Chapter 20

Psychometrics, Cognitive Ability, and Occupational Performance*

John Raven

Overview

In two previous articles (Raven, 1989b, 2000), I reviewed studies suggesting that, contrary to what Flynn (1987) would have us believe, the *Raven Progressive Matrices* measures psychological abilities of fundamental importance, and that the steadiness in the improvement in these abilities over time and the similarity in the norms obtained in many – but not all – cultures at any point in time reinforce this conclusion.

In this article I will summarise remarkable new evidence that the *Raven Progressive Matrices* is measuring an important aspect of cognitive functioning. Thereafter, I will return to the question of the extent to which it measures "intelligence" (and competence more generally). This will lead to a re-examination of the test's construct validity. This discussion has important practical implications because it underlines the need to situate eductive ability scores in the context of a yet-to-be-developed framework for thinking about the wider aspects of intelligence and competence. At the same time, it raises serious questions about the way we think about the procedures to be used to establish the validity of a test and the ethics of insufficiently *comprehensive* assessment – however *in*valid some of the necessary assessments may be. The article concludes by outlining some of the parameters which must be satisfied in seeking to develop a better framework for thinking about competence and its assessment.

***

Raven (1989b) argued that the reproducibility of the psychometric properties of the RPM across different socio-economic and ethnic groups,

* Earlier versions of this article were published in S. M. Wechsler, & F. S. R. Guzzo, (1999) Psychological Assessment: International Perspectives, Sao Paulo, Brazil: Casa do Psicologo, and in the (Croatian) Review of Psychology, Vol. 7, (2000).

the regularity in the increase in scores over time, and the similarity in the norms obtained in many different cultures at any point in time all suggested that the RPM measures something of fundamental psychological importance. This theme was developed further in Raven (2000) and Raven, Raven and Court (1998, updated 2003; 2000, updated 2004), where the increases over time and the new tests developed to restore their discriminative power are discussed more fully. In this section, I will muster evidence suggesting that the RPM directly taps one important aspect of psychological functioning and that this is what most psychologists refer to as "cognitive functioning". Later I will argue that, although this is indeed the case, this label misleads – for what is generally regarded as "cognitive" functioning is primarily affective and conative. It is therefore more appropriate to claim that the RPM measures "eductive" ability – at least in relation to one potentially valued set of activities.

That the RPM measures, and reveals something about, basic cognitive functioning actually follows from the application of Item Response Theory in its construction.

Item Response Theory (IRT) was developed in Britain in the early 1930s, used in the development of the RPM, translated into mathematical formulae by Rasch in the early 1940s (in the course of which he specifically tested his formulations by showing that they applied to the RPM [Rasch, 1980]), and popularised in the US and elsewhere by Wright and others (e.g. Wright and Panchapakesan, 1969) in the 1960s.

To establish the internal consistency of the RPM, graphs (Item Characteristic Curves, ICCs) were plotted (Raven, J.C., 1939) to show the way in which the probability of solving any one item related both to the probability of solving every other item and total score. To the extent that these graphs reveal that the probability of solving any one item does indeed increase in step with the probability of solving easier and more difficult items, it shows that, whatever the manifest content of the items, they are tapping some common underlying continuum.

Given that the manifest content of the items changes from simple perceptual ("Gestalt") items, through easy analogies, to complex analogies which seem to require considerable "analysis" to discern and isolate the "relevant" elements, this shows that "perception" and "reasoning" form part of the same psychological continuum. Put the other way round, perception is not an immediate, visual, process but involves *conceptual* activity. Such activity is required to discriminate figure from ground and relevant from irrelevant. It is therefore a mistake to regard the RPM as

a measure of "problem-solving ability" since, as Spearman (1927) was at pains to emphasise in his principles of noegenesis*, the capacity to identify and handle problems depends on *simultaneously* developing an understanding of the whole in order to know what to look for in the parts (i.e. in order to "analyse") *and* knowing which parts to discriminate from background "noise" in order to "see" the whole.

It follows from what has been said that the item analysis – the set of Item Characteristic Curves – for the RPM demonstrates (a) that something which might tentatively be named "general conceptual ability" does "exist", (b) that the RPM in some sense measures this ability, and (c) that the qualitatively different items of which the test is composed form part of a common continuum. These qualitatively different types of item are not measuring "different things". Just as the concept of "hardness" is not negated by the fact that it is different substances which display different degrees of the characteristic, so the fact that the items in the RPM differ in their manifest content does not invalidate the notion that their solution demands different levels of "cognitive ability". The ability to solve one type of item increases incrementally and in step with the ability to solve other types. There are no metamorphoses in thinking between the ability to solve one kind of item and the next (although this does not imply that there are not spurts and plateaux in individual development).

At this point it is necessary to counter the objection that factor analysts have isolated separate factors made up of these "perceptual", "reasoning", and "analytic" items. I have shown elsewhere (Raven, Ritchie & Baxter, 1971) that the correlation matrix obtained by intercorrelating the items of a perfect Guttman or Rasch scale can be fitted by neither a principal components analysis nor by any orthogonal or oblique rotation. The nature of the correlation matrix is determined by the properties of such scales. A knowledge of whether someone gets a very easy item right does not enable one to predict whether they will get a difficult item right. The correlation between very easy and very difficult items therefore tends to zero. On the other hand, items of similar difficulty are highly correlated: A knowledge of whether someone gets one item right or wrong is a good predictor of whether he or she will get the next most difficult one right or wrong. The correlation matrix obtained by intercorrelating the items after they have been arranged in order of difficulty thus has correlations tending toward unity around the diagonal and approaching zero in the

* The word noegenesis derives from the Greek word *noetic*, and thus means "mind creation".

distal corners. This correlation matrix cannot be re-created by multiplying and adding loadings on *any* set of factors smaller in number than the original items. If one forces data of this kind into a factor analysis one gets a series of "power" factors. These are made up of items of "similar" difficulty because adjacent items intercorrelate highly. (The average within-factor correlation is determined by the number of factors extracted.) But now comes the misinterpretation. Items of similar difficulty consist predominantly, though not exclusively, of items of the same manifest type. In fact, the factors contain some – in reality easier – items from the qualitatively different type which comes developmentally earlier, and some – in reality more difficult – items from that which comes developmentally later than the bulk of the items in the cluster. But these "non-conforming" items can easily be overlooked when naming the factor. Researchers have tended to name these factors to reflect their dominant manifest content when they are, in reality, power factors.

We can now return to our conclusion that the IRT-based item analysis of the RPM really does show that there is a *continuum* in "cognitive (actually 'conceptual') ability" and that this continuum can be assessed using a range of items running from easy "perceptual" items to difficult "analytic" ones. It involves the ability to discriminate figure from ground; the ability to discern order and meaning in (or make meaning out of) confusion; the ability to form high-level, usually non-verbal, concepts which enable one to make sense of the environment. Spearman used the Latin word *educere* – to draw out – to characterise and discuss this component of General Intelligence – ***g*** – and contrasted it with *reproductive* ability, the ability to reproduce already verbalised knowledge.

The conclusion that something which might be termed "general conceptual ability" or "eductive" ability "exists" has been reinforced, and its generalisabilty underlined, in a number of studies in which the RPM has been correlated with other tests.

However, both its existence and generalisability have been neatly confirmed in the study reported by Styles in an earlier chapter in this volume and in Styles (1999). This study, like the scaling procedures used in the development of the tests, was grounded in IRT.

Using a mathematical formulation of IRT, Styles mapped the levels of thought revealed by three Piagetian tasks – the Balance, Chemical Combinations, and Correlational tasks – onto the set of RPM ICCs.

What emerged was that the (Piagetian) level of answers given to these problems increases gradually and incrementally in step with the ability

to solve RPM problems of similar difficulty. It is again clear, therefore, that the ability to solve qualitatively different types of problem develops progressively and does not emerge from the kind of metamorphosis which has sometimes been said to lie behind development of the ability to solve the more complex Piagetian problems.

Styles and Andrich's study not only provides a further demonstration that the RPM is tapping a continuum of fundamental psychological importance, it also indicates that whatever is being measured cannot be dismissed as an ability of mere academic interest. It follows from their work that RPM scores reflect the ability to solve complex, "real-life", problems of an apparently very different character.

A quite different line of work showing that the RPM taps basic psychological abilities comes from researchers working with Reaction-time and Inspection-time (Jensen, Larson & Paul, 1988; Vernon, 1989, 1991, 1993; Deary, 1993, 1995).

Vernon and Deary independently concluded from their reviews of the work of a number of researchers that the RPM has significant, but not strong, correlations with:

- *Low* cerebral glucose metabolic rate. That is, those who get high scores on the RPM appear to work more efficiently.
- *Some* – but far from all – measures of "reaction-time". (Many measures of "reaction-time" do *not* correlate with the RPM, and the studies which have reported the highest correlations have included speed of response to difficult "IQ-type" questions among the measures composited. No *general* statement to the effect that the RPM and "reaction-time" measure the "same thing" is, therefore, justified.)
- *High* cortical response (averaged evoked potential) as measured by EEG, to unanticipated visual stimuli – but low cortical response to self-administered stimuli.
- Blood calcium level, which is itself associated with neural conductivity.

Most impressive, however, is Deary's demonstration that the RPM – and eductive ability more generally – is more highly correlated – at the level of about .6 – with the amount of time people require to be 85% accurate in perceiving which of two lines of very different length is the longer, i.e. "Inspection-time". (Deary makes a point of emphasising that the measurement of Inspection-time does NOT require the person being assessed to work at speed. Inspection-time is measured by varying the

amount of time the lines are exposed and finding the point at which respondents are unable to discriminate accurately between them.)

Inspection-time, like eductive ability, and unlike reproductive ability, "declines" with age ... i.e. has gone up with date of birth.

There is one other recent study which deserves to be singled out for mention here. Carpenter, Just, and Shell (1990) reported that 95% of a sub-set of verbally encoded *Advanced Progressive Matrices* items could be solved by a computer programme which was required to check for the presence or absence of only five rules which might govern the orderliness of the matrix and which, if present, collectively determine the characteristics of the element required to complete the pattern.

