

Are intelligence tests outmoded?

Most doctors know little about clinical and educational psychologists' areas of expertise; although most do know that they carry out intelligence tests. Now, just when doctors have accepted the intelligence quotient (IQ) as a measure of intelligence, psychologists seem reluctant to use the term. Why? Is something wrong with the concept of IQ? Are IQ tests outmoded?

The history of IQ tests

According to Eysenck, the ancient Greeks introduced the concept of intelligence.¹ They regarded it as the potential to think, reason, and solve mental problems, and they differentiated it from a person's observed behaviour. Centuries later tests of intelligence evolved, the first widely used one being Binet's, which was developed in France in the early 1900s.

Binet's original test still exists today, recently restandardised as the Stanford-Binet test of intelligence. The original purpose of the test, which was designed at the request of the French education authorities, was to select children who were having difficulties in learning so that they could have extra tuition. Binet himself believed that ability could change, but many of those who adopted the idea of testing IQ believed that intelligence was inherited and essentially unalterable. This belief was particularly strongly held by the eugenicists, who proposed that 'superior' people should be encouraged to procreate while 'inferior' people, such as criminals, epileptics, alcoholics, and imbeciles, should be discouraged from doing so. Indeed, in more than 30 American states in the early twentieth century there was legal provision for such people to be compulsorily sterilised.²

More recent IQ tests

Since Binet's time other IQ tests have been developed. They are all designed to measure intelligence, even though there are still major disagreements between psychologists about what intelligence is.³ The tests include series of tasks, both language based and language free. During the development of a test many people (children or adults, depending on the age range to be tested) are given the same tasks, and scoring is adjusted so that someone who obtains a raw score exactly in the middle of the range will be assigned a scaled IQ score of 100—that

is an average IQ. Most test scores have a standard deviation of 15 so that in theory, assuming a normal distribution of scores, about 67% of the population will achieve scaled scores between 85 and 115 (one SD below and above the mean) and about 95% will score between 70 and 130 (two SD below and above the mean), leaving about 2.5% at the extreme ends of the distribution curve. This is not, however, as straightforward as it sounds; for instance, IQs of given populations do not conform exactly to a normal distribution. In addition, the composition of the standardisation group is crucial: if, for example, only highly intelligent people are tested during the standardisation, when the rest of the (untested) population is assessed most of them will score below average. The converse also holds true, and this makes the interpretation of IQ scores difficult. Most importantly, it means that if groups of people are tested who were not represented in the standardisation sample their IQ scores can be interpreted only as indicating that they are above or below the mean of the standardisation sample, rather than above or below the mean of the population of which they are a part. The mean IQ of a population is thought to rise by about 0.3 IQ points each year. Thus if tests with out of date standardisation scores are used children's IQ are overestimated if their scores are thought to compare to those of the current population (the original mean development quotient score, on the Griffiths scale was 100, but the current mean is 110 or more).

What do IQ tests measure?

IQ tests measure a person's performance on a series of tasks, in a particular setting, on a particular day, with a particular tester. Whether the person's score reflects his or her underlying ability, (in comparison with others) to think, reason, and solve mental problems will depend on many factors. One of these is the degree to which such a score can be generalised for different days, settings, and testers for the person in question. Another important factor is the degree to which scores can be generalised for particular tasks. A third factor is the person's motivation.

Most well constructed IQ tests contain information about the short term reliability of scores in the test manual; this is in the form of test/retest checks to show the degree to which people's scores on one day correlate with their scores on a later day. Over short

periods the correlations are usually good—for example, in the Weschler intelligence scale for children (revised) (WISC-R) the test/retest reliability over four weeks for the full scale IQ was about 0.95. Conventional IQ test manuals also provide figures for the standard error of the test scores—that is, the error attributable to ‘imperfect measuring procedures.’ This allows the tester to state that, for example, the ‘true’ IQ score will lie between stated limits on 95% of occasions. Thus if a child aged 9½ has a score of 90 on the WISC-R it is 95% certain that his ‘true score’ will lie between 84 and 96 (his actual score \pm twice the standard error). Other tests, such as the McCarthy scales, Weschler preschool and primary scale of intelligence (WPPSI), Weschler adult intelligence scale (WAIS), and British ability scales (BAS), have similar margins of error. All these tests include specific instructions to try to reduce the variations between testers, and the level of error may be expected to be higher if the tester deviates from the procedure laid down in the manual—for instance, in how much practice he allows on a test item, or how much help he gives on an item.

The degree to which a person’s scores may be generalised for different tasks is more complicated. One way of examining this issue is to look at the correlations between people’s scores for different IQ tests. To the uninitiated the tasks in different IQ tests look similar, but the published correlations between IQs using different tests vary between 0.45 and 0.95. Thus even though those who construct the tests try to include a wide variety of tasks and do their best to avoid using tasks that depend on a person’s cultural experiences (something that many psychologists think is impossible), a person’s IQ score obtained with one IQ test can differ considerably from the score obtained with a different test.

