

## University of Groningen

### Biographies of a Scientific Subject

Mülberger Rogele, Annette

*Published in:*  
Oxford Research Encyclopedias

*DOI:*  
[10.1093/acrefore/9780190236557.013.694](https://doi.org/10.1093/acrefore/9780190236557.013.694)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2020

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*  
Mülberger Rogele, A. (2020). Biographies of a Scientific Subject: The Intelligence Test. In *Oxford Research Encyclopedias: Psychology* Oxford University Press.  
<https://doi.org/10.1093/acrefore/9780190236557.013.694>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

---

# Biographies of a Scientific Subject: The Intelligence Test

Annette Mülberger, University of Groningen

<https://doi.org/10.1093/acrefore/9780190236557.013.694>

**Published online:** 19 November 2020

## Summary

The intelligence test consists of a series of exercises designed to measure intelligence. Intelligence is generally understood as mental capacity that enables a person to learn at school or, more generally, to reason, to solve problems, and to adapt to new (challenging) situations. There are many types of intelligence tests depending on the kind of person (age, profession, culture, etc.) and the way intelligence is understood. Some tests are general, others are focused on evaluating language skills, others on memory, on abstract and logical thinking, or on abilities in a wide variety of areas, such as, for example, recognizing and matching implicit visual patterns. Scores may be presented as an IQ (intelligence quotient), as a mental age, or simply as a point on a scale. Intelligence tests are instrumental in ordering, ranking, and comparing individuals and groups.

The testing of intelligence started in the 19th century and became a common practice in schools and universities, psychotechnical institutions, courts, asylums, and private companies on an international level during the 20th century. It is generally assumed that the first test was designed by the French scholars A. Binet and T. Simon in 1905, but the historical link between testing and experimenting points to previous tests, such as the word association test. Testing was practiced and understood in different ways, depending not only on the time, but also on the concrete local (cultural and institutional) conditions. For example, in the United States and Brazil, testing was immediately linked to race differences and eugenic programs, while in other places, such as Spain, it was part of an attempt to detect “feebleness” and to grade students at certain schools.

Since its beginning, the intelligence test received harsh criticism and triggered massive protests. The debate went through the mass media, leading to the infamous “IQ test wars.” Thus, nowadays, psychologists are aware of the inherent danger of cultural discrimination and social marginalization, and they are more careful in the promotion of intelligence testing. In order to understand the role the intelligence test plays in today’s society, it is necessary to explore its history with the help of well-documented case studies. Such studies show how the testing practice was employed in national contexts and how it was received, used, or rejected by different social groups or professionals. Current historical research adopts a more inclusive perspective, moving away from a narrative focused on the role testing played in North-America. New work has appeared that explores how testing was taking place in different national and cultural environments, such as Russia (the former Soviet Union), India, Italy, the Netherlands, Sweden, Argentina, Chile, and many other places.

---

**Keywords:** mental measurement, intelligence, Binet-Simon scale, IQ controversies, APA, antitesting movement, differential psychology, human classification, school psychology, education

In the 1970s, public scandals about scientific fraud and personal dramas catapulted the history of intelligence testing into the mass media, leading to what sometimes is dubbed a “dirty history” or the “IQ test wars.” Public debates about the test’s worthiness and dangerous social consequences continue today. Thus, the reader can hardly be expected to be optimistic at this point.

The aim here is to give an overview of the different historical narratives that have been generated to tell the test’s history. Roughly, the bibliography on the history of intelligence testing can be subdivided into three types: first, the psychologists’ optimistic history of the test, framed within a narrative about social and scientific progress; second, the history of the negative consequences and rebellions against testing (the infamous nature–nurture debates); and, finally, the insights of historical research. This article follows these subdivisions in order.

The article starts by pointing out the very basic and often-repeated milestones in the test’s history. The emphasis is on the reasoning of such a “heroic narrative.” The story of the test’s success and drama is told in the literary format of a biography; thus, it is constructed as a story of the test as a “living thing disposing of agency.”<sup>1</sup> A scientific instrument can be viewed either the way it has been traditionally presented—namely, as an inanimate, passive object in the hands of good or bad scientists and obsessive administrators—or some kind of life can be attributed to the test itself. The second option is useful for making us aware of the fact that technology, such as a test, is not neutral but carries, incorporated in a condensed and often ambiguous way, a wide range of theories, hypotheses, and strategies of social (political) intervention.

The strategy of personification echoes a trend in philosophy and the history of science to blur the boundary between the human and the nonhuman.<sup>2</sup> Historians and philosophers of science have pointed out that instruments and technologies are power configurations acting through internalized rules and expectations oriented toward human management. Hacking (1992) connected technologies to specific styles of thinking, whereas others understood them as spaces of epistemic, symbolic, political, and material interests that are negotiated (Borck, 2006; Valverde, 2016). In a similar way, psychological instruments are supposed to be theory laden. Heidelberger (2003) asserted that there are always some theoretical assumptions involved in the functioning of the instrument, even if the instrument is a piece of white paper.

In the present historiographical overview, I will not evaluate to what extent a test itself constructs subjectivity or psychological knowledge. I will limit myself to an emphasis on the test’s action imposing certain tasks and questions, as well as an IQ score and ranking, which in many cases had wide-reaching personal and social implications for both the testers and the tested, as we will see.

Using the idea of a double personality, I will complement this narrative with a second part, dealing with the story about the test’s “other” (dangerous and harmful) agency. It is constructed on protests and criticism triggered by use of the test. The article does not have space for an extensive exposition; therefore, it is limited to some public debates and scandals surrounding testing in the United States.

---

Finally, in the third part, I will expose some contributions made by historians from outside the arena of dispute between testers (i.e., psychologists) and antitesters (activists). Their work has been helpful in correcting some myths, in nuancing certain global-historical statements, and in broadening the narrative on a cultural level.

## The Successful Intelligence Test

---

What is intelligence and what is an intelligence test? Both terms are so popular nowadays that probably every reader already has an idea in mind when reading the two terms. The American Psychological Association (APA) explained cautiously that *intelligence* refers to intellectual (cognitive) functioning. *Intelligence tests* “compare your performance with other people of your age who take the same test” (1). Although these definitions are rather vague and tests are manifold, for the moment they suffice to set the stage.

It is generally assumed that the intelligence test’s family tree stems from the phrenological and craniological measurements of the 18th and 19th centuries. The test is viewed as the direct offspring of Francis Galton’s and James Mckeen Cattell’s anthropometric measurements (Fletcher & Hattie, 2011; Gould, 1981). Both scholars used several physiological measurements (cranial circumference, precision of visual perception, muscular force, time estimation, and reaction time<sup>3</sup>) to evaluate a person’s intelligence. Finally, once Cattell and his student Wissler had found “that the ‘mental tests’ showed virtually *no* tendency to correlate with academic achievement” (Fanher, 1985, p. 48, emphasis in the original) of graduate students, they gave up, because they realized that these “were no real measures of intelligence” (Gould, 1981). Shortly after, in 1905 in Paris, the psychologist Alfred Binet (1857–1911), in collaboration with the physician Theodore Simon, was giving birth to the test. Their off-spring would soon be known under the name of Binet-Simon test.

By that time, Binet had already raised two daughters and produced a considerable number of works dealing with topics like magnetism, hypnosis, individual differences, and intelligence, among other topics (Wolf, 1973). In contrast to Galton and Cattell’s attempts, Binet’s success lay in the fact that his new test dealt directly with “higher and more complex mental functions, in situations that closely simulate[d] real life” (Fanher, 1985, p. 68).

French society was already eagerly waiting for the test’s arrival. The aim of issuing the intelligence test was to manage and control another scientific object, mental abnormality (which was also called feeble-mindedness or mental retardation). This elusive object had emerged in the mid-19th century and was supposed to be causing trouble in French public (primary) schools at the turn of the century. Children affected by it were not able to follow normal teaching and ended up disturbing the rest of the class. Therefore, the first task that the Binet-Simon test completed successfully was to identify abnormal children, children who needed a slower and more specific (practical) education within a smaller group (called *classes de perfectionnement*). Peterson (1926) described the process in the following way: “In the early fall of 1904 the [French] Minister of Public Instruction appointed a commission to study measures to be taken.” Binet, who by then was the leader of La Société Libre pur l’Étude Psychologique de l’Enfant, became the member of the commission in charge of offering a scientific (technical) solution to this problem in the educational system, because “it was

---

decided that [abnormal/ subnormal] children . . . should be eliminated from the normal work of the schools and . . . be taught in a special school. . . . It was specifically to meet this emergency that the first intelligence scale was constructed” (Peterson, 1926, p. 165).

