BJHS Themes (2026), 1–16
doi:10.1017/bjt.2026.10044

INTRODUCTION

Testing knowledge: validation and regulation in the health and human sciences

Lara Keuck[1] and Angela N. H. Creager[2]

[1]Institute for Studies of Science, Department of History, Philosophy and Theology, and Medical School OWL, Bielefeld University, Germany and [2]Department of History, Princeton University, USA
Corresponding author: Lara Keuck; Email: lara.keuck@uni-bielefeld.de

Abstract

This essay offers an overview of the importance of testing for forming objects of knowledge and decision making since the late nineteenth century, focusing on applications in the health and human sciences. We argue that scientific reductionism and bureaucratic priorities – as well as ideas of fairness, feasibility, and marketability – have shaped the development and institutionalization of standardized tests. Against this background, we zoom in on a key socio-epistemological question that has guided our collection: how do experts, patients and regulatory decision makers know that tests and their data are trustworthy? In the twentieth century, 'validation' became the term of art to discuss processes of showing that a test actually measures what it is intended to measure, and for certifying the reliability of test data for assessment and regulation. In this introduction, we draw on this issue's articles and further literature to offer a scaffold of three historical phases of validating tests and highlight some overarching themes that cut across this chronology: testing for equivalence; the politics of statistical validation; and the contestation of validity by users, regulators and critics.

> As science has moved beyond the walls of the laboratory, so have tests. They are mobile, but they are not immutable. Tests are materialized networks: they are the manifestation of practices, ideas, values, and institutions.[1]

The cover image of this issue of *Themes* is a scientific photograph that was published in 1978 in a pharmacological article (Figure 1). We see a rat in a long glass cylinder that is half-filled with water. It is difficult to ascertain whether the nose of the rat is sufficiently above the water line to let it breathe. Its voluminous white body, its mouth and most of its whiskers are below the water surface. The limbs are at rest, making the scene recall wet preparations of dead animals in natural-history museums. The vessel stands on a shiny clean ground that reflects the water and the lighting, adding to the scientific allure of the picture, while the black-and-white photograph emphasizes the contrast between the white rat and the black background, making it appear almost schematic, prototypical. In the article, the photograph was presented as depicting the so-called rat immobility

[1] Alexandra Hui, Mara Mills and Viktoria Tkaczyk: an introduction', in Hui, Mills and Tkaczyk, *Testing Hearing: The Making of Modern Aurality*, New York: Oxford University Press, 2020, pp. 1–19, 5.

© The Author(s), 2026. Published by Cambridge University Press on behalf of British Society for the History of Science. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

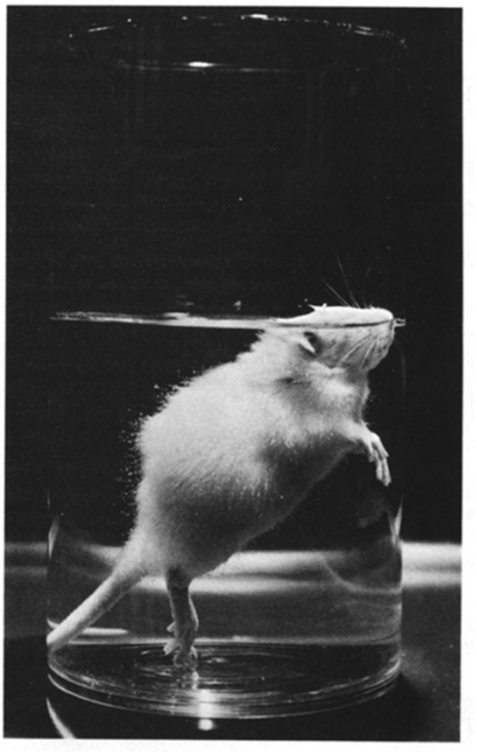


Figure 1. 'Rat showing typical posture of immobility after 10 min immersion in water'. Image and caption reprinted from Roger D. Porsolt, Guy Anton, Nadine Blavet and Maurice Jalfre, 'Behavioural despair in rats: a new model sensitive to antidepressant treatments', *European Journal of Pharmacology* (1978) 47(4), pp. 379–91, Figure 1, at 380. Copyright (1978), with permission from Elsevier.

state. The publication was entitled 'Behavioural despair in rats: a new model sensitive to antidepressant treatments'.[2]

As Yingying Han explains, this model of despair in rats was developed to help identify therapies for human depression.[3] Her contribution to our collection is a history of the rise and fall of using the forced swim test to validate antidepressants; that is, to generate sound evidence for deciding whether to test a drug candidate in clinical trials with human patients. Animal resignation was a stand-in for human depression. Pharmaceutical treatment that resulted in a rodent not giving up its (hopeless) effort to escape watery confinement was validated as effective. We have chosen to make it the cover image because it anchors the abstract notions of validation and regulation in the concreteness of an animal

[2] Roger D. Porsolt, Guy Anton, Nadine Blavet and Maurice Jalfre, 'Behavioural despair in rats: a new model sensitive to antidepressant treatments', *European Journal of Pharmacology* (1978) 47(4), pp. 379–91. The image is Figure 1 of the article.

[3] Han, this issue. On the epistemic challenges of modelling human behavioural disorders in animals see Nicole C. Nelson, *Model Behavior: Animal Experiments, Complexity, and the Genetics of Psychiatric Disorders*, Chicago: University of Chicago Press, 2018. Nelson discusses the forced swim test on pp. 102–4. The modelling of human depression in animals to test psychopharmaceuticals is tightly connected to the pharmaceutical industry's impact on transforming depression into a common disorder that general practitioners would diagnose and treat. See Lucie Gerber and Jean-Paul Gaudillière, 'Marketing masked depression: physicians, pharmaceutical firms, and the redefinition of mood disorders in the 1960s and 1970s', *Bulletin of the History of Medicine* (2016) 90(3), pp. 455–90.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

experiment. Han's article is one of several contributions in this issue to examine the shifting status of animal testing for assessing the potential effects of biologicals, drugs or chemicals on human health.[4]

The granularity of the picture reminds us today of a time before digital photography, thereby giving it the aesthetic allure of a historical source. The picture's ambition to strip off context to depict a transposable experiment conveys how standardized testing and model validation streamlined pharmacological research for drug approval by regulatory agencies. This instantiation of validity raises questions about the materiality and contingency of standardized tests in behavioural research. How did the amount of time until a rodent became immobile come to indicate 'behavioural despair', and how did this measurement become accepted as a valid model for testing antidepressants in humans?

