal for a behavioral performance-based computer test for assessing risk propensity. *Anales de Psicología*, 27(3), 862-870.
- Andrés, J. C., Aguado, D., & De Miguel, J. (2022). What's behind LinkedIn? Measuring the LinkedIn big four dimensions through rubrics. *Psychologist Papers*, 43(1), 12-20. <https://doi.org/10.23923/pap.psicol.2979>
- Armstrong, M. B., Landers, R. N., & Collmus, A. B. (2016). Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management. In D. Davis & H. Gangadharbatla (Eds.), *Emerging research and trends in gamification*. (pp.140-165). IGI Global.
- Bangerter, A., Roulin, N., & König, C. J. (2012). Personnel selection as a signaling game. *Journal of Applied Psychology*, 97, 719-738. <https://doi.org/10.1037/a0026078>
- Bell, I. H., Lim, M. H., Rossell, S. L., & Thomas, N. (2017). Ecological momentary assessment and intervention in the treatment of psychotic disorders: A systematic review. *Psychiatric Services*, 68(11), 1172-1181. <https://doi.org/10.1176/appi.ps.201600523>
- 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>
- Bogarín, A., Cerezo, R., & Romero, C. (2018). Discovering learning processes using Inductive Miner: A case study with Learning Management Systems (LMSs). *Psicothema*, 30(3), 322-329. <https://doi.org/10.7334/psicothema2018.116>
- Boyd, D., & Ellison N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460-502. <https://doi.org/10.1177/0013164410375112>
- Caballero-Hernández, J. A., Palomo-Duarte, M., & Doderó, J. M. (2017). Skill assessment in learning experiences based on serious games: A systematic mapping study. *Computers & Education*, 113, 42-60. <https://doi.org/10.1016/j.compedu.2017.05.008>
- Cattell, R. B. (1966). The 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.
- Cattell, R. B., & Warburton, F. W. (1967). *Objective personality and motivation tests*. University of Illinois Press.
- Collmus, A. B., Armstrong, M. B., & Landers, R. N. (2016). Game-thinking within social media to recruit and select job candidates. In N. R. Landers & B. G. Schmidt (Eds.), *Social media in employee selection and recruitment: Theory, practice, and current challenges* (pp. 103-124). Springer International Publishing.
- Delgado-Gómez, D., Carmona-Vázquez, C., Bayona, S., Ardoy-Cuadros, J., Aguado, D., Baca-García, E., & Lopez-Castroman, J. (2016). Improving impulsivity assessment using movement recognition: a pilot study. *Behavior Research Methods*, 48(4), 1575-1579. <https://doi.org/10.3758/s13428-015-0668-y>

- Delgado-Gómez, D., Sújar, A., Ardoy-Cuadros, J., Bejarano-Gómez, A., Aguado, D., Miguelez-Fernández, C., Blasco-Fontecilla, H., & Peñuelas-Calvo, I. (2020). Objective assessment of attention-deficit hyperactivity disorder (ADHD) using an infinite runner-based computer game: a pilot study. *Brain Sciences*, *10*(10), Article 716. <https://doi.org/10.3390/brainsci10100716>
- Donath, J. (2008). Signals in social supernets. *Journal of Computer Mediated Communication*, *13*(1), 231-251. <https://doi.org/10.1111/j.1083-6101.2007.00394.x>
- Ebner-Priemer, U. W., & Trull, T. J. (2009). Ambulatory assessment: An innovative and promising approach for clinical psychology. *European Psychologist*, *14*(2), 109-119. <https://doi.org/10.1027/1016-9040.14.2.109>
- Elliot, S., Lawty-Jones, M., & Jackson, C. (1996). Effects of dissimulation on self-report and objective measures of personality. *Personality and Individual Differences*, *21*, 335-343. [https://doi.org/10.1016/0191-8869\(96\)00080-3](https://doi.org/10.1016/0191-8869(96)00080-3)
- Elosua, P. (2022). Impact of ICT on the assessment environment. Innovations through continuous improvement. *Psychologist Papers*, *43*(1), 3-11. <https://doi.org/10.23923/pap.psicol.2985>
- European Commission (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence*. <https://op.europa.eu/en/publication-detail/-/publication/e0649735-a372-11eb-9585-01aa75ed71a1/language-en/format-PDF/source-205836026>
- Espinoza-Cuadros, F., Fernández-Pozo, R., Toledano, D. T., Alcázar-Ramírez, J. D., López-Gonzalo, E., & Hernández-Gómez, L. A. (2016). Reviewing the connection between speech and obstructive sleep apnea. *Biomedical engineering online*, *15*(1), 20. <https://doi.org/10.1186/s12938-016-0138-5>
- Fernández, S., Stöcklin, M., Terrier, L., & Kim, S. (2021). Using available signals on LinkedIn for personality assessment. *Journal of Research in Personality*, *93*, Article 104122. <https://doi.org/10.1016/j.jrp.2021.104122>
- Fokkema, M., Iliescu, D., Greiff, S., & Ziegler, M. (2022). Machine Learning and Prediction in Psychological Assessment: Some Promises and Pitfalls. *European Journal of Psychological Assessment*, *38*(3), 165-175. <https://doi.org/10.1027/1015-5759/a000714>