Progressive Matrices items were not constructed with a view to analysing the "problem-solving" strategies employed by respondents. As a result, the items of the classical series often have features which make it difficult to identify the operation of the rules Carpenter et al. sought to study. Likewise, the set of options from among which the correct answer has to be selected were not constructed in such a way as to make it possible to test hypotheses derivable from their theory about how the kinds of error which are made should relate to total score (although J.C.Raven did in fact find that type of error was directly related to total score).

Vodegel-Matzen (1994a&b) constructed a set of items which (a) contained no features extraneous to Carpenter and Just's framework that might influence their difficulty, (b) had theoretical difficulty levels which could be calculated from that theory, and (c) had distracters which differed systematically from the correct answer only in terms of the number and kind of rules omitted (and the probability of selection of each of which could therefore be expected, on theoretical grounds, to vary systematically with total score).

This new test had both excellent internal psychometric properties and a very high correlation with the RPM.

The results of the error analysis were as predicted. The most able of those who were unable to solve a given item selected answers which failed to take account of a single – the most difficult – rule governing the orderliness of the matrix. Less able respondents overlooked more rules. Thus the *type* of error made varied with total score in the way predicted by the theory. The finding thus gives us new insights into the causes of deficiencies in cognitive functioning.

Another factor determining item difficulty is the ease with which it is possible to identify the elements to which attention needs to be paid when trying to identify systematic variation between the cells of the matrix.

By making the elements of which the matrices were composed more "life-like" – i.e. using such things as hats, bananas, and faces instead of squares and triangles – while requiring respondents to apply the *same* rules in order to solve the problems, Vodegel-Matzen demonstrated just how important this factor really is. But what was most interesting was that the change to more life-like components made the items easier *for everyone* – not just for lower-scoring respondents. The rank-order of items and respondents remained virtually unchanged.

Use of pictorial elements may, however, result in cultural differences which are not found on the original test.

Going the other way – "hiding" the features that it is important to attend to – making "correspondence-finding" more difficult – makes the items harder for everyone. But it makes little difference to the order of difficulty of the items or the rank-order of respondents.

Moving away from specific studies to programmes of research, remarkable confirmation of the appropriateness of the eductive-reproductive framework for thinking about these abilities has come from an entirely unexpected quarter. Cattell (1963) and Horn and Cattell (1966) initiated a stream of research in the area by proposing that the basic distinction was between "fluid" and "crystallized" intelligence and further suggesting that the latter "differentiated out of" the former. Whereas Spearman argued that the natures of the two abilities were "trenchantly contrasting", Cattell and Horn viewed them as closely related and expected to find that they had a common neurological substrate which, they hoped, would be illuminated by research using "more fundamental" psychophysical and psychophysiological measures.

On reviewing the available material for Sternberg's (1993) encyclopaedia, however, Horn came to a series of conclusions which support Spearman's standpoint in virtually every detail: (1) the thousands of "more fundamental" measures which have been developed do *not* cluster into the eductive vs reproductive domains but generate some eight *additional* factors or components of "intelligence"; (2) none of these additional factors has anything like the explanatory power of eductive and reproductive ability; (3) "crystallized" intelligence does *not* "differentiate out" of "fluid" intelligence; the two are distinct from the beginning; (4) the two abilities have different genetic origins; (5) the two are affected by different aspects of the environment; (6) the two follow different developmental trajectories over the life cycle; and (7) the two predict very different types of performance.

Given this remarkable convergence between what were very different positions, it remains only to ask which terminology seems most appropriate and to suggest that the eductive-reproductive formulation seems less likely to mislead.

Construct Validity: A Measure of Problem-Solving Ability or a Measure of Intelligence?

The RPM was constructed neither as a measure of "problem solving ability" nor as a measure of "intelligence". Nevertheless, many researchers have treated it as if it were a measure of one or the other. In the next few paragraphs the conceptual difficulties involved in doing this will be discussed.

Problem-solving is a difficult and demanding activity. It requires people to be sensitive to fleeting feelings on the fringe of consciousness which indicate that something could be done better or merits exploration. It involves initiating, usually on the basis of "hunches" or feelings, "experimental interactions with the environment" to clarify the nature of a problem and potential solutions. Having used their feelings to initiate activity, people need to monitor the effectiveness of their actions in order find out what is working and what is not, and why. In this way they can learn more (not necessarily consciously) about the nature of the problem and the effectiveness of their strategies. They can then modify their behaviour and launch a further round of "experimental interactions with the environment".

Beyond what may be regarded as *process* components of problem-solving lie a set of social and personal beliefs – beliefs about society, how it works, and one's own place in it. These include the belief that one has a *right to* ask questions and to do such things as try to influence the way society works.

And, in addition to these internal components, effective problem-solving often also involves persuading other people to help, prising information out of other people's heads, and learning how to do things by imitating others.

It is important to note that all this implies that what is often thought of as "cognitive activity" is primarily affective, conative, and interpersonal. Without the use of feelings there would be no insights; without persistence (conation) there would be no testing of those insights; and without

actual behaviour (experimental interactions with the environment or "conversations with the problem") followed by feeling-based monitoring of the effects of that action, there would be a major failure in observation, "thinking", and learning.

It follows that it is not legitimate, except for purely conceptual purposes, to try to separate the cognitive elements of eductive activity from its other components. The process which is commonly described as "cognitive" is necessarily *primarily* dependent on affective, conative, and social processes. The attempt to develop "pure" measures of cognition is doomed to fail because the very basis of the attempt seeks to eliminate the processes on which effective cognition is most dependent. We will come back to the question of how effectively the RPM copes with these problems when we have completed our theoretical discussion.

No one is going to undertake any of the activities discussed in the last paragraph unless they, in some sense, care about the activity. It is difficult to formulate this statement more precisely because the kinds of things people are strongly motivated to do often seem to have much in common with compulsions. People do the things they are strongly motivated to do persistently and repeatedly despite punishment, despite their better judgement, indeed, "despite themselves". Of course, that is a circular statement.

The goals or contexts in relation to which people will undertake difficult and demanding activities like "problem-solving" vary enormously. One person will, for example, engage in them mainly while trying to advance scientific understanding. Another while trying to put others at ease. Another in the course of trying to advance him or herself in a career. And yet another while seeking to control others.

An important implication of this observation is that people will only develop important components of cognitive competence while they are working at tasks they are intrinsically strongly motivated to undertake. This point will not be developed here. A discussion will be found in Raven (1987).

The implication for assessment is that people's ability to carry out the kinds of activity that are needed to identify and solve problems is only likely to reveal itself when they are undertaking activities that are important to them.

This conclusion resonates with the views of "situated cognitionists" like Greeno (1989) and Brown, Collins and Dugoid (1989). However, our conclusion differs from theirs in that we are arguing that *the same*

psychological processes **do** occur in different contexts. They simply *look* different, just as copper looks and behaves differently when situated in the context of sulphur *and* oxygen as compared with, say, just oxygen.

What these observations mean is that the effective assessment of high-level competencies – including problem-solving ability – is dependent on the adoption of a *two-stage* measurement procedure. One must first find out what kinds of activity people find motivating (i.e. "engaging", "important", or in some other sense "valued"), and then, and only then, assess *how many* of the activities which make for effective "problem solving" they display while carrying out those activities.

People have too often been branded as "unable to think" simply because they do not "think" in a mathematics class or while undertaking tasks set by others in an Assessment Centre (or during a Piagetian experiment). Such people may be able think *very effectively* (i.e. make their own observations, learn without instruction, and make good judgements) on a football field, or when developing better materials for building the bridges which are to be assembled by the team in the course of leading which they are supposed to demonstrate their capacity to think in an Assessment Centre. The question which should be asked is, therefore, frequently *not* "How intelligent is this person?" but "*While undertaking which kinds of activity* does this person display his or her intelligence?" Only after that can one meaningfully ask: "Which of the competencies required for intelligent behaviour does he or she display in the course of these activities?"

This psychometric oversight has not only led to injurious and misleading assessments of individuals, it has also resulted in unjustifiable conclusions being drawn from research. These research conclusions have then often contributed to the introduction or perpetuation of damaging educational, occupational, and social practices. Insufficiently *comprehensive* assessments must be regarded as unethical: They have detrimental consequences for the individuals assessed and for others who would have benefited from the educational and other programmes which have been condemned. These detrimental consequences cumulate for society: Individuals who have been mis-assessed are often deprived of opportunities to contribute as they might to society and the cancellation of educational and social programmes which, in reality, have genuinely beneficial effects can have serious social consequences.

A series of seriously misleading "findings" arising from failure to employ appropriate measurement practices will be found in Raven

(1991). One of direct relevance to the deployment of the RPM is that cognitive development "plateaus" in adolescence. This conclusion stems from not having measured "intelligence" while those concerned were carrying out activities they cared about and in connection with which they had had opportunities to continue to develop their powers of reasoning. When the "ability to think" is assessed more appropriately, the available evidence suggests that it increases throughout life (Jaques, 1976, 1989; Kohn & Schooler, 1978).

Turning to the widely held view that the RPM measures "intelligence", one of the most fundamental difficulties is that qualities like "intelligence" and "enterprise" are, as Gardner (1987) and Deming (1993) have also argued, qualities which need to be studied and documented at *cultural*, rather than individual, levels. To advance understanding (i.e. to engage in intelligent activity) effectively, one really needs to proceed on a group basis. One needs a wide range of people who do very different things. Thus one requires some people who are good at each of the following: generating ideas; digging relevant information out of a diverse literature; getting people to work together effectively; discerning patterns in accumulating data; deciding what information to collected to test those insights; using their feelings to notice activities that are likely to succeed; capitalising upon whatever is discovered in the course of an "adventure" initiated on the basis of feelings or "hunches"; putting emerging understandings into words; communicating findings to others; and engaging in the political activities necessary to attract the funds needed to continue the work.

