Forthcoming in *The New Palgrave Dictionary of Economics*, Steve Durlauf ed.

Cognitive Ability

William T. Dickens Senior Fellow The Brookings Institution

ABSTRACT

Modern psychological theory views cognitive ability as multidimensional while acknowledging that the many different abilities are themselves positively correlated. This positive correlation across abilities has led most psychometricians to accept the reality of a general cognitive ability that is reflected in the full scale score on major tests of cognitive ability or IQ. This entry provides an introduction to the history of cognitive testing and some of its major controversies. Evidence supporting the validity of measures of cognitive ability is presented and the nature and implications of group differences is discussed along with evidence on its malleability.

Some people obviously and consistently understand new concepts quicker, solve unfamiliar problems faster, see relationships that others don't and are more knowledgeable about a wider range of topics than others. We call such people smart, bright, quick, or intelligent. Psychologists have developed tests to measure this trait. Originally called IQ tests (for Intelligence Quotient because the measures were constructed as the ratio of mental age to chronological age multiplied by 100) that name has fallen out of favor. Instead, such tests are now often referred to as tests of cognitive ability. Although the term IQ is still sometimes used to refer to what such tests measure, none construct a ratio.

History

Spearman (1904) first popularized the observation that individuals who do well on one type of mental task also tend to do well on many others. For example, people who are good at recognizing patterns in sequences of abstract drawings are also good at quickly arranging pictures in order to tell a story, telling what three dimensional shapes drawn in two dimensions will look like when rotated, tend to have large vocabularies and good reading comprehension, and are quick at arithmetic. This pattern of moderate to strong positive correlations across the whole spectrum of mental abilities led Spearman to hypothesize the existence of a general mental ability similar to the common notion of intelligence. A person's ability with any particular type of task would be equal to the sum of that person's general ability plus considerations unique to that particular task. Thus general ability could be measured by constructing subtests of a number of similar items (individual tasks of the same type such as arithmetic problems) of differing complexity. Each subtest would present items of a different type and individual scores across subtests could be aggregated. Task specific factors would average out leaving the final score as mainly a measure of general ability or "g." Using an approach like this Binet (1905) developed the first IQ test as a way of identifying student's academic potential. That test

was adapted for use in English by Terman and in 1916 became the Stanford-Binet IQ tests – still one of the most commonly administered tests of cognitive ability.

Spearman's hypothesis of a single general mental ability and many specific abilities was challenged by Thurstone (1935), who popularized the notion that people had a number of independent primary mental abilities rather than a single general mental ability. Both Spearman and Thurstone made contributions to the development of factor analysis as a way to identify the presence of unobserved variables (abilities) that affect a number of observable variables (sub-test or item scores). Today, the Spearman-Thurstone debate has been resolved with a compromise. The most common view among psychometricians who study cognitive ability is that there are a number of different abilities. Some people are better at solving problems verbally while others are good at solving problems that involve visualization. Some people who are good at both of these things may be only average at tasks that rely heavily on memory. However, there is a tendency for people who perform well in any of these broad areas to perform well in all others as well (Carroll 1993). Most modern tests of cognitive ability provide both a fullscale score that is most reflective of general intelligence, and a number of special ability specific sub-scores as well.

Validity

Binet's is considered the first successful test of cognitive ability in that it was able to accurately predict teachers' assessments of their long time students on the basis of a relatively short verbally administered test. Scores on tests of cognitive ability correlate well with common perceptions of how bright or smart someone is. They are also strongly correlated with measures of academic achievement such as achievement test scores, grades and ultimate educational attainment (typically .5 or better). They are less highly correlated (.5 or less) with many important life outcomes including reported annual income and job status. Performance on a wide range of jobs and work tasks is positively related to cognitive test scores with performance on more demanding jobs having higher correlations. Some have claimed that general cognitive ability is responsible for most of this explanatory power (Ree and Earls, 1992 and Ree et al. 1994). This was a major theme of the controversial best seller *The Bell Curve* (Herrnstein and Murray, 1994). Heckman (1995), in a review of that book, argues that even though g has significant explanatory power, many other factors, both cognitive and non-cognitive matter as well.