- Fonseca Pedrero, E., Pérez-Álvarez, M., Al-Halabí, S., Inchausti, F., Muñoz, J., López-Navarro, E., Pérez de Albéniz, A., Lucas Molina, B., Debbané, M., Bobes-Bascarán, M. T., Gimeno-Peón, A., Prado-Abril, J., Fernández-Álvarez, J., Rodríguez-Testal, J. F., González Pando, D., Díez-Gómez, A., García Montes, J. M., García-Cerdán, L., Osma, J., Peris Baquero, Ó., ... Marrero, R. J. (2021). Tratamientos psicológicos empíricamente apoyados para adultos: Una revisión selectiva [Evidence-based psychological treatments for adults: A selective review]. *Psicothema*, *33*(2), 188-197. <https://doi.org/10.7334/psicothema2020.426>
- Fonseca-Pedrero, E., Ródenas-Perea, G., Pérez-Albéniz, A., Al-Halabí, S., Pérez, M., & Muñoz, J. (2022). The time of ambulatory assessment. *Psychologist Papers*, *43*, 21-28. <https://doi.org/10.23923/pap.psicol.2983>
- Frasca, K. J., & Edwards, M. R. (2017). Web-based corporate, social and video recruitment media: Effects of media richness and source credibility on organizational attraction. *International Journal of Selection and Assessment*, *25*(2), 125-137. <https://dx.doi.org/10.1111/ijasa.12165>
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, *21*, 177-182. <https://doi.org/10.1177/0963721412445309>
- García-Garzón, E., Nieto, M. D., Garrido, L. E., & Abad, F. J. (2020). Bi-factor exploratory structural equation modeling done right: using the SLiDapp application. *Psicothema*, *32*, 607-614. <https://doi.org/10.7334/psicothema2020.179>
- Hall, M., Scherner, P. V., Kreidel, Y., & Rubel, J. A. (2021). A systematic review of momentary assessment designs for mood and anxiety symptoms. *Frontiers in Psychology*, *12*, Article 642044. <https://doi.org/10.3389/FPSYG.2021.642044>
- Heron, K. E., Everhart, R. S., McHale, S. M., & Smyth, J. M. (2017). Using obile-technology-based ecological momentary assessment (EMA) methods with youth: A systematic review and recommendations. *Journal of Pediatric Psychology*, *42*(10), 1087-1107. <https://doi.org/10.1093/JPEPSY/JSX078>
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced choice normative measures. *Psychological Bulletin*, *74*(3), 167-184. <https://doi.org/10.1037/h0029780>
- Hontangas, P. M., Leenen, I., & de la Torre, J. (2016). Traditional scores versus IRT estimates on forced choice tests based on a dominance model. *Psicothema*, *28*(1), 76-82. <https://doi.org/10.7334/psicothema2015.204>
- Jasper, F., & Ortner, T. M. (2014). The tendency to fall for distracting information while making judgments: development and validation of the Objective Heuristic Thinking Test. *European Journal of Psychological Assessment*, *30*, 193-207. <https://doi.org/10.1027/1015-5759/a000214>
- 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>
- Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*(6), 875-925. <https://doi.org/10.1037/a0033901>
- Kato, P. M., & de Klerk, S. (2017). Serious games for assessment: Welcome to the jungle. *Journal of Applied Testing Technology*, *18*(S1), 1-6.
- Klumper, D. H., Rosen, P. A., & Mossholder, K. W. (2012). Social networking websites, personality ratings, and the organizational context: More than meets the eye? *Journal of Applied Social Psychology*, *42*(1), 1143-1172. <https://doi.org/10.1111/j.1559-1816.2011.00881.x>
- Knight, R. G., & Titov, N. (2009). Use of virtual reality tasks to assess prospective memory: Applicability and evidence. *Brain Impairment*, *10*, 3-13. <https://doi.org/10.1375/brim.10.1.3>
- 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*, *6*, 17-23. <https://doi.org/10.25035/pad.2020.01.002>
- Kreitchmann, R. S., Abad, F. J., & Sorrel, M. A. (2021). A genetic algorithm for optimal assembly of pairwise forced choice questionnaires. *Behavior Research Methods*, *54*, 1476-1492. <https://doi.org/10.3758/s13428-021-01677-4>
- Kreitchmann, R. S., Abad, F. J., Ponsoda, V., Nieto, M. D., & Morillo, D. (2019). Controlling for response biases in self-report scales: Forced choice vs. psychometric modeling of Likert items. *Frontiers in Psychology*, *10*, Article 2309. <https://doi.org/10.3389/fpsyg.2019.02309>
- Kreitchmann, R. S., Sorrel, M. A., & Abad, F. J. (2022). On bank assembly and block selection in multidimensional forced choice adaptive assessments. *Educational and Psychological Measurement*. Advance online publication. <https://doi.org/10.1177/00131644221087986>
- Langenfeld, T., Burstein, J., & von Davier, A. A. (2022). Digital-First Learning and Assessment Systems for the 21st Century. *Frontiers in Education*, *7*, Article 857604. <https://doi.org/10.3389/educ.2022.857604>
- Landers, R. N. (2015). An introduction to game-based assessment: Frameworks for the measurement of knowledge, skills, abilities and other human characteristics using behaviors observed within videogames. *International Journal of Gaming and Computer-Mediated Simulations*, *7*, 4-8.