Empirical support for the central claim of the last paragraph comes from the work of Taylor, Smith, and Ghiselin (1963), who showed that effective advance of scientific understanding depends on having *teams* made up of people who are motivated to do very different things and who contribute in very different ways to the overall activity. In a similar vein, McClelland (1961) showed that enterprise and innovation stems from many people trying to do whatever they are doing in new ways. More generally, he found that what happens in a culture is primarily dependent on the shared values of the culture. Most important was whether its members would bring to bear multiple components of competence in order to undertake the kinds of activity they cared about effectively. Kanter (1985) has likewise shown that the innovativeness and survival of organisations depends on *everyone* contributing (through "parallel organisation" activity) in very different ways to a climate of innovation and improvement and on steps being taken to recognise and develop

their diverse talents. Based on his own observations, Deming (1993) has made a similar point. Roberts (1968) and Rogers (1962/83) have observed how innovativeness is dependent on "teamwork" and networks of contacts. Dalziel and Schoonover's (1988) research led them to a similar conclusion. Indeed, even Jaques (1989), while emphasising the need for a steep organisational hierarchy based on "cognitive ability", stresses the crucial importance of managers-once-removed devoting a considerable amount of time to thinking about the talents of subordinates-once-removed and how to place, develop, and utilise them.

"Eductive ability" contributes to the effective performance of each and every one of the activities mentioned.

These points can be reinforced, and additional implications highlighted, by reflecting on the way in which the word "intelligence" is used in phrases like "the (military) intelligence service". This reveals that, despite psychologists' inclination to adopt a reductionist definition of intelligence (as in the assertion that "intelligence is what intelligence tests measure"), what has actually happened is that psychologists have omitted from their measures a great deal that should have been there. They have done society a dis-service by leading parents, teachers, and managers to think that intelligence *tests* capture what they, as laymen, understand by the word "intelligence". They have led the members of these groups to overlook a great deal of what they should have been paying attention to as they sought to think about, nurture, and capitalise upon the talents of their children, pupils, or subordinates.

Generating new insights and understandings ("intelligence") through a military or industrial intelligence service clearly involves making sense of confusing and incomplete information. Intelligence officers frequently cannot know beforehand what to observe and report. They depend on their *feelings* ("intuition") and on recognising an emerging pattern to tell them what is significant. The qualities required to make sense of the incoming information include the ability to seek out, collate, re-interpret, and piece together scraps of unreliable and incomplete information in order to perceive something that has not been seen before and to use what is then perceived to tell them what to attend to and observe next and what to report. The qualities required to do well also include the ability to discern what further information would be required to test initial impressions and the determination to collect that information, perhaps through overt as well as mental "experiment".

What has been said so far amounts to nothing more than a statement that numerous components of what we have called "eductive" ability are

required to work intelligently. However, it also illustrates some of the things that have been missing from most previous attempts to assess eductive ability.

But much more is involved in intelligent activity. The qualities required to establish military intelligence also include the ability to prise information out of other people, the motivation and the ability to do such things as set up and manage networks of contacts to obtain information, the ability to make good judgements about who possesses the sensitivities and persistence to do well in the field, and the ability to supply those contacts with appropriate guidance concerning the kind of information to be sought.

The ability to carry out these activities clearly involves eductive ability. But it also involves many other motivational dispositions and abilities and the effective use of accumulated, specialist, knowledge of military operations, people, and systems.

It follows that, for a group to act intelligently, it is necessary to have a wide range of people who contribute in very different ways to establishing and running a network and who find ways of advancing a wide variety of activities. It is not possible for any one person to be motivated to carry out, and be good at performing, all the activities that are important.

Grid 20.1 has been prepared to make this way of thinking more concrete and to link what is being said here to the more general framework for reflecting about competence that will be summarised below and which has been published in full in *Competence in Modern Society* (Raven, 1984).

What has been said indicates that intelligent behaviour occurs when one has a range of people who are strongly motivated to carry out as many as possible of the activities listed down the left-hand side of the Grid and are capable of carrying them out in a co-ordinated way.

In the course of carrying out their chosen activities, each person needs to exercise as many as possible of the competencies listed across the top of the Grid (and others like them).

The first thing to be emphasised is that what is portrayed in this Grid is *not* a *culture* of intelligence. It is *intelligence* itself.

It follows that, while Jaques is right to emphasise the rarity of the motivation and the ability to carry out (and the societal and organisational importance of carrying out) organisational and societal management tasks involving such things as understanding and influencing opaque, international, socio-economic and socio-physical processes, his failure

to recognise at least some crucial components of the organisational arrangements required for intelligence and innovation has led him to some seriously misleading conclusions.

It also follows that, while Gardner is right to stress the importance of multiple talents, he may need to reformulate his theory of multiple intelligences. Our own theory suggests that the conclusion to be drawn from his observations is that there are many important activities which people may be strongly motivated to carry out and in relation to which they may develop and display high-level competencies. But, while there are also many more of these high-level competencies than psychologists have been inclined to acknowledge in the past, their number may still be relatively limited.

The framework developed here in some ways reinforces, but in other ways draws attention to limitations of, the observations of authors like Richardson (1991), Ogbu (1992), Tharp et al. (1984), and Gallimore (1985). These researchers argue that cognitive abilities will be revealed only when people are undertaking tasks which are meaningful and important to them and that their apparent ability to carry out these tasks depends on their prior opportunity to exercise, and thus develop, these abilities. Unfortunately, these authors mainly dwell on the *dominant* values of the cultural groups they studied and the kinds of "intelligence" evoked or called for in those contexts. They fail to note the variance in valued activities within all cultures. As a result they overlook two important things:

(1) That, if one is to nurture cognitive and other high-level competencies (in the way that effective parents nurture such qualities in their children and managers nurture them in their subordinates [Raven, 1980, 1984; Spencer & Spencer, 1993; Kanter, 1985]), it will be necessary to create individualised developmental programmes which engage with people's motives and thus enable them to practise (and thereby develop) these components of competence.

(2) That, if one is to measure eductive ability more effectively than in relation to a task which *most* people find inherently engaging (as in the RPM), one must first find out what kind of activity the person being assessed is strongly predisposed to undertake and then which of the components of competence that are needed to carry it out effectively are displayed while it is actually being carried out. (Instead of doing this, most critics of conventional measurement – such as Piaget, Vygotsky,

GRID 20.1

A MODEL OF INTELLIGENCE

Examples of Competencies Required to Carry out Activities Crucial to Intelligence (Observable only while activities which are personally engaging are being undertaken)

Examples of activities required to create a Culture of Intelligence or Enterprise but which people may or may not be strongly motivated to carry out.	Eductive ability (itself having cognitive, affective, and conative components and involving such things as the ability to initiate and learn from "experimental interactions with the environment").	Reproductive ability: The store of information and intellectual skills available from the past.	Ability to persist.	Ability to use feelings to initiate action, monitor the effects of the action, change one's behavior accordingly, and start a further cycle.	Ability to persuade others to help.	Ability to resolve value conflicts and to integrate values with each other and work toward their achievement over a long period of time.
Tendency to understand and influence the workings of society around the organization - including what is happening on the other side of the world.						
Tendency to generate new formal theories e.g. in connection with the workings of the organization or in connection with technology.						
Tendency to engage in organizational development activity.						
Tendency to notice new things that need to be done.						
Tendency to translate new theoretical understandings into a product.						
Tendency to get people to work together effectively.						
Tendency to think about, place, develop, and utilize the talents of subordinates.						
Concern to put others at ease.						
Tendency to soothe interpersonal tensions.						
Tendency to get together with others and set up indirect strategies to influence people higher up in the organization.						
Tendency to provide help and encouragement to those engaged in innovation.						
Tendency to initiate the collection of, seek out, sift, and come to good innovative decisions on the basis of forward-looking information						

and Richardson* – have simply confronted respondents with a set of problems geared to an alternative, but still single, value system.) The only researchers who have seriously addressed this problem are those who have worked in what may be termed the McClelland tradition.

We may now return to the question of whether, and how, the members of hierarchically differentiated occupational groups within organisations need to differ in motivation and eductive ability for the organisation to function most effectively. Reflection on Grid 20.1 suggests that it may be more important for people working at different levels in an occupational hierarchy to differ in the kinds of activity they are strongly motivated to carry out than in their eductive ability. As Hogan (1990) and Hope (1984) have shown, managers who apply their eductive ability mainly to advancing themselves in their careers (by, for example, getting rid of all personnel who are concerned with future development so as to present themselves as being able to run organisations which are "lean, mean, and profitable" in the short-term) can have disastrous effects on their organisations.

It is equally obvious from Grid 20.1 that, to carry out important valued activities effectively, many other components of competence besides eductive ability are required. The components of competence listed across the top of the Grid (and others like them) are unlikely to be highly correlated with each other. Instead, they contribute cumulatively and substitutively to effective performance, rather like the terms of a multiple regression equation. Competence is a value-based, internally heterogeneous, quality. Its measurement therefore cannot be assimilated to the internal-consistency model which dominates mainstream psychometrics.

Despite the implications of what has been said, it is obvious that we need to develop a more adequate descriptive framework to help us think about the components of competence listed across the top of the Grid. At present, for example, the use of feelings and persistence appear both as

* Unlike those of Vodegel-Matzen, Richardson's "real-life" matrices do not exhibit the same logical operations as the diagrammatic matrices with which they are said to be isomorphic. They cannot be of equivalent difficulty because they do not have as many transformations going on at the same time (cf. Jacobs & Vandeventer, 1968) and do not exhibit serial change of the same order of complexity in two dimensions simultaneously. They do not have the properties of mathematical determinants because the argument that applies in one direction does not apply in the other. Perhaps still more importantly, they do not require respondents to *simultaneously* attend to their emerging understanding of an overall pattern in order to discover what to pay attention to in the parts and to attend to the parts in order to discern the overall pattern, i.e. they do not require the same degree of *meaning-making* ability.

components of eductive ability and as qualities which make an important independent contribution to effective behaviour. We also need a better framework for thinking about the potentially valued styles of behaviour that appear down the side.

An examination of Grid 20.1 helps us to understand how the abilities assessed by the RPM contribute both to a wide range of occupationally relevant performance *and* to some occupationally dysfunctional behaviours. However, it also helps us to understand why the RPM:

- Does not necessarily reflect the level of eductive ability which people are capable of displaying *while carrying out tasks they care about*.
- Does not correlate more highly with occupational performance: Occupational performance is determined by whether an individual's values are aligned with those required to perform the job effectively, by the possession or otherwise of numerous other competencies, and by what other people do.
- Could probably, through a series of precisely targeted studies, be shown to be much more highly correlated than currently appears to be the case with the ability to carry out each of a wide range of important activities.
- Does not correlate more highly with level of job attained and retained. As things are currently organised in Western cultures, one would expect this to be more strongly determined by a valuation for personal advancement than by competence at doing the things which those employed in any position need to do to improve the overall effectiveness of the organisation.

It also helps us to see that intelligent behaviour involves an extended time dimension that is commonly overlooked, especially during assessment. To behave intelligently, one must organise one's life in a such a way as to be able to achieve one's valued goals effectively. To do this it is necessary to bring to bear relevant past experiences, imagine potential future scenarios, anticipate obstacles to their achievement, and find ways round the obstacles. It is necessary to resolve value conflicts, among other things by considering the probable consequences of alternative courses of action. The consequences to be considered run from personal (individual) consequences, through organisational consequences, to societal consequences. To consider the last two it is necessary to build up one's own understanding of social and ecological processes. To enact the conclusions of such reflections it is necessary to take a stand for what

one believes to be not only in one's own long-term best interests, but also those of one's family, organisation, community, society, and planet. It is these connections which result in cognitive ability being psychologically bonded to a valuation for such things as taking responsibility for others and taking one's own moral decisions. And they also explain why the adoption of reason-based discipline strategies results in the enhancement of eductive ability.

From a practical point of view, it is clear from Grid 20.1 that using the RPM as a selection and placement tool without the simultaneous use of more broadly based measures is inadequate because many people do not apply their eductive ability to doing what others need them to do. This observation underlines the importance of pressing, not only for developments in *assessment,* but also for more studies of what the short and long-term, personal, organisational, and societal consequences of people doing different things actually are. Given such information, we would be able to generate more meaningful job specifications.

It is also evident from Grid 20.1 that undue reliance on selection procedures which claim to identify "highly able" people may have the effect of absolving teachers and managers from two of their primary responsibilities. These are, on the one hand, to create developmental environments and, on the other, to introduce guidance, placement, and development activities which will help to develop, utilise, and recognise the contribution of, people who value, and are able to undertake, all of the activities revealed by considering each of the cells of the Grid.

To help society and organisations tackle this problem it is important for psychologists to engage in a number of different activities. They must help teachers, managers, and society to clarify the activities which may need to be carried out, develop the tools which are required if assessment systems which recognise and capitalise upon wider aspects of competence are to be introduced, clarify the organisational arrangements which are required if the results of staff and organisational appraisal activities are to be fed to audiences who will help to ensure that action is taken, and develop the understandings required if teachers and managers are to create developmental environments and climates of innovation which will enable society to develop and utilise all the human resources that are available. Developments in all these areas are vital if we are to reduce the most widespread and most serious misuses of tests highlighted by Raven (1991) and Moreland et al. (1995). Preliminary work to help fill some of them is summarised in Raven (1984, 1994).

Predictive Validity

Although the RPM was developed for research purposes, it is widely used in psychological practice for selection, guidance, and problem diagnosis and remediation. An examination of its predictive validity is therefore called for. As it happens, this will throw further light on its construct validity.

Educational Success

Numerous studies (see Court & Raven, 2001; Court & Raven, 1995) have shown that the RPM correlates with school performance, although, as the theoretical basis of both tests would lead one to expect, the correlations between school performance and the *Mill Hill Vocabulary* (MHV) test – a measure of *reproductive* ability – are generally higher than those with the RPM.

Unfortunately, these correlations do not exactly provide a cause for jubilation because, as will be suggested in the next paragraph and as the author has argued more fully elsewhere (Raven, 1991), most measures of educational performance themselves lack construct validity.

Consider the typical "science" test. There is no sense in which such a test assesses competence to function as a scientist, whether in a scientific career, in other fields of work, in the home, in politics, or in the community. The competencies required by scientists include the ability to problematise, the ability to invent ways of collecting relevant information, the ability to locate appropriate mathematics or other ways of summarising data, the ability to persuade others to collaborate, the ability to work with others, and the ability to communicate. In no sense does the typical science test assess such things. Instead, it measures the ability to present temporary knowledge of minuscule and arbitrary selections of out-of-date information (which also has little chance of relating to the assessee's current or future needs) in a way that meets the examiner's expectations (which themselves typically embody an inappropriate concept of science). Such tests, measuring neither scientific competence nor a knowledge of "science", clearly lack construct validity. (Given this understanding of what they actually measure, it is not surprising that they correlate more highly with measures of reproductive ability than with measures of the ability to make meaning out of confusion; the ability to perceive and think clearly.)

Similarly, tests of "English" (and, by implication the "ability to communicate") which ask students to do such things as underline the

verbs in sentences also lack construct validity. Effective communication involves the deliberate manipulation of structure to create and convey an impression, the use of allusion to evoke emotions, the use of innuendo and the evocation of feelings to elicit behaviour, and the ability to write with sensitivity to the values and prejudices of a target audience in order to induce desired action.

These examples highlight two of the problems that are inherent in conventional ways of thinking about the procedures to be adopted when establishing test validity. They illustrate the *criterion* problem in the very field – educational testing – in which testing is most widely applied. Yet, as McClelland (1973), Messick (1989), and others have shown, the problems become more numerous and more serious as one moves into the field of occupational testing.

Occupational Success

The technical and logistic problems involved in establishing the predictive validity of a test in occupational settings include:

a) *Problems associated with the Criteria of Success* (including their validity as indices of the construct being assessed): The qualities apparently required to perform a job "well" depend on the criteria adopted when evaluating performance. Different qualities are, for example, required to secure rapid advancement in an organisation, to secure the survival of that organisation through the invention of new products, to secure its growth through financial and/or political manipulation, and to secure the survival of society. Those who are best able to obtain the esteem of those above them are not necessarily best at releasing the energy and talents of their subordinates and, indeed, often advance themselves by applying their cognitive abilities to make their sections appear more "efficient" by getting rid of the personnel, the time, and the networks of contacts which are required for institutional development, and by eliminating those with alternative viewpoints who might challenge their views or compete for their position (Chomsky, 1987; Hogan, 1990, 1991; Hogan et al., 1990; Jaques, 1989; Nuttgens, 1988; Raven, 1984; Raven & Dolphin, 1978; Spencer & Spencer, 1993).

b) *Problems deriving from the use of Inadequate Job Analyses and Job Descriptions*: The activities required for the effective performance of a job may differ from those identified in the job

description and thus be overlooked when attempts are being made to validate selection procedures (Fivars & Gosnell, 1966; Klemp, Munger & Spencer, 1977; McClelland & Dailey, 1973, 1974; Raven, 1984; Spencer & Spencer, 1993; Taylor & Barron, 1963). Indeed, the notion of "effective job performance" is itself problematical. Thus bus driving can be construed as involving only such things as the ability to avoid accidents. Yet, as Van Beinum (1965) has demonstrated, the effectiveness and adaptation of a bus service is dependent on bus drivers sharing their insights with their managers and contributing to a climate of innovation. Kanter (1985) has generalised the point: the innovativeness and survival of organisations depends on people doing things which would never be suspected if one asked merely "What do they need to do to produce widgets?"

c) *Problems created by the use of Inappropriate Selection Procedures in the past*: Those best able to perform a job may have been (intentionally or unintentionally) eliminated from those admitted to the workforce. If this has happened it will be impossible to demonstrate the importance of the required qualities (Berg, 1973; Holland & Richards, 1965; Hope, 1984; McClelland, 1973; Raven, 1994; Raven, J. & Stephenson, J. (Eds.), 2001; Taylor, Smith & Ghiselin, 1963).

d) *Problems created by the Non-Attributable Nature of Outcomes*: In most organisations it is extremely difficult to attribute observable effects to any one person or group of persons (see Day & Klein, 1987). This is especially so when circumstances are continuously changing and the effects of actions may take many years to show up. This makes it difficult to collect accurate information about whose work genuinely benefits an organisation and distinguish those who confer important benefits from those who are only able to create a good impression and move on before their mistakes are discovered.

e) *External Constraints*: Organisational arrangements, and other people's expectations, may prevent people doing the things required for effective job performance.

f) *Change Over Time*: People do different things in the "same" job at different times. They may, for example, engage in routine activities for part of the day and in innovative ones at other times. They may develop technological innovations early in their careers

and engage with the political processes which control the funding for such innovations later in their lives.

Despite these problems, it has been shown that the RPM *does* relate to a variety of measures of managerial performance: Staff and financial turnover, profitability, and the ability of the firm to survive financial and other crises. Thus, data supplied by a transnational corporation which runs several thousand small retail stores shows that the RPM correlated .50 with the *Watson-Glaser Test of Critical Thinking* and .20 with assessments of work management performance, .13 with assessments of interpersonal skills performance, and .12 with planning and problem solving performance. Although these correlations are statistically highly significant (being based on a study of 1120 managers), their true significance only emerges as one realises that the correlations between the performance measures and most of the other tests used in the study were zero.

Ingleton (1990) found that while managers with high vocabulary test scores performed well in unchanging conditions, it tended to be those with high RPM scores who were best able to help their firms weather the crisis produced by the 1970s oil price increases. (It is for reasons like this that it is so important for the public service, in particular, to recruit, and to promote into positions having very different job descriptions, a wide range of people who have distinctive patterns of motivation and ability.)

But it is not only in management settings that the tests have been validated. Several studies (see Court & Raven, 1995) have shown that the RPM and MHV between them can predict about 10% of the variance in performance within a wide range of occupations. Validity generalisation analysis, which adjusts these figures for restriction of range and the unreliability of criteria, suggests that a "truer" estimate of the proportion of variance accounted for is 25%. However, since those concerned with personnel selection are necessarily operating in situations involving restricted range and unreliable criteria, it is not entirely clear that the adjusted figure conveys an appropriate impression of the benefits that can be obtained from testing.

One unpublished study illustrating the use of the APM in predicting non-managerial performance involved computer programmers. The data (supplied in tabular form) showed that the APM, administered without a time limit, was a particularly good predictor of success. This is perhaps because similar levels of attention to detail, checking, and persistence are required for success at both tasks.

More generally, meta-analyses (Ghiselli, 1966; Hunter & Hunter, 1984) show that tests of intellectual ability predict proficiency within at least the following types of work: managerial, clerical, sales, protective professions, service jobs, trades and crafts, vehicle operation, and simple industrial work.

All such studies yield what may be regarded as relatively low predictive validities. There is, however, another way of coming at the question of validity which yields a much more positive conclusion. Instead of seeking evidence for the predictive validity of the RPM *within* occupational groups, one can focus on its ability to predict the *level* of job an individual will attain and retain.

Before discussing this topic further, it is necessary to examine more carefully the nature of the activities which distinguish more from less effective performance both within and between jobs.

Critical-incident studies (such as those summarised in Raven, 1984, Spencer & Spencer, 1993, and Raven & Stephenson, 2001) have shown that effective performance in a wide range of jobs depends on doing such things as building up one's own understanding of the way in which the organisation in which one works functions, viewing one's own part in it in appropriate ways, taking initiative to intervene in organisational processes when necessary, building up one's own understanding of the workings of external political and economic systems and intervening in them for the benefit of one's organisation and society, and thinking about the motives and talents of subordinates and how best to place them so as to harness their motives and develop their talents.

Although, as shown in data summarised in Raven (1984), more effective workers in all occupations are distinguished from their less effective peers by the frequency with which they do such things, Jaques (1976, 1989) has argued that these high-level activities are more important in high-level jobs.

He has also argued that the ability to undertake many of these activities is primarily dependent on "cognitive ability". However, he defines "cognitive ability" to include the use of feelings to initiate action which is then monitored to learn more about the situation with which one is dealing and the effectiveness of the strategy one has adopted – together with the ability to take corrective action when these observations show that it is necessary. Such activities require great determination and persistence. Precisely because Jaques wishes to include these affective, conative, and "experimental action" components in his concept of "cognitive ability" he

denies (as we did above) that "intelligence" tests measure it. Nevertheless what he has in mind does seem to have much in common with "eductive ability" as identified by Spearman and as conceptualised here.

Note the problems Jaques' contentions pose for test validation. Even supposing we had a test which adequately measured what he means by "cognitive ability", we would need a collection of very sophisticated studies to validate it. To get high zero-order correlations between the test and criteria it would be necessary to find an organisation in which people were not constrained by day-to-day pressures to attend to matters that did not require them to exercise the maximum level of cognitive capacity of which they were capable. The organisation would also need to be one which *did* require them to apply their cognitive ability to undertaking the kinds of activity mentioned above and which discouraged them from applying it to such things as securing their personal advancement mainly by creating a good impressions on their superiors without doing the things that needed to be done. Alternatively one would have to make a series of detailed – almost ethnographic – studies of what individuals were actually doing in their jobs and relate test scores to conceptually crucial components of that performance.

These observations strongly reinforce the claims of McClelland (1973) and Messick (1989) that the validity of a test cannot be estimated directly. An impression of its validity can only be achieved by first making a theoretical analysis of what the test measures, the competencies required in particular types of job, and the organisational arrangements through which work is conducted, and thereafter reviewing studies – each imperfect in itself – which illuminate what the test measures and predicts and the factors which enhance or reduce the observed relationships. Thus test validation involves nothing less than applying (properly understood forms of) scientific method to illuminate a hidden reality (House, 1991). It is more than a little unfortunate that Barrett and Depinet (1991) do not seem to have understood this position when preparing their highly influential, but altogether misleading, paper.

With these reservations in mind, we will now review evidence suggesting that the RPM, and other measures of ***g,*** are better at predicting the *level* of job an individual is able to attain and retain than at predicting performance within any particular occupation.

Vernon and Parry (1949) summarised the results of testing 90,000 British naval recruits with a short, non-cyclical, version of the SPM during the Second World War. There were systematic differences in the mean

scores of men from 12 general classes of occupation: clerical, electrical workers, precision workers, woodworkers, sheet metal workers, machine operators, retail tradesmen, building workers, "mates", drivers, farm workers, and labourers.

Foulds and Raven (1948) tested the entire workforce of a photographic factory and found very large average differences in the SPM scores of workers at five different levels (Table 20.1).

de Leeuw and Meester (1984) showed that about 50% of the variance in occupational level can be predicted from RPM scores.

Fraser-Roberts (1943) likewise found that there was a marked correlation between RPM scores and level of job attained and retained.

If Jaques is right to argue that cognitive ability is closely related to level of job attained and retained, there ought to be an *optimal* range of scores – neither too high nor too low – for most jobs. The most convincing evidence on this point comes from the work of Hope (1984) which will shortly be reviewed in some detail. However, evidence supporting the argument that there is an optimum range of scores for each occupation comes from a study conducted by J.C. Raven and his colleagues (Crichton Royal, 1957) among telephone engineers.

The conclusion that the relationship between ability and performance is curvilinear may be reconciled with the finding of Hunter and Hunter (1984) that the relationships within all groups are linear by recalling the criterion problem. In Raven's study, the finding was not that higher scoring employees performed worse but that higher and lower scorers left the employment. This strongly supports Jaques' contentions.

But while Jaques' argument is plausible, Kanter's work and the previously mentioned studies showing that eductive ability is important at all levels – especially when the criteria applied in test validation include the survival of the organisation concerned or the society in which it is located – suggests that it is not the whole story.

Social Mobility

Despite the absence of a dramatic relationship between most psychological tests and measures of work performance, the RPM, and "intelligence" tests in general, do not do a bad job of predicting social mobility. Unfortunately, this is again not quite such a cause for celebration as might at first sight appear. The problem is that the link between "cognitive ability" and social mobility is not necessarily direct and may be via patterns of motivation. Given the limited data currently available, it is impossible to

decide whether the link is indeed direct (as Jaques would have us believe) or whether motivational predispositions are responsible for both the test scores people attain and their social mobility.

Perhaps the most impressive evidence of the power of "intelligence" tests to predict social mobility comes from the *Scottish Longitudinal Mental Development Survey* (Scottish Council for Research in Education, 1933, 1949, 1953; MacPherson, 1958; Maxwell, 1961, 1969; Hope, 1984).

Using these data and others, Hope (1984) showed that (a) some 60% of social mobility (both upward and downward) in both Scotland and the US can be predicted from 11 year olds' intelligence test scores; (b) that, by the time children are 11 years old, Scotland achieves (or did achieve) a degree of association between "intelligence" and socio-economic status (SES) that is not achieved in America until age 40; and (c) that, even when the effects of home background are partialled out, children's "intelligence" makes a major contribution to a variety of indices of their occupational success at 28 years of age. The contribution of intelligence is very much greater than that of educational achievement and, as the slow sorting process in America makes clear, is not a surrogate for sociological tracking by the educational system. Early success in the educational system predicts later educational success – but success in the educational system has very little predictive power outside. On the other hand, "intelligence" and, importantly, teachers' ratings (at age 11) of qualities like originality, creativity, determination, and persistence independently enable one to predict life success.

So far so good. The problem is that children from the same family vary almost as much in the kinds of activity they are strongly motivated to carry out (or can be said to value) as in their "intelligence" (Kohn & Schooler, 1978; Raven 1976, 1977), and the available evidence suggests that social mobility, both upward and downward, can be predicted every bit as well from a knowledge of the activities they are strongly motivated to carry out as from their "intelligence". Kohn (1969/1977; Kohn et al., 1986) among others demonstrated that people occupying high socio-economic status positions in several different societies embrace activities like thinking for oneself, originality, taking responsibility for others, and initiative. In contrast people occupying low socio-economic status positions stress toughness, strength, obedience, and having strict rules and moral codes to guide their lives. Kohn initially believed that these differences were a product of occupational experience (and, indeed, to some extent, they are). But, by sectioning the data we obtained from adolescents by

Table 20.1. Standard Progressive Matrices
Score Distributions for Five Classes of Employee in a Photographic Works

	Quartiles of score distribution			
	1	2	3	4
Directive and Executive	79	9	12	-
Highly skilled workers	48	23	19	10
Skilled workers	29	25	27	19
Qualified workers	18	26	28	28
Unskilled workers	12	15	28	45

From Foulds and Raven (1948)

origins and anticipated occupational destinations, we (Raven et al., 1975; Raven, 1976) were able to show that there was a great deal of variance in the concerns of children from similar backgrounds, and that this variance was related to the status of the jobs they expected to enter. This finding, like the finding that two thirds of the variance in "intelligence" test scores is within-family variance, raises serious questions about its origins. A somewhat similar finding was reported by Kinsey (1948). Kinsey found that there was huge variation in the sexual behaviour and attitudes of children who came from similar backgrounds and that this variation predicted where those children would end up. They *joined* others who thought and behaved similarly. Children could hardly have learned sexual attitudes and behaviours so different from those of their parents by modelling or formal instruction. So, where does the variance come from and how does it come about that personal attitudes and behaviour of the kind exemplified by sexual behaviour come to correspond to those of the socio-economic groups people eventually enter? The variance between children from the same family has often been attributed to genetic factors, and, in this context, we may note that Tellegan et al. (1988), Bouchard and McCue (1990), Bouchard (1991), and Waller et al. (1989) have shown that many values and beliefs – including religious beliefs – are as heritable as "intelligence". But, if these attitudes and behaviours are not learned at work and in society, how does it come about that, in the end, their attitudes and behaviours tend to be characteristic of the groups with whom they end up living and working?

