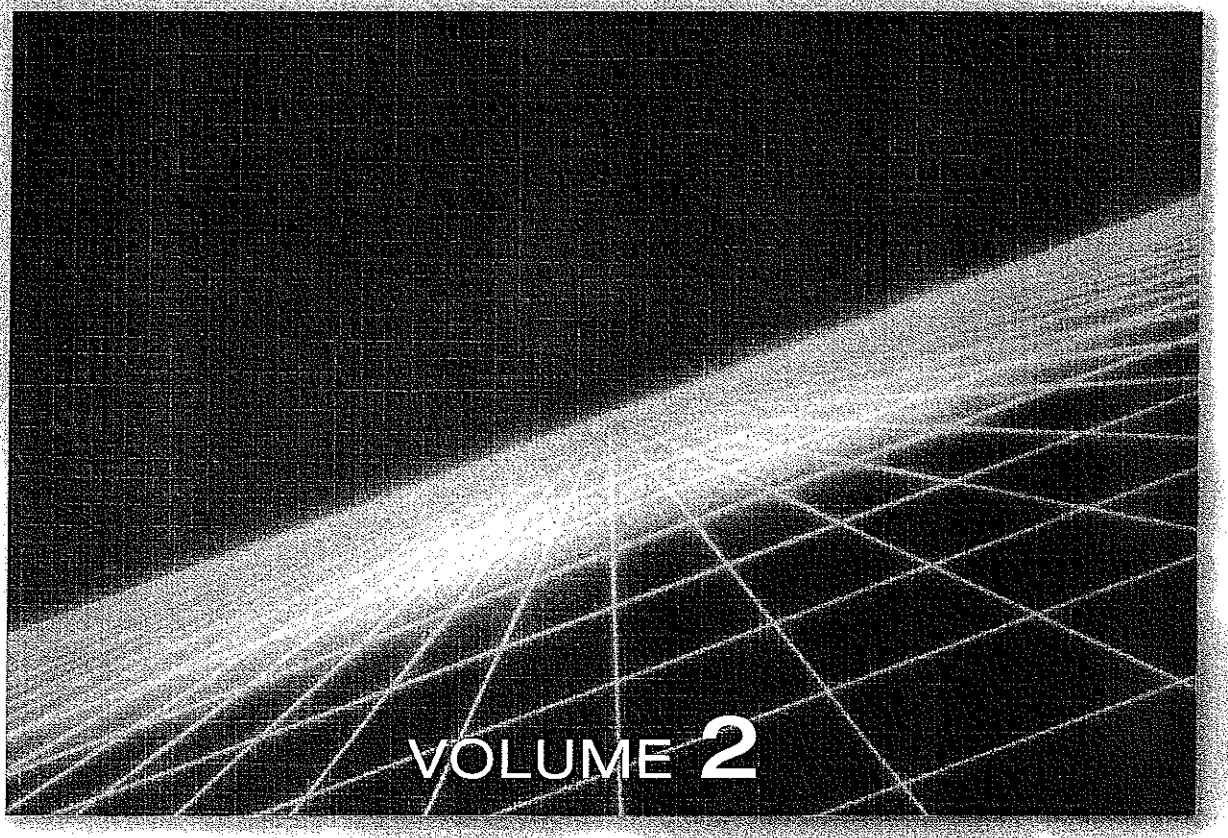
ENCYCLOPEDIA OF MEASUREMENT AND STATISTICS



VOLUME 2

EDITED BY

NEIL J. SALKIND

Further Reading

Anderson, T., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics, 20,* 46–63.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variable estimation when the correlation between the instruments and endogenous explanatory variables is weak. *Journal of the American Statistical Association, 90,* 443–450.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica, 55*(4), 765–799.

Stock, J. H., & Watson, M. W. (2003). *Introduction to econometrics*. Reading, MA: Addison-Wesley.

Wooldridge, J. M. (2005). *Introductory econometrics* (3rd ed.). Belmont, CA: Thomson.