The Binet-Simon test was considered to be scientific. Therefore, the assessment technique acquired high social status, equivalent to a collaborator with the government or at least a member of the elite. The test promised to set objective standards and to produce true statements about human intellectual capacity. Using 30 items, the test instructed the children to perform very different tasks. Here are some examples: the test urged subjects to state in what sense a poppy and blood are similar; the test instructed subjects to order objects by their weight; subjects were asked to define familiar objects, such as “house” or “fork”; repeat sentences, to fill in missing words in a sentence (“The weather is clear and the sky is \_\_”), and to construct a sentence with three given words. Subjects were asked to draw from memory a design they were exposed to for 10 seconds and to give an answer to simple questions, such as “What is the thing to do when you are sleepy?”. Moreover, they were challenged to imagine the design that would result if a piece of paper were folded in quarters, a triangular cut was made in it, and, after that, the paper was unfolded; and, in one of the most difficult test items, the subjects were asked to read the time on a clock that had the hands interchanged.<sup>4</sup>

The Binet-Simon test procedure was well accepted among contemporary psychologists because it fit with their view supposing that intelligence entails a wide variety of separate mental functions, such as memory, sensory discrimination, language, imagination, etc. , which all together were expected to help the individual to adapt to new situations or surroundings. As Elisabetta Cicciola (2019) has recently pointed out, Binet’s triple view of intelligence, which was based on understanding, judgment, and reasoning, changed with time. In 1909, he emphasized a functionalist approach, keeping only understanding while adding invention, direction, and correction. Greenwood emphasized the test’s efficiency, stating: “Binet and Simon’s test items represented what Spearman called a ‘hodgepodge’ of factors, but they did the job for which they were designed, namely, to provide an objective means of identifying children in need of remedial education”(Greenwood, 2015, p. 129). Moreover, it is important to point out that the intelligence test turned the scholars that designed, controlled, and used the test into experts (Ash & Sturm, 2005). For psychologists, the test instrumentalized intelligence, black-boxing ontological assumptions and conceptual clarifications. Now intelligence became what the test was measuring.

After a first successful job, in 1908, the test developed further, becoming a scale. Binet ordered the tasks by difficulty and the level of performance that could be expected at each age. Now the test ranked children and young adults by the stage of their intellectual development, what Binet termed *mental age*. The transformation broadened the test’s potential application and power: the test started to give human beings a score and their performance could be ranked and compared to those of others. If 75% or more of the children of a certain age passed a particular task, the task was assigned to that age level (Hergenhahn, 2009). This way, the test classified a 9-year-old child solving only the tasks meant for a 7-year-old or any child scoring 2 years lower than expected as being “retarded” (feeble-minded or abnormal).

---

Binet was internationally known among psychologists and pedagogues, and he was firmly connected to the teachers and education network in France. With the presentation of the intelligence test at international conferences and the help of visits to Binet's laboratory by foreign scholars, the test was able to start a successful international career. During the years that followed and despite Binet's sudden death after the second revision of his test in 1911,<sup>5</sup> the test encountered competition from an increasing number of similar tests and underwent several transformations. In 1912, with the help of the German psychologist William Stern, the test's scores were standardized by dividing the mental age by the chronological age of the child. Thereby, the test was able to produce a number. The number was the famous intelligent quotient (IQ). For many testers and tested, it represented the exact amount of innate intelligence of a person.

Around the same time, the test went with the North American scholars Henry H. Goddard (1866–1957) and Lewis Terman (1877–1956) to the United States. Both scholars became the test's main promoters. Goddard, who directed the Vineland Training School for feeble-minded boys and girls in New Jersey (founded in 1888), used a translated version of the French scale and was soon fascinated by the test's efficiency. In his paper "Four Hundred Feeble-minded Children Classified by the Binet Method," he stated: "No one can use the tests<sup>6</sup> on any fair number of children without becoming convinced that whatever defects or faults they may have, and no one can claim that they are perfect, the tests do come amazingly near [to] what we feel to be the truth in regard to the mental status of any child tested" (Goddard, quoted in Zenderland, 1987, p. 65).

Goddard's successful strategy was first to convince physicians of the efficiency of the test and then to plan its use in schools (Zenderland, 1987, 1998). Meanwhile, the test was standardized by Terman and was adapted to the new cultural surrounding. Therefore, in 1916, the most widely known form of the test gained U.S. citizenship under the new name of the Stanford-Binet test. This version would become the dominant instrument for assessing intelligence in the following 40 years (Benjamin & Baker, 2014). Thus, Terman and his test played a key role in shaping the early testing movement, promoting the idea of the IQ as being constant and not influenced by age, schooling, or environment (Chapmann, 1988).

With each revision, the test items steadily increased (the Binet-Simon test of 1905 had 30 tasks; the last version, of 1911, had 54; the first Stanford-Binet version had 90 items and the revised version had 129). With time, also, the test procedure became more and more regulated and standardized. After studying several test manuals, Derksen (2001) rightly pointed out that the individual and collective administration of a test consists of a complicated process in which various disciplinary mechanisms operate. Very precise instructions limit timing and vision and regulate the testing situation regarding what to say and how the tester and the tested should behave.

During World War I, the intelligence test was being prepared for a more efficient, collective assessment. This happened in 1917. With the assistance of Robert Yerkes (1876–1956), Terman, and Goddard, the test gave rise to twins: The Army Alpha and the Army Beta tests. The tests were multiple-choice paper-and-pencil tests<sup>7</sup> and they were not identical. The first used English, while the second was based on pictures meant for non-English speakers.<sup>8</sup> The two tests assessed around 1,750,000 U.S. Army recruits. With this mass testing, "IQ testing had truly come of age" (Fletcher & Hattie, 2011, p. 22).

The test pointed toward race differences and registered a very low average intelligence among the young male population (Gould, 1981). The study was commented on in the press, and yielded a tremendous impact, although official Army reactions to the testing were mixed (see also Samelson, 1977). Through the publicity about the testing, psychologists felt their science started to be “put on the map” (Cattell, 1922, cited in Greenwood, 2015, p. 132).

Civilian organizations like schools and businesses began to hire intelligence testing services for their own purposes, giving rise in the following years to what sometimes has been called the “testing craze” (Samelson, 1982). With the help of the Educational Testing Service and other institutions and companies, the intelligence test family grew, and many of the tests became famous (see Table 1). The testing developed into an economically efficient industry. Meanwhile, the test changed name and outlook, adapting to the kind of measurement testers needed, whether it was assessing a linguistic, logical-mathematical, musical, bodily-kinaesthetic, spatial, interpersonal, intrapersonal, or a naturalist type of intelligence (Fletcher & Hattie, 2011). Despite the changes, Fancher and Rutherford (2017) thought that “Binet would probably be pleased to see that his basic testing techniques have retained their usefulness, and that psychologists such as Wechsler have revived this original conception of intelligence as multifaceted assortment of abilities and aptitudes” (p. 514).

**Table 1. Some Early Intelligence Tests, Listed by the Year They Were Published<sup>9</sup>**

<b>Test</b>	<b>Year</b>	<b>What Did It Measure?</b>
Binet-Simon Method	1905	Mental abnormality (feeble-mindedness)
Binet-Simon Test (scale)	1908/1911	Mental age
Sante de Sanctis Test	1905	Mental abnormality (feeble-mindedness)
Rossolimo’s psychological profile Test	1909	Ability of children to detect nonsense in pictures
Yerkes, Bridges, Hardwick’s Point Scale	1915	Mental ability
Stanford-Binet Test (Terman)	1916	IQ
(J. C.) Raven’s (Standard) Progressive Matrices	1936/1938	Abstract reasoning and nonverbal intelligence
Wechsler-Bellevue Intelligence Scale	1939 1949	General cognitive ability
Wechsler Scale of Intelligence	1955	
Wechsler Adult Intelligence Scale (WAIS)		
(J. C.) Daniels’ Reasoning Test	1949	Nonverbal test used to assess IQ

---

Until this point, the account is a story of success and scientific progress, in which the test played the role of a useful scientific tool, arriving at the precise moment when the lack of a reliable and useful diagnostic system for feeble-mindedness was most pressing in society (Fancher, 1985, p. 69). The test stepped in as a helper who could aid in the ordering and sorting of children in schools and asylums. The test also selected and ranked human beings in other social settings: for example, it was used for diagnostics in decisions about placements in a psychiatric institution or for the selection of workers for a job.

