4 Intelligence Testing

Matthew J. C. Crump

2021-09-01

This chapter covers the development of intelligence tests, delving into early examples how the tests work. There is some discussion of the concept of intelligence in relation to these tests, along with historical examples of how the tests were deployed by the eugenics movement.

0.0.1 The intelligence test race

The Stanford-Binet test was among the first so-called "intelligence" tests to be widely adopted and used in America. In 1905, Alfred Binet (1857-1911), a French psychologist, published the Binet-Simon test with his student Theodore Simon (Binet & Simon, 1905b). This paper, published in French, described revised mental tests that Binet had been developing for over a decade prior (Nicolas et al., 2014). Lewis Terman was an American psychologist at Stanford University who helped popularize Binet's test in America (Terman, 1916), hence Stanford-Binet. Also in 1916, Psychologist Henry Goddard published an English translation (by Elizabeth Kite) of five of Binet's papers on intelligence testing in a book titled "The Development of Intelligence in Children" (Kite, 1916). The entire book is in the public domain and can be downloaded from the internet archive.

It is convenient to start with the Stanford-Binet test because Terman and Binet represent somewhat different progressive notions (for the time) about how psychological science could and should be used to improve society. These notions involve the

Author note: My training as a cognitive psychologist mostly skipped over the topic of mental testing, particularly the history and development of intelligence tests. This chapter is an ongoing attempt to organize some of that history and connect it with issues in cognition. I have found the history to be complex and often extremely fraught. Although mental testing is widespread and has numerous proponents and use cases, it is also intertwined with the eugenics movement. The practice of mental testing has negatively impacted marginalized groups. The scientific

nature/nurture debate about the heritability of mental abilities and the implications of that debate for enacting social policy that follow from taking different sides of the debate.

In the early 1900s, nations were finding ways to respond to social issues (crime, education, mental healthcare) with social institutions and policies, and academics were discussing how science could make responses more efficient and less resource demanding. As described in the last chapter, the eugenics movement, of which Terman was a strong proponent, billed itself as a progressive movement capable of fixing societal ailments. Social problems were attributed to hereditary influences and solutions included measures that would prevent "defective" people from breeding.

In the domain of mental health, people labelled "feeble-minded" were institutionalized and placed under care of the state. Constructing, maintaining, and staffing these buildings took resources; but, from the perspective of eugenics, the investment was worth the cost because the institutions allowed "defective" people to be segregated from society. Segregating and institutionalizing people for crimes or mental health issues was a widespread practice before the advent of mental tests. Proponents of eugenics envisioned mental testing as a new technology that could enhance the efficiency of sending unwanted people to institutions.

From Terman's perspective (1916), intelligence tests would have benefits for negative and positive eugenics policies. For negative eugenics, mental testing would make institutionalization even more effective at segregating unwanted people from society. For example, in envisioning how testing could be applied to discover hidden "defective" schoolchildren he wrote, "it is safe to predict that in the near future intelligence tests will bring tens of thousands of these high-grade defectives under the surveillance and protection of society. This will ultimately result in curtailing the reproduction of feeble-mindedness and in the elimination of an enormous amount of crime, pauperism, and industrial inefficiency". Furthermore, intelligence testing could also be used for positive eugenics, by identifying hidden geniuses and grooming them to become national leaders; Terman wrote, "The number of children with

very superior ability is approximately as great as the number of feeble-minded. The future welfare of the country hinges, in no small degree, upon the right education of these superior children."

Binet did not presume that mental abilities were fixed or inherited and he argued that such presumptions were premature in the absence of rigorous tests providing objective evidence on the matter. Binet was also hired by the French government to address pressing issues in education. For example, without objective mental tests, French schoolchildren were already being divided up into "normal" children, sent to regular school, and "defective" children sent to special schools or otherwise institutionalized. Binet described the existing methods for deciding the fate of children as deeply flawed, arbitrary, open to numerous forms of bias, and a drain on government resources. One of his concerns was child welfare. The lack of accurate testing meant that some normal children were accidentally being institutionalized for the wrong reasons. Relatedly, Binet speculated that education could improve mental abilities, even of "sub-normal" children. Another set of concerns involved efficiencies for society. Binet considered that educational resources would be wasted on children not capable of learning. As a progressive, he also envisioned a high-functioning utopian society where science was able to accurately test every person for their aptitudes, and then assign people to tasks in society with utmost efficiency (Binet & Simon, 1908). But, those lofty ideas could not be accomplished without a scientifically accurate intelligence test.

The race to develop intelligence tests more like a relay race. Researchers were vying to create intelligence tests that withstood scientific and public scrutiny. These tests would be handed off like batons to officials involved in decision-making at the level of social institutions and government. Researchers and practitioners took different sides on the nature/nurture debate and assumed that mental abilities were inherited and fixed at birth, or flexibly acquired over development. These assumptions tinted the interpretation of test results, and biased decision-making and implementation of social programs. For example, what should happen to children who are graded as mentally inferior according to an intelligence test? A eugenicist and hereditarian

might argue that these children should be institutionalized and refused education because they were genetically incapable of learning, would never contribute to society, and worse, would further pollute the gene pool if they had children. Alternatively, psychologists like Binet were open to the possibility that mental abilities could be developed and acquired with experience, and that education systems could be improved to meet the needs of children with low to high mental abilities. The next sections explore iterations of research that culminated in Binet's intelligence test, followed by examples of how the tests were applied in the United States.