Intelligence Quotient

Few terms generate as much anxiety, or as much controversy, as *intelligence quotient* (IQ). The term is emotionally loaded and carries with it the connotation of a fixed entity that resides in each person to varying degrees and that, to a large extent, determines a person's worth in the larger society. Yet, as so often happens, the connotation of IQ today, 100 years after the term was coined, represents something quite different from the original meaning and spirit behind the phrase.

Historical Background

Before Alfred Binet (1857–1911) and, to some extent, Francis Galton (1822–1911), gradations of mental worth were generally determined by physiological indices, such as cranial capacity. Galton extended the psychometric assessment of human qualities in many directions, including the assessment of intellect with various response time measures; however, Binet's work had the single most significant impact on the conceptualization of IQ.

Binet was commissioned by the French government to assist with the identification of students who were unlikely to benefit from ordinary schooling and therefore should be offered remedial or special education. Having become quite discontented with the utility of cranial measures, Binet searched for something more definitive. His early ideas for developing a test of intelligence drew heavily from one of his countrymen, Blin, who had developed a series of structured questions that were designed to assess the judgment abilities of the individual.

It is important to note that Binet was very explicit in stating that the scores derived from his tests were rough, that they were not intended for use in ranking normal children, and above all else, they were indicators of current functioning and did not speak to the past or future capabilities of the child. An educator at heart, Binet was a strong believer in cognitive modifiability, a view that suggests that intelligence is not a fixed quantity, but one that can be modified and enhanced. This view tends to sit in stark contradiction with many modern theories of intellectual ability, which suggest that intelligence is an innate and relatively fixed capacity.

As part of his remedial education programs, Binet advocated what he called exercises in "mental orthopaedics." These were based on the belief that one first needed to learn how to learn. He linked increased academic performance as a function of training to an increase in intelligence. Binet was also concerned that scores on his tests should not be misinterpreted, and he cautioned overzealous teachers against the temptation to use the test results to get rid of unruly or uninterested students.

The Problems of Measuring Intelligence: The Appearance of Mental Age

Binet recognized that when he added up the marks on his scales, the score in and of itself was unable to tell him very much about the ability of the individual. What was needed was some way to compare a child's score with some benchmark. Binet was particularly interested in disentangling native intelligence from the effects of schooling; thus, tests of educational achievement would not serve as an appropriate comparison. Binet recognized that this benchmark needed to be empirical, because he was rightfully cautious

about accepting the evaluations of parents and teachers too literally, believing they were susceptible to any number of biases (e.g., the protective parent might exaggerate the capacity of his or her child; a teacher wanting to minimize troubles in his or her classroom may provide an underwhelming evaluation of the child). Furthermore, without some clearly defensible and replicable criterion that could be agreed upon, meaningful assessment of change would not be possible.

With this in mind, Binet set out to establish a replicable, empirical criterion for grading intelligence. The first problem that needed to be addressed was the generally accepted observation that intelligent *behavior* tends to increase with age. The line of reasoning that Binet pursued to deal with this is very illuminating and, although subsequently modified and improved, serves as an exemplary model of the application of sound logic and experimentation to the investigation of human ability. Binet, along with his student Theodore Simon, administered his tests to children of different ages and collated their scores. They then ranked the tests according to difficulty. Out of this work, the term *mental age* was coined. Mental age was defined as the average age of a child of normal intelligence who could pass the test. It is significant that the chosen comparisons were normally functioning children. Hence, regardless of his or her age, a child who was able to pass a test that a normal child was able to pass at 5 years old was assigned a mental age of 5. This meticulous process of establishing benchmarks for comparing individuals still informs modern methods of test development, albeit in more sophisticated forms. We call this process *test standardization,* and the benchmarks *norms.*

Problems With Mental Age and the Advent of the IQ

Almost immediately, mental age ran into problems. Consider two individuals with the same mental age of 9 years; the first is 18 years old and the second is 6. The 18-year-old is likely to think qualitatively differently from the 6-year-old, even though they have the same mental age—their behaviors, judgments, and processes are simply different. The solution as to how to evaluate such differences meaningfully was to consider the ratio of mental age to chronological age—this ratio was called the intelligence quotient, or IQ. Using this transformation, an 18-year-old with a mental age of 9 would have an intelligence quotient of 0.50. The 6-year-old with the same mental age would have an intelligence quotient of 1.50. To remove the decimal and simplify reporting, the quotient was multiplied by 100, and hence, the 18-year-old was assigned an IQ of 50, and the 6-year-old an IQ of 150. The lay person's understanding of the IQ scale has generally stayed the same ever since.