The test suggested, legitimized, and fostered a grouping of minorities of people obtaining scores on the upper or lower end of the normal distribution. Low scorers were grouped in schools in special tracks or were interned in institutions for the retarded or feeble-minded, such as Goddard's training institution in Vineland mentioned above. Sometimes (although more rarely), high scorers were also selected and grouped in tracks and special schools to make sure they received the education needed to develop their talents fully and, thereby, to arrive to positions of high responsibility "serving the nation." National organizations appeared to coordinate and facilitate access to special schools and grants to the "right candidates", such as the National Association for Gifted Children (NAGC) created in 1954 in the United States (Gargiulo, 2012). Previously, in 1946, a high IQ society called Mensa International had been founded in the United Kingdom, as an aristocracy of the intellect, with members scoring 131 or more (in the upper 98th percentile) as measured by the Stanford-Binet Intelligence Scale.

Although traditional narratives give agency exclusively to the test makers and users, it is easy to adapt the oft-repeated episodes into a biographical format. A biographical narrative is instrumental in fostering awareness of the fact a tool like a test "does something" with the persons taking it. It is theory laden, working with a set of norms and values that assume a type of mental work a citizen should be able to do. Despite all the methodological care psychologists took in assessing its validity, reliability, and replicability, the test also triggered scepticism and fear. It was a dangerous tool and was criticized for being able to harm people and for not being objective.

## The Dangerous Intelligence Test

---

Now it's time to take a look at the controversial side of the test's history, passing from the biography of Dr. Jekyll to that of Mr. Hyde. It begins with the test's scoring and ranking, which identified intelligent and unintelligent persons, each of them receiving several levels and not very clearly defined subcategories. The fact that low scorers received labels like "idiot," "stupid," "imbecile," "moron," "retarded," "abnormal," "subnormal" or "feeble-minded" in academic publications (Zenderland, 1987, p. 65), shows the strong social implications of testing in terms of discrimination. It is easy to see that the test was extremely powerful, able to act as gatekeeper to job opportunities and careers, and, more generally, to support or strengthen the social status of some persons while questioning the status or even sanctioning social marginalization of others.

Meanwhile, the Binet-Simon scale proved effective in selecting gifted children. One first initiative in this direction was Terman's and Catherine Cox's (1890-1984) long-term research, often cited as example of the test's positive effect of creating opportunities (Tannenbaum,



---

2000). The guiding idea of their research was that, if gifted and talented children were selected early enough and cultivated in the right manner, they would eventually take their rightful place at the top of society. Of the 1,528 recruits, most completed higher education and some even became famous, such as Lee Cronbach and Ancel Keys (Richard & Hattie, 2011, p. 21). The psychologist and feminist Leta Stetter Hollingworth (1886–1939) also conducted substantial research in this area. Her quantitative and qualitative studies showed the variability among gifted children and, based on this research, she argued for further curriculum differentiation for them (Jolly, 2007).

On the other hand, the test also marked and identified a second group of persons obtaining low scores. Among their ranks were persons supposed to be ill adapted to society, and who were often from poor and immigrant family backgrounds. Members of this group are less recognized. We might cite here the degenerate “Kallikaks,” a fictional name of a family immortalized by Goddard (1912) with the help of retouched pictures and descriptions (see eugenic propaganda in Gould, 1981, and Zenderland’s critical comment, 1998).

Terman stated in 1922: “all the available facts that science has to offer support the Galtonian theory that mental abilities are chiefly a matter of original endowment” (Terman, cited in Minton, 1987, p. 95). The test’s massive uses and abuses within educational institutions and the association of some of the test’s promoters with racist, eugenic, and immigrant-hostile policies increased the test’s bad repute.

The test and the testers received strong attacks in the press. Around 1920, Terman’s “nativist technocracy” was contested by the journalist Walter Lippman. Lippman questioned whether the test measured what it was supposed to measure. He also warned about the social consequences of testing and ranking people, because teachers would start to classify students instead of investing their efforts in remedying the causes of mental retardation through proper education (Gould, 1981). Historically, this was not the first reaction. Chapmann (1988) cited publications appearing after 1911, warning against the mystification of testing scores. Previously, in the 19th century, Charles Cooley (1896) and others had already argued against the nature–nurture dichotomy established by Galton. They insisted on the need to prevent “abnormality” by offering a “good environment” and by establishing educational institutions. Nevertheless, these warnings did not hinder massive testing in the following decades, in which intelligence tests decided on the educability and genetic worth of persons.

Yerkes’ extensive report published 1921 on the test scores of 160,000 subjects (of the Army testing) had a great impact. The test had pointed toward race differences in intelligence and revealed the alarming result of an average mental age of 13 years for white men; a level of intelligence that was close to mental deficiency. Again, it was Lippman who, in a series of articles in *New Republic* magazine, argued that the age standards applied to the Army scores must have been “ludicrously inappropriate” (Fancher, 1985, p. 130). Nevertheless, writings like Carl Brigham’s *A Study of American Intelligence* (1923) echoed the social alarm of Yerkes’ report about race differences. Brigham, for whom intelligence was genetically determined, warned about the danger for the whole American population to become mentally degenerated. The latter development was attributed by Brigham to the test’s identification of low-scoring ethnical groups, such as African Americans and foreign immigrants.

---

Kamin (1974), Gould (1981), Greenwood (2015), and others pointed out two consequences of the Army testing. First, the testing consolidated psychology's status as a science and a profession.<sup>10</sup> Second, it led to the justification of racial discrimination and segregation in higher education, as well as to the wording and final approval of the Immigration Restriction Act of 1924.<sup>11</sup> Moreover, the combination of testing with eugenic purposes is thought to have led in 1927 to the infamous ruling by the U.S. Supreme Court, legalizing forced sterilization of citizens with developmental disabilities (persons labelled "feebleminded") on the basis of low IQ scores. The ruling resulted in around 70,000 coerced sterilizations of individuals<sup>12</sup> (Cohen, 2016; see also Greenwood, 2015). Furthermore, in the 1930s in Nazi Germany more than 200,000 "degenerates" were sterilized (Fancher, 1985). Despite the revulsion in the United States against the Nazi practices, compulsory sterilization on the basis of IQ or criminality continued formally until the second half of the 20th century.

Thus, the test had deep and far-reaching consequences for the lives of the tested persons.<sup>13</sup> No wonder the test's action awoke gravest concerns. The test produced winners and losers at the same speed that it generated a growing number of fans and testers, on the one hand, and skeptics and antitestors, on the other.<sup>14</sup>

In the 1960s and 1970s, criticism became widespread, leading to extensive and harsh debates in public media and a growing antitestors movement. In part it was provoked by a paper published by the Berkeley psychologist Arthur R. Jensen in 1969 in which he stated that there was substantial evidence pointing to the much greater importance of nature (in relation to nurture) for a person's level of intelligence.<sup>15</sup> He cited a long list of testing data in support of this conclusion. His highest praise was for the British test promoter Cyril Burt, whom he considered "probably the most distinguished exponent of the application of [quantitative genetic] methods to the study of intelligence" and provided "the most satisfactory attempt to estimate the separate [heredity and environmental] variance components of [IQ]" (Jensen quoted in Fancher, 1985, p. 195).

Jensen's article immediately triggered attacks by Black and liberal student groups against the journal (the *Harvard Educational Review*) for publishing a racist article. Some academic institutions reacted as well. For example, the Council of the Society for Psychological Study of Social Issues (SPSSI) published a statement denouncing the fact that Jensen had "once again evoked popular misconceptions about the relationship of race and intelligence" (SPSSI Council, 1969, p. 1). They acknowledged that intelligence tests often marked differences between "Whites and Negroes" but argued that "there is no direct evidence that supports the view that there is an innate difference between members of different racial groups" (SPSSI Council, 1969, p. 1). Thus, "as behavioural scientists, we believe that statements specifying the heredity components of intelligence are unwarranted by the present state of scientific knowledge" (SPSSI Council, 1969, p. 1).

