Measuring 'Intelligence': what can we learn and how can we move forward?

Dylan Wiliam, ETS[1]

Paper presented at the annual meeting of AERA, Montreal, Canada, April 2005.

Introduction

A quick reading of the reporting on intelligence in the popular media, or in much of the academic literature in education, would lead one to believe that:

- Intelligence is determined by the environment, and not one's genes;
- Intelligence tests measure only the ability to take intelligence tests;
- Intelligence doesn't matter in the real world; and
- There are several different kinds of intelligence, all independent of each other

One would also probably come to believe that this was the consensus in the scientific world generally—and that would be wrong. It turns out that there ***is*** a high degree of consensus on the science underlying intelligence, but it is that:

- Intelligence is determined by both environment and genetics, and the genetic influence is substantial;
- Intelligence tests correlate strongly with a range of other measurements of mental capability;

[1] Address: ETS, Rosedale Road (ms 04-R), Princeton, NJ 08541, USA. Email: dwiliam@ets.org

- Intelligence is strongly associated with success in a wide range of real world activities;
- There are several different aspects of intelligence, but most of them are strongly inter-related;

It is widely agreed in moral philosophy that one cannot deduce an 'ought' from an 'is' (Hume, 1739/1896)—just because things are the way they are, it doesn't mean we have to accept them that way. And saying that things are so doesn't mean that one supports the status quo, even though this exactly what is often assumed. That is why talking about intelligence is very difficult.

If we want to change the way things are, we should not be blind to the constraints on change. If an engineer wants to build a bridge, she doesn't say that the facts about the tensile and compressive strength of steel are wrong just because these facts mean she can't get the bridge she wants. Rather these facts about the world become the constraints within which the process of bridge design must take place.

In a similar vein, the purpose of this paper is set out an argument for accepting what we know about intelligence, and then working within these constraints to design education systems that can help all students reach the highest possible standards—I was tempted at this point to say something like "helping all students reach their potential" but as my colleague Randy Bennett points out, "Why be limited by your potential?".

Background: the research on intelligence testing

Much of the early research on intelligence was motivated by interest in eugenics, and this fact alone makes the idea of talking about intelligence anathema to some. However, the original observations—that people who are good on one mental task are also good on a whole range of others—are manifestly facts about the world. Egalitarian principles may lead us to hope that talent is 'spread out' amongst the population, but as long as we are talking about the kinds of skills valued in our education system, it just isn't so. While the idea of 'multiple intelligences' has attracted significant interest over recent years, the fact is that these different intelligences are in fact highly inter-related (Mackintosh, 2000), and, perhaps more importantly, a small number of these are disproportionately important for success at school.

The origin of the idea of intelligence is that the simplest model that would account for such a situation is that the tasks given to students in school require a certain amount of general ability, and a certain amount of special ability, particular to that task. The level of skill that an individual displays on a particular task depends on how much of the general ability and how much of the special ability for that task the individual possesses. By analyzing performance on a range of more or less different tasks, we can hypothesize the relative importance of the general component in carrying out the task. Intelligence is just the name given to the general component in the model—sometimes called the 'g' factor—but of course we have no idea what it is. We have an idea what it's like, because we know that some tasks seem to be very closely related to this general factor, while others seem to require as much, or more, special ability for that specific task. However, knowing what something is like is not the same as knowing what it is (see Coloma et al, 2004 for a summary of research on the nature of the g factor and its links to other aspects of mental capability).

IQ testing in England

Some detailed research on the link between IQ and school success has been provided by a natural experiment that has been taking place in England over the past 20 years. In the 1980s, the Conservative government led by Margaret Thatcher started publishing 'league tables' of school performance on the national school leaving examinations taken by 16-year-olds in England. As might have been expected, the performance of students in urban schools was lower than those in suburban schools, leading to invidious—and highly flawed—comparisons about the quality of education offered in different schools.

In response to this, many urban secondary schools decided to administer IQ-type tests to all incoming 11-year-old students so that they could use the results of these tests to contextualize the results obtained by these students on the national school-leaving examination five years later. In fact, it would have been much better to use an achievement test, rather than an IQ-type test for this purpose, since in the most deprived areas, the achievement of students at age 11 was already below that which would be expected given their IQ.

Nevertheless, the use of such tests spread rapidly so that now the vast majority of England's 4000 secondary schools administer the same test—the Cognitive Abilities Test—to their incoming 11-year-olds. Data on exactly how many students are tested every year is not available, but it would appear that at least half-a-million eleven year olds take this test each year in England. The test is a fairly traditional IQ type test with three batteries—verbal, non-verbal, and quantitative reasoning—with scores reported on the familiar IQ scale: mean of 100 and standard deviation of 15 (Lohman et al, 2003).

The consistent finding has been that the correlation between the score on a measure of IQ at age 11 and a range of measures of general scholastic achievement at 16 is around 0.7 (NFER-Nelson, 2004)—all the more remarkable given the unreliability in both predictor and criterion. This appears to be higher than is found in the US—Duncan et al (1972) found a correlation of around 0.54

between IQ and educational achievement—but this may reflect the fact that the outcome measure is strongly standards-based and more objective than measures typically used in the US, such as high-school GPA. Quite what this correlation means, of course, is problematic, since some schools may be using the results of these tests to place students into more or less demanding tracks, although it appears that these data are mostly used to identify students with moderate or high IQ that are not demonstrating the academic achievement that might be expected.

As might be expected, the highest correlation is between IQ and mathematics—0.75—and, again perhaps not surprisingly, the lowest is for Art & Design but this is still a respectable 0.44. This does not of course prove that 'multiple intelligences' do not exist, but it does indicate that the least, at least as they are defined in English secondary school, all subjects depend to a significant extent on intelligence.

The key performance indicator for English secondary schools is the proportion of the students enrolled at age 15 who achieve proficient performance on at least five subjects in the national school leaving examination at age 16. Therefore most interest has centered on the ability of the Cognitive Abilities Test to predict this.

About 55% of students with a score of 100 on the CAT achieve this benchmark, and around 93% of those scoring one standard deviation above the mean do so. However, not until the CAT score exceeds 123 (1.5 standard deviations above the mean) does more than 99% of the cohort achieve the same standard. What is perhaps more interesting is that of the students with CAT scores of 70 (i.e. two standard deviations below the mean) approximately 5% of students manage to achieve the key performance indicator five years later. This result is important because it suggests that measured IQ does not appear to place any kind of absolute ceiling on future performance, but may perhaps indicate how difficult a particular standard may be to reach.

The heritability of intelligence

There is considerable debate about the extent to which intelligence is inherited, is innate and is fixed. Almost all psychologists agree that intelligence is, to an extent, inherited, although it is frequently necessary to point out that all this means is that the proportion of variability in intelligence due to genetic factors is not zero. Intelligence is also, to an extent, innate, in that, non-genetic factors affecting the maternal womb environment can impact intelligence before birth (Devlin et al, 1997). Finally, intelligence is also to an extent fixed, in the sense that as one ages, one's IQ is less and less likely to change substantially (Brody, 1992)

That leads naturally to the question of how much of intelligence is inherited, but that is in fact an impossible question to answer, because the heritability of any trait can be measured only relative to a particular environment.

For example, one in approximately 10,000 children is born with a condition known as phenylketonuria in which a chemical called phenylalanine accumulates in the brain, poisoning neurons and causing mental retardation. The condition can be alleviated by adopting a low-protein diet, at least to the age of 12. The condition is definitely inherited (a child has this condition only if both parents pass on the gene for this, because the gene is recessive), and in an environment in which high-protein foods are ubiquitous, this will be obvious—everyone carrying two genes for this will develop the condition. A little research would then make the genetic basis of the condition obvious. On the other hand, in an environment in which protein is scarce, it is likely that the condition will never become apparent. Where environments are more variable, so that some children have high-protein diets and others do not, then phenylketonuria will appear to be partly genetic and partly environmental in origin.

As another example, consider the heritability of physical height, say at age 18. To find out how heritable height is, we would seek to establish the proportion of variability of height within the population of 18-year-olds due to genetic factors and that due to environmental factors. However, in an environment in which all children received the same standard of nutrition we would find that the inherited component was quite high, whereas in an environment in which children' standard of nutrition differed greatly, the inherited component would appear to be lower, because it would be swamped by the effects of environment. In other words, estimation of heritability makes sense only for a particular population under particular circumstances. The irony here is that if we want taller people, the optimal strategy is to ensure that everyone gets a nutritious diet, which actually makes the heritability greater, because the environment is less variable. This is important because the importance of environmental effects is often confused with the size of environmental effects. The size of the environmental effect of nutrition is maximized when the environment is most nutritious for everyone, but the variability of the environment is then at a minimum, and so it appears as if the environment doesn't matter.

Even for a specific population, quantifying heritability is difficult. The most popular way of teasing apart the effects of genetics and the effects of environment has been to compare at the IQs of monozygotic (i.e identical) twins with other siblings. However, since by definition, monozygotic twins share the same maternal womb environment, any similarities of monozygotic twins may be the result of the shared experience prior to birth rather than due to their common genes (Devlin et al, 1997).

Environmental influences on intelligence are certainly the best explanation of the huge rises that have been seen in IQ scores over the last 50 years (Neisser, 1998)—there is no way that people's genes could have changed so quickly. In a range of countries, and over a range of different kinds of IQ tests, IQ scores appear to be rising at the rate of one-fifth of a standard deviation every ten years. The fact that the IQs of African-American students lags those of white students by a standard deviation is used by some as evidence that the source of the difference must be genetic in origin. However, entirely environmental explanations appear much more plausible if we can accept the idea that the environment for African-American students today resembles the environment for white students in 1955.

There is also evidence that IQ seems to become more important as people get older. This, again, has been used by some as evidence that IQ is largely inherited, but as Dickens and Flynn (2001) show, this actually supports the strength of environmental influences on IQ, and that IQ is, in fact, highly malleable.

Dickens and Flynn show that the increasing importance of IQ as people age is best accounted for by the idea that people select (or have selected for them) environments that match their IQ. People with high intelligence, for example, engage in more of the activities that enhance intelligence, and so become more intelligent, whereas people with lower intelligence opt out of (or are denied) these intelligence-enhancing activities and so lose the opportunities to enhance their intelligence. This suggests that IQ and environment are mutually constitutive of each other: environment causes IQ and IQ causes environment. However, the model proposed by Dickens and Flynn also suggests that the impact of transient improvements in environment are themselves transient, and in particular, that adult IQ is influenced mainly by adult environment. In other words, programs like ‘Head start’ are not enough. We need to create situations in which individuals choose to expose themselves to challenging environments throughout the lifespan.

In summary, the research on intelligence suggests that:

- Intelligence is heritable, although to ask how much is an essentially meaningless question;
- It is stabilized early in life;
- It is a strong predictor of future success in a range of endeavors, especially school;
- It is malleable, depending on the quality of experiences, and in particular, closely linked to the challenge of the current environment.
- It places no ceiling on achievement, but may perhaps indicate how difficult success may be.

How can we move forward?

If we want to maximize the academic achievements of all students, the research on the nature of intelligence seems to indicate that the crucial measure is to maintain all students in high-demand environments. In the analogy of physical height described earlier, we maximized the height of everyone by making sure that everyone got the same highly nutritious diet. In the same way, we can maximize the intelligence of all students by ensuring that all students are exposed to maximally challenging environments from as early as possible, and for as long as possible.

This may appear to be a paradoxical result, since by ensuring that all students get the same challenging environment, we reduce the variability of the environment, thus increasing the heritability of intelligence. What is important however is that the *total* variability of intelligence would be reduced.

From this basic premise, a range of policy prescriptions might follow. The work of Hart and Risley (1995) shows conclusively how much richer are the environments experienced by children of middle-class parents from as early as six months, and so one policy prescription might be directed at enriching the linguistic environment of young children. The model of influences on intelligence proposed by Dickens and Flynn (2001) account for the success and importance of early intervention programs, but shows also how the effects begin to wear off if the stimulus is not maintained. Therefore, for the remainder of this paper, I want to concentrate on some changes that could be effected at much lower cost, but which nevertheless hold out the possibility of significant change.

The work of Carol Dweck has shown that students hold differing views on the nature of ability (see Dweck, 2000 for a summary), Perhaps the most important distinction is between a view of ability as fixed, and a view of ability as incremental. For students who see ability as fixed, challenging environments are to be avoided, since there is the risk of being exposed as not smart. It is safer to avoid such environments, on the grounds that it's better to be thought lazy than dumb. Students who see ability as incremental on the other hand see challenging environments as providing opportunities to get smarter.

One of the things we should seek to do in designing education systems, then, is to seek to inculcate in students a belief that ability is incremental, rather than fixed, and moreover, that what they choose to do can increase their intelligence. After all, we should seek to do this even if it were not true, since the consequences are likely to be so positive, but the work of Dickens and Flynn suggests that it is true as well.

However, what faces students in the public schools in the United States appears to be designed for the opposite effect. From third or fourth grade on, the work of students is graded on the same scale, so that the experience of most students is that they are a 'C' student or an 'A' student. In absolute terms, students make appreciable progress, but their experience is one of being classified more and more finely, culminating in a high-school GPA measured to two places of decimals.

It seems likely that the idea that ability is fixed will be even more widespread as the provisions of the No Child Left Behind Act come into force. From grades three through eight all students will be assessed as 'advanced', 'proficient', 'basic', or 'below basic'. The problem, given what we know about student growth trajectories, is that students who are classified as 'below basic' in third grade are likely to be 'below basic' in fourth, fifth, sixth, seventh and eighth grade too, reinforcing the idea that ability is fixed rather than incremental.

If we are to break into this vicious spiral of self-fulfilling prophecies, we need to build educational systems from the ground up to support the notion of ability as incremental. The impact of grading systems that re-inforce the notion that ability is fixed needs to be minimized, and instead, feedback should be task-involving rather than ego-involving (Black and Wiliam, 1998). The evidence from cohort data in England shows that, at any point in the child's education, it really is "still all to play for", even if one believes that IQ is fixed, and the models provided by Dickens and Flynn suggest that we can increase IQ provided we can keep students in challenging environments. The 'Cognitive Acceleration' work of Adey and Shayer (1993) has demonstrated that long-term gains in cognitive skill are possible, provided that students are persuaded that thinking is more important than remembering in classrooms. In this context, the relatively low correlation of IQ with school achievement in the US may be symptomatic of the problem rather than an indication of success.

The fact that IQ correlates only modestly with school achievement might be taken by some as indications of the success of the American education system in detaching school success from IQ. However, I believe the analysis above supports the contrary interpretation. The low correlation of IQ with school achievement in the US in my view supports that idea we have made our classrooms 'thought-free zones'. This low correlation does allow more students to be successful, because you don't have to have a high IQ to succeed. But the price to be paid is that students are successful at things that don't really matter, and worse, students are locked into unstimulating environments. Perhaps paradoxically, then, the best way to make our students smarter, may be to strengthen the link between school achievement and intelligence . Doing so will also increase the heritability of IQ, since we would be minimizing the effects of environment, but the achievement gaps would be reduced to the minimum determined by genetic factors.

References

Adey, P. S. & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention program in the high school science curriculum. *Cognition and Instruction,* **11**(1), 1-29.

Black, P. J. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice,* **5**(1), 7-73.

Brody, N. (1992). *Intelligence*. San Diego, CA: Academic Press.

Coloma, R.; Rebolla, I.; Palaciosa, A.; Juan-Espinosaa, M. & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence,* **32**(277–296).

Devlin, B.; Daniels, M. & Roeder, K. (1997). The heritability of IQ. *Nature,* **388**, 468-471.

Dickens, W. T. & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: the IQ paradox resolved. *Psychological Review,* **108**(346-369).

Duncan, O. D.; Featherman, D. L. & Duncan, B. (1972). *Socioeconomic background and achievement*. New York, NY: Seminar Press.

Dweck, C. S. (2000). *Self-theories: their role in motivation, personality and development*. Philadelphia, PA: Psychology Press.

Flynn, J. 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* (pp. 25-66). Washington, DC: American Psychological Association.

Flynn, J. R. (2000). IQ gains, WISC subtests, and fluid g: g-theory and the relevance of Spearman's hypothesis to race. In G. R. Bock, J. A. Goode, & K. Webb (Eds.), *The nature of intelligence* (pp. 202-227). New York, NY: Wiley.

Hart, B. & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Brooks.

Hume, D. (1739/1896). *A treatise of human nature: being an attempt to introduce the experimental method of reasoning into moral subjects*. Oxford, UK: Clarendon Press.

Lohman, D. F.; Hagen, E. P. & Thorndike, R. L. (2003). *Cognitive Abilities Test* (3 ed.). Windsor, UK: NFER-Nelson.

Mackintosh, N. J. (2000). *IQ and human intelligence*. Oxford, UK: Oxford University Press.

Neisser, U. (Ed.) (1998). *The rising curve: long-term gains in IQ and related measures*. Washington, DC: American Psychological Association.

NFER-Nelson (2004). *Cognitive Abilities Test (CAT) and GCSE grades: Autumn 2004*. Windsor, UK: NFER-Nelson.