0.0.1.1 Cattell's mental tests

Galton's research on individual differences in the vividness of mental imagery from 1880 was an early attempt to measure mental abilities associated with intelligence. Galton's work, including his eugenics ideas, inspired many psychologists to continue developing mental tests. For example, James McKeen Cattell (1860-1944) published "Mental tests and Measurements" in 1890 (CATTELL, 1890). Cattell made numerous other contributions to American psychology, and continues to be honored by the Association for Psychological Sciences (APS) through their "James McKeen Cattell Fellow Award" for lifetime achievement. Like Galton, Cattell was also a proponent of eugenics.

Here is the first paragraph from Cattell's paper:

> "Psychology cannot attain the certainty and exactness of the physical sciences, unless it rests on a foundation of experiment and measurement. A step in this direction could be made by applying a series of mental tests and measurements to a large number of individuals. The results would be of considerable scientific value in discovering the constancy of mental processes, their interdependence, and their variation under different circumstances. Individuals, besides, would find their tests interesting, and, perhaps, useful in regard to training, mode of life or indication of disease. The scientific and practical

> value of such tests would be much increased should a uniform system be adopted, so that determinations made at different times and places could be compared and combined."With a view to obtaining agreement among those interested, I venture to suggest the following series of tests and measurements, together with methods of making them."

Cattell was also corresponding with Galton to make the tests "meet with his approval" but he only indirectly references the application of mental testing to the eugenics movement (e.g., "indication of disease"), and he mentions several other reasons to pursue the creation of mental tests. In a footnote, he mentions that "the nationality (including that of the parents), and the age, sex, occupation and state of health" should all be recorded when participants take the test". The inclusion of these measurements is consistent with eugenical aims to show that different races had inherited different mental abilities.

Cattell proposed to measure each person on ten tests:

1. Dynamometer Pressure, (to measure squeezing hand strength).
2. Rate of Movement (how fast you can move your hand).
3. Sensation-areas (telling apart two pin-pricks)
4. Pressure causing Pain
5. Least noticeable difference in Weight
6. Reaction-time for Sound
7. Time for naming Colours
8. Bi-section of a 50 cm line
9. Judgment of 10 seconds time
10. Number of Letters remembered on once Hearing

These individual tests measure specific physical and mental abilities and none of them measure a complicated concept like human intelligence. At the end of the paper, Cattell lists 50 additional tests for sight (14 tests), Hearing (8 test), Taste and Smell (3 tests), Touch and Temperature (7 tests), Sense of Effort and Movement (4 tests), Mental Time (7 test), Mental intensity (2 tests), and Mental Extensity (5 tests), that he thought should be important for the incoming discipline of Experimental Psychology. Many of the individual measurements

later became useful for investigating how individual psychological processes work.

By 1896, Cattell had moved from Pennsylvania to New York City and was publishing measurements of Columbia University students' performance on his mental tests (Cattell & Farrand, 1896). By today's standards, Cattell's test might raise privacy concerns about what he planned to do with the data he collected. Cattell reported statistics on hair and eye color, height and weight, head size, breathing capacity, color blindness, vision, color preferences, hearing, pitch perception, skin sensation, hand strength, reaction time, perception of time and space, memory, and mental imagery. In addition, the examiner who administered the test made separate judgments of each student's qualities (physical goodness, good student, level of intellectual ability, strong-will etc.), based on their professional opinion. Finally, students were given a lengthy questionnaire to report on their family history, medical history, daily behaviors, and preferences (e.g., favorite novel, what gives them pleasure, etc.). Cattell concludes that science should proceed to determine the interrelations between his measurements, and establish how much knowing one thing about a person can predict something else about them. He also concludes that "we must use our measurements to study the development of the individual and of the race, to disentangle the complex factors of heredity and environment", and that the most important thing that science can do is guide the development of man.

0.0.1.2 Binet's critiques

Several other psychologists inspired by Galton were also publishing results from their own mental tests around this time, including Hugo Munsterberg (Munsterberg, 1891), J. Allen Gilbert (Gilbert, 1895), and Emil Kraepelin (Kraepelin, 1895). Alfred Binet was among the psychologists interested in mental testing, and well before he published his famous test he published critiques on the existing mental testing literature (Nicolas et al., 2014).

Binet pointed out that tests were measuring physical ability (like grip strength) and basic sensory abilities; but, rarely measured what he considered higher mental abilities. Binet

proposed that individual differences in higher mental abilities could be measured by simple tests, but only if the tests were mentally challenging. He proposed several tests, including the following:

Tests for memory: - drawing a geometric shape from memory - reproducing a sentence from memory - memory for musical notes - memory for a color - memory for 12 objects

Tests for mental imagery and imagination

Tests for attention - duration of attention - size of attentional field - performing multiple tasks at once

Tests for understanding - give a definition - criticize a sentence

Tests for suggestibility

Tests for aesthetic feeling - what are a persons preferences, are they same as artists?

Tests of Moral Feelings; muscular and will power; and motor skill.

