Raven, J. (5/17/02). Response to Flynn: Searching for Justice: The Discovery of IQ Gains Over Time
WebPsychEmpiricist. Retrieved (date) from: http://www.wpe.info/papers_table.html

WPE WebPsychEmpiricist

Response to Flynn: Searching for justice: The discovery of IQ gains over time

5/17/02

John Raven[1]

Edinburgh, Scotland

Abstract

This paper was prepared as a *commentary* on Flynn's "Searching for justice: The discovery of IQ gains over time" which was published in the *American Psychologist* in 1999. In that paper, Flynn essentially argued that it would be possible to advance humane ideals (justice) by, on the one hand, discrediting "IQ" tests and, on the other, establishing a just meritocracy. In this paper it is argued that route to advancing humane ideals lies instead in both broadening the range of talents that we are able to adequately conceptualize, measure, develop, utilize, and reward, and also discrediting faith in the effectiveness of hierarchy and the concept of "ability" which sustains it. To do either, it will be necessary to introduce radical paradigm shifts in the ways in which we think about psychological measurement, the nature of society, and the nature of science.

[1] Correspondence concerning this article should be sent to: Dr. John Raven, 30 Great King Street, Edinburgh EH3 6QH, Scotland Phone: ++ 44 131 556 2912 Fax: ++ 44 131 556 3754 e-mail: jraven@ednet.co.uk or WPEList@yahoogroups.com

Response to Flynn: Searching for justice: The discovery of IQ gains over time
5/17/02

The January 1999 issue of American Psychologist contained an article by Jim Flynn which merits response at a number of different levels.

I want to begin by acknowledging that I know Jim as an extraordinarily socially committed person and as an unusually thorough and serious research worker – but also as someone who does not take himself too seriously. This last characteristic allows him to do such things as attack socially sensitive issues obliquely by, in a sense, making fun of well established and meaningful differences between ethnic groups while drawing attention to others which, although probably of much greater importance, have been much less well investigated. This has enabled him to evade the storm which would otherwise have gathered around him for highlighting the latter. While this behavior has the advantage that one can respond in like vein, it, unfortunately, also means that anyone who wishes to argue both that the attempt to discredit one set of differences by making fun of the measures is not well founded and that the other differences, if properly investigated, are important lays themselves open to attack on both counts.

Flynn's article opens with an attack on the validity of the RPM and the way in which scores have been used to legitimize the assignment of second-rate status to certain ethnic groups. It continues with a demonstration that other important variables – such as creativity, motivation, and social support – have been neglected. But it then launches into a discussion of meritocracy. And here is the real problem with the article because Flynn fails to discuss either the legitimacy of the concept of hierarchy into which it fits or the forces which support hierarchy and require a single-factor concept of ability to legitimize the very unjust discriminations he wishes to ameliorate. Those discriminations compel participation in a system which is not merely inhumane but also relentlessly driving homo sapiens toward both its own extermination as a species and the destruction of the planet it inhabits.

This response is divided into two main parts. The first deals with Flynn's attempt to discredit the meaningfulness of the Raven's Progressive Matrices. The second part discusses five other issues pertinent to the advancement of humane ideals through psychology.

"The Tests Cannot Save Themselves": The Scientific Status of the Raven's Progressive Matrices

One of the impressions Flynn creates is that scores on the Raven's Progressive Matrices (RPM) test have little meaning. Yet he actually knows very well that there is abundant evidence that Raven's Progressive Matrices (RPM) and Vocabulary Scales (MHV) are among the best established measures of the two human characteristics whose scientific status is most secure. Almost uniquely, they measure these variables directly – that is without requiring the complex

calculations needed to derive latent trait scores from information gained using a cluster of oblique sub-test scores. Space constraints prevent me listing more than a fraction of the evidence supporting this assertion. Full summaries will be found in Raven, J., Raven, J.C., and Court (1998a,b,c,d&e) and Raven (2000).

The evidence that "general cognitive ability", or g, accounts for about 10% of occupational performance and that adding a battery of other tests purportedly measuring other human qualities improves one's ability to predict such performance hardly at all has been repeatedly summarized by Schmidt and Hunter (e.g. 1998) and Jensen (e.g., 1998). Less well known is that it predicts an astounding 60% of social mobility (see, e.g., Hope, 1984). In short, despite Flynn's efforts (in the tradition of such authors as McClelland, 1973) to demonstrate that "the tests cannot save themselves", whatever is being measured by tests of "general cognitive ability", or g, is important.

What, then, of the claim that the RPM and MHV are among the best measures of this ability?

First note that attempts (mainly by laboratory psychologists) to either generate "more basic" measures of "general cognitive ability" or to measure other important components of cognitive functioning have not met with great success (see, e.g., Carroll, 1993). The same has been true of applied psychologists who have tried to generate tools useful to educators and other practitioners (see, e.g., Matarazzo, 1990 for work on the minimal extra information contained in the sub-scales of full-length "intelligence" tests). Full-length educational attainment tests generally lack both construct and predictive validity (see, e.g., Flanagan, 1978; Raven, 1991).

The RPM and MHV were developed as theoretically-based tests for research into the very issues – the effects of genetics and the environment on general cognitive ability – that Flynn has so marvelously highlighted. It was precisely because the previously-mentioned problems involved in interpreting scores on multiple-component "intelligence" tests had already become apparent that J.C. Raven (1936) set about developing tests which would be both theoretically-based and yield scores which would be directly interpretable without having to engage in complex calculations to arrive at scores on "latent factors".

Raven was a student of Spearman's. As is well known, Spearman (1927) noticed that what are generally thought to be independent abilities tend to be intercorrelated and that the correlations between these different abilities could be accounted for by positing a single underlying general factor. This he named ***g*** to avoid the connotations of the word intelligence which are still causing havoc in Flynn's logic. It is less well known that Spearman saw ***g*** itself as being made up of two, very different, abilities. These he termed "eductive" and "reproductive" ability. It is important to note that he did not isolate these through factor analysis. They were functionally different abilities. The term "eductive" derives from the Latin word "educere", which means "to draw out"; hence eductive ability is the ability to make sense of confusion.

Much later, Cattell and Horn (e.g., 1978) popularized the terms "fluid" and "crystallized" intelligence. This nomenclature is misleading and has trapped endless psychologists – including Flynn – into useless, purely epistemological, argument. Not only does the retention of the word "intelligence" in these terms make (as Spearman anticipated) for confusion, so does the notion that one of the two components of ***g*** is a "crystallized" form of the other. Fortunately even Horn (1994) now recognizes this is not the case. The abilities are distinct from birth, have different genetic precursors, are affected by different aspects of the environment, and are associated with different types of life performance. Yet they are not factorially separable.

Spearman's (1927) observation that:

> To understand the respective natures of eduction and reproduction – in their trenchant contrast, in their ubiquitous co-operation and in their genetic inter-linkage – to do this would appear to be for the psychology of individual abilities, and even for that of cognition in general, the very beginning of wisdom.

appears to have been perspicacious.

Had they been in the habit of thinking in these terms, it would have been much easier for those who noted the increase in certain sub-scores of "intelligence" tests with date of birth [among whom were such researchers as Owens (1966), Thorndike (1977), Schaie and Willis (1986) and Schaie (1994) besides Flynn] to both make sense of their observations and avoid much unproductive argument. Scores on measures of reproductive ability have increased hardly at all while scores on tests which involve both eductive and reproductive ability (such as the sub-scales of multiple-component "intelligence" tests) have (as data published by e.g., Thorndike [1977], Schaie and Willis [1986] and Schaie [1994] reveal) been increasing at rates which are proportional to the extent to which they measure eductive rather than reproductive ability. (Note that this is true whether eductive ability is measured by verbal or non-verbal methods.)