Finally, test scores are correlated with a number of social behaviors including unwed motherhood, criminal activity, and welfare receipt (Jensen 1998, Chapter 9). While these correlations are substantial, and cognitive test scores are typically better predictors of most of these outcomes than any other single personal attribute, they still explain less than half the variance.

Individuals' scores on tests of cognitive ability also tend to be strongly correlated over time – much more so for adults than for children. A study of older adults found their full scale IQ scores to be correlated .92 when tested at two points in time three years apart (Plomin et al. 1994). In contrast, a study of children tested at two points in time roughly two years apart found correlations of only .46 for those who were less than 1 year old at first testing and .76 for those who were one at first testing (Johnson and Bradley-Johnson, 2002).

It is common to draw a distinction between tests of achievement and tests of ability. Achievement tests measure how much knowledge the test taker has accumulated in a particular area while ability tests endeavor to measure how quickly a person can solve unfamiliar problems. Typically scores on the two types of tests will be highly correlated. In fact, all tests of ability are, to some degree, tests of achievement as it is impossible to measure ability without also measuring the test taker's reading or verbal comprehension at least. Further, to the extent that the task being tested relies on knowledge of geometry, arithmetic, general knowledge, etc., the rolls of the achievement test and ability test are confounded.

Cultural bias has been a concern with knowledge based tests. Some knowledge is more accessible to some people than others. For example, we would expect that a child growing up with upper middle class parents in New York or Paris to find it easier to learn the distance between the two cities (a general knowledge question that was once on one of the popular IQ tests), than someone from the slums of St. Louis or a tribesman in the bush in Africa. For this reason a number of tests have been constructed that require a minimal amount of prior knowledge such as Catell's Culture Fair Test (Cattell, 1960) or Raven's Progressive Matrices (Raven, 1941).

Group Differences

No matter what test is administered, men and women of the same background tend to have very similar average scores on tests of cognitive ability, though they differ slightly on their performance on some subtests (Jensen 1998 pp531-536). However, there are large differences across ethnic groups and geographic areas. The difference that has generated the most controversy is the difference in average scores of US blacks and whites which is typically reported to be about one white standard deviation, though this gap has declined some in recent years (Dickens and Flynn forthcoming). Do these represent real differences in cognitive ability or do they reflect cultural bias in the tests?

Defenders of the tests offer several pieces of evidence suggesting that they are unbiased. Foremost is the evidence of "external validity" - that the same regression equation that predicts outcomes such as job performance, grades, or educational attainment for one group will typically do a similarly good job for any other group. Also, different groups find the same questions more or less difficult. Members of different groups with similar scores will have similar patterns of right and wrong answers. If some questions are more culturally biased than others, the disadvantaged group should find those items more difficult than the mainstream group does. But, researchers looking for such cultural bias have found no evidence of it (an exception occurs when one of the groups being compared is made up of non-native speakers of the language in which the test was administered in which case scores on questions requiring a better knowledge of the language will be lower). Surprisingly, to the extent that there are black-white differences across test items, blacks do worst on what seem to be some of the least culturally dependent items - those involving abstract or symbolic problem solving. Differences tend to be smaller on seemingly culturally rich items such as general knowledge. Herrnstein and Murray (1994) provide a review of the evidence on bias in appendix 5.

The best evidence that tests can be biased in at least some circumstances come from studies of a phenomenon called stereotype threat. It has been shown that reminding people of their group identity can cause them to perform in ways more consistent with stereotypes of the group's abilities. For example, blacks have been found to perform worse on some particularly difficult vocabulary items when asked to answer a questionnaire that asked them to state their race before taking the test or when the test was represented as a test of intelligence as opposed to a test of vocabulary. Women who were told that the difficult math test they were taking generally showed gender differences performed worse than those taking the same test who were told the test showed no differences. Men showed the opposite effect and performed better when told the test showed a gender difference (Steele, 1997). However, it has not been demonstrated that stereotype threat produces substantial bias on standard tests in standard test taking circumstances.

