Contested Summary Measures[1]

Robert Klitgaard
October 2007

It's a pleasure to be the warm-up act for James Flynn, a professor in the departments of political studies and psychology at the University of Otago, New Zealand. He's going to talk with you about the Flynn Effect, namely the increase across the generations in IQ. When he first documented this increase, his results were greeting with incredulity—how could IQ be increasing? Well, we know that height has increased over the past two or three generations in most countries around the world. Could the same be true for IQ? Professor Flynn showed that the answer is yes. By now the Flynn Effect is widely accepted and analyzed. True, tests of mathematics, general knowledge, and vocabulary have shown little Flynn Effect. But tests of abstract reasoning have shown large and persistent increases over time.

What's going on here, and what it means for education and more broadly for society, are the subjects of Professor Flynn's research, including his new book from Cambridge University Press, *What Is Intelligence?*[2]

As a kind of warm-up exercise for Professor Flynn's presentation, I would like to back us up one logical level to consider, or to deconstruct, the meaning and use of summary measures such as IQ. Consider another contested summary measure, GDP (gross domestic product).

GDP per capita is a measure of the total dollar value of a nation's domestic output in a given year divided by the number of citizens. As this figure grows or retrogresses, it provides a summary measure for the progress of the national economy. GDPpc has an advantage for international comparisons. Because it uses each nation's own prices in the calculations—prices that reflect the tastes and the value judgments of each country's citizens—comparing GDPs does not impose one country's set of tastes and values on another country.

In the 1980s, adjustments were made to GDPpc to take into account the relative costs of a basket of goods and services. The usual direction of this adjustment is that "poorer" countries have relatively larger GDPpc, and countries with high costs for nontraded goods and services, such as Switzerland and Japan, have their GDPpc adjusted

---

[1] Introduction to a lecture by James R. Flynn at Claremont Graduate University, October 10, 2007.

[2] James R. Flynn, *What Is Intelligence? Beyond the Flynn Effect.* Cambridge: Cambridge University Press, 2007.

downward. GDPpc adjusted for purchasing power parity has become the *de facto* benchmark of national progress.

But a contested measure. Consider this complaint by James Fox:

> I regard one of the biggest failings of economists in the development business today to be their lack of creativity in thinking about how to measure changes in human welfare in poor countries. Most economists treat GDP growth rates as the sole indicator of economic development, despite serious measurement problems, particularly in poor countries, and its lack of a distributional dimension. GDP simply does not resonate with many noneconomists.[3]

Sound familiar? Is there a "text" or template here that cuts across many summary measures?

- First, a measure or ranking or principal component in a factor analysis becomes "the" definition of something important. The Dow Jones industrial average becomes *the* measure of the health of the economy. The *U.S. News and World Report* ranking is treated as *the* gauge of a college. You can think of others: the Conference Board Consumer Confidence Index, a player's batting average, and so on. Some people behave as if "This measure is everything."
- Second, that summary measure is contested as inaccurate, incomplete, and overused or misused. Some critics seem to say, "This measure is nothing."
- And third, despite the critiques, the contested summary measure has interesting correlations, and sometimes predictive power. The correlations aren't one or zero; the measure isn't everything or nothing. "It's something."
- But what, exactly? Our task with regard to contested summary measures is to be attentive to both concept and evidence—and to avoid the extremes of saying the summary measure is everything or that it is nothing.

At the turn of the century I attended a remarkable small conference where people talked about what was meant by progress. I thought it might be fun to try to improve on GDPpc as that widely accepted measure of a country's progress. Income, I argued, does not measure such things as life, liberty, and the pursuit of happiness. Just for fun, I tried to create a new summary statistic building from these three concepts. *Life* included life expectancy and "health-adjusted life expectancy." *Liberty* included various indices of political and economic freedoms. *The pursuit of happiness* included measures of educational attainment and self-reported satisfaction with life.

---

[3] James W. Fox 1997. "What Do Economists Know that Policymakers Need To?" *American Economic Review*, Papers and Proceedings 87(2): 52.

The results? First, at the country level GDPpc correlates pretty highly, between 0.4 and 0.8, with each of the measures of life, liberty, and the pursuit of happiness (see Table 1 at the end of this talk).

Second, when I statistically combined the measures of life, liberty, and the pursuit of happiness into a single overall metric of a nation's progress, the correlation of that new index with a country's GDPpc was about as high as a correlation can be, given the underlying measurement error.

Let me now prepare the way for Professor Flynn by hypothesizing that a similar pattern can be observed for IQ. Like GDPpc, this contested summary statistic is criticized for being inaccurate, incomplete, and misused. Just as I argued that GDPpc is not conceptually the same as life, liberty, and the pursuit of happiness, so too Professor Flynn notes that IQ is distinguishable from such things as mental acuity (the ability to provide on-the-spot solutions to problems), the detachment and use of abstract thinking as opposed concrete thinking, attitudes about abstract reasoning, possession of knowledge and information, speed of information processing, and memory. Not to mention from other forms of "intelligence" that have been advanced.

And yet, IQ turns out to be correlated about as highly as it can be, given measurement error, with "g" (say, 0.9). IQ is highly correlated with such popular tests as the SAT and the GRE. Moreover, when measurement error and restriction of range are taken into account, it is fairly highly correlated with such things as later academic performance, with quickness to benefit from on-the-job training, with some thresholds of job performance, and with later-life success.

Professor Flynn, what might we make of this text or template, this pattern of argumentation and results? Your work has done much to help us understand what is useful and what is changing in a classic "contested summary measure," IQ. Your research can help us navigate the rhetoric-strewn road between "this measure is everything" and "this measure is nothing."

And we hope that from your work we may gain a sophisticated understanding of this particular contested summary statistic, given our hopes for improving education, enhancing labor markets, and creating a society that is even more open to what summary measures seem to miss.

Welcome, James Flynn.

Table 1

Correlations among Some Variables of Interest

	Life Ex	HALE	Politica	Civil lib	Econfree	Satv	Educ	lnGDPpc
Life Ex	1							
HALE	.96*	1						
Politica	.53*	.41*	1					
Civil lib	.62*	.50*	.89*	1				
Econfree	.65*	.57*	.73*	.80*	1			
Satv	.30*	.21	.34*	.46*	.49*	1		
Educ	.66*	.62*	.46*	.53*	.47*	-.06	1	
lnGDPpc	.86*	.79*	.64*	.73*	.76*	.40*	.69*	1

Note: N = 69 countries. Life Ex = life expectancy at birth. HALE = healthy life expectancy. Politica = a measure of political liberty on a 1 to 7 scale (7 best). Civil lib = a measure of civil liberties on a 1 to 7 scale (7 best). Econfree = a composite measure of economic freedoms on a 1 to 5 scale (5 best). Satv = self-reported happiness with life on a 1 to 10 scale (10 best). Educ = average years of education of the population 25 years of age and older. lnGDPpc = natural logarithm of per capita gross domestic product, adjusted for purchasing power parity.