We may turn now to the status of the RPM and MHV as measures of eductive and reproductive ability.

Factorial work generally agrees (see, e.g., Snow, 1989; Jensen, 1998) that the RPM is one of the purest measures of ***g***. Encouraging though this is, however, it falls short of demonstrating that the RPM measures *eductive* ability. Despite Spearman's exhortation to research the area, it turns out that there are only a handful of studies (reviewed in Raven, J., Raven, J.C. & Court 1998e) which bear directly on the scientific status of the eductive-reproductive, fluid-crystallized, distinction.

But, as soon as one starts to probe the question of the construct validity of the RPM and MHV, one encounters a problem. A little thought undermines the usual assumptions about how tests are to be validated and, in particular, faith in correlational studies. It turns out that, as Messick (1989, 1995) has argued, "thick", descriptive, even ethnographic, methodology is required.

The problem is this: Eductive ability is conceptualized as the ability to make meaning out of confusion; the ability to separate figure from ground; the ability even to perceive. This is a difficult and demanding activity in which no one is going to engage unless they are strongly intrinsically (and not extrinsically) motivated to carry out some task which demands it. It requires sensitivity to feelings, the initiation of feeling-based action, and learning from the effects of those actions – or "experimental interactions with the environment" – more about the nature of the problem one is attempting to solve and the effectiveness the strategy one has adopted to deal with it. Thus people will only develop, deploy, and display their eductive ability while carrying out particular – and very different – kinds of activity. One person will do so while putting people at ease or dealing with drunkards, another while seeking to advance a scientific theory, and another while seeking to create political chaos. One cannot expect people who demonstrate a high level of eductive ability in the course of carrying out one of these activities to display it if asked to carry out activities in which they are not interested.

A number of consequences follow from these observations. The most serious – to which we will return – is that, as several researchers (e.g., Raven, 1987; Tough, 1973; Kohn & Schooler, 1978) have demonstrated, "cognitive ability" becomes bonded to motivational dispositions. Here it is sufficient to note the implications for anyone who – like Flynn – seeks to validate a test that claims to be a measure of eductive ability by relating it to selected "real-life" performances like filing patents or performances (such as are measured by so-called academic attainment tests) which are known to be primarily dependent on reproductive ability.

So we, as Messick's writings would lead us to expect, are forced to seek *conceptual* attempts to establish the construct validly of the RPM and MHV.

Perhaps the most striking tour-de-force in the area is that of Snow and his colleagues (Snow, 1989; Snow, Kyllonen & Marshalek, 1984). These authors depicted the relationships between various tests that had emerged from numerous studies in the form of a Guttman radex. The RPM emerges as the best single index of eductive ability, while vocabulary tests (of which the MHV is but one example) emerge as indices of one of three types of reproductive ability. (The other two sectors comprise [a] tests of the ability to perform routine arithmetical calculations and [b] tests of the ability to perform routine spatial operations.) The relationships between measures of performance in these three sectors is determined by both the extent to which the abilities in question require related routine skills (i.e., the forms of reproductive ability shown around the circumference) *and the extent to which they depend on eductive ability rather than reproductive knowledge and skills* (i.e., their distance from the center of the circumplex).

But by far the most convincing evidence for the *construct* validity of the RPM and MHV comes from the item analyses of the tests themselves.

The RPM and MHV were constructed using Item Response Theory (IRT).

Item characteristic curves (ICCs) were plotted to establish that the capacity to solve any one item was systematically related to the capacity to solve every other item and to total score (see Raven, J.C., 1936; Raven et al., 1998a).

To illustrate the procedure in a different domain consider what one would have to do to establish a scale to measure "hardness". One would assemble a sequence of substances: cotton wool, cheese, chalk, granite, steel, diamond etc. and seek to show that, as standards or criteria of hardness, they bore a systematic relationship to each other and to the "hardness" of other substances. One would discard cheese as a suitable index because its position would vary with its maturity, temperature, etc. *Note that the qualitative differences between the substances has nothing to do with their value as standards, or indices, of hardness.*

Exactly the same was done with the items of the RPM and MHV. The items which were retained scaled – that is to say, the ability to solve any one of them bore a consistent and stable relationship to the ability to solve other items in the scale. They scaled in the same way for people of different ages and from different socio-economic backgrounds. The development of the abilities required to solve the items was thus shown to be cumulative and incremental. It is not possible to solve the more difficult items without being able to solve the easier ones.

Note that, as in the case of developing a scale of standards against which to index the hardness of all substances, the fact that the easier ("gestalt") items are qualitatively different from the more difficult items has nothing to do with the case. Or, put the other way round, "simple" perception is a *conceptual* process requiring the ability to make order out of confusion, to separate figure from ground. Thus the abilities required to solve the more difficult items require the same cognitive processes as those required to solve the easier ones. Only the *level* required is different. The processes required to analyze, abstract, infer, and check, to infer the nature of the whole from inspection of the parts and to understand what one should pay attention to in the parts as a result of having built up an understanding of the whole, do not differ as one moves from the easy to the more difficult items. The sequence of items reveals a *continuum* of cognitive ability which is unchanging in its essential character from beginning to end.

Going beyond this internal item analysis, Styles and Andrich (1997) have shown that ICCs for a series of Piagetian tasks map onto the ICCs for the RPM. In other words, development of the abilities required to solve these problems at different levels is incremental and in step with the development of the abilities required to solve RPM problems of similar difficulty. The tasks are simply widely spaced standards of performance. They do not reveal "stages" or metamorphoses in development. Vodegel-Matzen (1994) has shown that making the items more "realistic" (i.e., using hats, bananas etc. to make up the designs) while retaining their logic makes them easier for everyone but changes neither the order of the items nor the rank order of people (i.e., it does *not* benefit the "disadvantaged"). It does, however, change the nature of the task to be undertaken

since these revised items are less dependent on the ability to discover order in apparent confusion – i.e. eductive ability itself.

Two things follow from these observations: (i) the test is robust in the sense that it, like a good scale of hardness, "works" in the same way in different environments, and (ii) it cannot be true that, in any general sense, people from certain backgrounds are "unfamiliar with the way of thought required to solve the items". (That it may be *less* familiar to people from some backgrounds than others is not in dispute.)

Beyond these things, a number of researchers (see Deary, 1995, 2000 for reviews) have shown that scores on the RPM are systematically related to some "more basic" measures of cognitive functioning – such as the amount of time required to make accurate discriminations between the lengths of two lines presented tachistoscopically.