Finally, one of the variables that intervenes between a person’s underlying capacities and his behaviour (and hence IQ score) is motivation. Disaffected youths for example, may not be motivated to try to complete tasks, nor may children with whom the tester has poor rapport because of differences in class, race, culture, or personal affinity.

Changes in IQ over time

Do people have IQs in the same sense that they have blue eyes? It seems to be a common belief among the general population that a person’s IQ changes little over time. Research, however, suggests that this is not so. If people are given IQ tests on several occasions, years apart, the usual finding is that the scores correlate positively, with higher

correlations when the intervals between tests are short and the lowest correlations between tests carried out in infancy and those done in later childhood. The reasons for the low correlations between infant test scores and those later in childhood are controversial. A major factor is probably that most tests of infant ability concentrate on tasks assessing motor skills; more recently developed tests that concentrate on the infant’s ability to deal with information are better predictors of later IQ. Even when the correlations between tests are high, however, Hindley and Owen have shown that quite wide variations in IQ are not unusual.⁴ In their longitudinal British study they found that between the ages of 8 and 17 the scores of half of their subjects changed by 10 points or more, and one quarter changed by 17 points or more. The correlation between the scores at 8 and 17 years was 0.74. This, however, was a study of a sample of children in London and entailed two different IQ tests. Correlations for samples from more stable populations may change less, especially if the same or similar tests are used. Yule *et al* found a correlation of 0.86 between scores for children in the Isle of Wight at 5½ years old on the WPPSI and 16½ years old on the WISC-R.⁵

Genetics, environment, and IQ

The changes in IQ described above imply that an IQ score is not a constant that is associated with a person for life. What produces these alterations? Changes in a person’s environment seem to have major effects on IQ. Studies by Dennis, for example, showed that normal children placed in orphanages in Lebanon before 1956 had IQs of only about 50, presumably as a result of neglect. Those adopted into ordinary families before the age of 2, however, regained a normal IQ. Many other studies have shown that a variety of environmental factors can affect IQ, including social class, maternal responsiveness, the provision of toys, the arrival of siblings, and schooling. Which of these factors is most important is a matter for debate, but, even though there may be some inheritance of intelligence, the effect of environmental factors can be substantial.^{7 8}

The uses of IQ tests

One of the major uses of IQ tests in the past has been to assess children in ordinary schools who were having difficulties in keeping up with their classmates. At one time strict application of IQ tests consigned those with an IQ below a certain score to special schools. Psychologists have become increasingly unhappy about this because of the accum-

ulation of evidence on the degree of error in the tests, the importance of comparing like groups when interpreting the scores (particularly for ethnic minorities), the changes in IQ with time, the fact that different IQ tests are not comparable, and the educational and social effects of attending a special school.

Does this mean that IQ tests are not useful? This necessarily prompts the question: useful for what? There are, as yet no better tests of general cognitive function, so that for certain clinical purposes (such as distinguishing specific reading retardation from overall developmental slowness), and some research questions (such as whether a group of children with phenylketonuria, for example, are in the normal range of general cognitive development) IQ tests may continue to be used. For many other problems psychologists may well turn to tests of specific abilities rather than to general IQ tests. In either instance, what is most important is that those who interpret the scores should know what they mean, what range of error is possible, and how the scores may change over time.

References

- ¹ Eysenck HJ. Where does the concept come from? In: Eysenck HJ, Kamin L, eds. *Intelligence: the battle for the mind*. London: Pan Books, 1981, 11–15.
- ² Kamin L. Some historical facts about IQ testing. In: Eysenck HJ, Kamin L, eds. *Intelligence: the battle for the mind*. London: Pan Books, 1981, 90–7.
- ³ Sternberg RJ. *Beyond IQ: a triarchic theory of human intelligence*. Cambridge: Cambridge University Press, 1985.
- ⁴ Hindley CB, Owen CF. The extent of individual changes in IQ for ages between 6 months and 17 years, in a British longitudinal sample. *J Child Psychol Psychiatry* 1978;**19**:329–50.
- ⁵ Yule W, Gold RD, Busch C. Long-term predictive validity of the WPPSI: an 11 year follow-up study. *J Person Indiv Differences* 1982;**3**:65–71.
- ⁶ Dennis W. Children of the creche: conclusions and implications. In: Clarke AM, Clarke ADB, eds. *Early experience: myth and evidence*. London: Open Books, 1976:122–34.
- ⁷ Clarke AM, Clarke ADB. *Early experience: myth and evidence*. London: Open Books, 1976.
- ⁸ Rutter M. Family and school influences on cognitive development. *J Child Psychol Psychiatry* 1985;**6**:683–704.

GLYN MURPHY
Bethlem Royal Hospital,
Croydon CR0 8DR