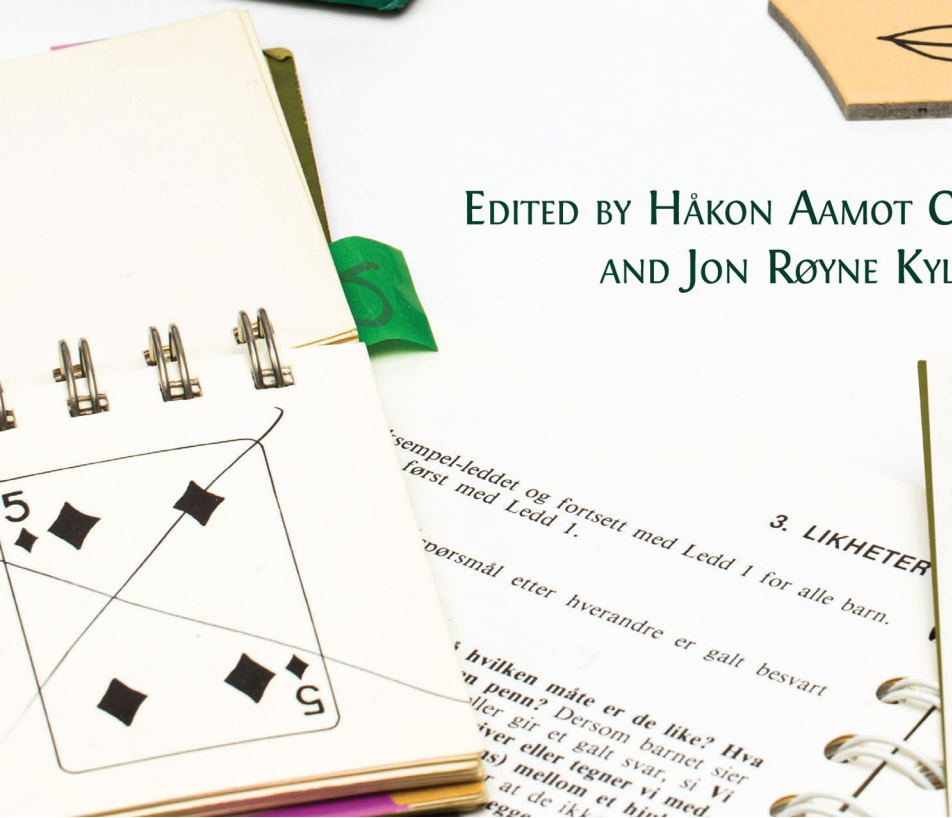
Historicizing IQ Testing

Intelligence Assessments and their Role in Norwegian Society from the 1900s to the Present

Edited by Håkon Aamot Caspersen and Jon Røyne Kyllingstad



https://www.openbookpublishers.com

Paperback: 978-1-80511-617-2
Hardback: 978-1-80511-618-9
PDF: 978-1-80511-619-6
HTML: 978-1-80511-621-9
EPUB: 978-1-80511-620-2
DOI: 10.11647/OBP.0471

Cover image: © Helge Brekke/MUV, UiO, CC BY-SA 4.0
Cover design: Jeevanjot Kaur Nagpal

The Cultural History Museum, University of Oslo, supported this Open Access publication.

Historicizing IQ Testing in Norway: Introduction

Håkon Aamot Caspersen and Jon Røyne Kyllingstad

The power of Intelligence Quotient (IQ) tests rests on their legitimacy as scientific instruments, and the fact that they are developed, controlled, and used by experts with legal, public, and academic authority. Intelligence testing, as it developed historically, has offered specific ways of visualizing and codifying individual performance into 'writing and numbers, quotients, scores, profiles', thus making 'the individual knowable, calculable and administrable' in new ways.[1] But there has rarely been a consensus about exactly what IQ tests measure, or clear agreements about their fairness (or lack thereof) towards different ethnically, socially, or racially defined groups.

IQ tests are deeply institutionalized in Norwegian society today, forming part of many official assessment, diagnostic, and selection procedures, often in cases that involve the allocation of specific rights and resources. They contribute to practices of constructing and categorizing the normal and abnormal, and have existential, pedagogical, social, legal, and economic consequences, both individually and societally.

Given their influence, it is crucial to develop a critical understanding of intelligence tests and the practices of testing. This book contributes towards realizing such an objective by historicizing IQ testing, asking questions about how these practices came into being, and how they achieved the functions they currently have in our society. Our aim is

1 Nikolas Rose, *Governing the Soul: The Shaping of the Private Self* (London: Free Association Books, 2nd ed. 1999), p. 143.

 https://doi.org/10.11647/OBP.0471.00

to open up historical and contemporary understandings of IQ testing, inviting a closer look into this particular socio-technological practice and its associated disciplinary boxing practices (to borrow phrasing from Susanne Bauer in chapter 8).

Throughout the chapters, themes that recur include the importance of distinguishing between how IQ testing and intelligence have been explained, and how IQ tests have been used—that is, the practical operationalization of IQ and intelligence. An IQ test, in this sense, is multi-faceted—a practical clinical and bureaucratic tool, a scientific method, as well as a communicative device that spans national borders, societal sectors, and disciplines. It is often in its practical applications, and in conversations across disciplinary fields, that one sees the paradoxical reification of IQ as an objective and frequently black-boxed technological measure in society, despite more nuanced understanding among the holders of test expertise. It is such themes, relationships, apparent paradoxes, and more, that we examine in this volume.

The Significance of a Test Score: An Example from Norway

In 2017, journalistic investigations brought media attention to multiple cases of misdiagnosis of intellectual disability, following the use of one specific IQ test—the third edition of the Wechsler Intelligence Scale for Children (WISC-III).[2] A low score on an IQ test is a key criterion for diagnosing intellectual disability, and the WISC tests are the most commonly used IQ tests for children in Norway and internationally. The investigation showed that recommendations of sterilization and cases of forced adoption grounded in mental disability misdiagnoses partly followed from a low IQ score caused by a flaw in the Norwegian version of the test. This testifies to the far-ranging consequences an IQ test may have. A disability diagnosis will inform and significantly shape lives, and people's legal rights.

2 See Asbjørn Øyhovden, Connie Bentzrud and Olav T. Hustad Wold, 'Metoderapport SKUP 2017: IQ Testing og Diagnosehelvete', in *SKUP: Stiftelsen for en kritisk undersøkende presse* (2017). Examples from the news coverage: Asbjørn Øyhovden and Connie Bentzrud, 'Slo alarm om IQ: Barn og unge kan ha blitt stemplet som utviklingshemmede etter feil' (TV2.no, 26.02.2017), Asbjørn Øyhovden, 'Strid om ansvaret for at barn og unge kan ha fått feil IQ-diagnose' (TV2.no, 01.03.2017), Asbjørn Øyhovden and Connie Bentzrud, 'Helseministeren vil rydde opp i IQ-testing' (TV2.no, 28.02.2017).

As early as 2005, two years after the introduction of the Norwegian version, experienced psychologists had warned of a mismatch between their own clinical judgment and scores produced by this test.[3] Despite warnings and advice of caution in interpreting the results, the test was widely used. The mismatch, later confirmed by frequent cases of misdiagnosis, indicates too much confidence in test results, and their institutionalized position as indispensable tools in diagnostic work. It was later documented that the Norwegian version of the test consistently indicated lower scores than its predecessor and comparable test batteries. An analogy employed to characterize the problem was that the scale showed the wrong 'weight'.[4] The critique did not question IQ as a valid measurement of ability, but located the problem in the technical process of adapting this specific test for use in Norway. That, in turn, produced an "error" in the IQ scale. However, these technical issues were related to the economic, legal, institutional, and societal framework within which this work was undertaken. The development of the Norwegian version was not influenced only by scientific considerations. The state rejected a request from the Norwegian Psychological Association for the funding of a Norwegian adaptation. Instead, the work was carried out by a newly established private company that aimed to finance the project through test sales. For this endeavor to be economically viable, taking the smaller scale of the Norwegian test market into consideration, they had to keep costs down.[5]

As several of the following chapters show, issues of adaptation, translation, and statistical standardization of a test—including the social work necessary to legitimate them as scientific tools of measurement—are intimately tied in with different historical trajectories and social and economic concerns that have influenced the development of IQ tests. Far from infallible instruments, IQ tests and corresponding ideas about intelligence, like other technologies of measurement, are constructed, and need to be understood as historically contingent 'scientific objects'.[6]

---

3 As reported by, for instance, Nina Strand, 'WISC-III skaper uro' (Psykologitidskriftet.no, 01.06.2025).

4 Thor Herman Andreasen and Øyvind Fallmyr, 'Kan vi leve med manglende norsk normering på WISC, WAIS og WPPSI?' (Psykologitidsskriftet.no, 01.05.2006).

5 *Tidsskrift for Norsk psykologforening*, 40.1 (2003), 52–53; 39.2 (2002), 130–131; and 40 Nr. 8 (2003), 704–705. *WISC-III Norsk versjon Administrasjonsveiledning* (Oslo: Assessio Norge 2003), pp. 5–7.

6 Lorraine Daston, 'Introduction: The Coming into Being of Scientific Objects', in

This entails discussions that go beyond internal epistemic critique or explorations of the validity of a test in relation to certain scientific criteria. The starting point of this book is that a test is a cultural object that participates in producing the very thing it purports to measure.

The Binet–Simon and the Wechsler Test Families: A Brief Introductory Sketch

The WISC-III belonged to one of two "families" of intelligence tests—the Binet–Simon test and its successors, and the Wechsler tests—that have been central to the international history of IQ, and which play key roles in this book. The history of psychological testing began in the nineteenth century, with the English polymath Francis Galton as a key contributor. Galton coined the term 'eugenics', and advocated the improvement of humankind through interventions in the biological reproduction of humans to ensure that the talented had more children than the less talented. Galton saw intelligence as closely linked to sensory capacities and speed of reaction, and developed a method for measuring intelligence based on non-verbal, sensor-motoric tests. It is, however, common and reasonable to localize the birth of IQ testing proper to Paris at the beginning of the twentieth century, arising from the work of the psychologist Alfred Binet and physician Théodore Simon. This was born partly in response to the need for a tool that the French school system could use to identify children who required special education measures. Binet and Simon developed an intelligence scale with standardized scoring procedures, published in 1908, which involved the individual child responding to a set of tasks, and test results were measured against the average test result among Parisian children of different age groups. By comparing these scores, the child's "mental age" was calculated. This opened the way for the understanding of intelligence as a uniform entity unevenly distributed between individuals. The first instances of intelligence testing in Norway were done using more or less direct translations of the French Binet–Simon test.

While gaining a limited foothold in France, the Binet–Simon test had

Biographies of Scientific Objects, ed. by Lorraine Daston (Chicago: University of Chicago Press, 2000). See also F. A. Hanson, *Testing, Testing: Social Consequences of an Examined Life* (Berkeley: University of California Press, 1993).

its greatest success in the USA.[7] It was first translated and propagated as an instrument in the scientific study of the "feebleminded" by psychologist Henry H. Goddard, the director of the Vineland Training School for Feebleminded Boys and Girls. In 1916, his colleague Lewis Terman transformed, adapted, and standardized the test with scoring norms adjusted to an American population. It was in this new US version—the so-called Stanford–Binet—that the Intelligence Quotient was introduced. The IQ, originally developed by German psychologist William Stern, was calculated by dividing mental age by biological age and multiplying by one hundred. In contrast to Binet, who had been agnostic about the effect of nature and nurture on measured intelligence age, Terman largely portrayed IQ as a direct measure of inherited 'general intelligence'.[8]

In Britain, Galton's early influence was carried forth by figures like Charles Spearman and Cyril Burt. Spearman observed that children's performance across seemingly unrelated school subjects was statistically correlated, and in 1904, he put forward the theory that these correlations were produced by a single, underlying, general intelligence. Cyril Burt—who eventually succeeded Spearman as professor of psychology at University College London in 1932—was appointed as psychologist to the London County Council in 1913. He adapted the Binet–Simon and created his own tests while promoting Galton's hereditarian stance on intelligence. From his institutional position and his strong interest in the psychology of individual differences, with support from Spearman, he played a significant role in institutionalizing both intelligence testing and educational psychology as scientific research programmes. Alongside other psychologists and educational thinkers of the time, including in Norway, Burt saw IQ testing as a tool that offered the combination of meritocratic ideas of educational selection, a scientific passion for measurement, and a commitment to child-centred education.[9]

---

7 John Carson provides an excellent in-depth account of these comparative developments in *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics 1750–1940* (Princeton: Princeton University Press, 2007).

8 See Jim Porter, chapter 13, for a further expansion on these developments, building on previous historical accounts.

9 See Adrian Wooldridge, *Measuring the Mind: Education and Psychology in England, c.1860–1990* (Cambridge: Cambridge University Press, 1994) and Gillian Sutherland, *Ability, Merit and Measurement: Mental Testing and English Education 1880–1940* (Oxford: Clarendon, 1983) for good overviews of ability measurement and educational

Despite multiple international adaptations of its French original, it is the US-based Stanford–Binet that perhaps can be counted as the most influential, and it became the model for many other tests widespread in the education system, healthcare, and armed forces in the US and beyond. It was also the first intelligence test that was successfully adapted and normed for and made mainstream in Norway in 1931. By 'normed', we here refer to how the calculation procedures for converting raw test scores into an IQ score were adjusted, based on the average distribution of test scores among Norwegian school children. It was only after World War Two that the Binet tests were challenged as the most influential way to measure intelligence, through the introduction of the Wechsler tests.

The first Wechsler test was launched by psychologist David Wechsler in 1939.[10] He did away with the practice of calculating the IQ score by dividing mental age by biological age, and introduced the method that dominates today. According to this method, an IQ score is calculated by statistically comparing the test scores of an individual with the average score of a "representative" group that stands for the population and age group that the individual belongs to. The scale is constructed on the principle that general intelligence, as measured by IQ, is normally distributed within the total population, forming a bell curve. An IQ score of 100 is set as the median, and 68 percent of the population scores between 85 and 115, defined as the normal range.

The Wechsler tests come in numerous versions for pre-school children, school-age children and adolescents, and adults, and the latest versions have been translated and adapted for use in nations around the world. Since at least the 1970s, they have been the most used intelligence tests in Norway. They are standard assessment tools used by the Educational Psychological Service (EPS) and in psychiatric healthcare. It is through being certified to use this test that students in psychology, educational psychology, and special needs education learn to use intelligence tests and psychometric tests more generally.

The Binet–Simon and Wechsler test families are only a few among a

psychology as it developed in the UK. While interesting parallels and differences certainly exist between the UK and Norwegian cases, it is the standard history from the US that takes centre stage in the following pages, as it was US versions of tests that dominated in the work of Norwegian historical translations.