As Binet might have predicted, Cattell's initial mental testing program flopped. Cattell tried to demonstrate he could predict students' grades in college from their scores on his choice of tests, but he wasn't able to reliably reproduce a predictive relationship. It seemed obvious that college grades should measure something about mental abilities. If Cattell was measuring mental abilities with his tests, then students with high scores should have high grades, and students with low scores should have low grades. Such a positive correlation would provide validity to Cattell's mental ability tests. However, multiple attempts to show that test performance positively correlated with grades instead showed no consistent correlation. One interpretation was that Cattell's tests were simply not measuring mental abilities. Binet would go on to implement his own suggestions in developing his mental tests.

0.0.2 Binet-Simon Test (1905-1911)

This section describes the Binet-Simon test in more detail, including summaries of the five papers translated to English in 1916 (Kite, 1916). Many variations of mental tests have been created since the Binet-Simon test, but this early test still provides a useful example of an "intelligence" test. I quoted "intelligence" because we can reserve judgment on what the test measures until after we look at it.

Binet and Simon published several papers in 1905. The first paper motivated the need for intelligence testing (Binet & Simon, 1905c). The second paper explained the new method (Binet & Simon, 1905b). And, the third paper provided empirical validation of the approach. This paper described results from the tests that measured in a population of children in school and another group of institutionalized children who had been identified as "subnormal" (Binet & Simon, 1905a).

0.0.2.1 Motivation

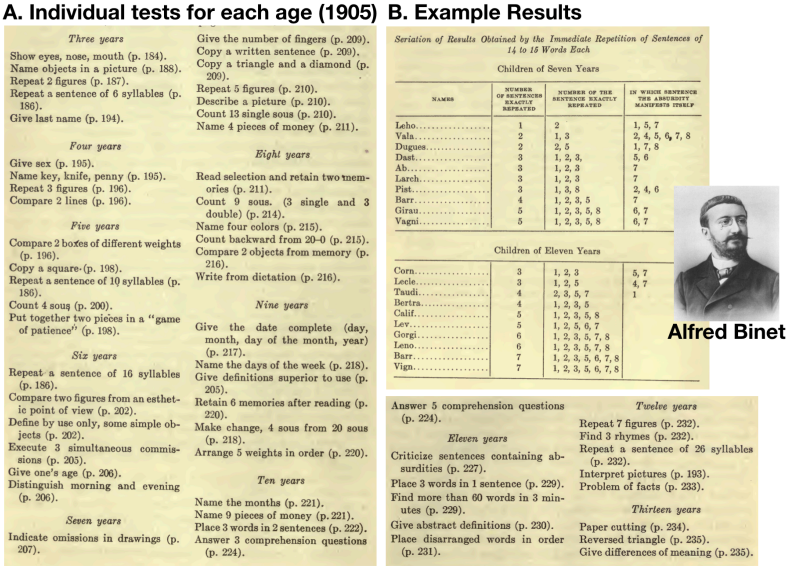
Decisions about child welfare were being made in France without the benefit of objective tests of mental abilities. Binet observed opportunities for bias in the procedures for making judgments about children that would determine their futures. He thought an objective test would be a valuable tool to guard against bias, and increase the efficiency of how the state spent resources on social programs like education. As a description of the entire research enterprise, Binet wrote, "When the work, which is here only begun, shall have taken its definite character, it will doubtless permit the solution of many pending questions, since we are aiming at nothing less than the measure of the intelligence; one will thus know how to compare the different intellectual levels not only according to age, but according to sex, the social condition, and to race; applications of our method will be found useful to normal anthropology, and also to criminal anthropology, which touches closely upon the study of the subnormal, and will receive the principal conclusion of our study." (Binet & Simon, 1905a).

0.0.2.2 Method

Binet created sets of psychological, pedagogical, and medical tests that he used to measure individual differences in children. On the whole, the approach was very similar to Cattell's, in the sense that both were measuring as many physical and mental features of children and adults as they could feasibly fit into a session where the measurements occurred. The major difference was that Binet had innovated on his psychological questions. Binet had more questions and they covered a wider variety of tasks presumed to involve higher order mental abilities. The questions were also tailored for children at different ages, from 3 to 13, in yearly increments of difficulty.

Figure 1 shows individual tasks that were determined to be appropriate for children in each year, from 3 to 13. Binet describes the individual tasks in great detail, along with instructions for administering each test, and scoring of each test.

A. Individual tests for each age (1905) **B. Example Results**



Three years

Show eyes, nose, mouth (p. 184).
Name objects in a picture (p. 188).
Repeat 2 figures (p. 187).
Repeat a sentence of 6 syllables (p. 186).
Give last name (p. 194).

Four years

Give sex (p. 195).
Name key, knife, penny (p. 195).
Repeat 3 figures (p. 196).
Compare 2 lines (p. 196).

Five years

Compare 2 boxes of different weights (p. 196).
Copy a square (p. 198).
Repeat a sentence of 10 syllables (p. 186).
Count 4 sous (p. 200).
Put together two pieces in a "game of patience" (p. 198).

Six years

Repeat a sentence of 16 syllables (p. 186).
Compare two figures from an esthetic point of view (p. 202).
Define by use only, some simple objects (p. 202).
Execute 3 simultaneous commissions (p. 205).
Give one's age (p. 206).
Distinguish morning and evening (p. 206).