The conclusions to be drawn out of what has been said so far are:

1) That something which might be called "General Cognitive Ability" "exists" and has a number of extremely important real-life correlates such as occupational performance and social mobility.
2) That ***g*** has a number of components of which the two most important are "eductive" and "reproductive" ability. These are every bit as "real" – have as secure a scientific status – as "hardness" (which is equally "intangible").
3) That the Raven's Progressive Matrices and Vocabulary Scales are almost unique among the best established measures of these qualities in that they measure them directly – that is to say in much the same way as one measures height and weight. It is not necessary to estimate them indirectly via complex calculations from scores on obliquely related measures.
4) That, if one employs direct measures of these two qualities in one's research, the relationships with other variables fall out much more neatly. Eductive ability *is* increasing at a dramatic rate in all cultures. By and large reproductive ability is not. The rate of increase on scores on other tests is determined by their "loading" on these two variables.
5) That any relationships which are established between scores on these tests – such as with date of birth and occupational or ethnic group – call for a careful search for an explanation. They cannot be summarily dismissed as "meaningless".
6) That Flynn's attempt to discredit changes in mean scores on the RPM and other measures of *eductive* ability over time by suggesting that such changes have not been accompanied by commensurate changes in knowledge or routinized skills (i.e., indices of *reproductive* ability), in such things as numbers of patent registrations (which can be expected to be heavily determined by institutional arrangements), and in such things as "creativity" (of which, as we shall shortly see, psychologists have failed to develop meaningful measures) is fundamentally misguided for the reasons indicated.

What the evidence so far reviewed does *not* do is:

1) Confirm that the interpretational framework used to give meaning to the variables which have been isolated is the most appropriate.
2) Confirm the validity of the predictions or generalizations commonly made from knowledge of an individual's scores on "intelligence" tests. (It is because the scientifically sound constructs we can measure with confidence are typically embedded in much broader popular concepts – such as "ability" – which lack any status whatsoever as scientific constructs and whose predictive power is grossly over-generalized [as in *The Bell Curve* {Herrenstein & Murray, 1994}] that Flynn's critique at first seems so devastating.)
3) Say anything about how the observed relationships come about.
4) Say much about the *relative* power of eductive and reproductive ability (taken separately or together) in comparison with other factors (such as motivation) to predict a variety of types of performance (although it *does* say that the addition of scores on most of the available tests which purport to measure these qualities adds little to our ability to predict these performances from ***g*** alone).
5) Consider uses to which it might be inappropriate to put test scores (such as to administer educational or occupational selection programs) or the forces which induce us to continue to try to order society using a (scientifically unsupportable) single-factor concept of "ability". (The use of test scores to reinforce the very concept of hierarchy which lies behind Flynn's notion of a meritocracy is particularly open to question because it actually supports an inhumane system of social differentials which have the effect of compelling participation in a incredibly dysfunctional and destructive [i.e. unintelligent] social system.)

It is precisely because psychologists have generally failed to address the issues briefly alluded to above that Flynn has been able to have such a field day. If his article, and the huge amount of work which lies behind it, (and this response) leads to the initiation of appropriate research, he is to be congratulated. And this is not only because it is by unpacking the issues he has brought to light that the route to the advancement of humane ideals will be found. It is also because the questions his article raises point to some of the most important unfinished (even unstarted) business to be addressed by psychologists. Unfortunately, even trying to get our minds round the issues not only presents even more serious conceptual problems than those we have already encountered. It also brings us into conflict with (i) organizational arrangements which have evolved to legitimize and perpetuate the hegemony of particular "scientific" viewpoints and methodologies (cf. Kuhn, 1962/70), and (ii) sociological arrangements which call for, and rely on, a single-factor concept of "ability" to (a) legitimize the power and reproduction of established elites (which implies also legitimization of discrimination against the "disadvantaged") and (b) legitimize

and perpetuate deep divisions in society, the function of which is to compel participation in our social system, however widely rejected that social system actually is.

In the remainder of this response I will discuss these five open questions.

1. The interpretational framework

The reader may be surprised to learn that, despite what we have said about the secure scientific status of ***g*** and eductive ability, we are about to suggest that an alternative interpretation of what the tests measure is possible – or, at least, that whatever is measured is so closely linked to other variables that the interpretation of the relationships which have been established between scores on measures of eductive ability and other variables (like occupational performance and social mobility) is seriously open to question.

We have already hinted that, for a number of reasons, the measurement of – and especially the validation of measures of – eductive ability is much more problematic than is commonly supposed. Among the problems are the following: (i) provided task performance is appropriately conceptualized and evaluated, eductive activity is needed for the effective performance of almost all types of task, yet (ii) the effective performance of those tasks also calls for numerous other components of competence; (iii) the types of activity which a cross-section of human beings have the potential to perform appear to be legion and no adequate classificatory framework (equivalent to atomic theory in chemistry) has been developed to help us think about them; and (iv) eduction is a difficult and demanding activity which is unlikely be displayed (let alone developed) if those concerned are not engaged in activities they find intrinsically motivating [i.e. unless they are engaged in one or other of the activities which are to be distinguished and classified using the yet-to-be-developed framework the need for which was noted under (iii)].

One of the implications of (iv) above is that the tests which we have so far accepted as measures of eductive ability in general may in fact be measures of the willingness and ability to think about only one type of problem. People who are strongly motivated to put others at ease, create a warm convivial atmosphere, craft metal panels into wonderful shapes, or make music may not reveal their capacity for eduction on “academic” tests.

The methodological problem hinted at in the last paragraph can perhaps be made more obvious by taking another example: people who show all sorts of creativity and self confidence when dealing with drunks may not display that creativity when asked to think of as many uses as they can for a brick. Nor may they display that self-confidence when asked whether they are confident about passing academic examinations.

To make the point in yet another way, we may note that “thinking” – cognitive activity – is primarily a feeling- and action-based activity which requires the use of feelings to suggest insights and initiate action (“experimental interactions with the environment”) to test those “insights”. Testing such insights and strategies requires extraordinary persistence. So the question we should

be asking may be closer to “In relation to what kind of activity does this individual display his or her capacity to make meaning out of confusion?” than to “Of what level of eduction is this person capable?”. Hence, tests currently thought to measure “eductive ability” may be more correctly described as measures of the motivational disposition to think about a particular type of problem.

It would follow from this that Sperry (1983) and Trevarthen (1990) may well be right to suggest that what is neurologically localized is not “cognitive ability” as such but the motivational disposition to engage with a particular type of problem. If we had a series of tests which measured the level of cognitive activity displayed while people are carrying out different types of activity, “cognitive activity” might well turn out to be located in many different parts of the brain.

Now, as it happens, this theoretical insight is not without empirical support. As we shall shortly see, it has long been known that ***g*** is psychologically bonded to certain motivational dispositions and that variance in these motivational dispositions is as strongly predictive of social mobility as are ***g*** and *eductive ability*. Furthermore, the relationships between social mobility and these other variables are so strong that it is impossible to choose between an explanation of social mobility which is grounded in motives or values and one grounded in “general cognitive ability”.

But, before looking at the data just mentioned in more detail it is useful, in assembling material which relates to Flynn’s paper and the wider pursuit of humane ideals, to first re-examine the nature and assessment of the outcomes which people most commonly wish to predict using measures of eductive and reproductive ability – namely educational and occupational performance.

2. The generalizability, or predictive power, of the measures

In his attempt to advance humane ideals by discrediting the measures, Flynn makes much of the increase in eductive ability over the years not having been accompanied by commensurate changes in performances that he, using, like most other people, an all-embracing concept of “ability” or “intelligence”, would have expected to have occurred.

In any response, it is therefore necessary to draw attention to the sloppy background thinking which has guided most empirical work in this area.

We have already seen that the RPM measures only one component of “general cognitive ability” and that that “general cognitive ability” has, in turn, never been presented by any serious researcher as a measure of all the qualities required for occupational or life success. Yet it is widely presented as a measure of the much more inclusive concept, “ability”.