Note the problems which these observations pose for the validation and interpretation of "intelligence" tests: Children from similar backgrounds, including members of the same family, vary enormously in both their motives and values and their "intelligence". The variance in their motives predicts their future position in society every bit as well as does their "intelligence". Which is the more basic set of variables? How does variance in "intelligence" come to be linked to variation in motives, values, and personal behaviour?

One study which throws light on the last question has been reported by Maistriaux (1959). Presenting his results in tabular, rather than correlational, form, Maistriaux documents a remarkable relationship between "intelligence" and the kind of activity that people say they want to carry out and enjoy carrying out. Those with higher RPM scores find "intellectual" activities more enjoyable while those with lower scores are more attracted by "practical" activities. In a sense, these results suggest that we may be dealing with different perspectives on "the same" psychological variable.

Other studies – such as those reported by Flynn (1987) and McClelland (1961) – do not, however, support this contention. These studies show that the differences in actual life performance of different ethnic and religious groups in America are very much greater than, and cannot be explained by, differences in their "intelligence". In other words differences in motives, values, and such things as social support, associated with ethnicity and religion are better predictors of "real life" performance than "intelligence".

The overall conclusion to be drawn from this material is that we do not, at present, know whether the portion of the variance in social position and performance that can be predicted from "intelligence" is the same as that which can be predicted from motivation and values or whether the two are additive. In the current state of our knowledge, one clearly has the option of concluding that we should be focusing on the variance in the kinds of behaviour to which people are attracted and their ability to undertake those behaviours effectively rather than on their "intelligence".

Further evidence that the link between "intelligence" and social status and social mobility may be mediated by the kinds of behaviour which attract people comes from two other sources.

The first of these comes from studies of the links between cognitive activity and values. In the first place, cognitive ability and activity is

not universally valued. Many parents do not want their children to ask questions or to be able to use books to find information for themselves (Raven, 1980). Secondly, cognitive ability is psychologically bonded to other personal characteristics, such as curiosity and independence. These may not be valued even if cognitive activity itself is valued (Maistriaux, 1959; Raven, 1987). Thirdly, nurturing cognitive ability depends on child-rearing, educational, and staff-development practices which may not be valued even if cognitive ability itself is valued. Thus the development of cognitive ability is facilitated by the adoption of democratic discipline strategies, encouragement of adventurousness and independence, and studying children's needs and responding to them (Raven, 1980, 1987, 1989a; Feuerstein et al., 1990; Sigel, 1986). It develops in the workplace if managers encourage their subordinates to participate in establishing, and finding ways of achieving, organisational goals and if they study subordinates' motives and talents in order to find ways of developing and utilising them (Kohn & Schooler, 1978, 1982; Jaques, 1976; Lempert, 1986; Lempert et al., 1990).

The second comes from neuropsychology. Trevarthen (1990, 1992) and Sperry (1983) have suggested that the most important psychological concomitants of neurological differences lie in the affective and motivational area. They suggest that the differences in cognitive performance that are associated with neurological locale (including the left and right brain) are merely expressions of more basic differences in motivational predispositions and that it is these which are neurologically located. If "cognitive ability" were assessed while people were undertaking a task that tapped very different motives, such as putting others at ease, not only would our estimates of the "cognitive ability" of those concerned be very different, those abilities would appear to have very different neurological locations. They suggest that the way to make sense of such results would be to recognise, as we have done here, that important components of competence (including eductive ability) will only be displayed while people are carrying out activities they care about. More consistently interpretable data would be obtained by attending to the neurological localisation of motivational predispositions. In that context, Trevarthen underlines the importance of developing a framework for thinking about what he terms the modules of motivation (cf. Murray and McClelland). He moves on to emphasise the need for engagement between the motivational predispositions of parents and children, teachers and pupils, and managers and subordinates, if the development of high-level, generic

competencies, including eductive ability, is to be facilitated (cf. Feuerstein et al. 1990; Vygotsky, 1981; Raven, 1989a).

We may now attempt to draw some tentative conclusions from this review.

1) It is impossible, on the basis of the evidence currently available, to decide whether to explain the allocation of people with different concerns and levels of "cognitive ability" to different socio-economic groups by reference to variations in patterns of value-based competencies or by reference to "cognitive ability".
2) It is impossible to discover whether the relationship between neuro-anatomy and psychological dispositions is to be attributed to differences in motivational predispositions or "cognitive ability" using tests developed within the dominant psychometric tradition.
3) It is particularly difficult to reconcile two sets of claims.
   On the one hand it is argued that:
   - People employed at different occupational levels differ markedly in both cognitive abilities and values.
   - To have an effective organisation it is necessary to have a steep differential in cognitive ability by occupational level.
   - The ability to understand, and find ways of intervening in, the operation of international socio-politico-economic systems for the long-term good of the organisation and the future of humankind calls for exceedingly rare levels of cognitive ability.

   On the other hand it is argued that:
   - The effective performance of low status jobs demands high-level competencies.
   - The culture of intelligence and innovation needed for the development and survival of an organisation or society requires those involved *at all levels* to exercise high-level competencies.

A Framework for Thinking About Competence

Having illustrated some of the limitations of the mainstream "ability" position, the problems associated with the psychometric and validation paradigm with which it is associated, and the vital need to develop a more comprehensive and psychologically appropriate form of assessment, it

is time now to present a brief outline of an attempt to develop a more fruitful way of thinking about competence and its assessment.

But have not numerous psychologists – such as Guilford (1977), Gardner (1985, 1991), Hatch and Gardner (1986, 1990), Sternberg et al. (1986), and Taylor (1971, 1976) – tried to develop such a framework, and have not people like Spearman (1927), Eysenck (1953), Hunter and Hunter (1984), Matarazzo (1990), Barrett and Depinet (1991), and Ree, Earles, and Teachout (1994) shown that all these abilities reduce to the very eductive and reproductive abilities we have been talking about and that no measures of other abilities are both sufficiently distinct from these abilities and sufficiently reliable in themselves to stand up to scrutiny? Indeed they have. Unfortunately, all of these researchers have approached the problem with what might, for the want of a better phrase, be called something approaching a classical psychometric mindset.

Fortunately, some other psychologists have come at the problem from another starting point. Instead of starting with psychometrics, they have studied the nature of occupational, civic, and parental competence. Following Flanagan (1949, 1954), those who have worked in the occupational area have asked supervisors, subordinates, and job encumbants to describe actual incidents of effective and ineffective behaviour – what happened, what led up to it, what the outcome was, what they were thinking and feeling and doing, what other people did, and how others reacted.

Spencer and Spencer (1993) summarise more than 350 studies of this sort, using them to guide their development of a "dictionary" of occupational competencies.

In seeking a way forward here we may first recall that we have seen that, in reality, we need to employ a two-stage measurement model to assess the wider aspects of both "intelligence" and "competence": We first need to discover what kinds of activity people are spontaneously motivated to undertake and then which components of competence they display when undertaking those activities. This means that it will be necessary to develop an agreed conceptual framework for describing the kinds of activity people may "value" and the components of competence they may display while undertaking those activities. [Attempts to develop such a framework have been published by Raven (1984), Huff et al. (1982), and Spencer & Spencer (1993), although Raven and Stephenson (2001) contains a critique of the latter.]

The framework we have ourselves constructed out of that developed by McClelland et al. (1958) for scoring their *Test of Imagination* may be

represented for heuristic purposes in the form of a two-dimensional grid – Grid 20.2 – which is a modified version of that published in Raven (1984, 1991). This lists a number of activities an individual may be strongly motivated to undertake across th e top and a number of the cognitive, affective, and conative components of competence he or she might utilise to carry out those activities effectively down the side.

To move toward a comprehensive assessment of an individual, one could insert ticks (or crosses) in the cells of an extended version of this Grid to show *which* components of competence he or she displayed whilst undertaking each of the activities he or she cared about. One could then reduce data overload by summing the ticks in each column, and compositing the totals for the columns belonging to the Achievement, Affiliation, and Power clusters. This would a yield 3-score, value-based, internally-heterogeneous, personal profile which would be isomorphic with McClelland's *need* Achievement, *need* Affiliation, and *need* Power "motive" profiles. These scores (which obviously have little in common with internally-consistent factor scores) can be understood as being something like multiple-regression coefficients predicting the success with which someone would be able to carry out activities he or she valued.

Despite the succinctness and value of such profiles, examination of the detailed information contained in a completed Grid is much more informative than a collection of scores . This is partly because people may value – or be somehow motivated to undertake – many activities which do not fall into the Achievement, Affiliation, and Power categories and partly because there are many more components of competence than are taken into account in McClelland's scoring system. From a completed Grid one can see *which* competencies the person being assessed tends to display whilst carrying out *which* valued activities.

If one follows the line of argument advanced here further, however, one finds oneself moving away from a concern with scores and *variables* and instead making *descriptive statements* about the people one is assessing. One starts using descriptors (analogous to those used by chemists) to record the activities people value and the competencies they display while undertaking those activities.

However, as soon as one starts to do this, one is forced to recognise that the competencies people will develop and display are in part determined by the extent to which the environment in which they have in the past lived and worked, and the environment in which they are now observed, engages with their values and has led them to develop, and now

leads them to display, the competencies they possess. As a result, one finds oneself attempting to write statements *about those environments* at the same time as making statements about the individual. One then finds oneself trying to say something about the transformations in competence which a change of environment would be likely to effect. One then finds that one has unexpectedly solved the problem – highlighted by Jackson (1986) – of modelling the *transformational* processes which occur in homes, schools, and workplaces.