Seven years

Indicate omissions in drawings (p. 207).
Give the number of fingers (p. 209).
Copy a written sentence (p. 209).
Copy a triangle and a diamond (p. 209).
Repeat 5 figures (p. 210).
Describe a picture (p. 210).
Count 13 single sous (p. 210).
Name 4 pieces of money (p. 211).

Eight years

Read selection and retain two memories (p. 211).
Count 9 sous. (3 single and 3 double) (p. 214).
Name four colors (p. 215).
Count backward from 20-0 (p. 215).
Compare 2 objects from memory (p. 216).
Write from dictation (p. 216).

Nine years

Give the date complete (day, month, day of the month, year) (p. 217).
Name the days of the week (p. 218).
Give definitions superior to use (p. 205).
Retain 6 memories after reading (p. 220).
Make change, 4 sous from 20 sous (p. 218).
Arrange 5 weights in order (p. 220).

Ten years

Name the months (p. 221).
Name 9 pieces of money (p. 221).
Place 3 words in 2 sentences (p. 222).
Answer 3 comprehension questions (p. 224).

Seriation of Results Obtained by the Immediate Repetition of Sentences of 14 to 15 Words Each

Children of Seven Years

NAMES	NUMBER OF SENTENCES EXACTLY REPEATED	NUMBER OF THE SENTENCE EXACTLY REPEATED	IN WHICH SENTENCE THE ABSURDITY MANIFESTS ITSELF
Leho	1	2	1, 5, 7
Vala	2	1, 3	2, 4, 5, 6, 7, 8
Dugues	2	2, 5	1, 7, 8
Dast	3	1, 2, 3,	5, 6
Ab	3	1, 2, 3	7
Larch	3	1, 2, 3	7
Pist	3	1, 3, 8	2, 4, 6
Barr	4	1, 2, 3, 5	7
Girau	5	1, 2, 3, 5, 8	6, 7
Vagni	5	1, 2, 3, 5, 8	6, 7

Children of Eleven Years

Corn	3	1, 2, 3	5, 7
Lecle	3	1, 2, 5	4, 7
Taudi	4	2, 3, 5, 7	1
Bertra	4	1, 2, 3, 5	
Calif	5	1, 2, 3, 5, 8	
Lev	5	1, 2, 5, 6, 7	
Gorgi	6	1, 2, 3, 5, 7, 8	
Leno	6	1, 2, 3, 5, 7, 8	
Barr	7	1, 2, 3, 5, 6, 7, 8	
Vign	7	1, 2, 3, 5, 6, 7, 8	

Alfred Binet

Answer 5 comprehension questions (p. 224).

Eleven years

Criticize sentences containing absurdities (p. 227).
Place 3 words in 1 sentence (p. 229).
Find more than 60 words in 3 minutes (p. 229).
Give abstract definitions (p. 230).
Place disarranged words in order (p. 231).

Twelve years

Repeat 7 figures (p. 232).
Find 3 rhymes (p. 232).
Repeat a sentence of 26 syllables (p. 232).
Interpret pictures (p. 193).
Problem of facts (p. 233).

Thirteen years

Paper cutting (p. 234).
Reversed triangle (p. 235).
Give differences of meaning (p. 235).

Figure 1: Examples of questions and results from Binet's mental testing procedure.

As an example of results, panel B shows the performance of seven and eleven-year-old children on the task "immediate repetition of sentences of 14 to 15 words each." In this task, a

sentence was read aloud to a child who then had to repeat the sentence as many times as possible. Multiple performance measures were taken, including the total number of sentences repeated and notes regarding "absurdities", occurrences where the child repeated something deemed absurd by the examiner.

Two key features of the results were common to all the tests and were crucial to Binet's method. First, there were noticeable differences in task performance among children of different age groups. For example, seven-year-old children recited fewer sentences than eleven-year-old children. From my perspective, the average seven-year-old in the table (in the middle of the group) repeated about three sentences, whereas the average eleven-year-old repeated approximately five sentences. Second, there were individual differences within children of the same age group. Some seven-year-old children could recite merely one or two sentences, while others managed to recite up to four or five sentences.

0.0.2.3 Quantifying mental ability with Age

Binet recognized that human intelligence was a large multidimensional concept, but he also sought methods to quantify intelligence in simplified terms that would be familiar and easy to use, similar to using a ruler to measure length. Binet wanted a test that used numbers like a ruler for intelligence: larger numbers would indicate greater intelligence, and smaller numbers would indicate less.

One of Binet's challenges was that he had many different types of tests, each providing multiple measures of performance. He wanted to amalgamate the test results into a single dimension of numbers that would be simple, like a ruler. Binet selected age in years. Like a ruler, age is straightforward, progressing in increments of one year at a time. Children also develop physically and mentally as they grow into adults. Binet's theoretical assumption was that, on average, children's mental abilities steadily increase each year until they reach adulthood. Therefore, seven-year-olds would have more general intelligence than five-year-olds, and eleven-year-olds would be more intelligent than ten-year-olds.

Binet invented a relative scale of measurement to capture his notion of human intelligence. The scale was relative to the average age of children. Both adults and children could be measured, and the test would return a number in years. A precocious seven-year-old might have a mental age of eleven, or an adult might have the mental age of a three-year-old.

0.0.2.4 Comparison to Norms

Binet was well aware that none of his specific tests measured anything as complicated as intelligence. He wrote, "One test signifies nothing, let us emphatically repeat, but five or six tests signify something. And that is so true that one might almost say, 'It matters very little what the tests are so long as they are numerous.' " (Binet & Simon, 1911). What mattered more than the individual tests was Binet's innovation to compare test results from one individual to a larger group of individuals. The large group of individuals who had already been tested provided the "norms" for comparison.

Binet used the concept of average children and assumed that on average children gain more intelligence every year throughout development. Binet's goal was to assess real children and determine whether they were above or below average for their age. However, average children do not actually exist, they are a theoretical and statistical construct. Binet's solution was to create real-world estimates of average children by measuring large groups of children at different ages. He measured hundreds of children of different ages on his tests. In this way, Binet obtained norms or standards that could be used for comparison. For example, he gave three-year-olds his tests and recorded how they performed. The data on three-year-olds demonstrated a range of results. Some three-year-olds scored lower, some were in the middle, and some scored higher on the tests. This range of results then became an empirical standard to judge the mental abilities of three-year-olds who would later take the test.

A parent could have their three-year-old measured on the tests, and Binet could tell the parent how their child performed relative to the other three-year-olds he had measured. Or, if a child was tested and scored as having the mental abilities of a seven-year-old, this would roughly mean that the child performed

similarly to the seven-year-olds that Binet had measured while he was developing his scale.

The measurement of mental abilities based on a relative comparison to empirical norms is still prevalent in modern aptitude testing like the SAT, ACT, GRE, etc., also known as standardized testing. An intriguing feature of this method is that the norms adjust over time as more measurements are taken. For instance, when Binet published his tests, he had collected a limited number of measurements from three-year-olds, which served as the norms for three-year-olds. However, these norms for three-year-olds would update themselves as more three-year-olds were measured in the future.

0.0.2.5 Computing intelligence scores

A remaining issue is how Binet ended up computing an individual child's mental age from their performance across many individual tests. Another way to state this question is to ask how Binet calculated average performance for the children in each age group, and how were the results from new children compared to the results from the existing children in Binet's database.

The answers are that Binet imposed rules for classifying children based on their performance on the tests, and these rules determined how test results were converted into an age in years. Binet experimented with different classification rules and discusses advantages and disadvantages to the scoring methods.

A potential rule was, "A subject has the intellectual development of the highest age at which he passes all the tests, with the allowance of one failure in the tests for that age. Thus young Ernest has passed all the tests at nine years, except one; he has also passed all the tests at ten years except one; therefore we attribute to him the mental level of ten years."(Binet & Simon, 1908)

By 1911, Binet settled on a more nuanced point system:

"Here is the rule to follow: take for point of departure, the age at which all the tests are passed; and beyond this age, count as many fifths of a year as there are tests passed. Example: a child

of eight years passes all the tests of six years, 2 of seven years, 3 of eight years, 2 of nine years, 1 of ten years; he has therefore the level of six years plus the benefit of eight tests or eight-fifths years, or a year and three-fifths, equaling a level of seven years and three-fifths, or more simply 7.6. This calculation permits the appreciation of the intellectual level by means of a fraction. But it must be well understood that this fraction is so delicate an appreciation, that it does not merit absolute confidence, because it varies appreciably from one examination to another." (Binet & Simon, 1911).

0.0.2.6 Binet's use cases

Binet developed a method that has been used worldwide to measure mental abilities. There are some strikingly simple features of the test that make the method quite compelling. The main idea was to create a vast database documenting how children of various ages perform mental tasks, and then compare new individuals assessed on the same tasks against this database. This would yield a "mental age," a simple figure that represented mental abilities in relation to a large and ever-expanding database.

One of Binet's use cases was to guard against bias when making decisions about child-welfare. Teachers might send children away to institutions just because they were unruly. Without an objective test of mental ability the teacher could claim the unruly students were mentally unfit for school. Other children may have been unjustly institutionalized, and without any objective way to show these children belong in a regular school, they would be forced to stay in the institution. Indeed, Binet tested institutionalized children (Binet & Simon, 1905c) as a part of his research process. On the one hand, he used results from institutionalized children to further validate his tests and show that institutionalized children had lower "mental ages" when compared to non-institutionalized children. On the other hand, Binet also found cases where institutionalized children were shown to be of "normal" intelligence relative to their non-institutionalized peers. Thus, scores on the tests could potentially be used to save some children from institutionalization.

0.0.2.7 Meaning of the measure

The previous section reviewed one of the first so-called "intelligence" tests developed by psychologists, but avoided giving a definition to the word intelligence. So, what is the definition of the word intelligence? Does the Binet-Simon test actually measure intelligence according to that definition? If not, what does it measure?

Operational definitions are a research tool commonly used in psychology that will can help an evaluation of these issues. Operational definitions are used by individual researchers to specify the meaning of their own terms. For example, Binet could employ an operational definition of intelligence in terms of his own test. Here, the word intelligence would cease to have any common everyday meaning, and would only be used as a shorthand term to refer to patterns of performance on the test. And of course, just calling a test an "intelligence" test, doesn't mean that it measures the everyday meaning of intelligence. Clear operational definitions can benefit research because they allow researchers to communicate effectively, with clarity, and with terms that are limited to the context of the research. Operational definitions can also cause confusion, especially when researchers choose terms that already have well-established meaning in everyday usage. For example, using "intelligence" as a name for a test could easily confuse people who might assume the name was actually meaningful with respect to the everyday concept of intelligence.