While most evidence is consistent with the view that tests provide a fair measure of the underlying concept of cognitive ability across ethnic groups, it is not conclusive. For example, since tests rarely explain as much as half the variance in the outcomes in studies of external validity, there is always the possibility that the tests underestimate black cognitive ability, but that other disadvantages pull down black performance. If true, the validity of the tests as predictors of practical outcomes is an artifact of offsetting biases. This could explain why it is that when regressions of white performance on white test scores fail to predict black performance they tend to predict better performance than is observed. Further, common sense notions that people from different cultural backgrounds probably have less opportunity to learn certain types of information or practice certain skills should be given some weight. If studies find that blacks do no worse than similarly scoring whites on highly culturally loaded items that could indicate that the poor scoring whites were similarly disadvantaged. If disadvantage is more common for blacks than whites due to discrimination, that disadvantage could still explain some of the score gap. However, the strong correlation of even the culturally reduced tests with performance, and the similar magnitude of the gap on those tests between groups, suggests that large parts of the measured gaps in ability between groups reflect real differences in average developed ability. This conclusion naturally leads to the consideration of the sources of those differences.

The question of whether individual, and particularly group, differences in cognitive ability are due more to nature or nurture has been enormously controversial for the last century. Dickens (2005) presents a summary of the evidence on the origin of black-white differences and concludes that they are most likely not substantially genetic in origin. Rushton and Jensen (2005) reach the opposite conclusion. Whatever the right answer, asking whether the black white gap has genetic origins is probably the wrong question. It seems that most of the reason why people are concerned with the issue is because they confuse having a genetic cause with immutability. While genes almost certainly play a large role in explaining individual differences in cognitive ability within ethnic groups raised in similar circumstances, it also seems that developed cognitive ability is highly malleable.

Malleability

A large amount of evidence has accumulated on the role of genes in explaining individual differences in cognitive ability. Several reviews of this literature conclude that differences in genetic endowment explain somewhere between 60 and 80% of the

variance in cognitive ability in representative samples of the adult population in developed countries. The percentage for children is lower than for adults with most estimates placing it around 40% for 6 year-olds (Plomin et al. 2001, Neisser et al. 1996). The figure is also estimated to be lower among disadvantaged populations (Turkheimer et al. 2003) though not consistently (Asbury et al. 2005). This figure is referred to as the heritability of cognitive ability. It is estimated by contrasting people with different degrees of relatedness raised in the same home or people with similar relatedness raised in different homes. For example, the correlation of the cognitive ability of identical twins raised in completely independent environments will be equal to the heritability of cognitive ability under the assumptions typically employed to make such estimates (see Behavioral Genetics). While this evidence establishes that genes play a large roll in determining individual differences, little is known about which genes are involved or how they influence cognitive ability (Plomin et al. 2001).

The high heritability of cognitive ability has led some to conclude that people's environments play little role in shaping their ability and that, therefore, individual differences are largely immutable and that group differences must be largely due to differences in average genetic endowment. It has been argued that if all of the observable differences in environment between people only produce 40% or less of the variance in cognitive ability then the large differences between blacks and whites could not result from the relatively small differences in environment between the average white and the average black. Thus differences in genetic endowment must play a substantial role. A formal version of this argument was first presented by Jensen (1973, pp. 135-139). A similar argument was made by Herrnstein and Murray (1994 pp. 298-299).

Yet despite the high heritability of cognitive ability, it does seem to be quite sensitive to environmental changes. In a review of the effects of early education programs, Lazar and Darlington (1982, p44) noted that "The conclusion that a well-run cognitively oriented early education program will increase the IQ sores of low-income children by the end of the programs is one of the least disputed results in educational evaluation." The gains they surveyed were often quite large, though they also tended to decline substantially after children left the programs. There is also evidence that being in a cognitively demanding environment can increase measured cognitive ability. Ceci (1991) surveys the evidence on the effects of school attendance on measured ability and finds it to be substantial.

Finally, the most profound changes in measured cognitive ability have taken place over time. James Flynn has documented huge gains in cognitive ability -- as much as a standard deviation or more a generation -- in over 14 countries. Numerous other authors have found gains on other tests and in other countries (Flynn 1987, 1998, and 2006). This phenomenon of large and pervasive gains has been dubbed "The Flynn Effect."

So how is it that large gains are possible in the face of high heritability estimates? The chief flaw in the argument that high heritability implies a limited role for environment is that it misunderstands what heritability is measuring. It ignores the possibility that genetic and environmental influences might be correlated. In particular, it ignores the possibility that genetic influences on ability are largely the work of environmental advantages that come about due to modest physiological advantages.