10 See David Wechsler, *The Measurement of Adult Intelligence* (Baltimore: Williams & Witkins, 1939).

huge number of differently composed IQ tests. These varied tests and test batteries are like complex, unsystematic kinship systems. Not only do different versions develop and build upon each other, but as they develop, they borrow and combine new elements and subtests from a range of other tests, whether IQ tests or otherwise.[11] The basic structure of a Wechsler test, however, remains the same as its predecessors. They are scored according to the number of right or wrong answers to questions of increasing difficulty on different subtests, understood today to relate to a set of cognitive factors associated with intelligence.

Intelligence tests are tests by proxy. It is not "intelligence" in and of itself that is measured, but purported—and historically changing—representations of it, through subtests that stand for, or index, underlying cognitive factors. Many IQ tests today will provide not only a full-scale IQ score, but also a cognitive profile meant to indicate specific strengths and difficulties.

The Historiography of IQ Testing as Viewed from the Periphery

There is extensive literature on IQ tests, their historical development, and their role in the discipline of psychology and broader society. While a comprehensive review of the historiography of IQ testing is beyond our scope here, excellent overviews can be found elsewhere, and relevant literature is also referred to in the individual contributions to this volume.[12] It is, however, important to note that much of the historical

11 As seen, for example, in Corwin Boake, 'From the Binet–Simon to the Wechsler–Bellevue: Tracing the History of Intelligence Testing', *Journal of Clinical and Experimental Neuropsychology*, 24.3 (2002); Jacy L. Young, 'Test or Toy? Materiality and the Measurement of Infant Intelligence', *History of Psychology*, 18.2 (2015).

12 See Anette Mülberger Rogele, 'Biographies of a Scientific Subject: The Intelligence Test', in *Oxford Research Encyclopedias: Psychology* (Oxford: Oxford University Press, 2020) for an excellent, concise and recent overview. See also Anette Mülberger, 'The need for contextual approaches to the history of mental testing', *History of Psychology*, 17.3 (2014), 177–186 and the special issue of the *History of Psychology* it introduces. John Carson, in turn, provides a thorough comparative analysis of intelligence testing in the US and France in the first part of the twentieth century. For good accounts of their UK expansion see also Gillian Sutherland, *Ability, Merit and Measurement: Mental Testing and English Education 1880–1940* (Oxford: Clarendon, 1983) and Wooldridge, *Measuring the Mind*. For a disciplinary "textbook account", see, for instance, John D. Wasserman's 'A History of Intelligence Assessment: The Unfinished Tapestry', in

literature is framed by critical discussions about IQ testing and its societal consequences. On the one hand, there are highly critical narratives that emphasize IQ tests as scientifically dubious and socially harmful.[13] This critical stance, that often roots itself in a Foucauldian critique of power, highlights the tests as tools for sorting and governing individuals and populations, and often points to the historical connections between IQ testing, scientific racism, and eugenics. On the other hand, within the field of psychometrics and in the disciplinary histories of psychology and IQ testing textbooks, the development of intelligence testing is often portrayed as a history of steady scientific progress, with "bad" testing practice largely presented as something of the past.

A uniting feature of many historical accounts, however, is that the history of IQ testing has often been narrated as a story where IQ tests grow, travel, expand, and conquer the world. As the historian of psychology Annette Mülberger has noted, this is a story with a lot of repetition and overlap, perhaps even showing 'a dangerous tendency toward creating [and recreating] myths'.[14] Furthermore, there is a propensity on both sides—critics and supporters—to overemphasize and even caricature the historical reaches of intelligence testing and the intentions of its cast of actors. The strong historiographic focus on intelligence testing and its significance in the development and cooperative triage of the psychological sciences, education, and governance risks blindness to important alternative influences. It may ignore the influence of other aspects of psychology and tests that historically have competed with IQ testing, or that came into being in cases where IQ testing was found

Contemporary Intellectual Assessment, ed. by Dawn P. Flanagan and Erin M. McDonough (Guilford, 2022). Sattler's historical account stands as an interesting comparative example of this genre, in *Assessment of Children: Cognitive Foundations and Applications* (J. M. Sattler Publishing Inc. 6th ed., 2018). Stephen J. Gould's well-known *Mismeasure of Man* (New York: W. W. Norton & Co, 1996) stands as a classic in the tradition of test critique, with *The Bell Curve: Intelligence and Class Structure in American Life* by Richard J. Herrnstein and Charles Murray (New York: Free Press, 1994) straddling the opposite side. For a recent insightful practice-based investigation of the social and material life of historical intelligence tests see Sasha Bergstrom-Katz, *On Intelligence Tests: Psychological Objects and Their Subjects*. (PhD thesis, Birkbeck, University of London, 2022). See also Kurt Danzinger, *Constructing the Subject: Historical Origins of Psychological Research* (Cambridge: Cambridge University Press, 1990).

13 See, for instance, Tom Axelsson, '*Att konstruera begåvning—Debatten om IQ* (Malmö University electronic publishing, 2012) for an overview of central voices in the critical debate on IQ.

14 Mülberger, 'The need for contextual approaches', p. 181

lacking. This is forcefully shown by Bonnie Evans in her account of the changing form and phenomenal rise and transformation of autism diagnoses.[15]

Another aspect that existing historical narratives around intelligence testing and IQ measurement often have in common is a US and Anglo-Saxon focus that is additionally characterized by what John Carson describes as a 'meta-narrative', where standardization processes, such as the rise of the IQ scale, are framed as an 'imperial story, in which a metropolitan institution defines [...] a standard, and then demands that provincials everywhere adopt that standard'. Carson points to an alternative approach which 'emphasizes less the co-optation of the periphery by the center, and more the adoption, adaptation, and repurposing of artifacts and ideas as they circulate between putative metropoles and provinces.'[16] This is our starting point—that a proper historical understanding of these tests necessitates an in-depth empirical focus, which can capture the diversity of the different local and transnational constellations that IQ testing has been and is part of.

This book is positioned in relation to and against "imperialist" narratives of the development and use of IQ tests. It is inspired by recent efforts to de-center this story, as called for by Carson and Mülberger, and to study the multiple historical trajectories in a range of places and nations, including France, Brazil, Spain, Italy, Denmark, Sweden, and the Soviet Union.[17]

Although several accounts explore important transnational and/or

15 Bonnie Evans, *The Metamorphosis of Autism: A History of Child Development in Britain* (Manchester: Manchester University Press, 2017).

16 John Carson, 'Mental Testing in the Early Twentieth Century: Internationalizing the Mental Testing Story', *History of Psychology*, 17.3 (2014), p. 254. See also Mülberger, 'The need for contextual approaches'; Mülberger Rogele, 'Biographies'.

17 For France, see Carson, *The Measure of Merit*; see also the individual contributions to Mülberger's edited special issue 'Mental Testing after 1905: Uses in Different Local Contexts', *History of Psychology*, 17.3 (2014) for examples from Spain, Italy, Brazil, and more. For Denmark, see Christian Ydesen, Bjørn Hamre, and Karen Egedal Andreasen, 'Differentiation of Students in the Early Danish Welfare State: Professional Entanglements Between Educational Psychologists and Psychiatrists', *Nordic Journal of Educational History*, 5.1 (2018). For Sweden, Tom Axelsson, *Rätt elev i rätt klass. Skola, begåvning och styrning 1910–1950* (Avh. Linköpings Universitet, 2007). For the Soviet Union, see Irina Leopoldoff, 'A psychology for pedagogy: Intelligence testing in USSR in the 1920s', *History of Psychology*, 17.3 (2014); Andy Byford, 'The Mental Test as a Boundary Object in Early-20-Century Russian Child Science', *History of the Human Sciences*, 27.4 (2014).

comparative perspectives,[18] historical studies of IQ testing have usually focused on the nation as their main unit of analysis.[19] Such studies highlight the diversity of locally anchored actors and processes, indicating that the content of what is meant by "intelligence" and the techniques and scales for measuring it may vary by context. While the nation is also a key analytical unit in this volume, the national framework is supplemented and challenged by demonstrating that sub-national local units such as municipalities, or international units such as Scandinavia, have been important in shaping the history of IQ testing in Norway, as have networks that cut across local, national, and regional boundaries. By tracing the historical movement of tests, test-related knowledge, and people within and outside Norway, and by exploring the role IQ tests have played across various sectors in Norwegian society, this volume offers novel perspectives on the project of decentring canonized IQ test narratives. It is ultimately a book about IQ testing in Norway, rather than an attempt at narrating the Norwegian history of IQ testing.