Educational performance is typically assessed by tests which lack both construct validity and the ability to predict performance outside the school system. That they lack construct validity can be seen by considering, first, that the word “educate” comes from the same Latin root – educere – as eductive ability. It means “to draw out”, not “to put in”. Thus tests which measure the ability to absorb and regurgitate information by definition lack construct validity. As Goodlad (1983) and the

author (Raven, 1991) have argued, they do not merit designation as measures of academic ability since they involve little judgment, reconciliation of different points of view, or competence as scientist, writer etc.

It is not only the word “academic”, but also the word “learning” which has been operationally re-defined to both retain little of its meaning and in such a way as to minimize possible correlations with eductive ability. In most current “academic” discussion, the word “learning” is used to refer only to acquiring temporary mastery of unrelated snippets of knowledge. It is rarely used to refer to such things as learning to lead, to invent, to put people at ease, to create social chaos, to influence social and political systems, or to intervene in sociological processes. All of these forms of learning would draw heavily on eductive ability (and the ability of appropriately motivated individuals to undertake such activities has probably improved over the years). On the other hand, the ability to master and regurgitate snippets of temporary knowledge amounts precisely to an operational definition of *reproductive* ability (which has increased hardly at all over the years). It would seem to follow that there is, after all, no reason for Flynn to be surprised by the observations he has made in this area.

Turning to occupational competence, one finds an even more appalling mess. Critical Incident studies (see, e.g., Klemp, Munger, & Spencer, 1977 or Huff, Lake, & Schaalman 1982, for examples of individual studies, and Raven, 1984/97, Spencer & Spencer, 1993, and Raven & Stephenson 2001 for summaries of the literature) of the behaviors which distinguish more from less-effective workers have demonstrated the vital importance of a wide range of generic high-level competencies. Examples include initiative, the ability to build up a personal understanding of how the employing organization works and one’s place in it and then use that understanding to take the actions necessary to do something about problems, the ability to get together with other people at the same level to influence the perceptions of those higher up in the organization, and the ability to model the, normally private, psychological components of competence which determine effectiveness so that subordinates can learn to do likewise. Most of these have escaped the attention of psychometricians and defied measurement along conventional lines.

The development and display of such competencies is very much determined by people’s beliefs about society and their organizations, how they work, and their place in them. This has a number of very disturbing consequences. For example, it turns out that, if one wishes to nurture and release competence it is necessary to influence these beliefs. Perhaps more importantly, if one wishes to *assess* competence it is necessary to assess these social and political beliefs. The thought of influencing or assessing societal beliefs as part of assessments of competence raises a host of dilemmas, in dealing with which it is important to hang onto the fact that this is precisely what schools are already doing ... but without any public scrutiny. Making the processes explicit could help to make them not only more effective but also, given an appropriate supervisory structure,

help to ensure that whatever is done is done in such a way as to advance humane ideals. But make no mistake about it, *failure* to nurture and assess these things renders invalid any claim to competence as an educator or psychometrician.

Occupational competence also involves such things as assembling teams of people who collectively have the talents required to create the emergent properties of groups which we label "enterprising" or "intelligent". Scarcely anyone has studied these group properties. But just as copper sulphate has properties which cannot be predicted by adding the properties of copper, sulphur, and oxygen, so the emergent properties of groups cannot be determined by adding the properties of the people who comprise them.

Such authors as Kanter (1985), Schon (1973), and Russ-Eft and Brennan (2001) have, however, shown that the structural arrangements required for cultures of enterprise and innovation (and organizational survival) are not usually dependent on having separate cadres of innovators but on setting aside time in the normal working day for everyone, from lavatory attendant to managing director, to work in what Kanter calls "parallel organization activity" concerned with innovation. During this time people work in fluid, flat, non-hierarchical, structures which call for and utilize diverse talents not normally called for in the job They share observations which could only have been made in their "main" roles about how to improve products, services, and organizational arrangements. And they get together with others – performing roles of ideas-person, publicist, fund-raiser, politician, technician, or prototype-developer – to do something about it. What these observations reveal is that both traditional concepts of hierarchy and the range of activities focused upon in typical job descriptions are too limited. And the job descriptions for, and staff appraisal procedures applied to, managers are the most inappropriate.

Note: (i) All the competencies so far mentioned call for eductive ability, (ii) people who display high levels of eductive ability while carrying out one of these tasks are unlikely to display it if someone orders them to carry out a different kind of task, (iii) the effective performance of each and every one of these tasks also calls for a wide variety of cumulative and substitutable components of competence going well beyond eductive ability, and (iv) all these competencies have eluded – even defied – measurement by traditional psychometrics whether applied to assess the predisposition to behave in these ways or in performance appraisal exercises.

Note, above all, that traditional criteria of performance – turnover, supervisors' ratings, number of patents filed, etc. – cannot be other than wide of the mark. Organizational performance bears no simple relationship to the competence of individual managers or subordinates. What any one person can do is very much determined by what others do. Contrary to what the Spencers would have us believe, different people performing a job with the same designation do, and need to do, very different things. One manager makes a success of his company by manipulating the stock market, another (temporarily) by creating a "lean mean" organization denuded of all the staff

and time that would be needed to adapt in advance to the future, and a third by harnessing the idiosyncratic creativity and initiative of all his subordinates. In short, Flynn's attempt to discredit measures of eductive ability by showing that the observed changes over time have not been reflected in patent registrations etc. is not sustained. The effects of eductive ability have to be unscrambled from the effects of other motivational predispositions and components of competence as well as from embedded thoughtways and institutional arrangements. The issues cannot be unpacked using simple correlational methods.

Flynn is, of course, not alone in making these errors: Barrett and Depinet (1991), while correct in their strictures on the competency movement, have utterly failed to grasp the problems involved in thinking about and assessing occupational competence.

We turn now to the nature of the *knowledge* that is required for effective occupational performance – because Flynn, yet again, by following conventional thoughtways, leads us into a trap.

If one takes managers as an example (and similar findings apply to other occupational groups), one finds (Klemp et al., 1977; Jaques, 1976, 1989) that they need to make sense of previously unarticulated political, economic, sociological processes. They have to build up this understanding *themselves.* It is not available to them in any explicit form in any textbook. What is needed is idiosyncratic, largely tacit, knowledge. The level of understanding that is required would have eluded most thinkers from the past. One has only to compare the level of insight displayed by previous generations of British kings and generals that is documented in the material on display in the Public Records Office in London with that which Klemp et al. (1977) have shown to be necessary among modern Naval officers to obtain the necessary evidence. Even Machievelli does not get near it. (Of course, as reference to Shakespeare shows, there are exceptions to all generalizations!)

3. How the relationships come about

Given what we have seen in the last two sections, how can it possibly be that, as previously mentioned, measures of ***g*** can predict about two thirds of social mobility? The most likely explanation is that, as we have seen, ***g*** is psychologically bonded to motivational dispositions or valued styles of behavior – such as a value for taking responsibility for others at the upper end and having strict rules to guide one's life at the other (Raven, 1987; Tough, 1973; Kohn & Schooler, 1978).

But what of alternative explanations of the relationship between birth cohort and RPM scores?

This is not the place to explore answers to this question in any detail, partly because Neisser (1998) has brought together commentaries from a number of authors. It is sufficient to note that

most of the common "explanations" are invalidated by (1) the fact that the increase has been documented on all measures of eductive ability, whether verbal or non-verbal (see, e.g., Schaie and Willis, 1986, 1994); (2) It has occurred in all countries for which figures are available; (3) the RPM norms are, at any point in time, similar in most countries with a tradition of literacy regardless of whether those countries have advanced educational systems or access to television; and (4) it has been documented among young children as well as among adults.

But there is one more thing to be said. The increase in eductive ability is not unique (although its rate of increase may be). Most human capabilities – height, athletic ability, etc. – as well as life expectancy have been improving. But reproductive ability – almost uniquely – has not – and this despite enormous investments in "education" at both pre-school and post-secondary levels and a dramatic growth in access to the media. It would therefore seem that Flynn has focused attention on the wrong problem!

4. The relative ability of *g* in comparison with other indices of motivation and context to predict occupational and life performance

In his efforts to find ways of reducing ethnic discrimination and thus advancing humane ideals, Flynn cast about for ways of discrediting the tests which were being used. He tried to do this in four ways: (i) by arguing that the changes in intelligence test scores over time which had been noted by others revealed that such scores were subject to huge – and generally unsuspected – environmental influences and that this rendered any genetically-based explanation of ethnic differences untenable (but does the world-wide increase in height render genetically-based explanations of ethnic differences in average height untenable?), (ii) by suggesting that the observed increase in eductive ability with date of birth was not accompanied by commensurate increases in capabilities which one might have expected on the basis of the usual, too-inclusive, concept of "intelligence", (iii) by projecting the increases backwards in time and suggesting that our grandparents (or the Greeks) could not have been that stupid (the problem here is that similar backward projection of height would lead to similar absurdities), and (iv) arguing that other variables – and in a seemingly somewhat paradoxical *volte-face*, ethnicity itself – do account for differences between ethnic group's real-life performances. It is on the paradoxes that this last implies that we will focus in this section: Why, despite the obvious and huge predictive power of gender, ethnicity, and religion, have psychologists been unable to account for anything like as much of the variance in educational and occupational performance (albeit over-simplistically measured) using supposedly more scientifically-based measures of ability, motivation, personality, and environment?

The answer which will be offered is that the psychometric model which has been employed has been even more inappropriate than the remarks already made would lead one to suspect and

that there has been an almost total failure to appropriately conceptualize and assess the relevant features of the "environment" – supportive climate, engagement with values etc.

It will then be argued that the way to advance humane ideals is, not to seek to discredit the measurement of general cognitive ability, but to question the framework into which the way it has in the past been measured and interpreted fits. Although Flynn actually accepts this framework, it seems to the author that one of the most important ways to advance humane ideals is to use the understandings briefly shared in this paper to generate meaningful measures of a wider range of other talents and then to use them to find ways of identifying, developing, utilizing, and rewarding all the human resources available to society. To this end, we need to think more carefully about the nature of developmental environments and to set about more systematically introducing them into schools and workplaces. We need to develop the tools which teachers and managers need to perform their roles effectively. And we need to change such people's job descriptions, the criteria of performance applied to them, and means of assessing performance (and especially the job descriptions and performance appraisal systems of public servants) – and find better ways of giving effect to appraisals of performance.

But we will then discover that there are a host of reasons why it is going to be extremely difficult to introduce the relevant changes ... even in our ways of thinking about ability and the psychometric processes we employ. In the end it will become clear why Flynn has been able to get an article claiming to advance humane ideals by criticizing measures of general cognitive ability and discussing meritocracy into *The American Psychologist* while those of us who have advocated a much more scientifically well-grounded and humane rationale for student and staff assessment, development, and deployment have been unable to obtain the resources required to legitimize and develop that position and, largely, denied publication.

We have already seen that most current attempts to measure such things as eductive ability, creativity, and self-confidence – not to mention such things as internal-external locus of control – are fundamentally misguided because everybody displays all these qualities to a greater or lesser extent while carrying out activities they are strongly motivated to undertake.

Note the suggestion that motivations are extremely diverse and the implication that it is essential to determine the specific motives of an individual prior to any attempt to assess such things as creativity or self-confidence.

These observations render most of our current measurement paradigm untenable. A fundamental paradigm and methodological shift is required if we are to move forward.

This is of more than academic importance. "Evaluations" of educational programs such as "progressive education" which assert that they fail to enhance reading, writing, and arithmetic skills – outcomes which such programs were never meant to promote and which are, in any case, assessed by measures which lack construct validity – while failing to record that they nurture self-

confidence, the ability to make one's own observations, and the ability to understand and influence social and economic processes (because there are no good measures of these outcomes) are incompetent, unscientific, and highly unethical. They are unethical partly because they lead to the closure of programs which were nurturing competencies which are of the greatest importance from the point of view of finding ways of changing the way we live in order that the species will survive. But they are also unethical because they deny individual pupils the opportunity to develop and get recognition for their talents and thereby assign them to social positions in which they will be subject to degrading, dehumanizing, and inhumane treatment by others. By the same token, evaluations of individuals which record, for example, that they lack "intelligence", mathematical ability, and the ability to read while failing to record that they are excellent at putting people at ease and making them feel comfortable – and thereby precluding them from a decent way of life – are likewise incompetent, unscientific, and unethical.

Note that it is not only the development of the individual assessment procedures which is problematical: there is also the question of how to generate comprehensive evaluations of people and programs. How can we try to ensure that *all* important outcomes, personal and social, short and long-term, are investigated and recorded? This query, in turn, raises fundamental questions about how understanding – science – is to be advanced ... because one can only move toward comprehensiveness if one makes resources available for some people to investigate questions which everyone else thinks are unimportant or not capable of being investigated.

Returning to the more specific question of the inappropriateness of the psychometric model which has guided most test construction, we may first draw attention to the fact that there does exist an alternative model – although it is widely misunderstood and has been corrupted back into the traditional internal-consistency framework. The framework in question was developed by David McClelland (McClelland, Atkinson, Clark, & Lowell, 1958) and his associates. It was described as a framework to be used for the assessment of "motivation", but close attention to the operational framework which lies behind it reveals that it is a framework to be used to guide the assessment of competence.

As explained in more detail in Raven & Stephenson (Eds.) (2001) McClelland's framework has these features: It sees motivation as an integral component of competence, not separably measurable. It sees competence (or ability) as being determined by numerous, cumulative and substitutable, components of competence which are drawn from the cognitive, affective, and conative domains – all three of them. It assesses the likelihood that someone will be able to effectively carry out activities he or she cares about by cumulating indices of the presence or absence of requisite components of competence. Since this procedure is in fundamental conflict with the internal-consistency based measurement model, it is important to note that it is entirely compatible with another way of thinking about the prediction of behavior: namely the calculation of

multiple-regression coefficients. In predicting success in carrying out different kinds of task, the McClelland procedure sums across a series of independent, cumulative and substitutable, components of competence, that is to say it behaves like a multiple regression equation giving unit weight to the (independent) predictors.

Here, it would seem, we have exactly the model we require to both develop better indices of important human qualities and to classify and order the data which have emerged from the previously mentioned critical-incident-based studies of occupational competence.

Here we have an appropriate framework for thinking about multiple talents, their assessment, development, and deployment.

Yet that framework has been widely ridiculed (e.g., by Barrett & Depinet, 1991) and corrupted back into internal consistency notions (e.g., Edwards' [1954] measure of n.Ach) and even by the Spencers who have, in some sense taken on the mantle of McClelland.

So here we have one set of reasons why we have not been able to do better than astute common observation of the predictive power of crude indices like ethnicity and religion with our "sophisticated" measures of creativity and motivation.

But, in discussing the predictive validity of Asian-ness, Flynn (like disparaged workers such as Greely, 1972 and McClelland, 1958) highlights the importance of climates of support, even compulsion.

When thinking about the environmental determinants of behavior it emerges that, we, as psychologists, have utilized unduly simplistic models. Just as the only aspects of the environment that it is relevant to record when discussing a chemical reaction are those which interact with the substance being transformed, so the only aspects of the environment which it is relevant to describe when considering the behavior of an individual are those which engage with his or her motives and lead to the development and display relevant components of competence. What is more, the chemical analogy highlights our neglect of the *transformational* processes which take place in homes, schools, and workplaces. Our modeling has been essentially linear. Yet as Jackson (1986) has been at pains to point out, effective education is essentially idiosyncratic and transformational. (The neglect of these features has rendered our modeling of most educational and developmental processes, and most of our educational evaluations, grossly incompetent.)

To round out this discussion, we may return to our earlier chemical analogy regarding the emergent properties of groups. The ability to advance understanding, to adapt our behavior to environmental imperatives in such a way that the species will have any hope of survival, in short, the ability to act intelligently, requires, as Kanter has shown, a huge range of people to contribute in very different ways. Yet, despite the fact that books dealing with the emergent properties of chemical compounds fill whole libraries, there has been virtually no discussion of the qualities which combine together to result in groups having different emergent properties or systematic

development of a framework which would be useful for conceptualizing and systematizing those properties.

We may highlight the points being made here by asking "Where would biologists have got to if they had tried to describe all the variance between species in terms of 1, 2, or 16 'variables'?" "Where would they have got to if they had tried to summarize the variance in which those animals lived in terms of 10 or 12 'variables' and then sought to discuss the interactions between the species and their environments by running regression-based 'heritability' studies?"

The first step to responding to the challenge which Flynn has set us thus lies in putting our psychometric house and scientific paradigm in order. Unfortunately, although we have tried to take some small steps toward doing this in our *Competence in the Learning Society* (Raven & Stephenson, 2001) we are about to see that this is only the beginning.

5. The uses to which test scores are put, the legitimacy of that use, and the quest to promote humane ideals through meritocracy

An Aside

It is more than likely that material included in this section will make many readers feel uncomfortable. Many will claim that what is said is pure conjecture, lacking scientific support. But this is to misunderstand the nature of science. The observations which will be made are supported by evidence. The problem is that the reality to which the evidence points is not directly visible. This disturbs many would-be social scientists because the sociologically-functional image of science they have acquired (i.e. the reductionist model) tells them that things should be otherwise. Yet most science proceeds – as House (1991) has noted – by piecing together a network of observations to forge a picture of a wider hidden reality. One does not observe "glaciation" across Europe. One observes hanging valleys, terminal moraines etc. and infers "glaciation". We arrived at the picture of the hidden sociological reality outlined below by piecing together a range of psychological data. Having understood that reality, the task of dealing with it requires developments which are quintessentially psychological.

Back to the point

The absence of concepts for thinking about, fostering, and assessing diversity does not arise simply from psychologists' incompetence, using that term in its narrower sense. To glimpse the wider constraints we may note that, in the so-called "educational" system, there have been numerous attempts to offer something which might (because the word "education" comes from the Latin root "educere" – which means "to draw out") actually merit the name "education". That is, there have been numerous attempts to devise ways of catering for diversity – of nurturing multiple talents. Some teachers become aware that many of their pupils have talents which are not recognized by the system. Many become so deeply disturbed by this observation that they leave the profession. Yet programs designed to nurture multiple high-level competencies have been

deliberately (Graham & Tyler, 1993) eliminated through the imposition of such things as national curricula and nationwide achievement testing. These national curricula, attainment tests, and "performance-based, outcome-oriented" evaluations of competency-development programs are not just concerned with dumbing-down education to focus on common low-level content, assessing the outcomes in single-factor terms, and discrediting – even rendering invisible – other abilities. They also teach students that others have the right to decide what they will do, to assess them, place them, and orchestrate their lives – indeed that they have the right to evaluate their worth on their – the assessors' – terms (Raven, 1994). The ostensible reason for preventing students embarking on self-determined programs of personal development is that it is necessary to know the outcomes to be achieved before one can legitimately allocate public money to the activity. But more important, covert, reasons include the fear of what people might discover if they were allowed to investigate, to follow where their enquiries lead, and to transform themselves and their society in the process. As Jackson (1986) has demonstrated, one cannot even allow teachers to create developmental environments because one cannot know beforehand what the outcome is going to be. Robinson (1983) has documented the vehemence of the campaign orchestrated by the National Association of Manufacturers to discredit the books of Harold Rugg because they were marginally successful in leading children to develop the motivation and the abilities required to investigate the way society works and their role in it. Mrs. Thatcher hit the nail on the head when she announced that such programs were a left-over from the 1960s, that pupils had to be taught their place, and that the National Curriculum and associated assessment process was the way to do it.

What I am arguing is that there are both deep-seated sociological and vested-interest based reasons for emphasizing a single-factor concept of "ability". The more arbitrary the methods used to assess it the better – because it is not intellectual ability which is being assessed but instead such things as gullibility, the willingness to accept double talk and double think, the willingness to accept and echo "authority", the willingness to do whatever is necessary to secure one's own advancement, and the willingness to accept surveillance by big brother and follow his directives whatever their ethics.

It is not surprising, then, that practitioners (who are, after all, to be counted among those who have been selected, advanced, and promoted through the system just described) do not (as Flynn notes) care if the measures they use do not predict what they are said to predict: If everyone but a few mavericks (e.g., Gould, 1996) says they're OK, they're OK. It is not, in reality, as strange as Flynn thinks that practitioners accept without comment dramatic over-night re-classifications of their students and the differential treatments that that implies: If authority says that's what's to be done, that's what's to be done. After all, there is very little connection between such things as assignment to special education classrooms, the IEPs that are supposedly implemented in those classrooms, and the touted benefits of those programs. Given that all

students are already being roundly abused by the so-called "educational" system, given that so much injustice is being to all students – and to "special education" students in particular – what is the point of objecting to one more illogicality? After all haven't we been taught that ethics belongs to the domain of the citizen, not the professional, still less the scientist?

Our next question must be why the myth of the efficiency of hierarchy has been so widely touted and promoted when its not supported by a single strand of evidence and all the available evidence, such as that provided by Kanter and summarized earlier, points to the opposite conclusion? The answer is that the supposed benefits of hierarchy and the selection and promotion practices associated with it serve to legitimize and reproduce a social order which gives pride of place to dominators. More than that, the hierarchical social structure of society legitimizes crushing differentials in power, prestige, respect, and income which threaten those who do not join the system with dire consequences and, in this way, induce most people to compete for advancement according to criteria set by authority despite the effect that so doing has on the quality of their lives. It also induces them to undertake endless activities which they know to be wrong: To go to war and to produce and sell junk foods, junk toys, junk insurance, junk bonds, and junk "defense" systems; to each day commit hundreds of unethical acts.