The working hypothesis of this special issue is that the twentieth century might be described as a century of standardized testing – and of the recurring question of a test's 'validity': does a test really measure what it intends to measure? We thus bring historical and sociological scholarship on the ubiquity of testing in modern societies into conversation with debates on how to define, implement and assess the quality of a test.[5] Whether it is a psychological aptitude test, a diagnostic test in psychiatry, a randomized controlled trial in evidence-based medicine, or an animal model in pharmaceutical or chemical regulation – this issue of *Themes* considers how routine testing in the health and human sciences has become a pervasive aspect of modern life, informing decision making in many spaces, including clinics, courtrooms and boardrooms. In looking at various settings and a diversity of scientific methods, from chemical assays to statistics, our volume features cases in which a given test was authorized or contested as a suitable, reliable and informative guide to what it is meant to measure. Our contributors focus on efforts to establish tests related to human health and subjectivity, where the stakes are particularly conspicuous, and where various kinds of variability and fallibility are inescapable.[6] In part due to these complexities, the instruments, media and protocols of tests in the health and human sciences play an important role, and often become objects of testing and calibration themselves. As the editors of another special issue that dealt with testing as sociotechnical assemblages noted, 'no testing without media – no media without testing'.[7]

This introduction explains, first, the collection's aims to unpack validation and to analyse standardized testing as a form and object of regulation. It then situates the contributions within a tentative framework of three historical phases of validation. Next, we describe three cross-cutting themes – testing for equivalence, the politics of statistical validation and contesting validity – along which we have clustered the articles in the table of contents. We conclude by reflecting on how our relation to the historical trajectories of *Testing Knowledge* shifted during the five years in which we worked on this collective research project.

[4] See the articles by Hüntelmann, Lanier-Christensen, Phillips, Creager, Thomas and Han, and the discussion in Nelson and Keuck, this issue.

[5] On the ubiquity of testing, we take particular inspiration from Trevor Pinch, '"Testing – one, two, three … testing!" Toward a sociology of testing', *Science, Technology, & Human Values* (1993) 18(1), pp. 25–41; and Alexandra Hui, Mara Mills and Viktoria Tkaczyk, who have argued that 'testing as such became an enduring and wide-ranging social practice in the modern period. Individuals engage with tests from the moment they are born. Tests are built into daily lives with astonishing pervasiveness'. Hui, Mills and Tkaczyk, 'Testing hearing', op. cit. (1), p. 3.

[6] On the role of human subjectivity see Li, this issue; on variability as a challenge to reproducibility see Nelson and Keuck, this issue.

[7] 'Kein Test ohne Medien – kein Medium ohne Test'. Sebastian Gießmann and Carolin Gerlitz, 'Test: Einleitung in den Schwerpunkt', *Zeitschrift für Medienwissenschaft* (2023) 15(2), pp. 10–19, 12. We are grateful to Viktoria Tkaczyk for bringing this point to our attention.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

Unpacking validation

The term 'valid', derived from the Latin *validus*, has long been used as a general synonym for 'good', 'accurate' and, more specifically, 'legally sound'.[8] Early in the twentieth century, the meaning and use of evaluative concepts became more differentiated: in psychology, 'validity' became a technical term of art for judging whether a test was informative about the thing that one wanted to test.[9] This meaning of 'validity' was tied to modern articulations of not directly observable traits that could only be captured with the help of tests: in 1904, the British psychologist Charles Spearman famously inscribed Francis Galton's mathematical theory of association between two things into experimental psychology, conceptualizing 'general intelligence' ('g') as a 'latent variable'.[10] The rapid uptake and harsh contestation of psychological tests for mental abilities were propelled by their abundant use in many different sectors, such as in assessments for job selection, in school entry exams and in recruitment to military programs, as well as in psychological research.[11] By the mid-century, practical applications of testing and managing 'human resources' in industrialized societies coalesced with theoretical research in the new field of psychometrics and advances in statistical methods.[12] The sophisticated differentiation of sub-forms of validity concepts were deeply inflected by the US American psychology profession's concerns about the costs and benefits of their published tests' commercialization.[13]

Institutions removed from the human sciences were struggling with analogous problems of safeguarding their scientific authority while expanding their economic and societal impact. What measurements – and *whose* measurements – would certify the quality of agricultural products, the efficacy of an immune serum or the safety of synthetic chemicals?[14] The term 'validation' began cropping up in these other regulatory contexts, as individuals and regulatory bodies sought ways to legitimate the tests they relied upon. Over the twentieth century, validation became a shorthand for a certified process that included assumptions about a test's informativeness, reliability and applicability, tasking committees and agencies to negotiate compromises between these evaluative dimensions. Our collection takes care to unpack the respective understandings of concepts and methods

[8] See 'valid', *Oxford English Dictionary*, at www.oed.com/dictionary/valid_adj?tab=meaning_and_use#16114511 (accessed 14 June 2026). In logic (a field beyond our purview), validity conveys that an argument or inference has a logical form such that if the premises are true then the conclusion must also be true.

[9] See the definition of educational psychologist Kelley: 'The problem of validity is that of whether a test really measures what it purports to measure, while the question of reliability is that of how accurately a test measures the thing which it does measure.' Truman Lee Kelley, *Interpretation of Educational Measurements*, Yonkers-on-Hudson: World Book Company, 1927, p. 14.

[10] Charles E. Spearman, 'The proof and measurement of association between two things', *American Journal of Psychology* (1904) 15, pp. 72–101; Spearman, '"General intelligence" objectively determined and measured', *American Journal of Psychology* (1904) 15(2), pp. 201–93. See also Joseph A. Buckhalt, 'A short history of g: psychometrics' most enduring and controversial construct', *Learning and Individual Differences* (2002) 13(2), pp. 101–14. For a discussion of the mathematical foundations of modelling latent variables and their influence on the further development of test validity theory in psychometrics see Keith A. Markus and Denny Borsboom, *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning*, New York: Routledge, 2013.

[11] Stephen Jay Gould, *The Mismeasure of Man*, New York: Norton, 1981; John Carson, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940*, Princeton, NJ: Princeton University Press, 2007; Daniel Koretz, *Measuring Up: What Educational Testing Really Tells Us*, Cambridge, MA: Harvard University Press, 2008; Robert W. Lissitz, ed., *The Concept of Validity: Revisions, New Directions, and Applications*, Charlotte, NC: Information Age Publication, 2009. Howard Wainer and Daniel H. Robinson, *Testing and the Paradoxes of Fairness*, Cambridge: Cambridge University Press, 2025.

[12] Grant Mongin, 'Who is the "one best man?" Taylorism and personality tests (1924–1955)', *Labor History* (2025) 66(1), pp. 77–89.

[13] See Kindel, this issue.

[14] These examples are discussed in the contributions of Lanier-Christensen, Hüntelmann and Creager, this issue.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

in different times, places and disciplines, being aware that 'valid' and 'validated' were and continue to be associated with distinct, though partly overlapping, vernacular, technical and epistemic genealogies and meanings.

Despite the polysemy of 'validation', the contributions to *Testing Knowledge* interrogate, collectively, a prevailing assumption of the twentieth century, namely that validated tests, however understood in detail, constitute actionable knowledge. Is a pharmaceutical safe and effective to administer? Should a chemical be banned because it may cause cancer? Is a veteran's mental-health diagnosis attributable to their war experience, entitling them to public benefits?[15] Throughout the twentieth century, experts and regulatory authorities turned to testing to provide answers to such difficult questions. In turn, whether people accepted the answers reflected, at least in part, their trust in the institutions and actors that vouched for test methods and data. The compiled case studies direct our attention towards a diverse set of practices and regulations that have been mobilized in the past 150 years to guard institutional trust even in the absence of successful test performance, for instance by distinguishing a test's scientific validity from the ethics of misuse or, more mundanely, the ability to use the right test in the right way for decision making. At the most basic level, the fact that a test has been validated can become a shortcut for arguing for its scientific status, and thereby for legitimizing its role in adjudicating individual, institutional and political decisions.

Standardized tests as forms and objects of regulation

While testing is certainly not specific to modernity, our collection suggests that standardized tests became a typical feature of twentieth-century bureaucratic rationality.[16] In part this reflected ambitions to standardize technologies and products to facilitate international trade as well as the 'trust in numbers' in public life.[17] Along these lines, and more broadly, this volume complements existing work on standardization by historians, philosophers and sociologists of technology and medicine.[18] Scholars from various disciplines have pointed to how standardization serves as a form of regulation, whether formal or informal.[19] For instance, in their typology of standardization, Stefan Timmermann and Steven Epstein discuss 'procedural standards', which 'delineate the steps that are to be taken when specified conditions are met'.[20] In many cases, such 'specified conditions' arise from tests, which tie scientific procedures into regulation via standards.

[15] These questions are taken up in this issue by Phillips, Creager, Thomas and Freeborn.

[16] For some examples of premodern testing see Benjamin A. Elman, *Civil Examinations and Meritocracy in Late Imperial China*, Cambridge, MA: Harvard University Press, 2013; Elaine Leong and Alisha Rankin, 'Testing drugs and trying cures: experiment and medicine in medieval and early modern Europe', *Bulletin of the History of Medicine* (2017) 91(2), pp. 157–82; Alisha Rankin, *The Poison Trials: Wonder Drugs, Experiment, and the Battle for Authority in Renaissance Science*, Chicago: University of Chicago Press, 2021.

[17] Joanne Yates and Craig Murphy, *Engineering Rules: Global Standard Setting since 1880*, Baltimore: Johns Hopkins University Press, 2019; Theodore Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*, Princeton, NJ: Princeton University Press, 1995.

[18] Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences*, Cambridge, MA: MIT Press, 1999; Stefan Timmermans and Marc Berg, *The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care*, Philadelphia: Temple University Press, 2003; Martha Lampland and Susan Leigh Star (eds.), *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Ithaca, NY: Cornell University Press, 2009; Oliver Schlaudt and Lara Huber (eds.), *Standardization in Measurement: Philosophical, Historical, and Sociological Issues*, London: Routledge, 2016.

[19] E.g. Nils Brunsson and Bengt Jacobsson, *A World of Standards*, Oxford: Oxford University Press, 2002, which includes a section on 'Standardization as regulation', pp. 10–13.

[20] Stefan Timmermans and Steven Epstein, 'A world of standards but not a standard world: toward a sociology of standards and standardization', *Annual Review of Sociology* (2010) 36, pp. 69–89, 72.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

Sociologists of knowledge were among the first to point to testing itself as a promising subject. In 1989, Donald Mackenzie argued that the testing of technologies was 'a major locus of the construction of facts'.[21] A few years later, Trevor Pinch singled out testing as a key site for sociological inquiry, 'because in testing something is at stake'.[22] Both of them viewed the analysis of testing as a fruitful way to extend the sociology of science into technology, where, as Mackenzie put it, 'it will often be found that test procedures rest on the assumption of the validity of certain scientific theories'.[23] He argued that test validity remains open to challenge, a point echoed by many of the cases in this collection.

Contemporary with Mackenzie and Pinch, some sociologists and anthropologists of medicine focused on testing, with more of an emphasis on its social implications. They worked against a backdrop of increasing medical testing generally, as well as the impact of new diagnostics and screens based on genetics.[24] In 1989, Dorothy Nelkin and Laurence Tancredi warned that the conjunction of these new kinds of tests with information technologies meant that employers, government authorities and insurers could effectively use personal data as a means of social control.[25] They drew examples from disease screening, workplace testing and standardized exams in schools, where more and more students were being diagnosed with learning or behavioural disorders. Rayna Rapp's *Testing Women, Testing the Fetus* offered a much more nuanced representation of the new world of genetic testing by looking at how women experienced prenatal screening.[26] She differentiated responses to amniocentesis among US American women from different ethnic backgrounds, religions and classes, while being sensitive to the tension between reproductive rights and disability rights when it comes to medical abortion.

Since the turn of the century, testing has featured in scholarly analyses of medical systems and markets. Shobita Parthasarathy's analysis of genetic testing for breast cancer in the US and the UK pointed to how different health care systems led to testing regimes of different kinds. As she argues, a testing system's structure 'articulates the rights, responsibilities, and authority of its users'.[27] Likewise, historians of science and medicine have highlighted the manifold interactions between users, producers and regulators, in particular regarding the research and development of pharmaceuticals, and the advertisement of drugs and diagnostic devices to doctors and patient–consumers.[28] Especially after the Second World War, the predominant political economy of pharmaceutical companies in

[21] Donald Mackenzie, 'From Kwajalein to Armageddon: testing and the social construction of missile accuracy', in David Gooding, Trevor Pinch and Simon Schaffer (eds.), *The Uses of Experiment: Studies in the Natural Sciences*, Cambridge: Cambridge University Press, 1989, pp. 409–35, 411.

[22] Pinch, op. cit. (5), p. 26.

[23] Mackenzie, op. cit. (21), p. 413. Mackenzie argued that technological testing is analogous to scientific experiment, in that technical considerations alone can never force closure of a controversy. Here he was building on Collins's classic work on the challenge of replication, tacit knowledge and the 'experimenter's regress'. H.M. Collins, *Changing Order: Replication and Induction in Scientific Practice*, London: Sage, 1985.

[24] While much of medical testing happens in specialized laboratories, this period also saw the emergence of medical tests developed for direct use by consumers, emblematized by pregnancy tests. Jesse Olszynko-Gryn, *A Woman's Right to Know: Pregnancy Testing in Twentieth-Century Britain*, Cambridge, MA: MIT Press, 2023; and, on the history of clinical testing, Stanley Joel Reiser, *Medicine and the Reign of Technology*, Cambridge: Cambridge University Press, 1978, Chapters 6, 9.

[25] Dorothy Nelkin and Laurence Tancredi, *Dangerous Diagnostics: The Social Power of Biological Information*, New York: Basic Books, 1989.

[26] Rayna Rapp, *Testing Women, Testing the Fetus: The Social Impact of Amniocentesis in America*, New York: Routledge, 1999.

[27] Shobita Parthasarathy, 'Architectures of genetic medicine: comparing genetic testing for breast cancer in the US and the UK', *Social Studies of Science* (2005) 35(1), pp. 5–40, 8; see also Parthasarathy, *Building Genetic Medicine: Breast Cancer, Technology, and the Comparative Politics of Health Care*, Cambridge, MA: MIT Press, 2007.