Operationalisation of the Measurement Model

Those who wish to go into the way in which this framework for thinking about competence and its assessment can be operationalised should refer to Raven (1984, 1988, 1991) or Raven & Stephenson (eds) (2001). Suffice it to say here that there are two main ways in which this can be done. The first involves creating developmental environments which enable people to undertake activities they care about and, in the process, develop and display high-level competencies. The second involves getting inside people's heads in order to find out what motivates them and which components of competence they bring to bear to achieve their valued goals effectively. The latter can be done using specific types of projective methodology, *Behavioral Event Interviewing*, or value-expectancy-instrumentality methodology. As a brief antidote to Barrett and Depinet's failure to examine such methods with any care, the next three paragraphs summarise what each involves.

a) *Observation*. Just as a chemist needs to be familiar with atomic theory to appreciate the significance of a precipitate in a test tube, so the interpretation of what is revealed by behaviour in particular situations is dependent on familiarity with an appropriate interpretative framework. A pre-requisite to eliciting behaviour which reveals which competencies an individual is able to display is the creation of a "developmental environment" (Raven, 1984, 1989a, 1991; Burgess & Adams, 1980, 1986; Stansbury, 1980) which taps the individual's values and leads him or her to display high-level competencies. Thereafter, thorough familiarity with an extended version of the competency framework developed above is necessary to guide the analysis of that behaviour and understand its significance. (It follows that the current drive for

"portfolio" and "authentic" assessments, well-intentioned though it is, is almost certain to founder because of the absence of an adequate descriptive framework for summarising the material.)

b) *TAT and BEI Methodology*. Those scoring McClelland's *Test of Imagination* and *Behavioral Event Interview* protocols follow a detailed and explicit procedure (McClelland, 1951; McClelland et al., 1958; Winter, 1973). Those scoring *Test of Imagination* protocols first ask themselves "What kind of activities is the person who wrote this story motivated to undertake? (i.e. which kinds of activity does he or she value, care about, or somehow feels internally driven to undertake?)", and then "*How many* of a specific and experimentally-derived list of cognitive, affective, and conative components of competence does this person tend to engage in spontaneously while undertaking these activities?" Actually, the process is somewhat circular since a person's motives or values are identified by examining the kinds of things he or she tends to turn thoughts, feelings, and effort into achieving. Nevertheless, the effect is to produce profiles of value-based, internally-heterogeneous, scores of the kind outlined above. *Behavioral Event Interviews* substitute accounts of real-life events for projective stories. One asks people to think of specific times when things were going well (or badly) for them, what led up to the situation, what they were trying to do, what they were thinking and feeling, what they did do, what others did, their reactions to what others did, and what the outcome was (McClelland, 1978; Spencer & Spencer, 1993). More pointedly, they can be asked the same questions about critical life-events identified using Flanagan's "critical-incident" methodology. In the course of these interviews, people's preoccupations (or values) and the competencies they bring to bear to undertake these activities effectively become very obvious.

c) *Value-Expectancy-Instrumentality Methodology*. In the course of a number of programme evaluations and cross-cultural studies, we have first asked people to say how important it was to them to undertake each of a large number of different sorts of activity and then how satisfied they were with their available opportunities to do each of the things they felt it was personally important for them to do. Thereafter they were asked to select the *most* important of their important sources of dissatisfaction and indicate what would

GRID 20.2. A MODEL OF COMPETENCE

Examples of Potentially Valued Styles of Behaviour

Examples of components of effective behaviour.	**Achievement**				**Affiliation**				**Power**		
	Doing things which have not been done before.	Inventing things.	Doing things more efficiently than they have been done before.	Developing new formal scientific theories.	Providing support and facilitation for someone concerned with achievement.	Establishing warm, convivial relationships with others.	Ensuring that a group works together without conflict.	Establishing effective group discussion procedures.	Ensuring that group members share their knowledge so that good decisions can be taken.	Articulating group goals and releasing the energies of others in pursuit of them.	Setting up domino-like chains of influence to get people to do as one wishes without having to contact them directly.
Cognitive											
Thinking (by opening one's mind to experience, dreaming, and using other sub-conscious processes) about what is to be achieved and how it is to be achieved.											
Anticipating obstacles to achievement and taking steps to avoid them.											
Analysing the effects of one's actions to discover what they have to tell one about the nature of the situation one is dealing with.											
Making one's value conflicts explicit and trying to resolve them.											
Consequence anticipated: *Personal*: e.g. "I know there will be difficulties, but I know from my previous experience that I can find ways round them. *Personal normative beliefs*: e.g. "I would have to be more devious and manipulative than I would like to be to do that." *Social normative beliefs*: e.g. "My friends would approve if I did that": "It would not be appropriate for someone in my position to do that."											
Affective											
Turning one's emotions into the task: Admitting and harnessing feelings of delight and frustration: using the unpleasantness of tasks one needs to complete as an incentive to get on with them rather than as an excuse to avoid them.											
Anticipating the delights of success and the misery of failure.											
Using one's feelings to initiate action, monitor its effects, and change one's behaviour.											
Conative											
Putting in extra effort to reduce the likelihood of failure.											
Persisting over a long period, alternatively striving and relaxing.											
Habits and experience											
Confidence, based on experience, that one can adventure into the unknown and overcome difficulties, (This involves knowledge that one will be able to do it plus a stockpile of relevant habits).											
A range of appropriate routineised, but flexibly contingent behaviours, each triggered by cues which one may not be able to articulate and which may be imperceptible to others.											
Experience of the satisfactions which have come from having accomplished similar tasks in the past.											

happen if they were to try to do something about the problem. The potential consequences studied were drawn from the range indicated by Fishbein (1967). They therefore included questions about whether they would be able to gain the satisfactions which they personally wanted, whether they would be able to live up to their personal – moral – self-images, and what reactions they anticipated from reference groups.

The results obtained from programme evaluations and organisational surveys conducted in this way have been extremely revealing.

One study (Raven, 1980) generated numerous new insights into parents' and teachers' child rearing behaviour. Mothers tend to create individualised, competency-oriented, developmental programmes for their children. But, although some teachers would like to do this, they do not know enough about each of their pupils to so. And the attempt to do so often confronts them with a host of moral dilemmas: Should they, for example, encourage independence and question-asking among children who live in dangerous environments and have parents who cannot manage independent children who are liable to question commands? In the course of the study a whole new set of issues bearing on parents' and teachers' competence in child rearing – and the tools they would need if they are to behave more competently – came to light. More specifically, exploration of teachers' and parents' competence to pursue their own lives and do their jobs effectively showed that many were in no position to provide appropriate role models for children. It follows that, if one wishes to facilitate the growth of competence in children, one extremely important starting point is by enhancing the competence of *their caregivers* to do the things *they* want and need to do. It also turned out that it was mothers' lack of confidence in their own competence as mothers which led them to hand their children over to other caretakers, but, paradoxically, those professionals were in no position to nurture the children's most important competencies.

In another study (Raven, Johnstone & Varley, 1985; Raven & Varley, 1984) the methodology was used to assess the effect that different teachers had on children's awareness of their motives, their values and priorities, and their competence to undertake activities they cared about effectively. It emerged that, contrary to the claim that schools make no difference, teachers had dramatic, and markedly different, effects on pupils' values and the consequences they expected if they were to set about tackling

problems they cared about. Most importantly, it emerged that previous evaluations of interdisciplinary, enquiry-oriented, project-based education had been entirely – and damagingly – misleading. Properly organised, project-based education has dramatic, positive, effects on children's confidence and competence. There is, however, a fundamental problem which prevents generalisation of the work. This is that there are no good tools to help teachers identify each child's motives, create individualised, competency-oriented, developmental programmes, and monitor each child's growth.

In a third study (Raven, 1984; Graham & Raven, 1987) it was found that, as in McClelland's (1961) work, there are dramatic differences between the pre-occupations of people who live in different societies and their willingness to do the things that are necessary – that is to say, their competence – to translate those values into effect. As far as can be judged, these differences are directly related to the kind of society which develops.

What these studies show is that the application of value-expectancy-instrumentality methodology guided by the framework for thinking about competence and test validation developed above does yield information which is more revealing, more valid, more comprehensive, and therefore more ethical, than that which would have been obtained had the studies been conducted only with tests of the kind which the *Joint Committee on the Evaluation of Educational Programs and Policies* (Stufflebeam, 1981) enjoin us to use – namely tests which have been shown to be reliable and valid in the conventional sense.

The application of value-expectancy-instrumentality methodology to the assessment of individual competence has proved more cumbersome than its application in programme evaluation. Nevertheless, computerised tools in this area are now available (Raven & Sime, 1994).

Summary and Conclusion

Three sets of conclusions – at different levels – emerge from what has been said in this article.

The first is that Spearman appears to have been right to emphasise the distinctive psychological nature of eductive and reproductive ability and to argue these abilities have different genetic and environmental determinants and different consequences for people's lives. These aspects of "intelligence" emerge as being among the most important variables psychology – whether pure or applied, whether "cognitive", "educational", or "occupational" – deals with.

The underlying *reasons* for – i.e. the interpretation to be placed upon – this now well established network of relationships, is, however, seriously open to question. At this point in time it is possible to attribute the entire observed pattern of relationships to variation, not in *cognitive* ability, but in motivational predispositions.

Even setting that disturbing thought on one side, the material reviewed shows that there is, at the very least, an urgent need to reconsider the way we think about and assess problem-solving ability, intelligence, and competence. At a minimum, we need more appropriate ways of conceptualising and assessing "cognitive ability". More basically, and more importantly, it is vital to broaden our framework for thinking about competence and social functioning so that we can situate our assessments of eductive ability in the context of assessments of (i) motives in the service of which eductive ability (as a component of competence) is applied, and (ii) other components of competence. Without such developments, our assessments of both individuals and educational programmes appear to be unethical. This is because they are insufficiently comprehensive and , as a result, lead to practices which are not in the best long-term interests of the individuals or programmes being evaluated – and therefore not in the long-term interests of society. The psychometric model which is required to come to terms with this problem differs markedly from that which has been pre-eminent in the past. A two-stage measurement process must be envisaged. We must first identify people's motives or valued styles of behaviour and then ask which of a range of cognitive, affective, and conative competencies they bring to bear in their efforts to undertake the activities they care about. A number of ways in which this model has been operationalised have been presented, but a great deal of further development work is required.