The nation is still an important analytical unit in most of the chapters, owing to the fact that testing practices in large part have been connected to processes related to national culture and administration. For example, IQ tests have usually been explicitly translated and adapted for use on a Norwegian population and marketed to a Norwegian market. Test expertise and certification, likewise, have come to be organized on a national level. Most importantly, IQ testing has been assigned tasks

18 Carson, *Measure of Merit*; Mülberger 'Biographies'; Mülberger, ed., 'Mental Testing'; Bjørn Hamre, Tom Axelsson and Kari Ludvigsen, 'Psychiatry in the sorting of schoolchildren in Scandinavia 1920–1950: IQ testing, child guidance clinics, and hospitalisation', *Paedagogica Historica*, 55:3 (2019); Christian Ydesen, *The Rise of High-Stakes Educational Testing in Denmark 1920–1970* (Bern: Peter Lang Verlag, 2011), Christian Ydesen, Kari Ludvigsen and Christian Lundahl, 'Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960', *European Educational Research Journal*, 12.1 (2013).

19 See also Paul Davis Chapman, *Schools As Sorters: Lewis M. Terman, Applied Psychology, and the Intelligence Testing Movement, 1890–1930* (New York: New York University Press, 1990); Hamilton Cravens, *Before Head Start: The Iowa Station & America's Children* (Chapel Hill: University of North Carolina Press, 1993); Clyde Chitty, 'Eugenics, Race and Intelligence', in *Education* (London: Continuum, 2007); Jim Wynter Porter, *A "Precious Minority": Constructing the "Gifted" and "Academically Talented" Student in the Era of Brown v. Board of Education and the National Defense Education Act* (Doctoral dissertation, Michigan State University, 2017); Laila Zenderland, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge: Cambridge University Press, 2001).

within nationally organized systems, such as the education system or the legal system.

From the outset, there are many reasons to consider Norway an interesting case to include in the international literature on the history of IQ testing. From early on, IQ testing has been connected to eugenics and used to identify "superior" and "inferior" individuals, races, and classes, often being criticized for contributing to the naturalization of social hierarchies. Norway is arguably of special interest in this regard. IQ testing was introduced and became institutionalized in the Norwegian school system, health services, and legal system during a period when Norway developed into what is often described as a Scandinavian-style social democratic welfare state, where the working class and the labor movement played leading roles. The rising welfare state was dominated by a strong ideal of social equality and integration, and a wide acceptance of state intervention in the socio-cultural sphere and the economy to achieve these social democratic ideals. This appears to be in sharp contrast to the social Darwinist ideas that IQ testing is—rightly or wrongly—often associated with.

Regarding IQ, "race", and eugenics, it is clear that Norway, like the USA and other western countries, saw the rise of a eugenics movement as well as the racialization of its minorities in the early twentieth century. The role of race in Norwegian history is very different from the colonial powers of Europe and the USA, with their histories of slavery and racial segregation. The period when IQ testing was introduced into Norwegian society was strongly characterized by an outspoken assimilation policy towards minorities, which aimed to create an ethnically uniform nation.

As elsewhere, the rise of IQ testing in Norway is closely related to the development of universal schooling, and a desire to classify and sort children according to ability levels. Norway was, from the late nineteenth century, characterized by a particularly strong and long-lasting drive towards a unitary school system—a public school open to children of all social backgrounds, giving everyone the same access to higher education. This ideal was increasingly brought to life during the interwar years, and the introduction of IQ testing in the same period was intertwined with the implementation of the unitary school. This trend was reinforced in the post-war era by the Labor Party, the dominating political force which saw education politics as a key element in their

struggle to eradicate class distinctions and enhance social integration. IQ tests were mainly imported from the US, but in contrast, IQ testing in Norway became part of a social democratic welfare state. This makes Norway an interesting case as compared to the history of IQ testing in "core" countries, such as the colonial powers of Europe or the US, the latter with its free-market capitalism and strong meritocratic ideals combined with racial segregation and a weak labor movement.

Finally, and as shown in many of the contributions in this book, scale matters. IQ tests are commercial products, and the size of the Norwegian test market had significant implications for the development and marketing of IQ tests in Norway. A recurring topic in the history of IQ testing is the interaction between the development of IQ testing as both a commercial product and an "objective" scientific instrument, and the institutionalization and legitimization of psychology as a profession with a monopoly on the use of this instrument. In pre-war Norway, in contrast, the "ownership" of IQ testing was shared between psychiatrists, pedagogues, and psychologists. This development may perhaps be explained by the small scale of the Norwegian academic community, which is likely is also a contributing factor to why psychology was only institutionalized as a profession after World War Two in Norway.

Existing Works on the History of IQ in Scandinavia

The last two decades have seen new studies dealing with the early twentieth-century history of intelligence and psychometrics in Denmark, Sweden, and Norway, including comparative and transnational perspectives.[20] Yet, in Norway this topic has so far been predominantly handled as a tangential aspect of other topics, such as the history of professions (psychiatry, psychology, pedagogy, or special needs education), education history, and the history of eugenics. These works focus on the period up to the 1950s within a primarily national and Scandinavian context.[21] In this book, in contrast, we have urged our

20 Axelsson, Rätt elev i rätt klass; Christian Lundahl, 'Making testers out of teachers: The work of a Swedish State Research Institute 1946–1956', *History of Education*, 48:5 (2019); Ydesen, 'The rise'; Ydesen et al., 'Differentation'; Ydesen et al., 'Creating'.

21 Froestad and Ravneberg, 'Education Policy'; Per Haave, *Sterilisering av tatere 1934–1977. En historisk undersøkelse av lov og praksis* (Norges forskningsråd, 2020); Kari Ludvigsen and Åsmund Arup Seip, 'The Establishing of Norwegian Child Psychiatry:

contributors to put IQ tests, testing practices, and related knowledge production at the centre of each case study, collectively beginning in the early twentieth century and continuing into the present. While the chapters will substantiate this project, we offer here an extended, but necessarily selective, summation of the chapters and their broad arguments. Alternatively, we invite you to delve straight into the individual contributions, all written as self-contained pieces.

The Chapters of the Book

Chapter 1, by historian of science Ageliki Lefkaditou, begins by guiding us through the early implementation of IQ testing within education and psychiatry in the 1920s and 1930s. Lefkaditou shows how a group of progressive pedagogues, psychologists, and psychiatrists formed around the special school issue and the call for a Norwegian standardized IQ test. Their IQ-testing advocacy was characterized by ideals about care and control. Identification and segregation into special schools were said to enhance the quality of life and the education of the weakly gifted child, while at the same time relieving ordinary pupils and their teachers of a burden. The chapter culminates in an account of the activities of Norway's first school psychiatrist, Johan Lofthus, who translated, adapted, and normed a Norwegian version of the Stanford–Binet, known as *Osloprøven* [the Oslo test], recognized as the standard Norwegian intelligence test from 1931 to the 1950s.