A more fundamental problem influenced the reliable application of IQ scores based on mental age. This problem was particularly apparent when, contrary to Binet's original intentions, the scale was modified for use in the United States to rank normal and superior children and adults. As mentioned earlier, performance on the cognitive tests of Binet increased with age. But intelligence does not keep increasing indefinitely. At some point, an increase in age will not contribute to any significant improvement in performance on the test. Hence, to apply the IQ to adults, the chronological age needed to be truncated at a suitable point to allow the IQ scale to remain meaningful. Determining the appropriate point to truncate was complicated by the fact that the mental age of a test is determined by its difficulty. An easy test will have a low mental age and a harder test a higher mental age. Researchers presented various arguments to support their chosen point of truncation, and for practical purposes, this seemed to address the problem. However, the IQ score has an even more troublesome limitation.

Problems With IQ and the Advent of the Deviation IQ

One of the proposed benefits of the IQ over mental age was that it facilitated comparison of individuals at different ages. It was recognized that the useful interpretation of IQ required a constancy of scores across ages. That is, the ratio of mental age to chronological age should be the same (i.e., 100) for a normal functioning child no matter how old he or she is. However,

Wechsler reports that in the 1937 version of the Stanford-Binet intelligence scale, there was considerable variability in mean IQ scores across different ages—the mean IQ for 2½-year-old children was 109.9, but only 100.5 at 14 years. Furthermore, Wechsler argued that the variability (standard deviation) in IQ scores also differed considerably at different ages. The standard deviation in IQ scores for 12-year-olds was 20 IQ points, yet the standard deviation was only 12.5 points for 6-year-olds. The reason for this variation might be interpreted to lie with the particular characteristics of the standardization sample chosen as the comparison group, and there is evidence that this was recognized by researchers at the time. However, the problem this presented remained a significant obstacle to the reliable interpretation of IQ scores and had the potential to render any comparisons of individuals over time or at different ages meaningless.

The solution was an act of statistical ingenuity. As Wechsler describes, scores on the test were just that—scores. The numbers used to represent the scale are always arbitrary in the sense that there is no fixed point of origin—zero on the scale did not imply no intelligence, and each test's mean score depended on its difficulty. Hence, mental age and IQ were just numbers on some arbitrary scale determined by the researcher. However, it is possible to mathematically transform the scores on one scale to any other scale without changing the rank ordering of individuals or the *relative* distance between them. This is what we do when we convert raw scores to standardized scores (i.e., z scores) in, for instance, the process of testing for statistically significant differences between two population means. z scores always have a mean of zero and a standard deviation of one, regardless of what the original raw score scale is. Furthermore, if we have two groups of people, say, 5- and 10-year-olds, and we convert the raw scores of each group separately to z scores, then the mean and standard deviation will be the same (0 and 1, respectively) for both groups. Herein lies the ingenuity. With this simple transformation, we now have a scale that we can set to have the same mean and standard deviation for any number of subgroups. Furthermore, it is a simple mathematical calculation to convert any z score to have any other mean and standard deviation we choose. In the context of assessing intelligence, the obvious choice was the one that practitioners had become accustomed to. Hence, raw scores were transformed so that the mean for each age subsample of the standardization group was set to be 100, and the standard deviation was set to be around 15 (different test developers set slightly different standard deviations). This new IQ is referred to as the deviation IQ. It does not depend on a concept of mental age, and chronological age is used only for grouping.

An important caveat to this applies. The appropriateness of the linear transformation to z scores and the subsequent interpretations are premised on the assumption that the original raw scores fall along an equal interval scale. Equal interval scaling between scores is required if we are to make meaningful comparisons of differences. This has continued to be a major controversy in psychological assessment generally and one that is rarely questioned in clinical applications of IQ.