[28] Harry M. Marks, *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900–1990*, Cambridge: Cambridge University Press, 1997; Daniel P. Carpenter, *Reputation and Power: Organizational Image and*

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

the global North has promulgated a testing regime based on screening innovations and gatekeeping intellectual property.[29]

In addition, and more closely aligned with our interests, the realm of medical research and diagnosis lends itself to examining the socio-epistemological dimensions of testing. As Ilana Löwy's work makes clear, this approach is indebted to Ludwik Fleck's 1935 analysis of the Wassermann test for syphilis, and his understanding of diagnostic stabilization.[30] While Fleck emphasized how the Wassermann test allowed for the generation of new questions and a new understanding of syphilis in the first decades of the twentieth century, Löwy adds two further layers to this story. For one, she shows how the test was incisive in shaping the concepts of a diagnostic test's specificity and sensitivity, despite its wrong assumptions about how exactly the serological test functioned. Tests not only developed into key scientific tools but also became objects of analysis. Many contributions in our volume can be read in this regard: the abundance of testing paved the way for elaborating a terminology and methodological toolkit of evaluative concepts and practices to test the test and guide the user. For another, Löwy reminds us in her more recent work on the diagnosis and management of the infectious diseases of syphilis and Zika in Brazil that if a test is not incorporated into an infrastructure that makes tests and treatments available for those who really need them, it will not succeed. Löwy shows that the failure to get a clear epidemiological picture of the 2015–17 Zika epidemic was due not only to the methodological challenges of testing (particularly those related to the heterogeneity of disease manifestation and the cross-reactivity of antibodies to dengue) but also to the lack of economic and political interests in solving the problem once it lost its allure as an international public-health threat.[31] Testing in the context of scarcity and injustice reminds us that we should not reduce the socio-epistemic realm to researchers and professionals alone.

Against this background, we might also revisit Peter Keating and Alberto Cambrosio's concept of a biomedical platform to describe how instrumentation and skilled practice are combined in the realm of biology-based medical diagnosis. They argue that biomedical platforms, such as the one they examine for immunophenotyping, require extensive regulation, much of it by professionals.[32] Immunologists must agree on common classification systems to standardize results and physicians must agree on how to interpret the readout of immune cell types in terms of a disease diagnosis. Processes of validation are woven into these systems of testing and classification at each stage. Our collection makes use of this analysis but puts more emphasis on the gaps and failures of coordination, and the alliances between scientists and political and economic stakeholders. Some contributions detail the personal and sometimes existential impact that tests can have on the tested individuals. In this regard, we want to also point readers to a book by Katrin Solhdju that shares the same title as our

Pharmaceutical Regulation at the FDA, Princeton, NJ: Princeton University Press, 2010; Jean-Paul Gaudillière and Volker Hess (eds.), *Ways of Regulating Drugs in the 19th and 20th Centuries*, London: Palgrave Macmillan London, 2012.

[29] Jean-Paul Gaudillière, 'Pharmaceutical innovation and its crisis: drug markets, screening, and the dialectics of value', *BioSocieties* (2021) 16(3), pp. 411–32.

[30] Ilana Löwy, '"A river that is cutting its own bed": the serology of syphilis between laboratory, society and the law', *Studies in History and Philosophy of Biological and Biomedical Sciences* (2004) 35(3), pp. 509–24; Koichi Kameda, Ann H. Kelly, Javier Lezaun and Ilana Löwy, 'Imperfect diagnosis: the truncated legacies of Zika testing', *Social Studies of Science* (2021) 51(5), pp. 683–706, esp. 686. As Löwy notes in the first article (p. 512), while Fleck's work has been cited in the literature of science and technology studies, especially since an English translation was published in 1979, scholars only began using his work for its attention to practice in the 1990s.

[31] Ilana Löwy, 'Absent data, diagnostic practices and the Zika conundrum', *Statistique et société* (2024) 12(3), pp. 1–14.

[32] Peter Keating and Alberto Cambrosio, *Biomedical Platforms: Realigning the Normal and the Pathological in Late-Twentieth-Century Medicine*, Cambridge, MA: MIT Press, 2003.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

collection.[33] Solhdju stresses the life-changing consequences of genetic testing in presymptomatic carriers of Huntington's disease and outlines alternative ways of meaning-making that are in conversation with, but also contest, the scientific authority of the test.

Lastly, *Testing Knowledge* can be read as a historical commentary on current complaints that much scientific knowledge itself does not stand the test of scrutiny, particularly in the health and human sciences. The controversial methodologist John Ioannidis has gone so far as to claim that 'most published research findings are false'.[34] We would situate these methodological concerns with the reproducibility of research, coming from psychologists and statisticians concerned with *p*-hacking on the one hand and pharmaceutical companies concerned with the irreproducibility of preclinical findings on the other, in a historical trajectory. In the US, earlier demands for accountability in research surfaced in the 1970s and 1980s at the same time as regulatory agencies developed procedures for certifying reliable data for their decision making.[35] Mark Rushefsky, Sheila Jasanoff and Sarah Vogel have referred to the valorization of such procedures in terms of 'regulatory science', which differs from basic research in its emphasis on standardized protocols and materials, strict record-keeping and laboratory inspections.[36] Our collection offers a genealogy of how these audit mechanisms for science emerged over time as part of an apparatus to establish validated knowledge in regulatory contexts, and puts the recent prioritization of traceability as the measure of good science into critical perspective.

Three phases of validation

The history of validating measurements, models and tests is closely connected to longer trajectories of the roles of quantification for the making of the modern state. However, the case studies in this special issue show that the problem of validity also reflects more specific developments within the twentieth-century history of the science–market–policy interfaces. Taking this perspective, we differentiate three interlocking historical phases that have given shape to validation. These phases should be seen as overlapping layers, not as clearly demarcated epistemes; the testing related to standardization and market coordination of the late nineteenth century continued, even as methodological and conceptual innovations regarding validating tests on humans became signifiers of professional differentiation in postwar societies, to which was added the institutional entrenchment of testing regimes in late modern regulatory science.

First, the increasing use of more or less standardized testing for scientific, medical, military and industrial purposes in the *late nineteenth century* introduced a set of methodological and organizational sociological problems: who should evaluate the tests' performance, and how should this be done? We argue that the establishment and social acceptance of committees and licensing bodies – many of which were forerunners of national or international regulatory agencies – were tightly connected to this new need for *evaluating testing methods*. Industrialization provided the context for the spread of new kinds of standards, which went hand in hand with technologies for measurement and testing, many from engineering, but

[33] Katrin Solhdju, *Testing Knowledge: Toward an Ecology of Diagnosis, Preceded by the Dingdingdong Manifesto*, Goleta, CA: Punctum Books, 2021.

[34] John P.A. Ioannidis, 'Why most published research findings are false', *PLoS Medicine* (2005) 2(8), e124. Our last essay in the collection, by Nelson and Keuck, takes up this issue explicitly in their notion of the 'regulatory ethos' that has been migrating into scientific research itself.