In narrating the intricate story of early Norwegian efforts to adapt the Binet–Simon and Stanford–Binet intelligence scales, the chapter sheds important light on the test itself, introducing the practical and statistical acrobatics involved in defining a representative sample and establishing a national standard. Lofthus, for instance, made significant changes in the test content while maintaining its identity as a Stanford–Binet test through this process. The chapter also historically demonstrates how supposedly stable cut-off points between categories of children were

Ideas, Pioneers and Institutions', *History of Psychiatry*, 20.1 (2009); Bodil Ravneberg, *Normalitetsdiskurserog profesjonaliseringsprosesser: en studie av den spesialpedagogiske yrkesutviklingen 1880–1990* (Doctoral dissertation, University of Bergen, 1998); Eva Simonsen, *Vitenskap og profesjonskamp. Opplæring av døve og åndssvake i Norge 1881–1963* (Doctoral dissertation, UiO, 2000).

in flux and debated. Decisions about the transfer of a child to special education could be based on complex deliberations that, in addition to an IQ score, included the teacher's professional assessment, practical, institutional, and economic considerations, as well as the social standing and authority of the parent.

Chapter 2, by historians of education Christian Ydesen, Brit Marie Hovland, and Emma Vikström, maps out the closely knitted networks of Scandinavian actors and their mutual influence on intelligence testing in the interwar years. Centring on the Nordic School Association and their journal, we learn how influential psychiatrists, psychologists, and educators operated across Scandinavia and beyond, drawing inspiration and learning from each other in the early institutionalization of IQ testing in education. The chapter points to both differences and similarities in the special school system and in the use of IQ testing across Norway, Denmark, and Sweden in a time that saw the emergence of the foundations of unitary education. The idea of one-school-for-all that today is encapsulated in the Nordic model of education, as the authors point out, was dependent on the exclusion of those deemed to be uneducable. The authors show how IQ testing was discussed as one of the most effective solutions to sort children into special schooling, for perceived benefit both to the individual children and society at large.

Beyond this shared goal, however, there was no consensus about the role of IQ testing. The chapter identifies a line between a biologically grounded scientific and medical approach—that is, intelligence as hereditary and unchanging—and a pedagogical and educational approach that emphasized practical pedagogical knowledge and educational measures. But despite some disagreements, the interaction in the association's journal is characterized by mutual recognition and 'co-operation between the guardians of science and the teachers of the special school'.[22] IQ testing appears as the preferred means and medium of such cooperation, offering a standardized way of charting the potentiality of learners, and adding scientific objectivity to conversations between different professional judgments.

Chapter 3, by historian of science and medicine Svein Atle Skålevåg and legal scholar Linda Gröning, continues this theme of IQ tests as implicated within distinct fields of professional judgment, albeit within the confines of psychiatry and law. Specifically, they trace the

22 Elgström, cited in Ydesen, Hovland and Vikström, chapter 2, this volume.

role of intelligence testing in the history of criminal unaccountability through key Norwegian criminal law reforms from 1902 up until our own time. These reforms show not only changing conceptions of unaccountability, from the absence of reason to medical pathology, but also an increasing embeddedness of intelligence testing in assessments of criminal unaccountability. In fact, the authors propose that it was intelligence tests that brought conceptions of intelligence into law—first through the notion of mental age and, much later, from the 1970s onwards, expressed through IQ, and through later integration of ICD (International Classification of Diseases) classifications into legal definitions of unaccountability.

Psychiatrists were the first to employ intelligence tests in Norway. The Binet method, alongside various diagnostic practices, became a cornerstone of forensic-psychiatric assessments of intelligence and intellectual disability from the 1920s onward. The introduction of intelligence testing into forensic psychiatry and law are here shown to be tempered by existing disciplinary expertise and structures. We see how there is an ambivalence about the use of tests from both sides, and how the use of intelligence tests is followed by the insistence that the result must be understood in relation to a broader assessment. Nevertheless, intelligence tests pose as seemingly indispensable tools to aid difficult legal decision-making, imbuing a sense of objectivity into this process. With reference to current legal practice, for instance, Skålevåg and Gröning note that there are 'cases where the guiding IQ limit still seems to be interpreted as an absolute demarcation, so that formalized IQ tests can determine who can be considered unaccountable and not.'

Chapter 4, by historian of medicine Per Haave and historian of science Jon Kyllingstad, traces how IQ testing and eugenics figured in the implementation of the Norwegian sterilization law from its inception in the 1930s up to the 1960s.

The 1934 Sterilization Act both legalized sterilization as a contraceptive method and paved the way for the involuntary sterilization of people with severe mental disabilities. Haave and Kyllingstad show how IQ tests, while not figuring in the letter of the law, functioned as a criterion in cases of the sterilization of people with a "mental deficiency" a historic precursor to the contemporary diagnosis of intellectual disability. The chapter presents the eugenic background of the act and lays out how IQ testing, together

with an assessment of what was called 'ethical level' and psychopathy, became important in psychiatric assessments of mental deficiency levels.

The deliberations involved in psychiatric discretionary assessments are highlighted through a court case that received much public attention in the mid-1950s. A woman who had been sterilized without her consent in 1942 went to court against the state and the psychiatrist who had diagnosed her as "imbecilic". She was later retested, and the contradictory higher new test results became a key bone of contention.

While the state and the psychiatrist were acquitted on the basis of "the progress of science", the case became a turning point in the practice of involuntary sterilizations, with the Sterilization Board becoming more restrictive. This was both due to the fear of new court cases, but also due to changed attitudes towards the rights of mentally disabled people and a declining support for eugenic sterilizations. Nevertheless, and at first sight paradoxically, it is concluded that this increased attention to the rights of people assumed to have a mental deficiency is likely to have led to an even greater emphasis on IQ scores in sterilization cases.

Chapter 5 by Isak Emberland, historian of professions, charts the role of psychotechnics and vocational testing in socially legitimizing psychological expertise, and cementing the psychological profession in Norway in the period 1925 to 1948, when a professional university exam in psychology was founded. Emberland offers important reminders of how IQ tests are part of other psychological tests and assessment practices. The Norwegian story is situated within the international development of applied psychology, and the chapter traces how this professional and popular movement contributed to the professionalization of what eventually became a clinical profession in Norway. Emberland shows how these international ideas were picked up and advocated by a diversity of Norwegian actors, including engineers. Psychotechnics was seen as a politically neutral scientific tool that could contribute to increased effectiveness, and help individuals find their place in society and work life. Psychotechnics, with its associated ideas, in turn created a market for psychological expertise.

A central point of analysis is the Psychotechnical Institute, part of the Oslo Vocational School, where we are also introduced to Helga Eng's work on the 'Goldsmith test' against a background of industrialization, and the emergence of the need for new forms of training beyond the

traditional master-apprentice relationship. The development of the test showcases how local needs inform test construction. We learn that intelligence testing played a role in psychotechnics, but only a supplementary one. While IQ scores were meant to predict school performance, psychotechnical testing explicitly probed for aptitudes relevant for specific crafts that 'traditionally [were] suppressed or devalued in schooling'. This was related to a wider interwar critique of the school system as not sufficiently catering to the needs of the employment sector.

