Introduction

Louis Tay, Sang Eun Woo, and Tara Behrend

Technology, that is, the output of human innovation, has always been central to human progress worldwide. Early on, the ancients developed the wheel, concrete, calculus, and paper, which led to advances in transportation, construction, and communication. Today, the incarnation of technology falls in the realm of the digital and computational, and its progress has been rapid, even arguably exponential. In his chapter, "The Law of Accelerating Returns," Ray Kurzweil writes, "An analysis of the history of technology shows that technological change is exponential, contrary to the common-sense 'intuitive linear' view. So we won't experience 100 years of progress in the 21st century – it will be more like 20,000 years of progress (at today's rate)" (Kurzweil, 2004, p. 381).

We are at a unique juncture of mass adoption of mobile and wearable technology, gobs of internet and communication behaviors among people, artificial intelligence that can rival and surpass human performance on many tasks, and storage capacity for data and computational power hitherto unheard of. The global phenomenon of big data has arrived, and this has led to disruptions across the world and throughout multiple fields. Indeed, the tools and approaches in psychological, organizational, and sociological research have changed with the advent of big data and related technologies (McFarland et al., 2015; Oswald et al., 2020; Woo et al., 2020). The field of human assessments is similarly affected and transformed.

While technology is pervasive, its spread is not consistent across the globe due to differences in wealth and development among countries. It has been noted in the *2020 World Social Report* by the United Nations that inequality continues to pervade, and the average income of people living in North America is 16 times higher than those living in sub-Saharan Africa. This affects the ability of countries to purchase and use advanced technologies. The issue is, of course, not merely one of economics but also culture. Culture affects not only whether technologies are adopted (Ashraf et al., 2014) but how technologies are used, and it can shape the link between attitudes and technology use (Dinev et al., 2009). These kaleidoscopes of dimensions can lead to unique differences between regions of the world in how technology is shaping human assessments.

We (Tay, Woo, and Behrend) had these macro technology trends in mind when conceiving a vision for this book. To this end, we convened a conference in the summer of 2020, hosted virtually at Purdue (www.purdue.edu/hhs/psy/tmag/) with the sponsorship, endorsement, and support of the International Test Commission and Consortium for the Advancement of Research Methods and Analysis, Society for Industrial and Organizational Psychology (SIOP) Foundation, and Purdue Psychological Sciences to generate discussions about various aspects of technology, measurement, and culture. The presentations and conversations from that conference became chapters for the book you are now reading.

This edited book takes stock of how technology is changing the face of assessments in its relentless march around and in different regions of the world. Past technological progress in human assessments, or *measurement*, has primarily been constrained to carefully designed data within surveys or tests. Even when there are advances in the mode of data collection (e.g., paper and pencil to online) or procedure (e.g., static to computerized adaptive assessment), these data are crafted within a system that fits our typical psychometric models (e.g., classical test theory, item response theory). With the rise of big data, we now draw on organic data arising from multiple sources to make inferences about human attributes; we also rely on new computational models for measurement that often do not even have the notion of a latent trait (D'Mello et al., 2022). These new modes of data and measurement models require us to think through the applicability of traditional psychometric issues of reliability, validity, and bias (Liou et al., 2022). For instance, the notion of internal consistency reliability with Cronbach's alpha is less applicable to computational models that use thousands of distinct (and potentially uncorrelated) features to score individuals. Further, the notions of measurement bias need to be expanded to incorporate machine learning models that are distinct from traditional psychometric models (Tay et al., 2022). At the broadest level, new interdisciplinary approaches to designing, implementing, monitoring, and auditing these technology-based assessments are needed (Landers & Behrend, 2023).

We structured the book into three major sections: foundations, global perspectives, and regional focus. In foundations, we consider the core issues of computational models, new passively sensed data, and privacy concerns. In global perspectives, we seek to identify some key ways technology has touched on how we are measuring human attributes across the world. Finally, in regional focus, we survey how different regions around the world have adopted technological advancements in measurement for their unique cultural and societal context.