Soon after, Cyril Burt was accused of fraud. Burt was a highly respected scientist who had tried to show the link between IQ and hereditary psychological traits by studying identical twins with the help of the Binet-Simon and other intelligence tests. In the 1970s, the psychologist Leon Kamin started a campaign of criticism summarized in *The Science and Politics of IQ* (1974). In his book he exposed methodological problems and statistical inconsistencies of Burt's widely cited studies. Two years later, the journalist Oliver Gillie learned through Kamin's critique and Burt's students about the latter's apparently fictitious

---

collaborators. Gillie was the first to voice the issue of fraud, in his article, “Crucial data was faked by eminent scientist” printed on the front page of the *Sunday Times* on October 24, 1976. Thus, it seems that in collaboration with the ambitious scientific aristocrat<sup>16</sup> (Burt), the intelligence test was able to produce a great amount of totally fictitious data, thereby greatly influencing the debate between hereditarians and environmentalists (before and after the discovery of the fraud).

Soon the test gained a central role in a horror story called the “mismeasure of man,” written by Stephen J. Gould. Gould’s book, published in 1981, reached a wide audience. It was based on a historical and a methodological critique of intelligence testing. He argued that, at the beginning, while in France, the scale pursued a good goal (i.e., to identify and help retarded children through mental orthopedics). However, when the test later worked for politically conservative psychologists like Spearman,<sup>17</sup> Terman, Goddard, Burt, and Brigham, testing was used to favor members of their own social class, while consciously discriminating against “the other”—namely, foreigners or members of the working class. This was done by having the test ask questions and demanding tasks that were culturally biased. Test items would pose more difficulty to members of the African American community and other cultural minorities than to individuals from better situated social groups. An item in the Beta Test where the tested person has to recognize an element that is missing in a picture is a good example. The picture shows a sketch of a tennis match in which the rackets are missing. Probably not all recruits had experience with (playing) tennis, and therefore it must have been difficult for them to detect the missing element. Additionally, the psychologists cited in this paragraph advocated a nativist interpretation of the IQ as a fixed label to categorize candidates. This attitude conferred full power to the test to rank persons for the rest of their lives and thereby to decide their eligibility for higher educational programs and certain jobs.

A decade after Gould’s critique, R. Herrnstein and C. Murray’s polemical comments in *The Bell Curve* (1994) in which they linked intelligence with race, social class, sexual behavior, and procreation planning, caused a new public outcry, reawakening the earlier debate about the implications of testing triggered by Jensen’s paper. In the same year, the psychologist Linda Gottfredson decided to save the test’s reputation (Gottfredson, 1994). She mobilized 52 colleagues to sign a public statement printed in the journal *Intelligence* as well in the *Wall Street Journal*. The text reads like a political manifesto, stating:

“2. Intelligence . . . can be measured, and intelligence tests measure it well. They are among the most accurate . . . of all psychological tests and assessments. . . . 3. While there are different types of intelligence tests, they all measure the same intelligence . . . 5. Intelligence tests are not culturally biased. . . . 9. IQ is strongly related, probably more than any other single measurable human trait, to many important educational, occupational, economic, and social outcomes. . . . Whatever IQ tests measure, it is of great practical and social importance.” (p. 18)

Additionally, the article included comments on race differences and pointed out that IQ “may be highly heritable,” a phrasing that, once more, triggered fierce public and academic reactions.

---

Despite such efforts to defend the test's reputation, the Jensenism controversy and the Burt scandal changed the way in which the test was viewed and new test versions were designed. The intelligence test lost its aura of "mainstream respectability" and even started to be banned in some institutions and regions in the United States. Some fierce critics engaged in a lifelong crusade against testing. For example, Kamin closed down his animal laboratories to become a "professional critic" (Fancher, 1985). He as expert in psychology and psychometry, reviewed studies on genetics and intelligence, the biology of mental illness, and other socially charged topics, aiming to strengthen the environmentalist stance. Additionally, non-academic activists like Mark Garrison (2009) used all media at hand to promote his idea that standardized intelligence and ability testing was invented by social reformers like Binet and Horace Mann to hand over the control of public education to the governments. Thus, in his view, "test data mean what those in power say they mean" (p. 104). He used the intelligence test critique to denounce a more general unease with the educational system, meritocratic, and neoliberal society.

Historians and psychologists of the 21st century openly admit that the test did harm people and that there was a misuse of test scores by psychologists in the past. Nevertheless, until today these multiple criticisms and attacks, only some of which are mentioned here, did not lead to a dismissal of the test. How do psychologists of today deal with the paradox of employing a tool that is historically linked to scandals? Basically, using three strategies. First, stating that the test supported by psychologists like Terman, Goddard, and Bigham was acting under the pressure of the political agenda of the time. Fletcher and Hattie (2011), for example, argued that "science often walks hand in hand with the dominant social ideology of the day" (p. 20). Therefore, the mistake was not due to the test's inherent evilness or the scientists' personal agendas but to a kind of "Zeitgeist". One of the "dominant ideologies" of the time, for example, demanded that immigration should be more tightly restricted. Thus, "armed with his new hammer (mental testing), Goddard set out on a crusade to try to fix what he saw as a serious adjunct to 'feeble-mindedness'—notably, the influx of poorer southern and eastern European immigrants to the United States" (p. 20).<sup>18</sup>

Second, the test changed its name. In the same way that new expressions appeared to refer to persons who obtained low scores on an intelligence test, much of what formerly sailed under the flag of intelligence testing is nowadays called research on "cognitive functioning" or other similar expressions. The test's malleability evidenced in its historical development, as the test repeatedly changed name, items, and outlook. Also this can be seen in the fact that the test steadily multiplied toward an ever-growing number of types, subtypes, and complex batteries that could be purchased on the market.

Third, psychologists' optimism with regard to the progress of their science made them think that the current new test versions manage to overcome the shortcomings of previous tests. Thus, Fletcher and Hattie (2011) insisted: "Researchers and practitioners most certainly have learned from the mistakes and abuses of the past, and contemporary IQ tests reflect the advances made in test construction, scoring and analysis" (p. 28). One of the key improvements is the fact that tests now deal more carefully with the differences in cultural background.<sup>19</sup>

---

Thus, today, intelligence testing is still very much alive, and psychologists consider it a significant scientific achievement. The APA characterizes the present situation as a glass half full because “more research is needed to try to ensure that tests are used to maximize learning opportunities for all students—rather than becoming an additional barrier to some children’s achievement. Much is being done, but there is much more left to do” (2). At the same time, the association celebrates the test’s creation as a milestone in the history of humanity: “When psychologist Alfred Binet developed a test to measure the intellectual skills of French schoolchildren in 1904, he could not have possibly imagined how his research would change the world. In the last century, IQ and achievement tests have changed the face of education and employment all over the industrialized world” (1).

Fletcher and Hattie (2011) went even further and used the fact that testing had been withstanding criticism as evidence for the intelligence test’s being a useful instrument for the production of scientific truth: “Despite this waxing and waning, IQ has emerged in many different forms and has found its rank among the plethora of theoretically and empirically based psychological constructs. Indeed, it stands out as being one of the most well-supported notions in educational and mainstream psychology, with many uses in education, industry and health, to name but a few” (p. 28). A theory or construct, they argued, with the “most evidence” and predictive power “will likely gain the foreground. . . . IQ and intelligence have done this, . . . whether you agree or disagree with it, its ability to withstand the many attacks says much about its staying power. Ignoring it ignores scientific evidence and we do this for many cultural and social reasons, but it rarely leads to advances in our understanding” (p. 29).

In summary, in the second decade of the 20th century the test was quickly established and gained many supporters. Intelligence measurement acquired fame, multiplied, and became big business for testers, psychologists’ associations (like the APA), and test companies. As Rose stated, it gave psychologists their “first taste of power” (Rose, 1979, p. 55). At the same time, the test’s fame became highly controversial. The public stage was repeatedly overwhelmed by psychologists and antitesters harshly debating the scientific and the social implications of testing. Despite the unpleasant tone on both sides and the seemingly endless repetition of arguments, in a modern democratic environment, public debates are generally welcome as a healthy indicator of freedom of expression. TV shows, radio programs, podcasts, blogs, and other public platforms turn into informative tribunes enlightening the audience’s judgment with pros and cons. Beyond that reasonable aim and much to the psychologists’ despair, the competition of the media for public attention often led to exaggerations and a purely sensationalist highlighting of the most dramatic aspects of the test’s story.

## **Views From Beyond: Historians on Intelligence Testing**

---

Beyond the battle between test(ers) and antitest(ers), historians, philosophers, and sociologists have published on the topic from different angles. Although critical of psychological testing as a scientific undertaking, their primary aim is not to polemicize or to side with one or the other part.

---

Their contributions can be subdivided into two not completely differentiated types. On the one hand, there is a historiography based on social critique and the adoption of a (de-)constructivist approach. These historical narratives share the antitesters' critique and follow the lead of Latour and Foucault, adopting a sociological-philosophical perspective. The aim of their research is to examine scientific practice in order to expose the power relations and dynamics in contemporary society. An example for this is the work of the sociologist Nikolas Rose (1979), who sought to disentangle the constitution of psychology's "regime of truth." In his book *Inventing Our Selves* (Rose, 1998), he further examined the role of psychological techniques in controlling the human mind by recording, classifying, and managing individual differences.

Also, the social psychologist and historian Kurt Danziger moved along these lines. He inserted the history of mental testing in one of several traditions of social interaction in psychological research called "the aggregate model" (Danziger, 1990). He and Maarten Derksen (cited above) adopted a constructivist view, insisting that the testing practice did not unravel some psychological capacity or characteristic, such as personality or intelligence, but was (and is), in itself, a constitutive process of these entities (Danziger, 1997; Derksen, 2001). Moreover, Danziger (1993, 2003) accused psychologists of falling into the trap of assuming too quickly that psychological objects like IQ have essential qualities, forever fixed by nature. Such an assumption is common among psychologists because "there are strong professional interests bound up with the belief in the rock-solid permanence of certain psychological objects" (Danziger, 2003, p. 31).

A second kind of research is more historically oriented, aimed less at condemning and more at understanding. An attempt is made to elucidate the ways mental testing has been performed in the past and the reasons for its design and actions. Thereby, the focus lies on the social functions and consequences intelligence measurements achieved at a certain time and place. The awareness of the situatedness of testing led historians to undertake research on how the test arrived and was used at places beyond the borders of the United States.

The historical works are often quite detailed, varied, and disconnected, and they are based on cumbersome analyses of primary sources. A full overview is still missing (and only a few examples are cited in the following). Additionally, the works' historical nature and limited focus make them less attractive to a broader readership than the previous histories. Nevertheless, this kind of research has been able to offer new insights, and it has been instrumental in moving beyond the previous, all too often repeated, test stories by correcting and enriching them in substantial ways through contextualization and a broadening of historical figures, testing initiatives, and local contexts. Finally, by adopting a critical, relativistic, stance toward science and society, their perspective bridges the psychologists' and antitesters' views.

Using historical documents, historians have been able to correct the overstatements of both narratives explicated previously in this article: the psychologists' narrative, which claims that their tests have "changed the world," and the antitesters' narrative, which promotes the idea that, through the testing, psychology shaped much of social policy in the field of education, employment, racial segregation, and immigration regulation. Both narratives exaggerate the test's historical role and real political influence. Snyderman and Herrnstein (1983), for example, showed that the claim made by Gould (1981, see endnote 13) and others that the

---

racially biased Immigration Act of 1924 was passed with the help of the intelligence-testing community is simply wrong. The U.S. Congress took virtually no notice of the testing (see also Sokal's critique of Gould's book, 1987). This said, it does not mean that testing had no effect at all on the racist attitudes and the public debates about immigration of the time. But there is a difference between attributing causality to the issuing and text of a law and just strengthening a position with empirical data within a general discussion.

In a similar vein, both, the traditional pro-testing and denunciation narratives are histories in which good and bad (powerful, White, male) scientists acted deliberately to design test items and testing procedures (see, for example, Gould's view of Binet and Terman). Thereby, they attribute too much agency to individual scientists. The drama unfolds through the agency of some caricature-like figures instead of well-informed biographies. Franz Samelson (1975) asserted that historical figures are often complex human beings. They could change their ideas over time and be influenced by social movements, such as eugenics. Eugenics itself was diverse, containing many strains, and was interpreted differently at the beginning of the 20th century (when it was connected to what was viewed as progressive and liberal attitudes). Thus, while Kamin (1974) believed in the neutrality and power of psychology's empirical facts and testing data, Samelson denounced this attitude as naïve, asserting that no scientific effort or point of view is value free.

Also, Michael Sokal (1984) rightly pointed out that the insight promoted by the antitest narrative of "science being part of the social world" is no secret to any historian or sociologist of science. Authors like Kamin and Gould revised the measurements and statistical analyses done by anthropologists and psychometricians in the past using statistical techniques and scientific standards of today, which clearly seems inappropriate from a historical point of view.

Can we expect scientists like Goddard or Terman to have escaped from their contexts and to have realized the limitations of their points of view? On the one hand, Sokal (1987) was correct that they should not be judged by the standards and values of today. But, on the other hand, they should not be absolved, either, because they had the possibility to listen to critical voices, which spoke out as early as 1911, mostly coming from outside their discipline (see previous comment on first criticisms). An example of a critic from inside the discipline is Otto Klineberg, a psychologist and student of Franz Boas. In his research, Klineberg administered intelligence tests to Yakima Indian children. He soon became aware of the importance of cultural factors, and he argued in his book *Negro Intelligence and Selective Migration* (1935) for the importance of the cultural environment in the assessment of intelligence of African Americans (see also Fancher, 1985).

The explanation for why Terman did not take into account these critical views can be found in the book *Psychological Testing and American Society (1890-1930)*, edited by Sokal (1987). The book contains several essays aiming at situating the testers in their historical context. Minton (1987) explored in depth Terman's political background, which is interesting because his case undermines the generally assumed correlation between a nativist (hereditarian) position and the politically conservative (nondemocratic) attitude assumed by Gould (1981) and others. Terman defended a political democratic ideal while adopting a hereditarian stance in his scientific work on intelligence. In explaining this curious combination, Minton's study pointed to the roots of Terman's thinking, which are found in the liberal reform movement of

---

the Progressive Era in the United States. At the same time, Darwinian evolutionary thought dominated among scientists. Minton showed that it was basically Terman's commitment to natural science that blinded him to the social-cultural origins of differences in intelligence.

Furthermore, John Carson's (2007) book, *The Measure of Merit*, contextualized intelligence testing by offering a historical narrative to explain why in the 1920s and afterward the intelligence test became so popular in the United States. Carson argued that, in this historical context, "intelligence proved to be an attractive concept with which to unify the democratic and meritocratic" (p. 5). It was used to regulate, in "objective" terms, the increasing demand for limited educational resources and occupational opportunities. The result was that members of privileged socioeconomic groups generally scored well on intelligence tests, while the door was kept open to exceptional members of disadvantaged groups, who historically had been excluded.

Thus, Carson's study (2007) worked out a confluence of political concerns about democracy, meritocracy, and immigration, together with social Darwinist notions as a contextual setting that oriented the appropriation and uses of the intelligence test among psychologists on the American continent. But the test arrived at many places across the globe (see monographic issue of *History of Psychology*, 2014, Volume 17, and Carson, 2014). So, what happened there? *The Measure of Merit* showed how differently the French and the Americans constructed their "systems of merit." In France, intelligence was associated with an elite class, sustained through a selective education system. Therefore, neither the test nor the attribution of an underlying genetic intelligence factor was necessary to secure and justify social differentiation.

Although in France the use of the test was limited, Schneider (1992) showed that Binet's concept of multifaceted and complex understanding of intelligence was followed in the 1920s and 1930s by Henri Pieron, Henri Laugier, and other French psychologists and psychotechnicians. Additionally, Nicolas, Andrieu, Croizet, Sanitioso, and Burman (2013) found in their research that the reason that Binet built a scientific instrument to measure feeble-mindedness was because he wished to limit the role of psychiatrists. But testing did not start with Binet. Historians have shown the historical link between psychological testing and experimentation (Mülberger, 2017), the indefinite boundary between tests and toys (Young, 2015), and the existence of other popular tests before, as well as at, the time the Binet-Simon scale appeared (Mülberger, 2014, 2017).

For example, the standard (success) story unjustly ignores the Italian contributions. In Italy, scholars initiated a strong tradition of anthropometry, and some members became quickly interested in psychological testing. After 1896, they started to use the expression "testi mentali" (Ceccarelli, 2002) and the physician Sante de Sanctis presented his intelligence test to psychologists in 1905, at the same time and place that the Binet and Simon test was presented (Ceccarelli, 2013; Ciccioia, Foschi, & Lombardo, 2014). The De Sanctis test (Table 1) required fewer language skills: it asked test subjects to complete basic tasks, such as, to test memory. For example after the child was presented with five glass balls of different colors, it received a simple request like: "Give me a ball". Then the child was presented with the same five colored balls and was asked: "Which ball did you give me?" Another task was to invite the child to identify among pieces of wood with different shapes the ones that were equal and to reason about the relation between size and weight. Once the experimenter had



---

the necessary material at hand, the administration of the six de Sanctis test tasks was straightforward. The test became a very popular tool among European physicians for detecting mental deficiency.

Another test that appeared only a couple of years later was designed by Gregori Rossolimo (1860–1928). His psychological profiles test (Table 1) consisted of several logically absurd images, such as showing a small kid with a moustache. The tested child had to detect what was wrong in the picture and to explain the reason. The test became a very popular way to test intelligence in the Soviet Union (Léopoldoff, 2014). In the 1920s, the Russian pedologist Alexandre Neçaev (1870–1948) melted the Binet-Simon test with the Rossolimo profiles for his testings.

The Binet-Simon, the de Sanctis, and the Rossolimo tests were circulating at the same time that traditional methods for diagnostic investigations—including a protocol of questions (not only about the child but also about the family), anthropometric measurements, and general medical examination—were still very much in use among school physicians. School physicians often did not adhere to one or the other method but mixed and complemented them.

After Binet's death in 1911, his colleague, the physician Theodore Simon (1872–1961), became the test's strongest supporter. Simon initiated an international campaign to advocate the use of the French intelligence test (Wolf, 1961). Soon other influential psychologists and pedagogues, such as Édouard Claparède and Ovide Decroly, advocated for testing children's intelligence with the Binet-Simon scale. Through their wide-ranging network, the news about the efficiency and reliability of the new technique spread quickly at a time when schooling became public and graded.

One of the first places where a translated version of the original Binet-Simon scale of 1908 was applied was Barcelona. In 1908, a group of physicians was hired by the City Hall to undertake an anthropometric and psychometric assessment of schoolchildren, to identify the physically, mentally, and socially feeble. The physicians tested 420 boys and girls. Whereas in other cases the label of "feeble" for low scorers would have had negative consequences, in this instance the contrary happened. Parents and children were eager to get into the selected group because this permitted them to take part in subsidized summer camps in the countryside (Mülberger et al., 2019).

Intelligence testing also was pursued in other institutions, cities, and nations throughout Europe, such as the Netherlands (see, for example, Amsing & de Beer, 2009; Bakker, 2017; van Drenth, 2007), Denmark (Ydesen, 2011), Germany (Gundlach, 2013; Meskill, 2015), and Scandinavia (Hamre, Axelsson, & Ludvigsen, 2019), as well as beyond European and American borders. For example, Setlur (2014) studied the way the French and American intelligence tests arrived and were introduced by C. Herbert Rice, Prasanta Chandra Mahalanobis, and Venkatrao Vithal Kamat in British India. He explored the use of these tests after 1919 and how the testing practice challenged the Indian social structure based on caste. Other research pointed out how some American missionaries who pioneered mental testing in India downplayed the significance of racial differences (Linstrum, 2016). They attributed the low scores to a litany of environmental ills, such as malnourishment, malaria, illiteracy, and caste discrimination. Following Linstrum, "this critique reflected the growing willingness of missionaries between the wars to castigate the failures of British rule" (2016, p. 89).

---

In Brazil, the early French and American versions of the intelligence test were introduced and adapted by Ulisses Pernambucano, Isaías Alves, Lourenço Filho, and the Russian pedagogue Helena Antipoff, who worked with Claparède, in connection with the establishment of psychological and pedagogical laboratories in the 1920s. The scholars' concern with the professionalization of psychology, the mental hygiene movement, and the psychological studies of Brazilian races fostered their interest in testing the population (Jacó, 2014; Vieira & Campos, 2011). With regard to the use of the test in the field of justice, Silvana Vetö (2019) documented the systematic administration of adapted versions of the Binet-Simon intelligence test to youngsters at Santiago de Chile's Juvenile Court in the 1930s and early 1940s.

Until this point, the history of intelligence testing has been mainly connected to initiatives and perspectives of psychologists. The moment schoolchildren were assessed, the testing process got linked to schools and thereby incorporated teachers, pedagogues, school inspectors, school physicians, and administrators. Although historical research has been done by Avanzini (1969, 1999), Depaepe (1993), Ryan (2011), Zenderland (1988), Zuza (1948), and others, the role of these professional groups in the testing movement still needs to be fully assessed. Among teachers, testing was not always welcome. Canales and Polenghi (2019) correctly pointed out that there was also resistance, but most of the time intelligence testing persisted. According to some educationists, "education was colonized by the field of psychology" (Richardson & Johanningmeier, 1998, p. 699).

The history of psychological testing acquires a different meaning and function when it includes the opinions and contributions of teachers and pedagogues. Many of them took advantage of their access to children and the availability of tests to undertake intelligence measurements. Nevertheless, they rarely went so far as supposing the existence of an innate entity. On the contrary, some historical cases show the considerable flexibility with which the test was used for different purposes, such as training working-class children to make them fit for class struggle in industrialized cities (see, for example, Mülberger, Graus, & Balltandre, 2014).

## Websites

Intelligence. <<https://www.apa.org/topics/intelligence/>>

Intelligence and Achievement Testing: Is the Half-Full Glass Getting Fuller? <<https://www.apa.org/research/action/intelligence-testing>>

## References

Amsing, H., & de Beer, F. (2009). Selecting children with mental disabilities: A Dutch conflict over the demarcation of expertise in the 1950s. *Paedagogica Historica*, 45(1-2), 235–250.

Ash, M., & Sturm, T. (2005). Roles of instruments in psychological research. *History of Psychology*, 8(1), 3–34.

Avanzini, G. (1969). *La contribution de Binet a l'élaboration d'une pédagogie scientifique*. [Binet's contribution to the development of a scientific pedagogy]. Paris, France: Vrin.

- 
- Avanzini, G. (1999). *Alfred Binet*. Paris, France: Presses Universitaires France.
- Bakker, N. (2017). A culture of knowledge production: Testing and observation of Dutch children with learning and behavioural problems (1949–1985). *Paedagogica Historica*, 53(1-2), 7–23.
- Benjamin, L. T., Jr., & Baker, D. B. (2014). *From séance to science: A history of the profession of psychology in America*. Akron, OH: University of Akron Press.
- Binet, A. (1911). Nouvelles recherches sur la mesure du niveau intellectuel chez les enfants d'école [New research on the measurement of the intellectual level in school children]. *L'Année Psychologique*, 17, 145–201.
- Binet, A., & Simon, T. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'Année Psychologique*, 12, 191–244.
- Binet, A., & Simon, T. (1908). Le développement de l'intelligence chez les enfants [The development of intelligence in infants]. *L'Année Psychologique*, 14, 1–94.
- Borck, C. (2006). Between local cultures and national styles: Units of analysis in the history of electroencephalography. *Comptes Rendus Biologies*, 329(5-6), 450–459.
- Brigham, W. C. (1923). *A study of American intelligence*. Princeton, NJ: Princeton University Press.
- Canales, A., & Polenghi, S. (2019). Classifying children: A historical perspective on testing and measurement <<https://doi.org/10.1080/00309230.2019.1611890>>. *Paedagogica Historica*, 55(3), 343–352.
- Carson, J. (2007). *The measure of merit: Talents, intelligence, and inequality in the French and American republics*. Princeton, NJ: Princeton University Press.
- Carson, J. (2014). Mental testing in the early twentieth century: Internationalizing the story. *History of Psychology*, 17, 249–255.
- Ceccarelli, G. (2002). *Il testing psicologico in Italia - Materiali per una storia delle origini* [Psychological testing in Italy - Materials for a history of the origin]. Urbino: Quattro Venti.
- Ceccarelli, G. (Ed.). (2013). *Alfred Binet e la misura dell'intelligenza*. [Alfred Binet and the measurement of intelligence]. Milan, Italy: Franco Angeli.
- Chapman, P. D. (1988). *Schools as sorters: Lewis M. Terman, applied psychology and the intelligence testing movement, 1890–1930*. New York, NY: New York University.
- Cicciola, E. (2019). *La scoperta dell'intelligenza: Alfred Binet e la storia del primo test*. [The Discovery of intelligence: Alfred Binet and the history of the first test]. Rome, Italy: Fefé.
- Cicciola, E., Foschi, R., & Lombardo, G. P. (2014). Making up intelligence scales: De Sanctis's and Binet's tests, 1905 and after. *History of Psychology*, 17, 223–236.
- Cohen, A. S. (2016). *Imbeciles: The Supreme Court, American eugenics, and the sterilization of Carrie Buck*. New York: Penguin.

- 
- Cooley, C. H. (1896). Nature versus nurture in the making of social careers. *Proceedings of the National Conference of Charities and Correction*, 23, 399-405.
- Danziger, K. (1990). *Constructing the subject*. Cambridge, UK: Cambridge University Press.
- Danziger, K. (1993). Psychological objects, practice, and history. *Annals of Theoretical Psychology*, 8, 15-47.
- Danziger, K. (1997). *Naming the mind*. London, UK: SAGE Publications.
- Danziger, K. (2003). Where history, theory and philosophy meet: The biography of psychological objects. In D. Hill & M. Kral (Eds.), *About psychology: Essays at the crossroads of history, theory, and philosophy* (pp. 19-33). Albany: SUNY Press.
- Daston, L. (2000). *Biographies of scientific objects*. Chicago, IL: University of Chicago Press.
- De Sanctis, S., & Bolaffi, E. (1914). La gradazione dell'insufficienza intellettuale con il metodo dei reattivi.[the ranking of intellectual insufficiency with the reactive method]. *L'Infanzia Anormale*, 7, 153-173.
- Depaepe, M. (1993). *Zum Wohl des Kindes? Pädologie, pädagogische Psychologie und experimentelle Pädagogik in Europa und den USA, 1890-1940*. [Caring for the kid? Pedology, pedagogical psychology and experimental pedagogy in Europe and in the US, 1890-1940]. Leuven: Leuven University Press.
- Derksen, M. (2001). Discipline, subjectivity and personality: An analysis of the manuals of four psychological tests. *History of the Human Sciences*, 14(1), 25-47.
- Esping, A., & Plucker, J. A. (2015). Alfred Binet and the children of Paris. In S. Goldstein, D. Princiotta, & J. Naglieri (Eds.), *Handbook of intelligence* (pp. 153-161). New York, NY: Springer.
- Fancher, R. (1985). *The intelligence men: Makers of the IQ controversy*. New York, NY: W. W. Norton.
- Fancher, R., & Rutherford, A. (2017). *Pioneers of psychology: A history*. New York, NY: W. W. Norton.
- Fletcher, R., & Hattie, J. (2011). *Intelligence and intelligence testing*. London, UK: Routledge.
- Gargiulo, R. (2012). *Special education in contemporary society: An introduction to exceptionality*. London, UK: SAGE.
- Garrison, M. (2009). *A measure of failure: The political origins of standardized testing*. Albany: State University of New York Press.
- Goddard, H. H. (1912). *The Kallikak family: A study in the heredity of feeble-mindedness*. New York, NY: McMillan.
- Gottfredson, L. (1994, December 13). Mainstream science on intelligence. *The Wall Street Journal*, p. 18.
- Gould, S. (1981). *The mismeasure of man*. New York, NY: W. W. Norton.

- 
- Greenwood, J. D. (2015). Intelligence defined: Wundt, Cattell, Thorndike, Goddard, and Yerkes. In S. Goldstein, D. Princiotta, & J. Naglieri (Eds.), *Handbook of intelligence* (pp. 123–136). New York, NY: Springer.
- Gundlach, H. (2013). La ricezione del metodo di Binet per la misura dell'intelligenza in Germania [the reception in Germany of Binet's method for measuring intelligence]. In G. Ceccarelli (Ed.), *Alfred Binet e la misura dell'intelligenza* [Alfred Binet and the measurement of intelligence] (pp. 158–179). Milan, Italy: Franco Angeli.
- Hacking, I. (1992). The self-vindication of the laboratory sciences. In A. Pickering (Ed.), *Science as practice and culture* (pp. 29–64). Chicago, IL: University of Chicago Press.
- Hamre, B., Axelsson, T., & Ludvigsen, K. (2019). Psychiatry in the sorting of schoolchildren in Scandinavia 1920–1950: IQ testing, child guidance clinics, and hospitalization. *Paedagogica Historica*, 55(3), 391–415.
- Heidelberger, M. (2003). Theory-ladenness and scientific instruments in experimentation. In H. Radder (Ed.), *The philosophy of scientific experimentation* (pp. 138–151). Pittsburgh, PA: University of Pittsburgh Press.
- Hergenhahn, B. R. (2009). *An Introduction to the History of Psychology*. Belmont, CA : Wadsworth Cengage Learning
- Herrnstein, R., & Murray, C. (1994). *The bell curve*. New York, NY: Free Press.
- Jacó-Vilela, A. (2014). Psychological measurement in Brazil in the 1920s and 1930s. *History of Psychology*, 17, 237–248.
- Jolly, J. L. (2007). The research legacy of Leta S. Hollingworth. *Gifted Child Today*, 30(3), 57–64.
- Kamin, L. J. (1974). *The science and politics of IQ*. New York: Laurence Erlbaum.
- Klineberg, O. (1935). *Negro intelligence and selective migration*. New York, NY: Columbia University Press.
- Latour, B. (2000). On the partial existence of existing *and* nonexistent objects. In L. Daston (Ed.), *Biographies of scientific objects* (pp. 247–269). Chicago, IL: University of Chicago Press.
- Léopoldoff, M. I. (2014). A psychology for pedagogy: Intelligence testing in USSR. *History of Psychology*, 17, 187–205.
- Linstrum, E. (2016). *Ruling minds: Psychology in the British Empire*. Cambridge, MA: Harvard University Press.
- Meskill, D. (2015). Psychological testing and the German labor market, 1925 to 1965. *History of Psychology*, 18(4), 353–366.
- Minton, H. (1987). Lewis M. Terman and mental testing: In search of the democratic ideal. In M. Sokal (Ed.), *Psychological testing in American society (1890–1930)* (pp. 95–112). London, UK: Rutgers University Press.
- Mülberger, A. (2014). Special Issue: Mental Testing after 1905: Uses in Different Local Contexts *History of Psychology*, 17(3).

- 
- Mülberger, A. (2017). Mental association: Testing individual differences before Binet. *Journal of the History of the Behavioral Sciences*, 53(2), 176–198.
- Mülberger, A., Balltandre, M., & Graus, A. (2014). Aims of teachers' psychometry: Intelligence testing in Barcelona (1920). *History of Psychology*, 17, 206–222.
- Mülberger, A., Gómez, C., Cervantes, M., Cañas, A. M., & Anglada, L. (2019). Testing the intelligence of Barcelona's schoolchildren in 1908. *Revista de Historia de la Psicología*, 40, 1, 2–11.
- Nicolas, S., Andrieu, B., Croizet, J.-C., Sanitioso, R., & Burman, J. (2013). Sick? Or slow? On the origins of intelligence as a psychological object. *Intelligence*, 41, 699–711.
- Peterson, J. (1926). *Early conceptions and tests of intelligence*. Yonkers-on-Hudson, NY: World Book Company.
- Richard, B., & Hattie, J. (2011). *Intelligence and intelligence testing*. Oxon, UK: Routledge.
- Richardson, R., & Johanningmeier, E. (1998). Intelligence testing: The legitimization of a meritocratic educational science <[https://doi.org/10.1016/S0883-0355\(98\)00007-X](https://doi.org/10.1016/S0883-0355(98)00007-X)>. *International Journal of Educational Research*, 27, 699–714.
- Rose, N. (1979). The psychological complex: Mental measurement and social administration. *Ideology and Consciousness*, 5, 5–68.
- Rose, N. (1998). *Inventing our selves: Psychology, power, and personhood*. Cambridge, UK: Cambridge University Press.
- Rossolimo, G. (1911). Die Psychologischen Profil [The psychological profile]. *Klinik für Psychische und Nervöse Krankheiten*, 1(3-4), 249–326.
- Ryan, A. M. (2011). From child study to efficiency: District administrators and the use of testing in the Chicago public schools, 1899 to 1928. *Paedagogica Historica*, 47(3), 341–354.
- Samelson, F. (1975). On the science and politics of the IQ. *Social Research*, 42, 467–488.
- Samelson, F. (1977). World War I intelligence testing and the development of psychology <[https://doi.org/10.1002/1520-6696\(197707\)13:3274::AIDJHBS23001303083.0.CO;2-K](https://doi.org/10.1002/1520-6696(197707)13:3274::AIDJHBS23001303083.0.CO;2-K)>. *Journal of the History of the Behavioral Sciences*, 13, 274–282.
- Samelson, F. (1982). Intelligence and some of its testers <<http://search.proquest.com/docview/733240280?accountid=15292>>. *Science*, 215(4533), 656–657.
- Samelson, F. (1987). Was early mental testing (a) racist inspired, (b) objective science, (c) a technology for democracy, (d) the origin of multiple-choice exams, (e) none of the above? (mark the right answer). In M. Sokal (Ed.), *Psychological testing and American society (1890–1930)* (pp. 113–123). New Brunswick, NJ: Rutgers University Press.
- Schneider, W. H. (1992). After Binet: French intelligence testing, 1900–1950. *Journal of the History of the Behavioral Sciences*, 28(2), 111–132.
- Setlur, S. (2014). Searching for South Asian intelligence: Psychometry in British India, 1919–1940. *Journal of the History of the Behavioral Sciences*, 50(4), 359–375.
-