[35] See Colleen Lanier-Christensen, 'Between science and administration: regulatory epistemology, corporate influence, and endocrine-disrupting chemicals', *DePaul Law Review* (2025) 74(2), pp. 399–460.

[36] Mark E. Rushefsky, *Making Cancer Policy*, Albany: State University of New York Press, 1986; Sheila Jasanoff, *The Fifth Branch: Science Advisers as Policymakers*, Cambridge, MA: Harvard University Press, 1990; Sarah A. Vogel, *Is It Safe? BPA and the Struggle to Define the Safety of Chemicals*, Berkeley: University of California Press, 2013.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

also drawing on methods from analytical chemistry and bacteriology.[37] Although standards for unit measurement were associated with national governments, many industrial standards (like that for the ohm) were set by private bodies, either scientific or industrial.[38] In the US, a non-profit Underwriters Laboratory (UL) certified the electrical safety of products and materials, which received a UL listing and label.[39] This was emblematic of industrialized nations' reliance from 1870 to the 1920s on industry associations to set building-code, fire protection and other safety standards. Testing was a key part of this system of voluntary standard-setting, as evidenced by the establishment of an American Society for Testing and Materials in 1898.[40] Through the twentieth century, the expansion of government regulation, particularly over the safety of new technological systems, did not supplant the role of private bodies in regulatory activities. Legal scholar Timothy Lytton discusses examples from fire safety and kosher food to argue that private certification plays an important regulatory role and that it 'compare[s] favorably with government regulation'.[41]

In our collection, the contributions by Colleen Lanier-Christensen and Axel Hüntelmann study the establishment of national and international committees at the intersection of academic science and industrial stakeholders in this period. Lanier-Christensen looks at the success of the Association of Official Agricultural Chemists (AOAC) in the US in validating tests for agricultural products since the 1880s, ranging from the amount of nitrogen and phosphorus in fertilizer to the content of butterfat in milk. This de facto regulatory body took up similar problems in toxicological testing in the 1970s, without success. Hüntelmann's article underlines the challenges of method validation when working on biological materials, which are notoriously variable. As he shows, the serum evaluation method of the Royal Prussian Institute for Serum Research and Serum Testing enabled the international certification of biologicals, but only by standardizing and regulating a long chain of procedures, tests and subjects, including bacteria, guinea pigs, horses and technicians. Hüntelmann highlights that validation of methods involved moral authority and trust, which in his case came to be grounded in an institutionalized internationalism rather than in national bodies.

In Jingwen Li's article, the variable test subjects are humans themselves. She focuses on a contested glasses prescription in 1926 in Shanghai where the authority of a new machine for measuring nearsightedness could not override the ultimate arbiter of effective vision correction: consumer experience. Eram Alam also deals with the complexities of tests involving human subjects, moving beyond agriculture and health to the labour market. She traces the

[37] Edward W. Constant, 'Scientific theory and technological testability: science, dynamometers, and water turbines in the 19th century', *Technology and Culture* (1983) 24(2), pp. 183–98; Amy E. Slaton, *Reinforced Concrete and the Modernization of American Building, 1900–1930*, Baltimore: Johns Hopkins University Press, 2001; Roland Wittje, 'Absorption, transmission, reflection: testing materials in the laboratory', in Tkaczyk, Mills and Hui, *Testing Hearing*, op. cit. (1), pp. 243–69. On analytical chemistry and bacteriology see José Ramón Bertomeu-Sánchez, 'Managing uncertainty in the academy and the courtroom: normal arsenic and nineteenth-century toxicology', *Isis* (2013) 104(2), pp. 197–225; and Lanier-Christensen's and Hüntelmann's essays in this issue.

[38] Yates and Murphy, op. cit. (17); Simon Schaffer, 'Late Victorian metrology and its instrumentation: a manufactory of ohms', in Robert Bud and Susan E. Cozzens (eds.), *Invisible Connections: Instruments, Institutions, and Science*, Bellingham, WA: SPIE Optical Engineering Press, 1992, pp. 438–74; Bruce J. Hunt, 'The ohm is where the art is: British telegraph engineers and the development of electrical standards', *Osiris* (1994) 9, pp. 48–63; Graeme J.N. Gooday, *The Morals of Measurement: Accuracy, Irony, and Trust in Late Victorian Electrical Practice*, Cambridge: Cambridge University Press, 2004.

[39] Scott Gabriel Knowles, 'Private standards, public goods: underwriters laboratories and American safety engineering in the consumer era', *Historia Scientiarum* (2017) 27(1), pp. 42–62.

[40] Yates and Murphy, op. cit. (17), p. 47.

[41] Timothy D. Lytton, 'Competitive third-party regulation: how private certification can overcome constraints that frustrate government regulation', *Theoretical Inquiries in Law* (2014) 15(2), pp. 539–71, 542.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

development of US American certification policies for assessing the equivalence of immigrants' medical education throughout the twentieth century. Here the standardization had to do with human capital, and the validity of licensing tests encountered the racial politics of post-1965 immigration. These four authors show how political, economic and scientific interests – and the tensions among them – could complicate efforts to set standards and quality assurance mechanisms. When these efforts succeeded, testing was valued for facilitating market coordination.[42]

Testing in the health and human sciences gained a new theoretical momentum in the postwar period. The second phase of validation is marked by an increasing *theorization of evaluative concepts*, and their strategic employment in a time of increasing *professional differentiation* and scaling up of research infrastructures in the *mid-century*.[43] Human bodies and capacities had not been excluded from the materials testing of the nineteenth century (as emblematized by Francis Galton's attempts to measure and study human variation), but the emergence of psychology as an experimental discipline focused greater attention on testing people.[44] In a pragmatist move, unsettled questions about measurement of qualitative states and mental abilities had been transferred to a subsidiary question in the first decades of the twentieth century: were the tests fit for their purpose in assisting the selection of humans in military and educational contexts?[45] This implied that the epistemological question of the informativeness of a test was coupled to political ideology and societal structures that constrained the identification of purposes that testing should serve.

Alex Kindel's paper on psychological test standards examines how academic psychologists sought to protect their field's reputation against the misuse and misinterpretation of tests. The commercial success of psychological tests, which were published and in widespread use by the mid-twentieth century, prompted the American Psychological Association to form a committee to set recommendations on test validity. Contrary to the examples discussed in the first phase, this committee did not mean to authorize standards. Rather, it used the spectre of poor practices outside the academy to make the question of a test's validity methodologically demanding, so much so that it even annoyed a doyen of the field. This strategy ultimately helped to safeguard the exclusive role of professional psychological and psychometrics expertise in identifying malpractice in the lucrative world of educational testing. The work of this committee resulted in a paper authored by Lee Cronbach and Paul Meehl that became a heavily cited milestone not only in the human sciences but more broadly among practitioners theorizing validity. They offered a refined technical vocabulary including the new concept of construct validity that resonated with deep theoretical and epistemological problems that had already been raised by nineteenth-century attempts to make abstract human qualities measurable.[46]

In postwar science, the high stakes and many possibilities of misuse and error provided a justification for cultivating and differentiating professional testing expertise, which in turn led to the proliferation of profession-specific concepts. For example, in psychiatry, 'diagnostic validity' referred at the same time to theories about *what* a diagnosis should

[42] See also Christoph Gradmann and Jonathan Simon (eds.), *Evaluating and Standardizing Therapeutic Agents, 1890–1950*, London: Palgrave Macmillan, 2010.

[43] Regarding the expansion and professional differentiation in the human sciences, we take inspiration from Andrew Scull, 'Contending professions: sciences of the brain and mind in the United States, 1850–2013', *Science in Context* (2015) 28(1), pp. 131–61.

[44] Kurt Danziger, *Constructing the Subject: Historical Origins of Psychological Research*, Cambridge: Cambridge University Press, 1990.

[45] Gould, op. cit. (11); Nicholas Lemann, *The Big Test: The Secret History of the American Meritocracy*, New York: Farrar, Straus and Giroux, 1999; Carson, op. cit. (11); Wainer and Robinson, op. cit. (11).

[46] Michele Luchetti, 'Epistemic circularity and measurement validity in quantitative psychology: insights from Fechner's psychophysics', *Frontiers in Psychology* (2024) 15, article 1354392, DOI: 10.3389/fpsyg.2024.1354392.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

capture and to theories about *how to* measure and evaluate the degree to which a given psychiatric classification fulfils these ideals.[47] In the case of schizophrenia diagnosis, this was actually intended to test the psychiatrists (rather than the patients) and regulate how they diagnose – turning an old controversy about psychiatric schools into a problem of incoherent epidemiological data.[48] The advent of computers, digitalization and big data came with its own set of assumptions regarding data quality, which had repercussions for evaluating psychiatric diagnosis.[49] Yet, in the courtroom, the usefulness of a diagnostic category could be assessed quite differently, as Alfred Freeborn's contribution to our collection shows. It looks at a British law that gave veterans the benefit of the doubt in their war-related disability pension appeals and put pressure on courts to develop procedures for granting claims despite unsettled questions regarding causality and the nature of psychiatric diagnoses. In several cases in the late 1960s, the new medico-legal diagnostic category of stress-induced schizophrenia was put to use, allowing psychiatric experts to flag uncertainty (rather than reliability in the case of psychiatric epidemiology), while allowing judges to base their decisions on psychiatric expertise.

Several contributions in this volume probe the co-development of test methods, the theorization of evaluative concepts, and the contestation of what this implied about the measured constructs. Michael McGovern's contribution looks at how scientific testing was explicitly called upon by civil rights activists to identify unjustness and racial discrimination in US employment. He highlights how the emerging practice and technical vocabulary of statistical validation were mobilized in important ways. Eventually, the activists' grounds of evidence were rejected on the basis that one could not infer motive from statistical correlation: racial discrimination, the judges argued, was a category restricted to cases of intent. In disrupting the connection between test results that demonstrated unequal treatment and the normative definition of the concept of discrimination, the government could disregard scientific test data as being statistically significant but legally uninformative.

In contrast, Christopher Phillips's and Sarah Marks's contributions highlight how statistical validation became deeply embedded in conceptualizing, testing and interpreting a treatment's effectiveness, the assessment of costs and benefits, and evidence-based allocation of health care expenditures in the second half of the twentieth century. At the same time, both authors stress that statistics served a cause rather than dictating what to do. Sarah Marks shows, for instance, that the postwar British institutionalization of clinical trials that were randomized was not a mere methodological issue, but a pragmatic solution that was compatible with socialist ideals of fair treatment of all patients. Likewise, in the 1990s the 'what-works' policy of the New Labour years was perfectly suited to reform the management of the National Health Service (NHS) in a similar spirit, and established the National Institute for Health and Care Excellence (NICE) as the British iteration of evidence-based medicine. Phillips shows how statisticians at the US National Institutes of Health (NIH) developed methods to forge regulatory consensus about extrapolating toxicology experiments with laboratory animals to low-dose exposures of humans. Marks's and Phillips's case studies form a nexus to the *third phase of validation* as they are situated in two newly founded institutions, the NHS in the UK and the NIH in the US.

Postwar politics contributed to the successive transformation of *validation into a means for regulation*. From the *1960s onwards* regulatory sciences and agencies gained importance as gatekeepers of test validity and application. This is emblematized in the emergence of

[47] Eli Robins and Samuel B. Guze, 'Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia', *American Journal of Psychiatry* (1970) 126(7), pp. 983–87.

[48] Alfred Freeborn, 'Testing psychiatrists to diagnose schizophrenia: crisis, consensus, and computers in postwar psychiatry', *History of the Human Sciences* (2025) 38(2), pp. 18–39.

[49] Andrew S. Lea, *Digitizing Diagnosis: Medicine, Minds, and Machines in Twentieth-Century America*, Baltimore: Johns Hopkins University Press, 2023.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

environmental regulation in the US. Laws such as the Clean Air Act and Clean Water Act mandated environmental sampling and monitoring for pollutants.[50] In 1962, new safety requirements for pharmaceuticals went into effect in the US, followed by more stringent requirements for pesticides and oversight of all commercial chemicals.[51] These regulatory regimes rested on extensive toxicological testing, including new kinds of *in vitro* laboratory tests that sought to obtain information without more costly animal experiments. Validation – showing that certain tests are reliable and predict select end points reasonably well – was (and remains) an essential component of this kind of regulatory science. The heavy emphasis on safety testing transmuted public concerns about health and pollution from new chemicals into manageable technopolitical procedures.[52] Fleshing out this historical trajectory concerning the control of chemicals, Colleen Lanier-Christensen, Valentin Thomas and Angela Creager each examine how different tests were validated, and expose the challenges with stabilizing a test for regulatory use. If the first phase of validation was about standard-setting for marketable products, and the second phase about professional differentiation and negotiating expertise, the third phase is characterized by a preoccupation with regulatory procedure and managing risk.

The final paper extends the third phase by showing how the boundary between regulatory science and academic research has corroded over time. Nicole Nelson and Lara Keuck argue that a 'regulatory ethos' has been extended into mainstream science, as illustrated by the introduction of 'preregistration' schemes into academic research. This regulatory device, in which scientists must declare at the outset how they will collect and analyse their data, was instituted in clinical trials to ensure the integrity of drug efficacy reports. As with other audit mechanisms that are increasingly used in the public sector, there are questions as to who benefits from this kind of accountability, and what kinds of costs they entail for basic research.

The structure of the issue

We have introduced the papers in this issue along a three-phase but overlapping chronology. Here we want to highlight some themes that cut across these three time periods. The articles in this issue are organized to highlight these three areas of dispute around validity.

Papers in the first section, 'Validating "good enough"', feature cases in which testing was used for establishing equivalence among products, professionals or psychological

[50] Jongmin Lee, 'CHESS lessons: controversy and compromise in the making of the EPA', in James Rodger Fleming and Ann Johnson (eds.), *Toxic Airs: Body, Place, and Planet in Historical Perspective*, Pittsburgh: University of Pittsburgh Press, 2014, pp. 127–51.

[51] James T. O'Reilly, *Food and Drug Administration*, Colorado Springs: Shepard's/McGraw-Hill, 1993, pp. 3-19–23; John Wargo, *Our Children's Toxic Legacy: How Science and Law Fail to Protect Us from Pesticides*, New Haven, CT: Yale University Press, 1996, Chapter 5; Angela N.H. Creager, 'To test or not to test? Tools, rules, and corporate data in U.S. chemicals regulation', *Science, Technology, & Human Values* (2021) 46(5), pp. 975–97. As Lee Vinsel has shown, the performance standards imposed on automobiles between the 1960s and 1977 to improve emissions and safety resulted in other forms of testing to satisfy federal regulation. Lee Jared Vinsel, 'Designing to the test: performance standards and technology change in the U.S. automobile after 1966', *Technology and Culture* (2015) 56(4), pp. 868–94.

[52] David Demortain, *The Science of Bureaucracy: Risk Decision-Making and the US Environmental Protection Agency*, Cambridge, MA: MIT Press, 2020; and the set of papers by Colleen Lanier-Christensen, Emmanuel Henry, Angela Creager and Valentin Thomas (with an introduction by Emmanuel Henry, Valentin Thomas, Sara Angeli Aguiton, Marc-Olivier Déplaude and Nathalie Jas) in the collection 'Beyond the production of ignorance: the pervasiveness of industry influence through the tools of chemical regulation', *Science, Technology, & Human Values* (2021) 46 (5), pp. 911–1135. For a straightforwardly political interpretation of the validation of a new carcinogenicity testing for pharmaceuticals see John Abraham and Rachel Ballinger, 'The neoliberal regulatory state, industry interests, and the ideological penetration of scientific knowledge: deconstructing the redefinition of carcinogen in pharmaceuticals', *Social Studies of Science* (2012) 37(5), pp. 443–77.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

assessments. In three of these historical cases, the demand for validated tests came directly from the economic sector, to enable agricultural products to be graded (Lanier-Christensen), biologicals to be evaluated for potency (Hüntelmann), and physicians to be certified as licensable to practise in another country (Alam). In Kindel's paper on psychological test standards, psychologists sought to protect their reputation against those administering tests in commercial settings. As the work of Joanne Yates and Craig Murphy illustrates, the role of standards in facilitating economic integration and coordinating markets is well studied, but there is less attention to how stakeholders and scientists come to agreement on what constitutes a good measure.[53]

Our second set of papers, 'The politics of statistical validation', takes up a different challenge in testing, namely how government agencies and courts could use test data as evidence for efficient or fair decisions.[54] The authors of the papers in this section study the tensions and co-dependencies of the scientific or medical ways of assessing evidence with political ideology. Here we include the emergence of evidence-based medicine (Marks) and the process for 'invaliding' veterans seeking war disability pensions (Freeborn) in Britain. Picking up a key case in US civil rights law, McGovern probes how debates over the validity of standardized tests were deployed in lawsuits about racial discrimination in employment. Phillips focuses on the American regulatory system, tracing the impact of the framework that government statisticians developed for extending dose–response curves for carcinogens to low-dose exposures, an important extrapolation for Food and Drug Administration decisions about the safety of food additives and drugs. While the statisticians were aware of the limitations of their methods, both bureaucrats and industry representatives embraced these tools as part of the expanding world of risk assessment for governing chemical safety. By the 1980s, risk assessment became the dominant framework for using toxicology tests in regulatory decision making. As Phillips shows, one element of the politics of risk assessment was the black-boxing of statistical methods to try to prevent scientific uncertainty from destabilizing regulatory legitimacy.

The third set of papers, 'Contesting validity', focuses on the critics of validation who call into question tests and evidence-based standards. Li takes us to a Shanghai courtroom, where a medical malpractice case over a prescription for glasses challenged the reliability of a new refraction-determining machine (the retinoscope) and the trustworthiness of the optician against the subjective experience of the patient. Her article points to the inescapability of subjectivity in diagnostic tests on humans, particularly tests of sensory perception. Angela Creager and Valentin Thomas take us into the heart of the world of regulatory science in the 1970s and 1980s, as national agencies and international bodies struggled to establish toxicology standards to be used in regulating products of the chemical industry. The title of Thomas's essay, 'In the absence of adequate data on humans', explains the demand for laboratory tests that could supply regulators with reliable data on chemical safety. The challenges with validating those tests for regulatory use, such as at the US Environmental Protection Agency (EPA), opened up technical issues that could be exploited by stakeholders, particularly in the private sector. As both Thomas and Creager observe, raising questions about the scientific legitimacy of the rodent carcinogenicity test could

[53] Yates and Murphy, op. cit. (17).

[54] There is a literature on forensic testing, though it does not connect directly to this issue. See Ken Alder, 'To tell the truth: the polygraph exam and the marketing of American expertise', *Historical Reflections/Réflexions historiques* (1998) 24(3), 487–525; Ian A. Burney, 'Testing testimony: toxicology and the law of evidence in early nineteenth-century England', *Studies in History and Philosophy of Science* (2002) 33(2), pp. 289–314; Michael Lynch, Simon A. Cole, Ruth McNally and Kathleen Jordan, *Truth Machine: The Contentious History of DNA Fingerprinting*, Chicago: University of Chicago Press, 2008; Jennifer L. Mnookin, 'The uncertain future of forensic science', *Daedalus* (2018) 147(4), pp. 99–118; and Mitra Sharafi, *Fear of the False: Forensic Science and the Law of Crime in Colonial South Asia*, New York: Columbia University Press, 2026.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

lead government agencies to make policy changes favouring toxicology tests that industry preferred. We have already mentioned the rodent test that is at the centre of Han's essay. The 'forced swim test' was developed for use in identifying promising candidates for treating human depression. Its validity was contested not only by sceptical psychologists but also by animal welfare groups. In each of these articles, historical actors contested the validity of a test in the service of a particular (sometimes explicitly political) agenda.

Concluding reflections

Overall, this collection shows that the validation of tests for regulation has been pertinent in a wide range of health and human sciences, from serology and psychiatry to civil rights law and environmental protection, from the late nineteenth century to the present. Many of our papers refer to classic works of Theodore Porter, Sheila Jasanoff and Lorraine Daston and Peter Galison to discuss test validation in relation to their histories of quantification and coordinating trust, of technocracy and regulatory science, of epistemic fears and forms of objectivity, including regulatory objectivity.[55] While these frameworks remain useful points of departure, our authors also complicate and nuance them. Our collection shows that test validation is not just a passive outcome of broader characteristics of modernity such as standardization and institutionalization, but that relevant ideals, assumptions, practices and regulations – as well as a new class of technical specialists – have shaped the role of testing in late modern science, policy, legal debates and everyday life.