Shiva (1998) has argued that the processes described above are part of a still deeper network. She notes that the promotion of monocultures of mind (in both education and in the range of scientific perspectives [theories] that are deemed acceptable) seems to be somehow linked to the promotion of monocultures in society, to the promotion of monocultures in agriculture, and to the unthinking acceptance and promotion of reductionist science. The latter permits and encourages, nay requires, scientists to focus on single variables – such as the relationship between short-term crop yields in response to a particular artificial fertiliser – and to ignore longer term effects on yield, never mind effects on the fertility of the soil, the food chain, the health of those who eat the "foodstuffs", and ecology more generally. The effect is to promote a vision of science which is both deeply unethical (in the sense that it results in the destruction of the habitat our species requires for its survival) and lacking any form of objectivity worth the name.[2]

To summarize: the hegemony of a vastly over-generalized concept of "ability" arises in part because of the function it plays in a network of socio-cybernetic forces which not only perpetuate the status quo but contribute to the process whereby our species heads itself toward its own self-extermination, carrying all known life with it. It is one cornerstone in a network of myths (which include those associated with the concept of hierarchy) which legitimize, support, and perpetuate a culture of dominance. It arises in part because of the hegemony of established psychometric theories which not only render alternative abilities invisible but legitimize and perpetuate the

[2] The first draft of an article entitled *Engineered Invisibility and the Destruction of Life* is available from the author.

authority and power of established position holders in psychology. Research proposals based on alternative perspectives cannot attract funding. Lacking research support, perspectives which do not fit into the way of thinking that currently dominates science become increasingly unthinkable.

Worst of all, these hand-in-glove supports operate to perpetuate in power, not "able" people, but those who are most concerned with their own advancement (cf. Hogan, 1990, 1991) and least interested in the well-being and continuance of the species. Not only do they promote those who are least ethical, they legitimize an individualistic concept of ethics which emphasizes individual morality when what is actually needed is an ethical framework which enjoins pro-active, collective, system-oriented, action to overcome social pressures which force us all to act individually in highly unethical ways which we cannot, by taking individual moral decisions, escape.

It follows from these observations, taken singly and together, that, even if we had the concepts and tools which would make it possible for practitioners – teachers, managers, and selection, placement, and development agencies – to implement procedures based on a multiple-talent framework of abilities, they would not be used.

Conclusion

It would appear from the material we have reviewed that we have uncovered one reason why Flynn has been able to get his proposal to pursue humane ideals through meritocracy published in one of the highest-impact journals in psychology while we have been unable to publish proposals for so-doing by re-conceptualizing the nature of talent and nurturing and rewarding diverse contributions in much less prestigious places. Flynn's proposal challenges neither the sociological functions performed by the system nor any vested interests. On the contrary, it unquestioningly supports and reinforces hierarchical models of ability, motivation, and organizational and political systems.

Unfortunately, the stakes are much higher than those involved in resolving an academic argument about the means to be adopted to advance humane ideals. They involve the very survival of the species and the planet. The personal dilemmas to be confronted amount to extreme versions of the classic moral/ethical dilemma of acting in the long-term interests of the species when these interests are sharply in conflict with short-term personal consequences. The problem is exacerbated by the fact that a *personal* stance against the system is useless: the system would crush such stances taken individually or even by masses. The ethical imperative is therefore to induce pro-active collective action guided by new insights (of the very kind we have been under pressure to eliminate from this response) into the systems processes which constrain and distort social action. It can only be taken in the context of new insights into the societal learning and management arrangements (of the kind discussed in Raven, 1995 and Raven & Stephenson, 2001) which are required to manage society in the long-term *public* interest as distinct from the short-term interests of dominators.

We need to think seriously about ways of improving the quality of life of all. We need to find ways of distinguishing between important and unimportant work. Examples of the former include quality of life enhancing activities while the latter include the previously mentioned production and marketing of junk food, junk toys, and abusive education. We need to make better arrangements for getting important work done in ways which offer better quality working life. As Jaques (1989) has urged, we need to distinguish between ways of getting work done and ways of legitimizing pay differentials and reinforcing hierarchy. We need to find effective ways of reducing the differentials which compel participation in the frenetic activity which leads relentlessly toward the destruction of the species and the planet. We need to find ways of investigating and intervening in the hidden systems processes which compel us all to do things we know to be wrong and which are, in fact, incompetent, unethical, and unhealthy (in the sense that they are relentlessly driving our species toward its death). Yes, Jim, we need to exercise our collective intelligence to understand and influence processes which were previously beyond the grasp of the human intellect. If there has not yet been evidence of benefits commensurate with the gains you have documented, they had better come along quick!

In conclusion, the route to the advancement of humane ideals lies, not through the discreditation of measures of eductive ability, but through broadening the range of talents that we are able to adequately conceptualize, measure, develop, utilize, and reward. More fundamentally, it is to be found, not through devising a just meritocracy, but through discrediting faith in the effectiveness of hierarchy and the concept of “ability” which sustains it.

If we are to find ways of thinking about, assessing, developing and utilizing multiple talents, it will be necessary to change the way we think about, present, and enact our roles as scientists, educators, and social reformers. If we are to enact humane ideals it will be necessary to identify and develop social arrangements which enable us to introduce more intelligent and humane policies which will secure our future as a species. To do this, it is urgent that we find ways of controlling, not only the dominators who are heading us toward our self-destruction, but the pervasive Western dominance mindset which lies behind this. In other words, we have scientific, competence, and ethical obligations as psychologists to think out and introduce new forms of public management: new forms of democracy and bureaucracy – that is to say the institutional arrangements – the key aspects of the environment which so much determine our behavior. We need new job descriptions enjoining ourselves, our colleagues, our public servants, and our politicians to take innovative action in the long-term public interest. And we need new staff appraisal systems to find out whether both ourselves and our colleagues are doing these things. All of these are quintessentially jobs for psychologists. Flynn has a point: Ethically, scientifically, and socially we have not been living up to the standards which society could reasonably expect from us.

References

Barrett, G.V. and Depinet, R.L. (1991). A reconsideration of testing for competence rather than intelligence. *American Psychologist, 46(10),* 1012–1024.

Carroll, J.B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies.* Cambridge: Cambridge University Press.

Cattell, R.B. and Horn, J.L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational-Measurement, 15(3),* 139–164.

Deary, I.J. (1995). Auditory inspection time and intelligence: What is the direction of causation? *Development Psychology, 31*, 237–250.

Deary, I. J. (2000). *Looking down on Human Intelligence*. Oxford: Oxford University Press.

Edwards, A.L. (1954). *Edwards' Personal Preference Schedule*. New York: Psychological Corporation.

Flanagan, J.C. (1978). *Perspectives on Improving Education from a Study of 10,000 30-Year-Olds*. New York: Praeger Publishers.

Flynn, J.R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist, 54(1),* 5–20.

Flynn, J.R. (2000). *How to Defend Humane Ideals*. Nebraska: University of Nebraska Press.

Goodlad, J. (1983). *A Place Called School*. New York: McGraw Hill.

Gould, S. J. (1996). *The Mismeasure of Man*. New York: W. W. Norton.

Graham, D. and Tyler, D. (1993). *A Lesson for Us All: The Making of the National Curriculum*. London: Routledge.

Greeley, A.M. (1972). *That Most Distressful Nation.* Chicago: Quadrangle Books.

Herrnstein, R.J. and Murray, C. (1994). *The Bell Curve*. London: MacMillan.

Hogan, R. (1990). Unmasking incompetent managers. *Insight,* May 21, 42–44.

Hogan, R. (1991). An *Alternative Model of Managerial Effectiveness*. Mimeo: Tulsa, OK: Institute of Behavioral Sciences.

House, E.R. (1991). Realism in research. *Educational Researcher, 20,* 2–9.