Foundations

Led by computer and data scientists seeking to apply new technologies, data, and computational models for human assessments, there is a significant need for understanding the intersection between these newer machine learning models and psychometrics. In Chapter 1, Song et al. provide an overview of machine learning algorithms and applications, along with how they advance psychological measurement. Drawing from their expertise, they provide recommendations and resources to researchers and practitioners seeking to apply these new methods in measurement.

A leading edge of pervasive technology is passive sensing. The application of this technology for measurement exemplifies and reflects many of the issues, opportunities, and challenges that newer technology-enabled assessments face. In Chapter 2, De Choudhury illustrates a range of passive sensing measurements (e.g., smartphones, wearable devices, and social media) in the context of the workplace to elucidate the types of assessments that are possible. She also discusses the strengths of such data along with the need to address limitations of generalizability and modeling of such data in future work. In addition, she calls for cross-disciplinary teams to address foundational issues of construct validity, theoretical grounding, research design, and ethics when using passive sensing for assessments.

With big data, privacy and security issues come to the fore as these types of technology-enabled assessment data become pervasive. In Chapter 3, Xu and Zhang discuss the opportunities and challenges of organic data collection from a privacy lens. They show how inferences about individuals can be made even under anonymity due to the abundance and interconnectedness of data. Given these practical limitations, there is a need to rethink and redefine what privacy means and to evaluate it across applications and time. In closing, they describe some state-of-the-art privacy-preserving techniques and their limitations.

Global Perspectives

Social media data, which includes content posted to social networking sites, friend/contact network information, photos, and usage behavior, has been used extensively to make inferences about the characteristics of users. Much has been made of these big data and machine learning approaches to modeling constructs such as personality and cognitive ability. In Chapter 4, Min and Gonzalez offer words of caution and advice for researchers who wish to conduct cross-cultural studies using social media data, as well as advice about how to measure culture-related constructs using data collected from social media.

Games have captured the global public's attention as a means of increasing engagement and increasing applicant reactions, as well as a means of conducting "stealth assessment" by measuring the behavior of game players in order to make inferences about relevant constructs. In Chapter 5, Landers et al. explore the phenomenon of games from a privacy and legality perspective, exploring game vendors' attitudes about the challenges of administering game-based assessments. This chapter also outlines the cross-cultural design and implementation factors that game developers must consider when deploying global game-based assessments.

Like games, mobile sensing has dramatically changed the way that assessment can be done without relying on self-reports. Mobile sensors use devices, including smartphones, badges, RFID (radio frequency identification) tags, or other devices, to capture information about the movements and behaviors of a person. When deployed in groups, this data can be used to generate information about social networks and communication patterns as well. The implementation of these methods has far outpaced research, though, and researchers using mobile sensing risk introducing bias or overlooking cultural differences in how the data should be collected and interpreted. In Chapter 6, Phan et al. provide a thoughtful and thorough approach to thinking about the deployment of mobile sensing in a responsible and ethical manner that accounts for cultural variation.

Regional Focus

In Chapter 7, Song et al. give a helpful overview of current applications and challenges associated with technology and measurement. After describing five major technologies used for psychological measurement (i.e., smartphones, wearable devices, social media, computerized adaptive testing, and game-based assessment), the authors discuss issues of legislation and regulation (e.g., data privacy and security) that are specific to Asian countries such as China and South Korea, and cultural/economic differences within Asia that may influence technology acceptance within a specific region.

In Chapter 8, Chan takes a deeper dive into the country of Singapore. He provides an insightful analysis of various factors surrounding Singapore's quest to become a "Smart Nation," at multiple levels of analysis. Organized by three C's (i.e., contexts, changes, and collaborations), these factors range from global- and industry-level demands and infrastructure development to individuals' well-being and socio-cognitive biases, which are all crucial considerations when designing and implementing technology-enabled measurement tools. The explicit focus on national-level innovation and collaboration driven by strong people-centric values is noteworthy and serves as a useful benchmark for other parts of the globe.

In Chapter 9, Guenole et al. discuss how technology affects measurement practices in Europe. Most noteworthy are the survey results from 182 professionals in Europe who utilize measurement technologies in talent management. Results indicated that emerging assessment technologies such as game-based assessment, text parsing (e.g., resumés), chatbots, digital footprint scraping, and Internet of Things assessment technology are not as prevalent compared to more conventional techniques such as questionnaires, interviews, assessment and development centers, and situational judgment tests. On the other hand, more privacy-related concerns were expressed around the newer technologies. The authors describe more fine-grained trends within each of the emerging technologies and conclude by noting some new challenges associated with them, such as finding and establishing the bridge between scientific (psychometric) principles and "innovative and tech-driven" assessment practices that are increasingly popularized in industry.

Focusing on (but not limited to) the North American region, Munson, in Chapter 10, makes acute observations about legal, ethical, social, and practical challenges related to privacy, accommodations vs. accessibility, and opt-out-of-testing trends. Like Guenole et al., Munson also calls for further development of "computational psychometrics" where traditional psychometric ideas are well integrated with innovations in computer science and newer sources of data. This requires interdisciplinary work that was hitherto the domain of only psychometricians.

In Chapter 11, De Fruyt et al. discuss a couple of technological innovations in developing personalized assessment for educational and workplace applications in the South American region. Similar to Chan, De Fruyt et al. mention the need for technological infrastructure that allows for large-scale (remote) assessment that overcomes discrepancies in accessibility due to economic inequalities in the region. The authors also discuss challenges with user familiarity with technology, which have been noted in other chapters (e.g., Song et al.; Guenole et al.).

Conclusion

In the final chapter, the book concludes with integrative comments and conclusions from Oswald and Behrend, who reflect on themes including privacy, transparency, justice, and emerging technologies and offer future directions for researchers working on the intersection of assessment and technology, placing special emphasis on testing and the policy and societal implications of testing.

REFERENCES

Ashraf, A. R., Thongpapanl, N., & Auh, S. (2014). The application of the technology acceptance model under different cultural contexts: The case of online shopping adoption. *Journal of International Marketing*, *22*(3), 68–93. https://doi.org/10.1509/jim.14.0065

D'Mello, S. K., Tay, L., & Southwell, R. (2022). Psychological measurement in the information age: Machine-learned computational models. *Current Directions in Psychological Science*, *31*(1), 76–87. https://doi.org/10.1177/09637214211056906

Dinev, T., Goo, J., Hu, Q., & Nam, K. (2009). User behaviour towards protective information technologies: The role of national cultural differences. *Information Systems Journal*, *19*(4), 391–412. https://doi.org/10.1111/j.1365-2575.2007.00289.x

Kurzweil, R. (2004). The law of accelerating returns. In C. Teuscher (Ed.), *Alan Turing: Life and legacy of a great thinker* (pp. 381–415). Springer.

Landers, R. N., & Behrend, T. S. (2023). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. *American Psychologist*, *78*(1), 36–49. https://doi.org/10.1037/amp0000972

Liou, G., Bonner, C. V., & Tay, L. (2022). A psychometric view of technology-based assessments. *International Journal of Testing*, *22*(3–4), 216–242. https://doi.org/10.1080/15305058.2022.2070757

McFarland, D. A., Lewis, K., & Goldberg, A. (2015). Sociology in the era of big data: The ascent of forensic social science. *The American Sociologist*, *47*(1), 12–35. https://doi.org/10.1007/s12108-015-9291-8

Oswald, F. L., Behrend, T. S., Putka, D. J., & Sinar, E. (2020). Big data in industrial-organizational psychology and human resource management: Forward progress for organizational research and practice. *Annual Review of Organizational Psychology and Organizational Behavior*, *7*(1), 505–533. https://doi.org/10.1146/annurev-orgpsych-032117-104553

Tay, L., Woo, S. E., Hickman, L., Booth, B. M., & D'Mello, S. (2022). A conceptual framework for investigating and mitigating Machine-Learning Measurement Bias (MLMB) in psychological assessment. *Advances in Methods and Practices in Psychological Science*, 5(1). https://doi.org/10.1177/25152459211061337