Considering how operational definitions work, I can not give a clear answer about whether or not Binet's test actually measures intelligence. There are two problems. First, not everyone agrees on what intelligence is in the first place. Second, without an agreed upon definition, it is difficult to justify why each component of Binet's test actually measures some component of intelligence in a meaningful way.

Wikipedia gives the following definition of intelligence:

"Intelligence has been defined in many ways: the capacity for logic, understanding, self-awareness, learning, emotional knowledge, reasoning, planning, creativity, critical thinking, and problem-solving. More generally, it can be described as

the ability to perceive or infer information, and to retain it as knowledge to be applied towards adaptive behaviors within an environment or context."

From my perspective, the meaning of intelligence is large, varied, and has different meanings for different people, just like the word cognition. As a result, it is difficult for any test to claim that it measures something as diffuse as intelligence. An analogy might be to music. Imagine someone claiming they had a music test that measured whether a given recording had more or less music. Music is so highly diverse that it would be ridiculous (in my opinion) to try to measure all of music on a single scale with a "music test". Plus, who get's to define which music is counted as good music? The surrounding issues that would emerge from debates about a such a fictional music test, also surround debates about what intelligence tests mean.

Although I won't commit to a definition of intelligence, and I won't claim that the Binet-Simon test measures intelligence, it's clear that the Binet-Simon test can be inspected and evaluated. It is possible to interpret the meaning of the results based on the testing procedures described by Binet. For example, Binet used a process of trial-and-error to create mini-tests of different abilities that were age-appropriate. Binet specifically designed the tests so that the majority of children in a specific age group could successfully complete their assigned tests as well as those assigned to younger age groups. However, these same children would struggle when attempting tests assigned to older age groups. Also, Binet chose tests that produced variation within an age group–some children of the same age would do better or worse on the same test. Binet measured many children of all age groups on the tests. Then he proposed to compare new children, measured on the same tests, to his growing database.

So what does the Binet-Simon test measure? It measures how a child's performance on Binet's chosen mini-tests compares to performance on the same mini-tests by groups of children of different ages. This is not a straightforward measure like a ruler, where one inch on the ruler is one inch in the world. Although the Binet-Simon test produces a number in years, that refers to "mental age"; the number is deceiving because it depends on several shifting components. The components

include the actual mini-tests that Binet chose, the children that Binet tested to form the empirical comparison group, and the classification rules that Binet used to assign years to children based on their test performance. Changes to any or all of these many components would change how a child would be scored.

0.0.3 Mental testing and eugenics in America

After the Binet-Simon test was translated to English, it was popularized in America as the Stanford-Binet test. Many of the American psychologists who would use the new intelligence test were also advocates, members, or leaders in eugenics societies. As a result, intelligence testing was used in America as a tool to further the cause of the eugenics movement.

0.0.3.1 The Alpha-Beta test

In 1917, the same year that America entered World War I, the APA appointed committees to study the situation and prepare for action (Yerkes, 1918) and the National Research Council created a Psychology Committee to examine similar issues (Yoakum & Yerkes, 1920). Many psychologist committee members were proponents of eugenics, including: Robert Yerkes, Madison Bentley, Edward Thorndike, John B. Watson, Walter D. Scott, Robert Woodworth, and Carl Seashore (all of whom would take a turn as APA president).

War was a topic of considerable debate among eugenics societies (Kühl, 2013). It was clear that many people would perish during war. On the one hand, according to the logic of eugenics, if "low-quality" people tended to perish, then war could be positive for eugenics because those people would no longer be around to breed. On the other hand, the loss of "high-quality" individuals would be a negative outcome as their genes would also be lost. Furthermore, there was no way to measure the eugenic quality of soldiers and then use that information to determine who would live or die during war. Eugenicists were advocates of using intelligence tests on soldiers to help make

personnel selection decisions, such as who would become an officer, and who would have a higher probability of being killed by being sent to the front.

American Psychologist Robert Yerkes (APA president in 1917) wanted to establish a "mental census" of Americans, and then use that information to improve American society from a eugenics point of view. Yerkes advocated for the testing of all Americans and organized the largest mass mental testing of American men for the draft, administering tests to 1.75 million adults (Yerkes, 1923). There were two versions of the test (inspired by Stanford-Binet). The "alpha" test was created for soldiers who could read, and the "beta" test was created for soldiers who could not.

In his 1923 report, "Eugenic Bearing of Measurements of Intelligence in the United States Army," Yerkes describes the methods and results from the alpha-beta testing efforts and lists these five main reasons to conduct such widespread testing:

- In the discovery of men whose superior ability recommends their advancement.
- In the prompt segregation in the Development Battalions of intellectually inferior men whose inaptitude would retard the training of the unit.
- In building organizations of equal or appropriate strength.
- In selecting suitable men for various army occupations or for special training in the technical schools.
- In eliminating the feeble-minded.