Contemporary Perspectives on IQ: A Cautionary Concluding Note

IQ is an aggregated score. To the extent that it is meaningful to aggregate scores across disparate tasks, such as Binet presented and as used in modern tests, such a single score is potentially appropriate. However, the current dominant theories suggest that intelligence is multifaceted and composed of distinct, though related, classes of ability. For instance, using sophisticated statistical techniques, McArdle, Ferrer-Caja, Hamagami, and Woodcock investigated the developmental trajectories of a range of different cognitive abilities. They interpreted significant deviations in the trajectories of these separate abilities from the *IQ-equivalent* trajectory to suggest that a description of the cognitive system with only a single factor is overly simplistic. Hence, although the notion of IQ has become deeply ingrained in modern language, great care is required with interpretation.

—Damian P. Birney and Steven E. Stemler

See also Intelligence Tests

Further Reading

Binet, A., Simon, T., & Terman, L. M. (1980). *The development of intelligence in children* (1916 limited ed.). Nashville, TN: Williams Printing.

Gould, S. J. (1996). *The mismeasure of man* (Revised and expanded ed.). New York: Penguin.

Jensen, A. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger/Greenwood.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology, 38*(1), 115–142.

Michell, J. (1999). *Measurement in psychology: Critical history of a methodological concept.* New York: Cambridge University Press.

Plucker, J. A. (Ed.). (2003). *Human intelligence: Historical influences, current controversies, teaching resources.* Retrieved from http://www.indiana.edu/~intell

Stankov, L. (2000). The theory of fluid and crystallized intelligence: New findings and recent developments. *Learning and Individual Differences, 12,* 1–3.

Sternberg, R. J. (1997). *Successful intelligence: How practical and creative intelligence determine success in life.* New York: Plume.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore, MD: Williams & Wilkins.

Intelligence Tests

An intelligence test is a structured situation designed to elicit information about the cognitive abilities of an individual. The test may be administered individually or in a group. Scores are usually reported on a scale in which 100 indicates average intelligence. Scores are scaled so that about the top 16% of the population will receive scores of 115 or above, the top 2.5% will receive scores of 130 or above, the bottom 16% will receive scores of 85 or below, and the bottom 2.5% will receive scores of 70 or below.

The typical intelligence test will have a variety of items designed to tap different aspects of the person's cognitive abilities. Some of the items may ask for specific pieces of information, such as how many years there are in a decade or how much change you would receive if you bought an article costing $18.67 and you gave the clerk a $20 bill. Other questions might ask about objects missing or out of place in a picture; still others would be tests for memory, such as repeating a list of 5 digits that have been read, or tests of reasoning such as finding the right pattern piece to complete a design. In an individually administered test, the examiner asks each question, records the answer, and makes a judgment as to the answer's correctness or quality. Testing stops when the examinee has failed to answer a specified number of questions correctly. When the test is administered to a group, the questions are often in multiple-choice format, and responses are usually recorded by filling in bubbles on the answer sheet. Answers are compared to a key, so judgment as to correctness is avoided.

History of Intelligence Testing

Alfred Binet is generally given credit for creating the first modern intelligence test in 1905. In the 1908 version of his test, Binet introduced the idea of "mental level" as a way to express the cognitive ability of a child. The mental level of an item was the age at which the average child could solve that particular problem. An item that could be solved by the average child of age 7 or above, but not by a child of age 6, was given a mental level of 7 years. Items were grouped by mental level, and testing ended at the first level where a child could not answer any items correctly.

Henry Goddard popularized Binet's 1908 test in the United States. Several English-language versions of the Binet scale were quickly developed by Goddard and others. In 1916, Louis Terman published an American edition that came to be called the Stanford-Binet and soon replaced all competitors. This test popularized the term *intelligence quotient,* or IQ, because scores were expressed as the ratio of mental level or mental age, divided by actual or chronological age. A child who tested "at age" received an IQ of 1.00. This ratio came to be multiplied by 100 to remove the decimal point, resulting in a scale where the average IQ is 100, the reference point still in use.

During World War I, American psychologists under the leadership of Robert Yerkes produced two new group-administered tests for screening Army draftees: