g, Jobs, and Life: Honoring Arthur R. Jensen

Linda S. Gottfredson
International Society for Intelligence Research
Nashville, December 7, 2002

[PPT Slide 1] Arthur Jensen has reinvigorated and redirected the study of human intelligence in major ways. Perhaps the most important has been to turn the field's attention back to Spearman's *g*, the general intelligence factor. The discovery that the same *g* factor emerges from diverse batteries of mental tests in diverse populations, together with the consequent option to derive scores for individuals on this common factor, has allowed intelligence researchers to make some crucial advances.

[2-a] The first is to free the concept of intelligence from IQ. It does so by distinguishing "intelligence" (that is, *g*) from the vehicles of its measurement (including, test format and content) and by allowing us to use a common working definition of intelligence—*g*—despite often relying on different tests of mental ability.

[2-b] Second, the g construct provides a common yardstick for comparing *tests* in terms of how well they measure g, that is, how g loaded they are. Jensen has put this to excellent use with his method of correlated vectors.

[2-c] This second advance means, in turn, that *any* human task, any environment, can be assessed for the mental demands it makes on people if we correlate IQ with performance on that task. These correlations tell us how much of an edge higher-g people will have in such situations.

This third advance provides a wonderful conceptual tool for tracing the consequences of g in real life, what Jensen (1998) calls the horizontal aspect of *g*. Jensen himself has focused mostly on the vertical aspect of *g* (its biology), but his insights on the nature and measurement of g have been extremely helpful to sociologists of intelligence.

[2-d] For instance, they have prompted Robert Gordon (1997) to analyze the psychometric properties of daily life as an intelligence test. Bob shows, for example, how the degree to which daily life mimics rather than departs from the properties of a reliable, valid test of intelligence helps to explain the pattern both of *g*'s impact across life as well as the likelihood that people will subjectively *perceive* that impact.

[3-a] I would like to build on that work today by asking 6 questions about how our lives do and do not resemble an intelligence test battery. My aim here is not to provide answers to all these questions, but to try out some new ways of tracing g's effects in our lives, both individual and collective. And it's certainly not to argue that g is all that matters in life, because conscientiousness, emotional stability, opportunity, and much else clearly does. I'll quickly go through the list of questions and then go back to say a bit about each one.

[3-b] 1. The first question is, What is the distribution of *g* loadings across life's many tasks? For instance, which broad arenas of life—say, school, work, family life, health—are most *g* loaded and thereby most advantage the bright and most hobble the dull relative to the rest of the population?

[3-c] 2. To what extent do we all take the same subtests in life's long test battery—or do do we mostly get to pick and choose the ones we want, say, by picking different life styles?

[3-d] 3. To what extent does how bright we are affect which life subtests we end up taking, whether by choice or not?

[3-e] 4. To what extent are life's tests standardized, say, in the conditions under which we take them—when and where, how much *time* we can take, how much *help* we get, and so on? To the extent we decrease their standardization in daily life, perhaps they allow us to get around or at least mute the effects of individual differences in g.

[3-f] 5. Do life's myriad little tasks behave like mental test items with regard to the Spearman-Brown formula? That is, if most if not all daily tasks have at least some faint g loading, might these small effects pile up over time to create some surprisingly highly g-loaded life outcomes? And, in fact, might this not be how g produces some of its biggest, least escapable consequences in real life?

[3-g] 6. And sixth, how do a society's members, wittingly or not, shape the mental test battery that faces current and future generations? Is the battery getting harder, if so why, and with what social consequences?

[4-a] Question 1: How *g* Loaded are the Different Arenas of Life?

Tests are constellations of tasks where performance is judged against some standard of correct or incorrect, better or worse, including faster or slower. We use many such yardsticks in our lives for judging each other's success and well-being. I'll show you two sets of outcomes, the first with continuous and the second with dichotomous outcomes.

[4-b] These correlations with IQ can be interpreted as g-loadings for the outcomes in question—in this case mostly ones relating to education and work. They range from .2 to .8, illustrating, not surprisingly, that life's major outcomes vary more in their demands for g than do IQ subtests, whose correlations with g seldom dip below .4-.5. What may be more surprising is that many of these life outcomes—such as income, occupation, and performance on moderate to higher-level jobs—are at least as g loaded as IQ subtests usually are.

[5] g-related risk varies widely across these dichotomous life outcomes too. This can seen in the odds ratios for the different outcomes, which I have calculated here to compare the odds of experiencing an unfavorable outcome if you are somewhat below average in IQ rather than somewhat above average in IQ. For example, you can see that the odds of living in poverty are

six times as high—the odds ratio is 6.2—for young white adults of IQ 75-90 compared to ones of IQ 110-125. Once again, relative risk for dull compared to bright people varies widely across the different outcomes, with odds ratios ranging from just over 1 (which would be parity) to over 100.

Social scientists refer to individual differences in these sorts of outcomes as social *inequalities*. Inequality is by definition a problem to eradicate. What these g loadings show is that inequalities in life outcome vary systematically in how tightly they are tied to differences in g. The question of course is what explains this stable, highly regular pattern of g-related *differences* in risk across different spheres of life. Social scientists have attempted to explain away each of the individual correlations with g, usually invoking income and other social class variables, but none of their explanations can cope with the full pattern of results, with the systematic *differences* in how g relates to various social inequalities. We will know a lot about how g and other variables operate in life if we can explain this pattern.

[6-a] Question 2: How different are the test batteries that we each take in life?

Life differs from a mental test battery in that we tend to choose somewhat different subtests to undertake, when given the chance. Such choice allows us to create niches more compatible with our talents and interests, but, as Bob Gordon describes, the resulting noncomparability in forms of expertise we develop—you in tennis and computer sales and me in gardening and social work—also makes it harder for us to see g at work in everyday life. Once we start talking about adult intelligence in terms of specialized forms of knowledge or expertise, say, on the job, we have shifted our attention to explaining competence on the different tests we take, not on the ones we take in common.

[6-b] Many of life's yardsticks are common, however, and they are the ones that tend to most concern policy analysts and those status-conscious Joneses living next door to us. For example, the law requires that we all attend elementary and secondary school, surely two of life's most relentlessly public IQ tests. Some of the adult outcomes I showed you earlier, such as getting married, being employed, and staying out of jail, are often treated like minimum competency tests for adulthood because they are generally easily passed, when attempted, by all but the mentally retarded. All the outcomes I showed you earlier are part of the common decathlon of adult life. The tests are inescapable because those activities yield public signals by which others—including social scientists—rate and rank us.

Other subtests of daily life are more private but no less escapable for being so. I described two examples on Wednesday. One is daily self-maintenance in a highly literate society, where it is taken for granted that citizens will routinely be able to independently and effectively fill out forms, read posted notices, order from menus—including those on ATMs. Such tasks are part of the minimum competency test for mental normalcy, as revealed so poignantly by the great effort that many mildly mentally retarded adults make to hide their inability to do them so they can pass as normal in public settings. The second example was guarding one's health and safety, including being able to read medicine labels and understand simple spoken instructions on caring for one's chronic disease. We treat such tests as discretionary and refuse to face them only at our peril.

[6-c] Turning to post high-school education, training, and paid employment, both are highly organized realms of activity where our test performance tends to be officially graded, so to speak. But both are also life arenas where we tend to take different tests—I train to be a dental hygienist and you go for an MBA. Adult life does—and must—provide great variety in this

regard to accommodate the intellectual variety among us. This becomes clear when you look at the occupational ladder.

[7-a] You see on the left, that the higher you go up the occupational ladder, the more g loaded jobs are. That is, higher level, more complex jobs would be expected to function as IQ tests were they to recruit randomly from the population. They don't, of course, which is the point of this slide. It brings up the third question.

[7-b] 3. How does our own *g* level affect which tests we end up taking in life? This table shows you the IQs of the middle 50% of people *applying* for these jobs. It shows that applicants to any job range widely in IQ, but they tend to cluster higher on the IQ continuum when the job they are applying for is more complex and prestigious. (Jobs overlap less in IQ when you consider just the people hired, because they tend to come from the top half of the applicant pool.)

Researchers have also found that when people are not as bright as the typical worker in their job, they tend to

[7-c] gravitate over time to cognitively easier work. When they are brighter than the typical worker,

[7-d] they tend to move into more cognitively demanding jobs. This may help explain why the correlation of IQ with both occupational prestige and income level goes up during early to mid career.

Movement along this hierarchy of jobs—of our economy's set of occupational tests—can be seen as a metaphor for how we and others go about identifying the most congenial social niches for ourselves. In fact, it's a bit like computer adaptive testing—we try a few items, see how we do, and then move up or down on the difficulty scale till we zero in on a congenial level of difficulty. Schools and employers informally do this all the time when assigning us our next

task. But we also do it ourselves everyday. We do it when sizing up other people and figuring out how intellectually compatible we might be with them—we start with comments or questions of low-average difficulty and then, depending on their answers, we gradually zero in—whether it takes minutes or months—on where they stand intellectually, especially relative to ourselves. This may seldom be a conscious process and there are many social norms surrounding it, including the merits of announcing our conclusion—we are supposed to be tactful, for instance—but the process is ubiquitous. I also think this is partly how we set our trajectories in life—how we gradually wend our way to the niches we end up in. We all work to find a set of life activities—our personalized life test battery—that makes us feel competent—which means one *neither* too hard nor too easy.

[8] 4. The fourth question concerns how standardized life's different tests are. Mental test scores are hard to interpret correctly unless the tests are standardized. Good standardization means using the same or equivalent sets of items to measure the skills in question, measuring them under comparable conditions for everyone, scoring the answers in the same way, and interpreting the scores within the appropriate norm groups or against clear standards of mastery. Life's subtests are rarely as standardized as are IQ test batteries, of course. In fact, we encourage in real life what testers prohibit in the testing situation—namely, getting and giving help, or taking extra time if we need it.

Does this mean that life's tests often won't provide good signals of g? That g doesn't really matter much in the end? Yes and no. As Bob Gordon points out, much of daily life is structured—on *purpose*—to degrade signals that we differ in mental competence. Habits, rituals, routines, tact, surreptitious help, cultivating personal areas of expertise—all help reduce the invidious distinctions in mental competence that g is constantly threatening to expose.

Sometimes, however, it is the very non-standardization of the life tests that signals g level. Recall that degree of mental retardation is sometimes defined, not in terms of what people can do unassisted, but in terms of the amount of help they need to do it. And so it is in daily life too. We would be happy to see all our employees or our co-workers eventually get their assignments done well, but we would surely rate as more competent those who did so in half the time and with no special help or extra resources, especially from us.

[9] Question 5. Do low-g loaded life tasks produce highly g loaded life outcomes? This general issue has been brought up many times during the conference. As Buzz Hunt said, even small predictive validities can have huge dollar effects when they involve very, very big numbers of people. Others have alluded to the fact that as long as you use enough test items, you can create a very good test of g from items that individually hardly measure g at all as long as you have many of them. With enough items, the small bits of g-related variance that each item contributes to the total score will add up while the many less consistent influences on performance will cancel each other out. I have begun to suspect that everyday life often operates in this way too.

[10-a] Consider this short test, and assume that its items all have low correlations with the total test score. It won't give you a good measure of g. The more items you add, however, the more g loaded the total test score will become.

[10-b] Add enough items, and you'll eventually end up with a test that measures virtually nothing but g if g is the only *consistent* source of covariance in the test.

[10-c]

[10-d] Imagine now that this is a calendar and that each day is an item in the life test called controlling your diabetes. Let's pick the task of not letting your blood sugar swing above

300 for more than 24 hours. And assume for the sake of argument that it does a wee bit of damage to your retina if you do. Now, whether your blood sugar is too high on any particular day will not likely be related much to your IQ—because a lot of unexpected and uncontrollable things can push it up, such as having an infection, a friend cooking a surprisingly sugar-laden meal for you, being distracted and taking your smaller night dose in the morning, giving in to temptation, or perhaps even taking bad insulin, all of which I've seen with insulin-dependent friends. Whether blood sugar stays high and whether it often swings into the high range is quite another matter, and I suspect is meaningfully g related—yielding our hypothetical .1 validity for our imaginary diabetic population. If you think that is too high, shrink it to .01. Add enough days, however, and you start to get a g-loaded test. Then add the option for what's called tight control of diabetes, which requires more judgment and intensive monitoring, you boost the g loading further because bright people will be better able to implement it and more often opt for it.

I am currently working on this idea of cumulating small effects to explain a major puzzle in health epidemiology, which is that rates of morbidity and mortality for virtually all chronic diseases are higher in less educated populations, regardless of disease, whether it is treatable or not, or the decade or country you consider, including whether it provides free health care or not. Moreover, the more that nations do to equalize outcomes—such as making health information and health care more widely available, the bigger the inequalities become. Mean health improves, but more so among the more educated classes. Health scientists have decided that only something very very general can explain this pattern, but they have struggled unsuccessfully for decades to figure out what it could possibly be. An increasingly popular hypothesis, including within parts of the CDC, is that social inequality itself is hazardous to health in some manner.

The Spearman Brown prophecy formula applied to weakly g loaded daily health behaviors seems a more plausible hypothesis to me.

No matter how you measure social class, rates of morbidity and mortality are usually at least 2-3 times higher in the lower social classes. Among the measures of social class, education—the most g loaded of the measures—is virtually always most strongly related to them. When IQ itself is measured—as it rarely is—it outpredicts education, suggesting that education is just a rough stand-in for g.

You also find the same social class pattern with all but one form of accidental death, dying in a private plane. The SES-mortality gradient differs greatly from one cause of accidental death to another but, with that one exception, it always disfavors the lower classes, even when it is hard to see how differences in material resources or occupation could possibly be involved. The SES-mortality gradients that would most seem to implicate differences in *mental* rather than material resources include accidental death from suffocation, choking on food or objects, and neglect or exposure, which occur mostly among infants and the elderly, and drowning and being struck by lightning, which occur mostly among young males. The risk ratios range widely, from just 1.3 for suffocation, when you compare people living in very poor neighborhoods compared to those in middle-income areas, to over 7 for dying from exposure and neglect. For purposes of comparison, the relative risk associated with low social class is about 2.1 for dying in a motor vehicle and 2.5 for dying in a fire.

My point is that IQ probably *won't* explain much of the variance in individual differences in health and accidents, but that it may be the most *consistent* influence on health and accident-related behavior, relentlessly contributing its bits of variance over the many days of our lives, across the many millions of people in each social class.

[11] Question 6. I'll just say a few words about the sixth question, which is How do the members of a society shape the life test battery that the current or future generations must take? My guess is that advancing technology is driving up complexity in many life arenas, which portends greater g-related social inequality.

To take an obvious example, daily activities are being computerized in many ways, creating a digital divide between individuals and groups that, I suspect, is at least as much mental as material in origin. Perhaps a less well-known example, which I spoke about on Wednesday, is that the ever increasing complexity of health care is demanding more learning and problem solving on our part.

[12] What we need to do, then, is ask what forces push these g-related risk gradients up or down over time, and in which arenas of life. I have shown complexity as one factor here

I suspect that another factor is any sort of social change that gives individuals more latitude in making their own choices, because that will force people to fall back more upon their own reasoning abilities, which include moral reasoning, as David Lubinski mentioned the other day. We might ask, for example, whether the progressive destigmatization of having children out of wedlock has increased the tightness of its link to g over time. As I recall, it was educated women who led this charge for more personal freedom, but they are better able to calculate the risks of exercising that freedom, which is perhaps why they seldom do. Most of us would laud efforts to advance technology and increase personal freedom, but both come at a social cost—greater variance in outcomes.

I will conclude by saying that Jensen's work opens up entirely new ways of examining the horizontal effects of g, from understanding how we deal with the minutiae of daily life and their hidden consequences to the biggest social issues of the day. In so doing, it also reveals why

intelligence will inevitably be a controversial topic in societies that wish to mute intellectual and social distinctions, perhaps *especially* as they go about increasing them. It is no wonder that many people are discomfited by Jensen's drawing our attention to this incredibly general force in social life. His passion for empiricism, his scientific acumen, and his unwavering integrity are an inspiration, however, to follow the new paths wherever they lead.

Thank you, Art.

[13] In this regard, I would be happy to send anyone who wants one a free copy of the *Intelligence* issue devoted to honoring Art. I still have plenty of extra copies.

References

Gordon, R. A. (1997). Everyday life as an intelligence test: Affects of intelligence and intelligence context. *Intelligence, 24*(1), 203-320. **(Slides 2, 6, 8)**

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24*(1), 29-132. **(Slides 6, 7)**

Gottfredson, L. S. (2002). *g*: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence* (pp. 331-380). Mahwah, NJ: Erlbaum. **(Slides 4, 5)**

Gottfredson, L. S. (in press). *g*, jobs, and life. In H. Nyborg (Ed.), *The scientific study, of general intelligence: Tribute to Arthur R. Jensen.* New York: Pergamon. **(Slide 9)**

Gottfredson, L. S. (submitted, 2002). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? University of Delaware. **(Slide 10)**

Humphreys, L. G., & Lubinski, D. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence, 24*(1), 159-201. **(Slide 9)**

Jensen, A. R. (1998). *The g factor.* Westport, CT: Praeger. **(Slide 2)**