Chapter 6, by political scientist Kari Ludvigsen, examines the role of IQ testing in an urban Norwegian Educational Psychology office between 1953 and 1980 through a close analysis of yearly reports. During this period, the Norwegian school system went through a significant change. Reforms in the 1950s, characterized by interwar principles of segregated special education, led to a strong expansion of municipal special schools and classes for "slow learners", as well as state schools for the disabled. This development reached a peak in the latter part of the 1960s. In parallel, the Educational Psychological Services developed into a nationwide service, and its most significant task from the start was the assessment of individual children for transfers to special education. In 1969 and 1976, there were school reforms that led to the segregation principle being replaced by a principle of integration—that all children should have access to education within the normal primary school based on their abilities.

The chapter sheds light on how such a testing practice was described, legitimized, and related to other tasks, priorities, and regulations. Early reports emphasized efforts to equip the office with sufficient testing instruments and expertise, as well as the undertaking of individual and group testing. An ever-increasing demand for individual assessments created a great deal of work pressure, and during the 1960s there was more emphasis on preventive measures and closer cooperation with schools and teachers. Reports from the late 1960s and early 1970s included some expressions of concern about the potentially stigmatizing effect of individual assessments. Throughout the period, IQ and personality testing were held up as important tools in difficult individual cases, but reports show greater emphasis on the role of the school and other environmental conditions in the child's development.

Ludvigsen concludes, however, alongside other previous studies, that

the changes in the school system did not lead to any significant breaks in test use. Throughout the period, IQ tests were used pragmatically, along with other counselling and assessment tools, to meet growing demand from teachers and parents, and to solve changing tasks provided through national and local guidelines.

Chapter 7 by Håkon Caspersen, a social anthropologist, offers glimpses into the contemporary use of IQ tests, or rather *evnetesting*—tests of general ability, as they are frequently referred to—in the EPS today. While historically grounded, it builds in large part upon observational and interview-based material, and revolves around the use of the Wechsler Intelligence Scale for Children (WISC). This is a test that continues to act as an example of best practice in IQ testing internationally, and, since the late 1960s, it has been the most used IQ test in the EPS of Norway.

Throughout the chapter, we get a sense of how practitioners speak and learn to speak of, use, critically discuss, and understand these tests today. Here it is, perhaps surprisingly, not necessarily the full-scale IQ score that is of most interest. For some, test scores figure as an initial mapping tool in a wider assessment, while for others, these scores function as a last resort when other assessment procedures fail to elucidate the issue at hand. But regardless of frequency, the WISC remains a central tool of the trade, both a professional marker and a communicative device. Practitioners note the value of the test not only for its results—given perceived external expectations of testing from the schools or the psychiatric services—but also for the information gleaned in meeting with a child. The testing situation can become a time-efficient way to assess a child's difficulties firsthand even beyond the test.

Of special note is how in Norway, it is not only psychologists or psychiatrists who conduct IQ tests. Most employees in the EPS today come from a background in education or special needs pedagogy. The chapter points to regional differences, and the likely influence of disciplinary backgrounds on test practice—although the explicit question of whether the test is used differently by a psychologist or a special needs pedagogue is left unanswered.

Chapter 8, by the science and technology (STS) scholar Susanne Bauer, reveals the use of one of the Wechsler tests as part of a clinical intervention trial in a large interdisciplinary study on kindergarten

children. The study sought to assess the potential positive effect on cognitive abilities of eating fatty fish. It was funded and supported by the aquacultural industry, together with the Norwegian Research Council. The chapter shows how the test is pulled in as one element of a much larger set of techniques for quantification of a range of variables. The employment of a Wechsler test, Bauer shows, needs to be understood on the basis of the historical existence of this test system as a well-researched and validated tool, and, not least, one that has been standardized and translated into Norwegian. It becomes the best choice because it is readily available and understood as scientifically reliable, and hence easily plugged in to larger systems of knowledge production.

Bauer analyzes the function of tables and the assorted scientific boxing practices associated with statistics in this study, and how the Wechsler tests here join the "web of causation" at the core of statistical rationality and evidence-based reasoning. Importantly, the chapter notes the significance of the institutional placing of the research, and how educational institutions have long functioned as the "go-to field" for psychometrics and other research. This type of knowledge practice—intervention studies in educational settings—habituate public institutions to specific modes of evaluation, Bauer argues, that in sum enact an ableist project where enabling better performance has become part of pedagogy, care, and the public good. This is a dynamic that has been a feature of social policy and the welfare state from its beginning, and the employment of metrics such as IQ tests continues to lend credibility to such projects.

Chapter 9, by the historian of technology and business Gard Paulsen, takes us back to the post-World-War period and traces the travels, adaptation, translation, and use of the test that, in 1954, would be published as *The Norwegian Standard of Terman–Merrill's Revision of the Binet Tests*. IQ tests are not just "scientific technology"; for most of their history, the chapter notes, they have also been commercial (and copyrighted) products, existing as books and record booklets that dictate procedural action. Here, it is the "shelf life" that is explicitly taken up, and in the Norwegian case, the longevity and popularity of a product that Norwegian publishers first considered unsellable. In fact, its very advertisement upon its publication emphasized restrictions on its sale—it was only medical doctors or other qualified personnel

that were meant to get their hands on the product. Furthermore, this was a product that had been partly standardized with the help of the work of students in pedagogy at the Pedagogical Research Institute, and that in its original American form was peddled to both teachers and psychologists alike.

Importantly, the chapter also brings us close to the materiality of the test, and the changes and logistical efforts—which even required the importing of a specific set of toys—necessary to complete this product and the work deemed necessary for it to deserve its title as "standard". This process and its end product were nevertheless determined to be unsatisfactory by those involved. But once it was released, and in the context of a developing educational pedagogical service, Paulsen shows how the test quickly established itself as a standard test, not necessarily because it was especially good, but due to lack of other suitable alternatives.

Chapter 10, by historian of science Jon Kyllingstad, uncovers the conceptualizations of culture and ideas of representability at play in the Norwegian adaptation of WISC-R in the 1970s. The background of the Norwegian standardization was the EPS requirement for an instrument to measure general ability when assessing a child's need for special education. "General ability" was here understood to serve as a measure of the child's previous learning that, in turn, was based on both the child's innate predispositions and their general life experiences. The chapter describes the criteria used for constructing a "representative" sample of pupils, on whom the Norwegian version of the test was tried out and normed (the chapters by Lefkaditou and Paulsen provide interesting comparative takes on earlier standardization processes of different tests). It shows how the adaptation process was based on the expectation that intelligence, as measured by IQ, would be lower among working-class children than among children of well-educated parents. However, while socio-economic status was a key criterion in the construction of the representative sample, cultural or linguistic differences related to ethnicity or language minority backgrounds were not taken into consideration.

This is conspicuous, because the original US version of WISC-R came into being against the backdrop of tense scientific and public debates about the cultural, social, and racial bias of IQ testing, and the Norwegian test developers had close relations with their American colleagues. Still, these controversial issues received no attention in the

Norwegian adaptation, and it seems to have been taken for granted that these debates had no relevance for Norway.

The chapter ends with a reference to studies published in the early 2000s of EPS assessment practices that highlight cases where minority-language children were tested with WISC-R, without taking into consideration the fact that they did not belong to the population that the test was adapted for. In several cases, this led to children with limited Norwegian language skills being misdiagnosed as having general language difficulties.

Chapter 11, by historian of professions Fredrik Thue, delves into a public and political controversy in the late 1950s regarding intelligence testing of schoolchildren in Oslo. The dispute was not about the role of IQ tests as a sorting mechanism for special education, but about the general use of group tests on first graders. While proponents emphasized the use of the tests as an instrument to aid the adaptation of teaching to the individual abilities of the child, critics called it an attack both on the child's personal integrity and parental rights.