The interweaving of normative, methodological and socio-epistemological issues in notions of validity epitomizes the modern effort to both express and corral concerns about uncertainty, legitimacy and relevance within the technical realm. For instance, ethics concerning human subjects or protection were negotiated in and through setting up committees that enforce technical guidelines (which may be remote from considerations of actual well-being).[56] Committees are granted the power to limit or enforce testing. As our contributors make clear, such implicit normativity of validation plays a regulatory role on several levels: first, it is referred to when assessing the soundness of specific test procedures: is a test good enough? Second, the implicit normativity of validation is called upon in cost–benefit assessments of testing: is the testing worth its costs and side effects? Third, the normativity becomes foregrounded in meta-level discussions on distinguishing good from bad; that is to say, when validation is deployed in disputes over standards dividing 'good' research from 'misconduct'.[57] While situated mainly in twentieth-century North America and Western Europe, the contributions offer a critical perspective on the hegemony of Western actors and imperial thought in setting the methods that certify knowledge for international policymaking – or market.[58]

When we started to work on this collection, the COVID pandemic was in full swing, and the connection between trust in institutions and trust in science, diagnostic tests and

[55] Porter, op. cit. (17); Jasanoff, op. cit. (36); Lorraine Daston and Peter Galison, *Objectivity*, New York: Zone Books, 2007. Peter Keating and Alberto Cambrosio contributed the category of regulatory objectivity in *Biomedical Platforms*, op. cit. (32).

[56] Laura Stark, *Behind Closed Doors: IRBs and the Making of Ethical Research*, Chicago: University of Chicago Press, 2012; Noortje Jacobs, *Ethics by Committee: A History of Reasoning Together about Medicine, Science, Society, and the State*, Chicago: University of Chicago Press, 2022.

[57] For related concerns see the papers in this issue by Kindel, Marks, Han and Nelson and Keuck.

[58] See the papers in this issue by Li, Hüntelmann, Thomas and Alam. Here we are in conversation with the work of others that have taken a similar approach, e.g. on the use of clinical test subjects in postcolonial settings. Adriana Petryna, *When Experiments Travel: Clinical Trials and the Global Search for Human Subjects*, Princeton, NJ: Princeton University Press, 2009.

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

medical interventions received much public attention. Our articles do not take up the development of COVID testing or distrust in vaccination, though the vaccine trials and validation of diagnostic tests we all witnessed were deeply indebted to the processes and institutions discussed in this issue. Perhaps more to the point, our contributions, being more historical, do not address the ways in which scientific and regulatory institutions backed by state authority have been put under new pressures in the 2020s.[59] Nonetheless, we cannot help but comment on the timing of our endeavour. During the five years that we spent on this project, from its inception in 2021 to publication in 2026, the present has changed. We are witnessing the mainstreaming of pejorative stances towards institutions and distrust in scientific authority in a way that made us ask at times whether we had reached a tipping point in our positionality to the past: is our collection of papers still a genealogy of the contemporary?

In some cases, surely it is – for instance with respect to understanding the reproducibility crisis as an outcome of the century of testing. Even the Trump administration's EPA appeals to validation to authorize their intention to eliminate animal testing in toxicology. A recent EPA press release explained how clarity 'drives innovation' with the following assertion: 'A robust toolbox of validated methods gives industry regulatory certainty, rewards safer product design, and strengthens American scientific leadership'.[60] (Unmentioned is that the new methods save industry a lot of money.) On a similar track, Trump's Health and Human Services trumpets its reliance on 'gold standard science' in its funding and regulatory policies, even as many scientists argue that Robert Kennedy's decisions on vaccination and other issues are at odds with medical consensus.[61] But to the degree that technical certification no longer persuades a wary public, we might have already flipped from a genealogy into a historical reconstruction that gives us access to a past world view differing radically from our social reality today. Just as we begin to see the granularity of our black-and-white cover image of the rat in the glass vessel as being antiquated, a world in which a combination of bureaucracy and science was imagined to achieve social equality and fairness might need more explanation today.

Acknowledgements. Most of our chapters emerged from a pandemic-era call for papers from scholars whose work addresses how tests and the processes of validating them are tied to formal and informal regimes of regulation. Over the first half of 2022, about two dozen participants, including historians, sociologists and an artist, gathered through a regular webinar series to discuss twelve papers. Revised and new papers were discussed at an authors' workshop held in Berlin in September 2022. We would like to thank all contributors to and commentators on the Validation and Regulation Working Group for participating in this endeavour. We gratefully acknowledge audience members at the history-of-science seminar of Leopoldina, Germany, and the International Congress for the History of Science and Technology in Dunedin, New Zealand, for their valuable comments on early versions of this introduction. In the autumn of 2024, at the Institut d'études avancées, Paris, Mathilde Cohen, Benjamin Hegarty and Anne Le Goff provided excellent feedback on a draft, and more recently Jean-Paul Gaudillière, Gideon Manning and Viktoria Tkaczyk kindly reviewed a pre-final version of the paper. We also wish to thank the team members of the Max Planck Research Group Practices of Validation in the Biomedical Sciences for organizational and editorial support in the first phase of this project; Princeton University, the Max Planck Society and Bielefeld

[59] For some inspiring examples see Dan Bouk, 'Materializing COVID', *Isis* (2020) 111(4), pp. 783–6; Ilana Löwy, 'Ludwik Fleck where are you now that we need you? Covid-19 and the genesis of epidemiological facts', *Somatosphere*, 16 October 2020, at http://somatosphere.net/2020/ludwik-fleck-where-are-you-now.html (accessed 11 June 2026); Claire Beaudevin, Luc Berlivet, Soraya Boudia, Catherine Bourgain, Maurice Cassier, Jean-Paul Gaudillière and Ilana Löwy, '"Test, test, test!" Scarcity, tinkering, and testing policy early in the Covid-19 epidemic in France', *Medicine Anthropology Theory* (2021) 8(2), pp. 1–31.

[60] US EPA Office of Chemical Safety and Pollution Prevention, 'Trump EPA takes new action to eliminate animal testing', 2 June 2026, at www.epa.gov/newsreleases/trump-epa-takes-new-action-eliminate-animal-testing (accessed 7 June 2026).

[61] US Department of Health and Human Services, *Gold Standard Science at HHS*, at www.hhs.gov/gss/index.html (accessed 6 July 2026).

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press

University for funding our research; and the *BJHS Themes* editorial team, in particular Trish Hatton and Jahnavi Phalkey, for steering us so deftly through the peer review process and editing of this collective volume. Lastly, this endeavour is indebted in several ways to the work of Ilana Löwy, who also attended our initial discussions. We dedicate this essay and the special issue to her.

Competing interests. The authors declare none.

Cite this article: Lara Keuck and Angela N. H. Creager, 'Testing knowledge: validation and regulation in the health and human sciences,' *BJHS Themes* (2026), pp. 1–16. https://doi.org/10.1017/bjt.2026.10044

https://doi.org/10.1017/bjt.2026.10044 Published online by Cambridge University Press