0.0.3.2 Scientific Racism

Eugenics ideology typically included racist beliefs about the superior or inferior eugenic qualities of different ethnic groups (Turda, 2010). The Alpha-beta tests of 1.75 million American men produced results that fit existing eugenic ideology about inherent differences in intelligence between ethnic groups. For example, psychologist Carl Brigham, wrote an entire book analyzing the results of the Alpha-Beta tests (Brigham, 1922). He concluded that white Americans had superior intelligence to black Americans and immigrants. He also created dire warnings

about the future of America, suggesting that American intelligence was rapidly declining. He warned that although American deterioration was imminent, it could be prevented through public action and laws. For example, eugenically inferior immigrants could be kept out of the country. And, increased segregation of whites and blacks, along with laws against intermarriage could prevent further mixing of the races.

As discussed earlier, it is not entirely clear what these so-called intelligence tests measure, or what intelligence itself actually refers to. Nevertheless, proponents of eugenics were quick to claim that results from the tests really did legitimately measure supposedly intrinsic and genetically inherited qualities of humans that made some superior and others inferior. Alternative interpretations, such as the tests measured culturally acquired aptitudes, took additional time to be seriously considered. Racist motivations would continue to daunt intelligence testing in America, and an extended history is beyond the current scope of this chapter.

0.0.3.3 Mental health

The eugenics movement deeply impacted public policy and stigma around mental health. The last chapter mentioned that American eugenics proponents successfully petitioned for laws to legalize forcible sterilization of people deemed to be "feeble-minded". The invention of intelligence tests was heralded as a new scientific tool for the identification of "feeble-minded" people so that they could be segregated and/or sterilized. Psychologist Henry Goddard provides a case example of connecting intelligence testing to the eugenics agenda for treating mental health issues.

Goddard was director of research at the Vineland Training School for Feeble-Minded Girls and Boys in Vineland, New Jersey. He also arranged for the English translation of Binet's work in 1916 (Kite, 1916). Robert Yerkes visited Goddard and used his facilities at Vineland during the development of the Alpha-Beta test. Goddard was heavily involved in eugenics, and one illustrative example is from his 1927 article, "Who is a Moron?" (Goddard, 1927).

Goddard advocated the use of new terms for categorizing levels of "feeble-mindedness" based on intelligence tests. For example, "idiots" had the mental age of two-year old children, "imbeciles" had the intelligence of three to seven-year-olds. Goddard invented the term "moron" to describe people with the same supposed intelligence as eight to twelve-year-olds. The rest of his article describes eugenic ideology in the form of panic about how society is in danger unless it acts to solve the moron problem.

According to Goddard, morons were a problem because they appeared normal, and might only be identified with an intelligence test. As a result, hidden morons passing as normal people were running amok in society, and they also had many children, so they were potentially deteriorating the gene-pool by breeding. He considered extreme eugenic solutions, and wrote, "perhaps our ideal should be to eventually eliminate all the lower grades of intelligence and have no one who is not above the twelve-year intelligence level", but he also cautioned that eliminating half of society would be impossible and even undesirable. Instead, Goddard proposed that morons could be cured through education, and become very useful to society as workers who would very happily do the jobs they were trained to do.

As with Yerkes, Goddard forwarded intelligence tests as legitimate scientific measures of human quality that should be used to make decisions about the welfare of American citizens and their position in society.

0.0.3.4 Education and The Black Psychologists

The methods involved in intelligence testing became widespread in education in the form of standardized testing. Similar to previous examples, proponents of eugenics were involved in these efforts. For example, the SAT was created by Carl Brigham shortly after he published his book on "A study of American Intelligence" (Brigham, 1922). The eugenic purpose of the tests was to sort children in terms of their quality, and then give more resources to the education of superior children, and fewer resources to the education of inferior children.

Testing of American children showed achievement gaps between different ethnic groups. And, to fast forward into the 1960s, there was growing concern among Black Psychologists (only 1% of a primarily white discipline at the time) that educational testing and decision-making policy was harming outcomes for Black children. In 1968, the Association of Black Psychologists (ABPsi) was formed as a national organization during the San Francisco meeting of the American Psychological Association (APA) (for a more complete history see, R. Williams, 1974). The ABPsi was formed both to address needs within the small community of Black Psychologists that were not being met by the APA, and to petition the APA to address many broader concerns. For example, the ABPsi adopted the following statement on mental testing:

"The Association of Black Psychologists fully supports those parents who have chosen to defend their rights by refusing to allow their children and themselves to be subjected to achievement, intelligence, aptitude and performance tests, which have been and are being used to:

1. Label black children as uneducable;
2. Place black children in special classes;
3. Potentiate inferior education;
4. Assign black children to lower educational tracks than whites;
5. Deny black children higher educational opportunities; and,
6. Destroy positive intellectual growth and development of black children.

The ABPsi's calls for a moratorium on testing (Graves & Mitchell, 2011; see also, R. L. Williams & Mitchell, 1978) in 1969 were not supported by the APA, whose membership also had financial interests in the large educational testing industry. Although there were efforts to maintain a relationship with the APA, the ABPsi became a distinct professional organization and has been publishing its own journal since 1974. ABPsi has been on the forefront of modern research on the impact and legacy of eugenics and racism in Psychology. Finally, although testing is still widespread in America, a moratorium on intelligence testing of Black children was accomplished in

1979 through advocacy of Black Psychologists in California (Frisby & Henry, 2016), where the practice remains illegal.