- 
- Snyderman, M., & Herrnstein, R. J. (1983). Intelligence tests and the Immigration Act of 1924 <<https://doi.org/10.1037/0003-066X.38.9.986>>. *American Psychologist*, 38(9), 986-995.
- Sokal, M. (Ed.). (1987). *Psychological testing and American society (1890-1930)*. New Brunswick, NJ: Rutgers University Press.
- Sokal, M. M. (1984). Approaches to the history of psychological testing <<https://doi.org/10.2307/368017>>. *History of Education Quarterly*, 24, 419-430.
- SPSSI Council. (1969). SPSSI Council statement on race and intelligence. *Journal of Social Issues*, 25(3), 1-3.
- Strenio, A. (1981). *The testing trap*. New York, NY: Rawson Association.
- Tannenbaum, A. J. (2000). A history of giftedness in school and society. In *International handbook of giftedness and talent*. Oxford, UK: Elsevier.
- Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford Revision and extension of the Binet-Simon intelligence scale*. Boston: Houghton Mifflin.
- Valverde Pérez, N. (2016). Meanings of waves: Electroencephalography and society in Mexico City, 1940-1950 <<https://doi.org/10.1017/S0269889716000223>>. *Science in Context*, 29(4), 451-472.
- Van Drenth, A. (2007). Mental boundaries and medico-pedagogical selection: Girls and boys in the Dutch "school for idiots," The Hague 1857-1873. *Paedagogica Historica*, 43(1), 99-117.
- Vetö, S. (2019). Child delinquency and intelligence testing at Santiago's Juvenile Court, Chile, 1929-1942 <<https://doi.org/10.1037/hop0000123>>. *History of Psychology*, 22(3), 244-265.
- Vieira, R., & Campos, R. (2011). Notas sobre a introdução, recepção e desenvolvimento da medida psicológica no Brasil [notes with regard to the introduction, the reception and the development of the psychological measurement in Brasil]. *Temas em Psicologia*, 19(2), 417-425.
- Williams, W. M. (2000). Perspectives on intelligence testing, affirmative action, and educational policy. *Psychology, Public Policy, and Law*, 6(1), 5-19.
- Wolf, T. H. (1961). An individual who made a difference. *American Psychologist*, 16(5), 245-248.
- Wolf, T. H. (1973). *Alfred Binet*. Chicago, IL: University of Chicago Press.
- Ydesen, C. (2011). The international space of the Danish testing community in the interwar years. *Paedagogica Historica*, 48(4), 589-599.
- Yerkes, R. M., Bridges, J. W., & Hardwick, R. S. A. (1915). *Point scale for measuring mental ability*. Baltimore: Warwick and York.
- Young, J. L. (2015). Test or toy? Materiality and the measurement of infant intelligence <<https://doi.org/10.1037/a0038862>>. *History of Psychology*, 18(2), 103-118.

---

Zenderland, L. (1987). The debate over diagnosis: Henry Herbert Goddard and the medical acceptance of intelligence testing. In M. Sokal (Ed.), *Psychological testing in American Society (1890-1930)* (pp. 46-74). London, UK: Rutgers University Press.

Zenderland, L. (1988). Education, evangelism, and the origins of clinical psychology: The child-study legacy <[https://doi.org/10.1002/1520-6696\(198804\)24:2152::AID-JHBS23002402033.0.CO;2-6](https://doi.org/10.1002/1520-6696(198804)24:2152::AID-JHBS23002402033.0.CO;2-6)>. *Journal of the History of the Behavioral Sciences*, 24, 152-165.

Zenderland, L. (1998). *Measuring minds: Henry Herbert Goddard and the origins of American intelligence testing*. Cambridge, UK: Cambridge University Press.

Zuza, F. (1948). *Alfred Binet et la pédagogie expérimentale*. [Alfred Binet and experimental pedagogy] Paris, France: Vrin.

## Notes

---

1. The author is aware that there are different options for redefining a test. The intention here is not to fix it, but to leave the definition of its nature up to the reader's intuition.
2. The title of this article mirrors that of the book *Biographies of Scientific Objects*, published in 2000 by Lorraine Daston. The book aimed to overcome the opposition between realism and constructivism by affirming that scientific objects should be considered simultaneously real and historical. Latour inspired the project, stating the need to assume the historicity of (nonhuman) objects and their "partial existence," which is based on his argument that "for any entity to gain definitive access to existence, a deep rearrangement in space and time has to be worked out practically" (Latour, 2000, p. 267). A number of historians of science contributed to the volume, exploring the "coming into being" and/or "passing away" of scientific objects, such as mathematical entities, dreams, self-consciousness, society, ether, cytoplasmic particles, and medical testing.
3. The first two are part of Galton's measurements and the last three were used by both (see Sokal, 1987, p. 29).
4. For the complete list of instructions, see Binet and Simon (1905, 1908), Binet (1911), and, for a good summary in English, Fancher (1985).
5. In the 1911 version, Binet and Simon made some rearrangements of questions and tasks and extended the age range of the test to adulthood.
6. Goddard used plural because, at that time, *test* referred to each of the items or tasks included in the assessment.
7. Samelson (1987) pointed out that it was Arthur Otis who came up with the idea of the multiple-choice test format.
8. The recruits unable to answer either the alpha or the beta test were submitted to individual testing with tasks based on the Binet-Simon test (Gould, 1981).
9. Binet and Simon (1905, 1908), Binet (1911), De Sanctis and Bolaffi (1914), Rossolimo (1911), Terman (1916), and Yerkes, Bridges, and Hardwick (1915). Regarding Daniel's Reasoning Test, see Ash and Sturm (2005). Fancher and Rutherford (2017) stated that with periodic updates and a version for children, "The WAIS has remained the gold standard for assessing . . . intelligence ever since" (p. 512).
10. See the citation of Cattell saying that psychology had been "put on the map" (Cattell, 1922, cited in Greenwood, 2015, p. 132) thanks to the massive testing in the U.S. Army during World War I.
11. Gould (1981) wrote, citing Kamin and others: "But the timing, and especially the peculiar character, of the 1924 Restriction Act clearly reflected the lobbying of scientists and eugenicists, and the army data formed their most powerful battering ram" (p. 231). On the next page he insisted on the important role that psychological testing played



---

in the formulation of the new rule, asserting: “Congressional debates leading to the passage of the Immigration Restriction Act of 1924 continually invoke the army data” (p. 232). More recently, Esping and Plucker (2015, p. 160) also stated that the number of immigrants who were deported “increased exponentially” as a result of Goddard’s testing with the Binet-Simon scale in Ellis Island. But historical research offers a different picture, as is explained in the last part of this article.

12. For more information about this and *Buck v. Bell*, see Greenwood (2015).

13. For more information about how the testing led to wrong decisions and harmed personal life, see, for example, Strenio (1981).

14. Interestingly, psychologists like Fletcher and Hattie (2011) connected both levels, stating that antitesters are against intelligence testing because they are unable to obtain good (high) scores. Among the arguments for and against testing, this one does not make any sense at all.

15. He estimated IQ heritability at around 80%.

16. Sir C. Burt was knighted in 1946 for his scientific work.

17. In 1904, Charles Spearman published his two-factor theory, a paper that attracted much attention. According to this theory, the performance of any intellectual act requires *g* (a general intelligence factor including all intellectual acts) and *s* (a factor specific to the act and that varies in strength from one to another). For more information about this, see Fancher (1985).

18. Psychologists’ responsibility in the issuing of stricter immigration rules is returned to in the last part of this article.

19. For a collection of views about testing, see Williams (2000) and the numerous articles included in the monograph issue dedicated to this topic in *Psychology, Public Policy, and Law*, 6(1).

## Related Articles

Forensic Psychology in Historical Perspective