- Law, A. S., Logie, R. H., & Pearson, D. G. (2006). The impact of secondary tasks on multitasking in a virtual environment. *Acta Psychologica*, *122*, 27-44. <https://doi.org/10.1016/j.actpsy.2005.09.002>

- López-Mora, C., Carlo, G., Roos, J., Maiya, S., & González-Hernández, J. (2021). Perceived Attachment and Problematic Smartphone Use in Young People: Mediating Effects of Self-Regulation and Prosociality. *Psicothema*, 33(4), 564-570. <https://doi.org/10.7334/psicothema2021.60>
- Lejuez, C. W., Read, J. P., Kahler, C., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D.R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8, 75-84. <https://doi.org/10.1037/1076-898x.8.2.75>
- Lumsden, J., Skinner, A., Woods, A. T., Lawrence, N. S., Munafò, M. (2016). The effects of gamelike features and test location on cognitive test performance and participant enjoyment. *PeerJ*, 4, Article e2184. <https://doi.org/10.7717/peerj.2184>
- Maass, W., Parsons, J., Puraos, 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, Article 12, <https://doi.org/10.17705/1jais.00526>
- Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research*, 30, 457-500. <https://www.aaai.org/Papers/JAIR/Vol30/JAIR-3012.pdf>
- Martínez, A., & Salgado, J. F. (2021). A meta-analysis of the faking resistance of forced choice personality inventories. *Frontiers in Psychology*, 12, Article 732241. <https://doi.org/10.3389/fpsyg.2021.732241>
- Myin-Germeys, I., & Kuppens, P. (2021). *The Open Handbook of Experience Sampling Methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies*. The center for Research on Experience sampling and Ambulatory methods Leuven (REAL).
- 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>
- Nieto, M. D., Abad, F. J., & Olea, J. (2018). Assessing the Big Five with bifactor computerized adaptive testing. *Psychological Assessment*, 30(12), 1678-1690. <https://doi.org/10.1037/pas0000631>
- Nikolaou, I. (2014). Social networking web sites in job search and employee recruitment. *International Journal of Selection and Assessment*, 22(2), 179-189. <https://doi.org/10.1111/ijsa.12067>
- Ortner, T. M., & Proyer, R. T. (2015). Objective personality tests. In T. M. Ortner & F. J. R. Van de Vijver (Eds.), *Behavior-based assessment in psychology* (pp. 133-149). Hogrefe.
- 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.
- Palmier-Claus, J. E., Myin-Germeys, I., Barkus, E., Bentley, L., Udachina, A., Delespaul, P. A., Lewis, S. W., & Dunn, G. (2011). Experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatrica Scandinavica*, 123(1), 12-20. <https://doi.org/10.1111/j.1600-0447.2010.01596.x>
- Parsons, T. D. (2012). Virtual simulations and the second life metaverse: paradigm shift in neuropsychological assessment. In N. Zagalo, L. Morgado, & Boa-Ventura, A. (Eds.), *Virtual worlds and metaverse platforms: New communication and identity paradigms* (pp. 234-250). IGI Global.
- Pavlov, G., Shi, D., Maydeu-Olivares, A., & Fairchild, A. (2021). Item desirability matching in forced choice test construction. *Personality and Individual Differences*, 183, Article 111114. <https://doi.org/10.1016/j.paid.2021.111114>
- Powers, M. B., & Emmelkamp, P. M. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, 22, 561-569. <https://doi.org/10.1016/j.janxdis.2007.04.006>
- Roulin, N., & Bangerter, A. (2013). Social networking websites in personnel selection. *Journal of Personnel Psychology* 12(1), 143-151. <https://doi.org/10.1027/1866-5888/a000094>
- Roulin, N., & Levashina, J. (2019). LinkedIn as a new selection method: Psychometric properties and assessment approach. *Personnel Psychology*, 72(2), 187-211. <https://doi.org/10.1111/peps.12296>
- Russell, S., & Norvig, P. (2021). *Artificial intelligence: a modern approach*. Pearson.
- Ryan, A. M., & Deros, E. (2019). The unrealized potential of technology in selection assessment. *Journal of Work and Organizational Psychology*, 35(2), 85-92. <https://doi.org/10.5093/jwop2019a10>
- Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology*, 65, 693-717. <https://doi.org/10.1146/annurev-psych-010213-115134>
- Salgado, J. F. (2016). A theoretical model of psychometric effects of faking on assessment procedures: Empirical findings and implications for personality at work: A theoretical model of faking psychometric effects. *International Journal of Selection and Assessment*, 24(3), 209-228. <https://doi.org/10.1111/ijsa.12142>
- Salgado, J. F., Anderson, N., & Tauriz, G. (2015). The validity of ipsative and quasi-ipsative forced choice personality inventories for different occupational groups: A comprehensive meta-analysis. *Journal of Occupational and Organizational Psychology*, 88(4), 797-834. <https://doi.org/10.1111/joop.12098>
- Santamaría, P., & Sánchez-Sánchez, F. (2022). Open questions in the use of new technologies in psychological assessment. *Psychologist Papers* 43(1), 48-54. <https://doi.org/10.23923/pap.psicol.2984>
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS one*, 8(9), Article 73791. <https://doi.org/10.1371/journal.pone.0073791>
- Schwab, K. (2017). *The fourth industrial revolution*. Crown Publishing Group.
- Sedano-Capdevila, A., Porras-Segovia, A., Bello, H. J., Baca-García, E., & Barrigon, M. L. (2021). Use of ecological momentary assessment to study suicidal thoughts and behavior: a systematic review. *Current Psychiatry Reports*, 23(7), Article 41. <https://doi.org/10.1007/s11920-021-01255-7>
- Spence, M. (1973). Job market signalling. *Quarterly Journal of Economics*, 87(3), 355-374.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise-preference model. *Applied Psychological Measurement*, 29(3), 184-203. <https://doi.org/10.1177/0146621604273988>
- Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, 16, 199-202. <https://doi.org/10.1093/abm/16.3.199>
- Stoughton, J.W. (2016). Applicant reactions to social media in selection: Early returns and future directions. In N. R. Landers & B. G. Schmidt (Eds.), *Social media in employee selection and recruitment: Theory, practice, and current challenges* (pp. 249-263). Springer International Publishing.
- Stoughton, J.W., Thompson, L. F., & Meade, A.W. (2015). Examining applicant reactions to the use of social networking websites in pre-employment screening. *Journal of Business and Psychology*, 30, 73-88. <https://doi.org/10.1007/s10869-013-9333-6>

- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1), 24-54.
- Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory Assessment. *Annual Review of Clinical Psychology*, 9, 151-176.
<https://doi.org/10.1146/ANNUREV-CLINPSY-050212-185510>
- Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology*, 129(1), 56-63.
<https://doi.org/10.1037/ABN0000473>
- Van Iddekinge, C. H., Lanivich, S. E., Roth, P. L., & Junco, E. (2016). Social media for selection? Validity and adverse impact potential of a Facebook-based assessment. *Journal of Management*, 42(7), 1811-1835.
<https://dx.doi.org/10.1177/0149206313515524>
- van Roekel, E., Keijsers, L., & Chung, J. M. (2019). A review of current ambulatory assessment studies in adolescent samples and practical recommendations. *Journal of Research on Adolescence*, 29(3), 560-577.
<https://doi.org/10.1111/JORA.12471>
- von Davier, A. A., Mislevy, R. J., & Hao, J. (Ed.) (2021). *Computational psychometrics: New methodologies for a new generation of digital learning and assessment into psychometrics*. Springer.
<https://doi.org/10.1007/978-3-030-74394-9>
- Weiner, J., Herff, C., & Schultz, T. (2016). Speech-based detection of Alzheimer's disease in conversational German. In International Speech Communications Association (Eds.), *Understanding speech processing in human and machines* (pp. 1938-1942). Curran Associates, Inc.
- Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, 29(1), 64-77.
<https://doi.org/10.1080/1359432X.2019.1681401>
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036-1040.
<https://doi.org/10.1073/pnas.1418680112>
- Zide, J., Elman, B., & Shahani-Denning, C. (2014). LinkedIn and recruitment: How profiles differ across occupations. *Employee relations*, 36(5), 583-604.
<https://doi.org/10.1108/ER-07-2013-0086>
- Ziegler, M., Schmukle, S., Egloff, B., & Bühner, M. (2010). Investigating measures of achievement motivation(s). *Journal of Individual Differences*, 31, 15-21. <https://doi.org/10.1027/1614-0001/a000002>