Thue proposes that the IQ controversy should be seen in the context of a built-in tension in the Norwegian unitary educational system—one that, since its inception in 1884, has been seen as under democratic control from the municipality and parents, and at the same time as a tool for national integration and social equalization. Comprehensive Labor-Party-led reform work in the 1950s and 1960s led to parents' rights being overshadowed by macro-reform perspectives. In this light, the IQ controversy can be partly understood as a protest against centralization, technocracy, and Labor dominance. This may help explain why much of the criticism of intelligence testing in Norway came from the right and addressed the rights of the family and the individual vis-à-vis the state and scientific expertise. This contrasts with the US, where the criticism against IQ in the 1960s and 1970s often came from the left and focused on the racial and social biases of IQ testing.

Chapter 12, by the educational historian Kim Helsvig, explores the role of psychometric testing in the development of Norwegian school policy, comparing the rise of the social-democratic unitary school after World War Two with the OECD-inspired adaptation of the Norwegian school system at the turn of the millennium.

As the dominating political party of the 1950s, Labor sought to

prolong the unified mandatory public-school requirement from seven to nine years. The goal was both to capture the talent reserve in the population and to enhance national integration. The opposition feared a lowering of the overall level if all students were to attend the same secondary school. The last three years of school were therefore to be divided into tiered tracks. In 1954, the Institute of Educational Research at the University of Oslo was entrusted with the task of developing a test that could help advise a track in line with the students' abilities. Helsvig argues that this "maturity" test, and related psychometric expertise, became a decisive factor in legitimizing the reform. It soon lost significance, however, when the Parliament in 1963 abandoned the plan of a tracked system. During the next decade, school-policy-related psychometric research became strongly marginalized in Norway.

However, psychometrics made a strong comeback after 2001, when Norway's average score on the first Programme for International Student Assessment (PISA) survey indicated a Norwegian school system in crisis. This, combined with the rise of the ideal of an education system geared towards the production of competencies that could ensure the nation's competitiveness in the emerging knowledge economy, led to the revival of psychometry.

Like the maturity test of the 1950s, the PISA survey was meant to measure supposedly culturally neutral school-relevant skills. Both tests are very similar to IQ tests and were mainly developed as tools for political and bureaucratic control. Helsvig argues, however, that the maturity test was also developed with the pupils' interests in mind, to help them find an educational path in accordance with their abilities and interests. The psychometric wave half a century later was in contrast, mainly about governance and the aim of enhancing the population's future skills, as well as the economic competitiveness of the nation.

Chapter 13, by historian of science Jim Porter, takes Norway as a vantage point for a comparative view on the history of IQ testing in US schools and, based on insights from the chapters of this volume, discusses similarities and differences between the nations.

Around the start of the twentieth century, both nations saw an expansion of public schooling, along with increasing demands for more efficient and science-based schools, and the rise of segregated special education for students who had difficulty following the normal progression in school. Progressive educational reformers in both

countries embraced IQ tests as scientific instruments to help develop an education more adapted to the abilities of the individual child, and to single out children for special education, thus defining the boundary of normality. IQ testing in the Norwegian school system continued throughout the century to be mainly linked to special education, and only from the 1950s onwards on a nationwide scale.

In the US, however, IQ testing expanded much more rapidly and attained far more extensive functions. Porter connects this US breakthrough to three interconnected factors: 1) Racist and classist beliefs about the unequal educability of different social groups, which led to demands for differentiation; 2) educational policies which, in turn, led to organizational differentiation in schools ("tracking"), and thereby stimulated a demand for testing; and 3) market forces which affected educational policy and the rise of a testing business. The chapter shows how the mutually reinforcing relationships between these factors developed over time in the US and helped increase the significance of IQ testing in US schools, from the 1920s and into the decades after World War Two. It also discusses its difference from the Norwegian trajectory due to basic societal, economic, and political factors. Norway was too small to become a profitable market for a testing industry. The societal relevance of "race" and class were very different in Norway, with a less heterogeneous population and a strongly assimilationist policy towards minorities, coupled with a social democratic version of capitalism, and a unitary public school as part of a welfare state that developed very differently from the US.

Chapter 14 continues this comparative effort, standing as a companion piece to this introduction and the individual chapters by further elucidating a few overarching points regarding the history of IQ testing in Norway, partially glimpsed throughout this book. These include the comparatively limited breakthrough of testing in Norway versus the US, and the relative absence of discussions of "race" in the Norwegian history of IQ testing. Here, we also consider the oft-noted theme of the pragmatic and practical use of these tests, and their role in the reification of intelligence, as seen from Norway but with similarities to other countries.

The volume concludes with an Afterword by Annette Mülberger, a historian of psychology.

A Multidisciplinary Approach

Given the historically rooted but multidisciplinary nature of the book, the reader will notice different, even divergent, analytical takes on IQ testing, their scientific legitimacy, and social value. Despite some coordination, we have approached IQ testing in Norway from different starting points that together provide a multifaceted portrayal of the topic. In addition to being a contribution to international literature on the history of IQ testing, the individual chapters can be read as contributions to several other specialized and interdisciplinary discussions—including the role of IQ testing in criminal law today, the medicalization of schooling, as well as more general discussions on testing and assessment procedures in education and school psychology. Other themes that weave through the volume are the relationship between commodification and science, the legitimization of professional knowledge and power, social aspects of statistical categorization, changing conceptions of care, bureaucratization, the welfare state, and technological implementation. In spite of this far-ranging approach, there are still many gaps left to be filled in the mapping of this history, such as post-war testing in the military, the use of such tests in work and business, and, not least, the different usages of IQ tests in the healthcare system, in population-level statistics, or in scientific knowledge production more broadly (although Bauer, in chapter 8, explores the latter to some extent).

A Note on Historical Terms

The reader will come across a range of historical terms that today might be considered offensive, or at least problematic, due to their negative social connotations. They were used as descriptive markers of individuals or groups of people for categorization; examples include "backward", "idiot", "imbecile", "debile", "mentally retarded", "laggards", "mentally deficient", and "handicapped", among others. Their Norwegian equivalents are left out above, as are usages that get lost in translation. The word *rase* [race], for instance, can silence a seminar room of first-year Norwegian anthropology students ('Can we even use that term?'), while prompting politically charged discussions

amongst their counterparts in Britain or the US.[23] Even the meaning of terms that seemingly remain consistent over time, such as "IQ", change and develop, as seen above. While one speaks of IQ scores in the 1920s and the 2020s, they do not refer to the same thing across time. It is in recognition of such differences and possible contentions that the contributors have taken pains to use the terms employed historically, to ensure context and accuracy of meaning in line with changes over time. Such terms are indicated by double quotation marks upon first use in each chapter, and the original Norwegian term is often included alongside its English translation.

Bibliography

American Psychiatric Association, *Diagnostic and Statistical Manual: Mental Disorders* (Washington DC: Mental Hospital Service, 1952).

Andreasen, Thor Herman and Øyvind Fallmyr, 'Kan vi leve med manglende norsk normering på WISC, WAIS og WPPSI?' (Psykologitidsskriftet.no, 2006).

Axelsson, Tom, *Rätt elev i rätt klass. Skola, begåvning och styrning 1910–1950* (Doctoral dissertation, Linköpings Universitet, 2007).

Axelsson, Tom, *Att konstruera begåvning—Debatten om IQ* (Malmö University electronic publishing, 2012). https://doi.org/10.24834/educare.2012.1.1189

Bergstrom-Katz, Sasha, *On Intelligence Tests: Psychological Objects and Their Subjects* (PhD thesis, Birkbeck University of London, 2023).

Boake, Corwin, 'From the Binet–Simon to the Wechsler-Bellevue: Tracing the History of Intelligence Testing', *Journal of Clinical and Experimental Neuropsychology*, 24.3 (2002), 383–405, https://doi.org/10.1076/jcen.24.3.383.981

Byford, Andy, 'The mental test as a boundary object in early-20th-century Russian child science', *History of the Human Sciences*, 27.4 (2014), 22–58, https://doi.org/10.1177/0952695114527598

Carson, John, 'Mental Testing in the Early Twentieth Century: Internationalizing the Mental Testing Story', *History of Psychology*, 17.3 (2014), 249–255, https://doi.org/10.1037/a0037475

23 For a detailing of the history of race as a scientific category and its specificity of use in Norway, see Jon Kyllingstad, *Rase: en vitenskapshistorie* (Oslo: Cappelen, 2023).

Carson, John, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics,1750–1940* (Princeton: Princeton University Press, 2007).

Chapman, Paul Davis, *Schools As Sorters: Lewis M. Terman, Applied Psychology, and the Intelligence Testing Movement, 1890–1930* (New York: New York University Press, 1990).

Chitty, Clyde, *Eugenics, Race and Intelligence in Education* (London: Continuum, 2007).

Cravens, Hamilton, *Before Head Start: The Iowa Station & America's Children* (Chapel Hill: University of North Carolina Press, 1993).

Danziger, Kurt, *Constructing the Subject: Historical Origins of Psychological Research* (Cambridge: Cambridge University Press, 1990).

Daston, Lorraine, 'Introduction: The Coming into Being of Scientific Objects', in *Biographies of Scientific Objects*, ed. by Lorraine Daston (Chicago: University of Chicago Press, 2000).

Evans, Bonnie, *The Metamorphosis of Autism: A History of Child Development in Britain* (Manchester: Manchester University Press, 2017).

Froestad, Jan and Bodil Ravneberg, 'Education Policy, the Norwegian Unitary School and the Social Construction of Disability', *Scandinavian Journal of History*, 31.2 (2006), 119–143.

Gould, Stephen J., *Mismeasure of Man*, (New York: W. W. Norton & Co., 2nd ed., 1996).

Haave, Per, *Sterilisering av tatere 1934–1977. En historisk undersøkelse av lov og praksis* (Norges forskningsråd, 2000).

Hanson, F. Allan, *Testing, Testing: Social Consequences of an Examined Life* (Berkeley: University of California Press, 1993).

Hamre, Bjørn, Tom Axelsson and Kari Ludvigsen, 'Psychiatry in the sorting of schoolchildren in Scandinavia 1920–1950: IQ testing, child guidance clinics, and hospitalisation', *Paedagogica Historica*, 55.3 (2019), 391–415, https://doi.org/10.1080/00309230.2019.1597137

Herrnstein, Richard J. and Charles Murray, *The Bell Curve: Intelligence and Class Structure in American Life* (New York: Free Press, 1994).

Kyllingstad, Jon Røyne, *Rase: en vitenskapshistorie* (Oslo: Cappelen, 2023).

Leopoldoff, Irina, 'A psychology for pedagogy: Intelligence testing in USSR in the 1920s', *History of Psychology*, 17.3 (2014), 187–205, https://doi.org/10.1037/a0035954

Ludvigsen, Kari and Åsmund Arup Seip, 'The Establishing of Norwegian Child Psychiatry: Ideas, Pioneers and Institutions', *History of Psychiatry*, 20.1. (2009), 5–26, https://doi.org/10.1177/0957154X08089847

Lundahl, C., 'Making testers out of teachers: The work of a Swedish State Research Institute 1946–1956', *History of Education*, 48.5 (2019), 646–663, https://doi.org/10.1080/0046760X.2019.1565422

Mülberger, Annette, 'Biographies of a Scientific Subject: The Intelligence Test' in *Oxford Research Encyclopedias: Psychology* (Oxford: Oxford University Press, 2020), https://doi.org/10.1093/acrefore/9780190236557.013.694

Mülberger, Annette, 'The need for contextual approaches to the history of mental testing', *History of Psychology*, 17.3 (2014), 177–186, https://psycnet.apa.org/doi/10.1037/a0037487

Mülberger, Annette, ed., 'Mental Testing after 1905: Uses in Different Local Contexts' (Special Issue), *History of Psychology*, 17.3 (2014).

Porter, Jim Wynter, *A "Precious Minority": Constructing the "Gifted" and "Academically Talented" Student in the Era of Brown v. Board of Education and the National Defense Education Act* (Doctoral dissertation, Michigan State University, 2017).

Ravneberg, Bodil, *Normalitetsdiskurser og profesjonaliseringsprosesser: en studie av den spesialpedagogiske yrkesutviklingen 1880–1990* (Doctoral dissertation, University of Bergen, 1998).

Rose, Nikolas, *Governing the Soul: The Shaping of the Private Self* (London: Free Association Books, 2nd ed. 1999).

Sattler, Jerome M., *Assessment of Children: Cognitive Foundations and Applications*, 6th revision (Sattler Publishing Inc., 2018).

Simonsen, Eva, *Vitenskap og profesjonskamp. Opplæring av døve og åndssvake i Norge 1881–1963* (Doctoral dissertation, UiO, 2000).

Strand, Nina, 'WISC-III skaper uro' (Psykologitidskriftet.no, 2025).

Sutherland, Gillian, *Ability, Merit and Measurement: Mental Testing and English Education 1880–1940* (Oxford: Clarendon, 1983).

Wasserman, John D., 'A History of Intelligence Assessment: The Unfinished Tapestry' in *Contemporary Intellectual Assessment: Theories, Tests, Issues*, ed. by Dawn P. Flanagan and Erin M. McDonough (Guilford, 2022).

Wechsler, David, *The Measurement of Adult Intelligence* (Baltimore: Williams & Witkins, 1939).

Wooldridge, Adrian, *Measuring the Mind: Education and Psychology in England, c. 1860–1990* (Cambridge: Cambridge University Press, 1994).

Ydesen, Christian, *The Rise of High-Stakes Educational Testing in Denmark 1920–1970* (Bern: Peter Lang Verlag, 2011).

Ydesen, Christian, Bjørn Hamre and Karen Egedal Andreasen, 'Differentiation of Students in the Early Danish Welfare State: Professional Entanglements Between Educational Psychologists and Psychiatrists', *Nordic Journal of Educational History*, 5.1 (2018), 73–96.

Ydesen, Christian, Kari Ludvigsen and Christian Lundahl, 'Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960', *European Educational Research Journal*, 12.1 (2013), 120–138, https://doi.org/10.2304/eerj.2013.12.1.120

Young, Jacy L., 'Test or Toy? Materiality and the Measurement of Infant Intelligence', *History of Psychology*, 18.2 (2015), 103–118, https://doi.org/10.1037/a0038862

Zenderland, Laila, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge: Cambridge University Press, 2001).

Øyhovden, Asbjørn, Connie Bentzrud, and Olav T. Hustad Wold, 'Metoderapport SKUP 2017: IQ Testing og Diagnosehelvete' in *SKUP: Stiftelsen for en kritisk undersøkende presse* (2017).

Øyhovden, Asbjørn and Connie Bentzrud, 'Slo alarm om IQ: Barn og unge kan ha blitt stemplet som utviklingshemmede etter feil' (TV2.no, 2017).

Øyhovden, Asbjørn and Connie Bentzrud, 'Helseministeren vil rydde opp i IQ-testing' (TV2.no, 2017).

Øyhovden, Asbjørn, 'Strid om ansvaret for at barn og unge kan ha fått feil IQ-diagnose' (TV2.no 2017).