0.0.4 Cognition: Testing abilities vs. testing theories

This textbook is aimed at providing an overview of research into cognition. One piece of that history involves psychologists who were advocates of the eugenics movement and who created tests of cognitive abilities with the purpose of deploying those tests on society to further the aims of the eugenics movement. The role of eugenics in motivating the need to create tests of cognitive ability, and in spreading the use of those tests across society is not commonly discussed in introductory textbooks. However, the historical context raises important questions about how the enterprise of scientific research contributes to societies that fund it, and we should be mindful of these issues throughout our course. For example, it is clear that psychologists researching human cognition can produce tools of questionable merit that become widely adopted in society and that continue to have positive and negative outcomes for different groups of people in society.

The next chapter begins our transition into more conventional areas in cognitive psychology, like the domains of learning and memory. We will encounter more examples of cognitive tasks that are very similar, if not essentially the same, as the mini-tests used as part of intelligence scales. However, rather than testing for the purpose of measuring and classifying the people, we will discuss examples where cognitive tasks are used to test theories and claims about how cognition works.

0.0.5 Appendix

0.0.5.1 * References

Binet, A., & Simon, T. (1905a). Application of the New Methods to the Diagnosis of the Intellectual Level among Normal and Subnormal Children in Institutions and in the Primary Schools. *L'Année Psychologique*, *12*, 245–336.

Binet, A., & Simon, T. (1905b). New methods for the diagnosis of the intellectual level of subnormals. *L'Année Psychologique*, *12*, 191–244.

Binet, A., & Simon, T. (1905c). Upon the Necessity of Establishing a Scientific Diagnosis of Inferior States of Intelligence. *L'Année Psychologique*, *12*, 163–191.

Binet, A., & Simon, T. (1908). The Development of Intelligence in the Child. *L'Année Psychologique*, 1–90.

Binet, A., & Simon, T. (1911). New investigations upon the measure of the intellectual level among school children. *L'Année Psychologique*, 145–201. https://doi.org/c4b7b9

Brigham, C. C. (1922). *A study of American intelligence*. Princeton: Princeton University Press; London: Oxford University Press ….

CATTELL, J. McK. (1890). V.—MENTAL TESTS AND MEASUREMENTS. *Mind*, *os-XV*(59), 373–381. https://doi.org/dhn9nc

Cattell, J. M., & Farrand, L. (1896). Physical and mental measurements of the students of Columbia University. *Psychological Review*, *3*(6), 618. https://doi.org/ckms9q

Frisby, C. L., & Henry, B. (2016). Science, Politics, and Best Practice: 35 Years After Larry P. *Contemporary School Psychology*, *20*(1), 46–62. https://doi.org/gkzfj6

Gilbert, J. A. (1895). THE MENTAL AND PHYSICAL DEVELOPMENT OF SCHOOL CHILDREN.—(I). *The Journal of Education*, *42*, 291–292.

Goddard, H. H. (1927). Who is a moron? *The Scientific Monthly*, *24*(1), 41–46.

Graves, S., & Mitchell, A. (2011). Is the Moratorium Over? African American Psychology Professionals' Views on Intelligence Testing in Response to Changes to Federal Policy. *Journal of Black Psychology*, *37*(4), 407–425. https://doi.org/d56z38

Kite, E. S. (1916). *The development of intelligence in children (The Binet-Simon Scale)*. Williams & Wilkins Company.

Kraepelin, E. (1895). Der psychologische Versuch in der Psychiatre. *Psychologische Arbeiten*, *1*, 1–92.

Kühl, S. (2013). *For the Betterment of the Race - The Rise and Fall of the International Movement for Eugenics and Racial Hygiene*. Palgrave Macmillan. https://www.palgrave.com/gp/book/9781137286116

Munsterberg, H. (1891). Zur individual Pscyhologie. *Centralblatt Fur Nervenheilkund Und Psychiatrie*, *2*, 196–198.

Nicolas, S., Coubart, A., & Lubart, T. (2014). The program of individual psychology (1895-1896) by Alfred Binet and Victor Henri. *LAnnee Psychologique*, *Vol. 114*(1), 5–60. https://www.cairn.info/revue-l-annee-psychologique1-2014-1-page-5.htm

Terman, L. M. (1916). *The Measurement of intelligence: An explanation of and complete guide for the use of the stanford revision and extension of the Binet-Simon Intelligence Scale*. Houghton Mifflin. https://books.google.com?id=qtmdqoFx55YC

Turda, M. (2010). Race, science, and eugenics in the twentieth century. *The Oxford Handbook of the History of Eugenics*, 62–79.

Williams, R. (1974). A History of the Association of Black Psychologists: Early Formation and Development. *Journal of Black Psychology*, *1*(1), 9–24. https://doi.org/gg3hq4

Williams, R. L., & Mitchell, H. (1978). What Happened to ABPsi's Moratorium on Testing: A 1968 to 1977 Reminder. *Journal of Black Psychology*, *4*(1-2), 25–42. https://doi.org/gg6btt

Yerkes, R. M. (1918). Psychology in relation to the war. *Psychological Review*, *25*(2), 85–115. https://doi.org/dhdj4j

Yerkes, R. M. (1923). Eugenic Bearing of Measurements of Intelligence in the United States Army. *The Eugenics Review*, *14*(4), 225–245.

Yoakum, C. S., & Yerkes, R. M. (1920). *Army mental tests*. H. Holt.