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

PART I

Foundations

Cambridge University Press & Assessment
978-1-316-51528-0 — Technology and Measurement around the Globe
Louis Tay, Sang Eun Woo, Tara Behrend
Excerpt
More Information

CHAPTER 1

Machine Learning Algorithms and Measurement

Q. Chelsea Song, Ivan Hernandez, Hyun Joo Shin, Meaghan M. Tracy, and Mengqiao Liu

MACHINE LEARNING AND ITS UNIQUE CONTRIBUTION TO MEASUREMENT

Machine learning (ML) is a subfield of artificial intelligence that utilizes data to optimize predictions and discover underlying patterns (e.g., Mitchell, 1997). Compared to traditional approaches for measuring human attributes (e.g., classical testing theory, item response theory), machine learning uniquely contributes to measurement by being better suited to (1) utilize organic data and (2) capture complex relations. Organic data are naturally occurring digital footprints that are collected without reliance on a specific research design or measurement scale; examples include online search records, Twitter posts, and location data collected from fitness trackers (see Groves et al., 2011; Hickman, Bosch, et al., 2022; Xu et al., 2020). Such data convey rich behavioral and psychological traces embedded in everyday contexts, providing valuable information for measurement. However, due to their complexity and lack of structure, they were rarely utilized in psychological measurement – until the introduction of machine learning. With machine learning, we are now capable of analyzing a diverse and complex range of data, from self- and other-reports to audiovisual footprints. To name a few, machine learning is used to measure personality from interview videos (e.g., Hickman, Bosch, et al., 2022), stress and emotions from social media posts (e.g., Wang et al., 2016), and interpersonal relationships from proximity data obtained from wearable sensors (e.g., Matusik et al., 2019). Such capability allows for increased accuracy in measurement as well as ecological momentary assessment of human behavior and cognition – enabling a more comprehensive measurement of human psychology and behavior.

Machine learning can also capture complex relations (e.g., nonlinear relations and interactions), potentially uncovering new insights into psychological phenomena. Recent works utilized machine learning to study personality nuances at the facet and item levels, extending the theoretical and practical understanding of personality traits (e.g., Putka et al., 2018). Unlike traditional parametric methods, ML algorithms do not rely on a priori specification of dimensions, allowing for a more flexible examination of the data (e.g., Jiang et al., 2020). Finally, ML algorithms are capable of handling high-dimensional data (where the number of features is large relative to the sample size), enabling the integration of multiple and complex data types while seeking to maintain the accuracy and generalizability of measurement.

In general, machine learning contributes to measurement in two crucial ways: conceptualization of a construct (via unsupervised learning) and empirical keying (via supervised learning). Unsupervised learning aims to find structure or patterns within data, and it could be used to explore the structure of a construct (e.g., to identify depressive symptoms among a wide variety of mental health symptoms). Similar to factor analysis in classical testing theory, unsupervised learning contributes to the conceptualization of a construct, providing the foundation for measurement. Yet, compared to traditional measurement methods, unsupervised learning has the advantage of identifying patterns from a large set of unstructured data, potentially contributing to broadened conceptual understandings. Supervised learning aims to estimate certain psychological constructs (e.g., ability) from a set of features (e.g., interview transcripts, event logs). Supervised learning, when used in measurement, is effectively an empirical keying method to convert features (or variables) into construct estimates – similar to the empirical keys used in traditional measurement, yet with improved accuracy, scalability, and consistency.[1] Together, machine learning contributes to measurement through conceptualization and empirical keying.[2] In the sections below, we discuss common ML algorithms used in measurement.

[1] A construct estimate could be continuous (e.g., ability level) as well as categorical (e.g., depressive symptoms). When supervised learning is used to measure a continuous construct, the process is called *regression*; when it is used to measure a categorical construct, the process is termed *classification*.

[2] The use of unsupervised (contextualization) and supervised learning (empirical keying) ML algorithms in measurement are guided by different approaches, which vary on the theory-data spectrum. Hickman, Song and Woo (2022) provides a systematic discussion of these approaches, which include (1) the theory-driven, hypothetico deductive approach, (2) the construct-driven, data-flexible approach, and (3) the data-driven, construct-informing approach.