Consider a sports analogy. Two identical twins raised apart have a shared genetic endowment that tends to make them notably taller than their peers. As such they are both

better basketball players. Even though they are raised apart, both are likely to spend more time playing basketball than other children their age. They are good at it and thus enjoy it more than other activities in which they do not naturally excel. Consequently they both get more practice at basketball than their peers and that makes them better at the game. Being better players than their peers they are more likely to be picked by coaches for high school teams and more likely to receive yet more practice and more intensive coaching. If this leads to them playing in college they will both be enormously better players than the average person. A small physiological difference, that would make only a very modest difference in their performance on the court if they were untrained and inexperienced, has mushroomed into a huge difference in performance because it was reinforced by the environmental influences of practice and coaching.

It is not hard to imagine the same thing happening with cognitive ability. Someone who is slightly quicker or has an emotional disposition amenable to thought and contemplation will be more likely to spend more time in intellectual pursuits. Such a person will likely receive positive reinforcement from teachers, be more likely to be tracked into more demanding classes, and more likely to develop friendships with other similarly disposed children. Such a child will have much more opportunity to practice intellectual work and receive more "coaching" on intellectual pursuits. A small initial physiological difference could mushroom into a large difference in ability through a process by which the advantage leads to a better environment which improves ability and gives access to even better environments.

If such reciprocal causation is at work in the development of cognitive ability then small persistent exogenous differences in environment could produce large differences in cognitive ability. Dickens and Flynn (2001) lay out a formal model of such a process. If in a cross section of people in the same ethnic group most exogenous environmental differences are transient then they will not accumulate through reciprocal causation and will not explain much variance across individuals. However, small persistent differences between groups or generations could cause large differences if they drive the engine of reciprocal causation. Similarly, preschool programs which enrich children's cognitive environment can have large effects, but once the children are removed from the program the process can work in reverse and unravel the gains. The exogenous decline in the quality of the environment from the removal of the program's stimulation sets off a downward spiral of poorer performance leading the child into poorer environments and yet poorer performance and so on.

Conclusion

Modern psychology views cognitive ability as having a number of dimensions; all of which seem to be correlated with one another. Many interpret this correlation as reflecting an underlying general cognitive ability, or g, that is measured by the full-scale scores on the major tests of cognitive ability or IQ. General cognitive ability is an important predictor of a wide range of economic and life outcomes, with similar predictive validity across groups with different average levels of ability. Still, cognitive test scores typically explain far less than half the variance in life outcomes so cognitive ability is only one important factor among many that explain success.

Adult differences in cognitive ability within representative samples of ethnic groups raised in similar circumstances are subject to substantial genetic influence, but

this does not mean that group differences are genetic in origin. Despite the large role played by genetic differences in explaining adult variance in cognitive ability, there is considerable evidence that intelligence is highly malleable and the life outcomes influenced by intelligence even more so.