Hope, K. (1984). *As Others See Us: Schooling and Social Mobility in Scotland and the United States*. New York: Cambridge University Press.

Horn, J.L. (1994). Theory of fluid and crystallized intelligence. In R.J. Sternberg (Ed.), *Encyclopaedia of Human Intelligence* (443–451). New York: Macmillan.

Huff, S., Lake, D. and Schaalman, M.L. (1982). *Principal Differences: Excellence in School Leadership and Management.* Boston: McBer and Co.

Jackson, P.W. (1986). *The Practice of Teaching*. New York: Teachers College Press.

Jaques, E. (1976). *A General Theory of Bureaucracy*. London: Heinemann.

Jaques, E. (1989). *Requisite Organization*. Arlington, VA: Cason Hall and Co.

Jensen, A.R. (1998). *The **g** Factor: The Science of Mental Ability*. Westport, CN: Praeger.

Kanter, R.M. (1985). *The Change Masters: Corporate Entrepreneurs at Work*. Hemel Hempstead: Unwin Paperbacks.

Klemp, G.O., Munger, M.T. and Spencer, L.M. (1977). *An Analysis of Leadership and Management Competencies of Commissioned and Non-Commissioned Naval Officers in the Pacific and Atlantic Fleets*. Boston: McBer.

Kohn, M.L. and Schooler, C. (1978). The reciprocal effects of the substantive complexity of work and intellectual flexibility: A longitudinal assessment. *American Journal of* Sociology, 84, 24–52.

Kuhn, T.S. (1962/1970). *The Structure of Scientific Revolutions* (Second Edition). Chicago: University of Chicago Press.

McClelland, D.C. (1973). Testing for competence rather than for "intelligence". *American Psychologist, 28,* 1–14.

McClelland, D.C, Atkinson, J.W., Clark, R.A. and Lowell, E.L. (1958). Appendix. In J.W. Atkinson (Ed.), *Motives in Fantasy, Action and Society*. New York: Van Nostrand.

McClelland, D.C., Baldwin, A.L., Bronfenbrenner, U. and Strodtbeck, F.L. (1958). *Talent and Society*. Princeton, NJ: Van Nostrand.

Matarazzo, J. D. (1990). Psychological assessment versus psychological testing. *American Psychologist, 45,* 999–1017.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18(2),* 5–11.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50(9),* 741–749.

Neisser, U. (Ed.) (1998). *The Rising Curve.* Washington, DC: American Psychological Association.

Owens, W.A. (1966). Age and mental abilities: A second adult follow-up. *Journal of Educational Psychology, 57,* 311–325.

Raven, J. (1984/1997). *Competence in Modern Society: Its Identification, Development and Release*. Unionville, New York: Royal Fireworks Press (1997); Oxford, England: Oxford Psychologists Press (1984).

Raven, J. (1987). Values, diversity and cognitive development. *Teachers College Record, 89,* 21–38

Raven, J. (1991). *The Tragic Illusion: Educational Testing*. New York: Trillium Press; Oxford, England: Oxford Psychologists Press.

Raven, J. (1994). *Managing Education for Effective Schooling: The Most Important Problem is to Come to Terms with Values*. Unionville, New York: Trillium Press; Oxford, England: Oxford Psychologists Press.

Raven, J. (1995). *The New Wealth of Nations: A New Enquiry into the Nature and Origins of the Wealth of Nations and the Societal Learning Arrangements Needed for a Sustainable Society*. Unionville, New York: Royal Fireworks Press; Sudbury, Suffolk: Bloomfield Books.

Raven, J. (2000). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No.3 (Third Edition): A Compendium of International and North American Normative and Validity Studies Together with a Review of the Use of the RPM in Neuropsychological Assessment by Court and Drebing*. Oxford, England: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41*, 1–48.

Raven, J., Raven, J.C. and Court, J.H. (1998a). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General Overview*. Oxford, England: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.

Raven, J., Raven, J.C. and Court, J.H. (1998b). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: The Coloured Progressive Matrices*. Oxford, England: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.

Raven, J., Raven, J.C. and Court, J.H. (1998c, 2000). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3: The Standard Progressive Matrices*. Oxford, England: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.

Raven, J., Raven, J.C. and Court, J.H. (1998d). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4: The Advanced Progressive Matrices*. Oxford, England: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.

Raven, J., Raven, J.C. and Court, J.H. (1998e). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 5: The Mill Hill Vocabulary Scale*. Oxford, England: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.

Raven, J. and Stephenson, J. (Eds.). (2001). *Competence in the Learning Society*. New York: Peter Lang.

Raven, J.C. (1936). *Mental Tests Used in Genetic Studies: The Performances of Related Individuals in Tests Mainly Educative and Mainly Reproductive*. M.Sc. Thesis, University of London.

Robinson, D.W. (1983). *Patriotism and Economic Control: The Censure of Harold Rugg*. D.Ed.

Dissertation, Rutgers University, New Jersey. Also available in University Microfilms International, Ann Arbor, Michigan (1984).

Russ-Eft, D. & Brennan, K. (2001). Leadership Competencies: A Study of Leaders at Every Level in an Organization (Chapter 7). In J. Raven & J. Stephenson (Eds.), *Competence in the Learning Society*. New York: Peter Lang.

Schaie, K.W. (Ed.). (1983). *Longitudinal Studies of Adult Psychological Development*. New York: Guilford Press.

Schaie, K.W. (1994). The course of adult intellectual development. *American Psychologist, 49(4),* 304–313.

Schaie, K.W. and Willis, S.L. (1986). *Adult Development and Ageing* (2nd edition). Boston: Little Brown.

Schmidt, F.L. and Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124(2),* 262–274.

Schon, D. (1973). *Beyond the Stable State*. London: Penguin.

Shiva, V. (1998). *Biopiracy: The Plunder of Nature and Knowledge.* London: Green Books.

Snow, R.E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher, 18(9),* 8–14.

Snow, R.E., Kyllonen, P.C. and Marshalek, B. (1984). The topography of ability and learning correlations. In R.J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence,* Volume 2, (47–103). Hillsdale, NJ: Lawrence Erlbaum Associates.

Spearman, C. (1927). *The Nature of "Intelligence" and the Principles of Cognition* (Second Edition). London, England: MacMillan.

Spearman, C. (1927). *The Abilities of Man*. London, England: MacMillan.

Spencer, L.M. and Spencer, S.M. (1993). *Competence at Work*. New York: Wiley.

Sperry, R. (1983). *Science and Moral Priority: Merging Mind, Brain, and Human Values*. Oxford: Blackwell.

Styles, I. and Andrich, D. (1997). Faire le lien entre variables psychometriques et variables cognitivo-developpementales regissant le fonctionnement intellectuel. *Psychologie et Psychometrie, 18*(2/3), 51–69.

Thorndike, R.L. (1977). Causation of Binet IQ decrements. *Journal of Educational Measurement, 14,* 197–202.

Tough, J. (1973). *Focus on Meaning: Talking with Some Purpose to Young Children*. London: Allen and Unwin.

Trevarthen, C. (1990). Growth and education of the hemispheres. In C. Trevarthen (Ed.), *Brain, Circuits and Functions of the Mind: Essays in Honor of Roger W. Sperry*. Cambridge, England: Cambridge University Press.

Vodegel-Matzen, L.B.L. (1994). *Performance on Raven's Progressive Matrices*. Ph.D. Thesis, University of Amsterdam.