References

- Asbury, Kathryn, Theodore D. Wachs and Robert Plomin, (2005) "Environmental Moderators of Genetic Influence on Verbal and Nonverbal Abilities in Early Childhood," *Intelligence*, 33 (November-December, 6) pp. 643-661.
- Binet, Alfred, (1905) "New Methods for the Diagnosis of the Intellectual Level of Subnormals," first published in L'Année Psychologique, 12, 191-244.
 English translation by Elizabeth S. Kite, The development of intelligence in children. Vineland, NJ: Publications of the Training School at Vineland (1916).
- Carroll, J.B. (1993) *Human Cognitive Abilities: A Survey of Factor-analytic Studies*, UK: Cambridge University Press.
- Cattell, Raymond B. (1960) *Measuring Intelligence with Culture Fair Tests*, Institute for Personality and Ability Testing,
- Cecci, Steven J. (1991) "How Much Does Schooling Influence General Intelligence and its Cognitive Components? A Reassessment of the Evidence," *Developmental Psychology*, 27, pp. 703-722.
- Dickens, William T. (2005) "Genetic Differences and School Readiness," *The Future of Children*, 15 (Spring, 1) pp.55-69.
- Dickens, William T. and James R. Flynn (2001) "Heritability Estimates Versus Large Environmental Effects: The IQ Paradox Resolved," *Psychological Review*, 108 (2) pp. 346-369.
- Dickens, William T. and James R. Flynn (forthcoming 2006) "Black Americans reduce the racial IQ gap: Evidence from standardization samples," *Psychological Science* (October).
- Flynn, James. R. (1987). "Massive Gains in 14 Nations: What IQ Tests Really Measure," *Psychological Bulletin*, 101 pp. 171-191.
- Flynn, James. R. (1998). "IQ Gains Over Time: Toward Finding the Causes," in U. Neisser (ed.), *The Rising Curve Long-term Gains in IQ and Related Measures* Washington DC: American Psychological Association, pp. 25-66.
- Flynn, James. R. (2006). "Efeito Flynn: Repensando a inteligência e seus efeitos" [The Flynn Effect: Rethinking intelligence and what affects it]. In C. Flores-Mendoza & R. Colom (eds.) *Introdução à Psicologia das Diferenças Individuais* [Introduction to the psychology of individual differences] (pp. 387-411). Porto Alegre: ArtMed. (English trans: jim.flynn@stonebow.otago.ac.nz)
- Heckman, James .J. (1995) "Lessons from the Bell Curve," *The Journal of Political Economy*, 103 (October, 5) pp1091-1120.

Herrnstein, Richard J. and Charles Murray (1994) The Bell Curve, New York: Free Press.

- Jensen, Arthur R. (1973) *Educability and Group Differences*, New York:Harper and Row.
- Jensen, Arthur R. (1998) *The g Factor: The Science of Mental Ability* Westport Connecticut:Praeger.
- Johnson, C. Merle, and Sharon Bradley-Johnson, (2002) "Construct Stability of the Cognitive Abilities Scale- Second Edition for Infants and Toddlers," *Journal of Psychoeducational Assessment*, 20 (2) pp.144-151.
- Lazar, I. and R. Darlington (1982) "Lasting Effects of Early Education: A Report for the Consortium for Longitudinal Studies" *Monographs of the Society for Research in Child Development*, 47 (2-3).
- Neisser, U., G. Boodoo, T.J. Bouchard Jr., A.W. Boykin, N. Brody, S.J. Ceci, D.F. Halpern, J.C. Loehlin, R. Perloff, R.G. Sternberg, and S. Urbina (1996), "Intelligence: Knowns and Unknowns," *American Psychologist*, 51 pp. 77-101
- Plomin R, Defries JC, McClearn GE, McGuffin P. (2001) Behavioral Genetics, 4th ed. New York: Worth Publishers.
- Plomin, R., N.L. Pedersen, P. Lichtenstein, and G.E. McClearn (1994) "Variability and Stability in Cognitive Abilities are Largely Genetic Later in Life," Behavior Genetics, 24 (May, 3) pp207-215.
- Raven, J.C. (1941) "Standardization of Progressive Matrices," *British Journal of Medical Psychology*, 19 pp.137-150.
- Ree, Malcom James and James A. Earles (1992) "Intelligence is the Best Predictor of Job Performance," *Current Directions in Psychological Science*, 1 (June, 3) p86-89.
- Ree, Malcom James, James A. Earles and mark S. Teachout (1994) "Predicting Job Performance: Not Much More than g." *Journal of Applied Psychology*,79 (August) pp.518-24.
- Rushton, J. P., & Jensen, A. R. (2005) "Thirty years of research on race differences incognitive ability." *Psychology, Public Policy, and Law 11*, 235-294.
- Spearman, C. (1904). General Intelligence, Objectively Determined and Measured, *American Journal of Psychology*, 15 201-293.

Steele, Claude M. (1997) "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance," *American Psychologist*, 52(6) 613-629.

Thurstone, L.L. (1935). The Vectors of the Mind. Chicago: University of Chicago Press.

Turkheimer, Eric, Andreana Haley, Mary Waldron, Brian D'Onofrio, and Irving I. Gottesman, (2003) "Socioeconomic Status Modifies Heritability of IQ in Young Children," *Psychological Science*, 14 (November, 6) pp.623-628.