But what has been said also appears to have implications at a quite different level. We need to fundamentally reconsider the way in which seek to establish the validity of tests. On the one hand, the validity of the *criteria* as indices of the underlying construct we are seeking to assess is a much more serious problem than it has usually been taken to be. Behaviour is a poor index of psychological constructs since what people do depends on very many things, some arising from personal and value conflicts, some from environmental constraints. To find out what people are doing one needs somehow to get inside their heads. Setting them an alternative task – "performance assessment" – does not solve the problem. Thereafter one needs to somehow to examine the way in which motivational dispositions, eductive ability, and other components of competence contribute to that performance. What is required is fundamentally a conceptual, rather than a statistical, exercise … although path analysis certainly has a role to play.

References

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

Berg, I. (1973). *Education and Jobs: The Great Training Robbery*. London: Penguin Books.

Bouchard, T.J. (1991). A twice told tale: Twins reared apart. In W.Grove and D.Ciccehetti (Eds.). *Thinking Clearly about Psychology: Essays in Honor of Paul Everett Meehl: Personality and Psychopathology*, (Volume 2).

Bouchard, T.J. and McGue, M. (1990). Genetic and rearing environmental influences on adult personality: An analysis of adopted twins reared apart. *Journal of Personality*, March.

Brown, J.S., Collins, A. and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, January-February, 32-42.

Burgess, T. and Adams, E. (1980). *Outcomes of Education*. London: MacMillan Education.

Burgess, T. and Adams, E. (1986). *Records of Achievement at 16*. Windsor, Berks: NFER-Nelson.

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22.

Carpenter, P.A., Just, M.A. and Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97(3)*, 404-431.

Chomsky, N. (1987). *The Chomsky Reader*. London: Serpent's Tail.

Court, J., & Raven, C. J. (2001). *A Researcher's Bibliography for Raven's Progressive Matrices and Mill Hill Vocabulary Scales*. Obtainable in hard copy and disk format from Susan Middleton, Harcourt Assessment, 19500 Bulverde Rd., San Antonio, Texas 78259, USA. <susan_middleton@harcourt.com>

Court, J.H. and Raven, J. (1995). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No.7, Research and References, Summaries of Normative, Reliability, and Validity Studies and References to all Sections*. Oxford, England, San Antonio, TX: Harcourt Assessment.

Crichton Royal, Dumfries. (1957). *A Comparative Study of Two Psychological Tests*. Department of Psychological Research. Unpublished manuscript.

Day, P. and Klein, R. (1987). *Accountabilities: Five Public Services*. London: Tavistock Publications.

Dalziel, M.M. and Schoonover, S.C. (1988). *Changing Ways*. New York: American Management Association.

de Leeuw, J. and Meester, A.C. (1984). Over het intelligente – onderzoek bij de militaire keuringen vanaf 1925 tot henden. [Intelligence – as tested at selections for the military service from 1925 to the present]. *Mens en Maatschappij,* 59, 5-26.

Deary, I.J. (1993). Inspection time and WAIS-R IQ subtypes: A confirmatory factor analysis study. *Intelligence, 17*, 223-236.

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

Deming, W.E. (1993). *The New Economics for Industry, Government, and Education*. Cambridge, MA: Massachusetts Institute of Technology.

Eysenck, H.J. (1953). *Uses and Abuses of Psychology*. Harmondsworth, Mddx: Penguin Books.

Feuerstein, R., Klein, P., and Tannenbaum, A. (Eds.). (1990). Mediated learning experience: Theoretical, psychosocial, and educational implications. *Proceedings of the First International Conference on Mediated Learning Experience*. Tel Aviv: Freund.

Fishbein, M. (Ed.). (1967). *Readings in Attitude Theory and Measurement*. New York: Wiley.

Fivars, G. and Gosnell, D. (1966). *Nursing Evaluation*: *The Problem and the Process*. Pittsburg PA: Westinghouse Learning Corp.

Flanagan, J.C. (1949). Critical Requirements: A new approach to employee evaluation. *Personnel Psychology, 2*, 419-425.

Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-358.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171-191.

Foulds, G.A. and Raven, J.C. (1948). Intellectual ability and occupational grade. *Occupational Psychology, 22*, 197-203.

Fraser-Roberts, J.A. (1943). *Further Observations on the Efficiency of the Progressive Matrices Test*. London: War Cabinet Expert Committee Report.

Gallimore, R. (1985). *The Accommodation of Instruction to Cultural Differences*. Los Angeles: University of California, Department of Psychiatry.

Gardner, H. (1985). *The Mind's New Science*. New York: Basic Books.

Gardner, H. (1987). Developing the spectrum of human intelligence. *Harvard Education Review, 57*, 187-193.

Gardner, H. (1991). Assessment in context: The alternative to standardized testing. In B.R. Gifford and M.C. O'Connor (Eds.), *Changing Assessments*: *Alternative Views of Aptitude, Achievement and Instruction*. New York: Kluwer Publishers.

Ghiselli, E.E. (1966). *The Validity of Occupational Aptitude Tests*. New York: Wiley.

Graham, M.A. and Raven, J. (1987). *International Shifts in the Workplace – are we becoming an "Old West" in the Next Century?* Provo: BYU, Department of Organizational Behavior.

Greeno, J.G. (1989). A perspective on thinking. *American Psychologist, 44(2)*, 134-141.

Guilford, J.P. (1977). *Way Beyond the IQ*. New York: Creative Education Foundation and Creative Synergetic Associates.

Hatch, T.C. and Gardner, H. (1986). From testing intelligence to assessing competencies: A pluralistic view of the intellect. *Roeper Review, 8*, 147-150.

Hatch, T.C. and Gardner, H. (1990). If Binet had looked beyond the classroom: The assessment of multiple intelligences. *International Journal of Educational Research*, 415-429.

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.

Hogan, R., Raskin, R. and Fazzini, D. (1990). The dark side of charisma. In K.E. Clark and M.B. Clark (Eds.), *Measures of Leadership*. West Orange, NJ: Leadership Library of America.

Holland, J.L. and Richards, J.M. (1965). Academic and non-academic accomplishments. *Journal of Educational Psychology, 56*, 165-175.

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. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 72*, 242-259.

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

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

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

Hunter, J.E. and Hunter R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Ingleton, C.C.P. (1990). Personal communication, based on the study reported by C.C.P. Ingleton, *The Use of Tests in Graduate Selection* (61-76). In K.M. Miller (Ed.), (1975), *Psychological Testing in Personnel Assessment*. Epping, UK: Gower Press.

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

Jacobs, P.I. and Vandeventer, M. (1968). Progressive Matrices: An experimental, developmental, nonfactorial analysis. *Perceptual and Motor Skills, 27*, 759-766.

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

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

Jensen, A.R., Larson, G.E. and Paul, S.M. (1988). Psychometric ***g*** and mental processing speed on a semanitic verification test. *Personality and Individual Differences, 9(2)*, 243-255.

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

Kinsey, A.C. (1948). *Sexual Behavior in the Human Male*. New York: Saunders.

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. (1969/77). *Class and Conformity: A Study in Values (Second Edition)*. Chicago IL: Chicago University Press. (1st edition: Dorsey Press.)

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.

Kohn, M.L. and Schooler, C. (1982). Job conditions and personality: A longitudinal assessment of their reciprocal effects. *American Journal of Sociology, 87*, 1257-86.

Kohn, M.L., Slomczynski, K.M. and Schoenbach, C. (1986). Social stratification and the transmission of values in the family: A cross-national assessment. *Sociological Forum*, 1.

Lempert, W. (1986). Sozialisation und Personlicckeitsbildung in beruflichen Schulen, dargestellt am Beispiel der Entwicklung moralischer Orientierung. *Die berufsbildende Schule, 38*, 148-160.

Lempert, W., Hoff, E.H. and Lappe, L. (1990). *Occupational Biography and Personality Development: A Longitudinal Study of Skilled Industrial Workers*. Berlin: Max Planck Institute for Human Development and Education.

McClelland, D.C. (1951). *Personality*. New York: Sloane, Dryden, Holt. Reprinted by Irvington Publishers, New Jersey.

McClelland, D.C. (1961). *The Achieving Society*. New York: Van Nostrand.

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

McClelland, D.C. (1978). *Guide to Behavioral Event Interviewing*. Boston: McBer.

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

McClelland, D.C. and Dailey, C. (1973). *Evaluating New Methods of Measuring the Qualities Needed in Superior Foreign Service Workers*. Boston: McBer and Co.

McClelland, D.C. and Dailey, C. (1974). *Professional Competencies of Human Service Workers*. Boston: McBer and Co.

MacPherson, J.S. (1958). *Eleven Year Olds Grow Up*. London: University of London Press.

Maistriaux, R. (1959). *L'Intelligence et le Caractere*. Paris, France: Presses Universitaires de France.

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

Maxwell, J.N. (1961). *The Level and Trend of National Intelligence: The Contribution of the Scottish Mental Surveys*. London: University of London Press.

Maxwell, J.N. (1969). *Sixteen Years On*. Edinburgh: Scottish Council for Research in Education.

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

Moreland, K.L., Eyde, L.D., Robertson, G.J., Primoff, E.S. and Most, R.B. (1995). Assessment of test user qualifications. *American Psychologist, 50*, 14-23.

Nuttgens, P. (1988). *What Should We Teach and How Should We Teach It*? Aldershot: Wildwood House.

Ogbu, J.U. (1992). Understanding cultural diversity and learning. *Educational Researcher, 21*, 5-14.

Rasch, G. (1947). In G. Rasch (1980), *Probabalistic Models for Some Intelligence and Attainment Tests*. Chicago, IL: University of Chicago Press.

Raven, J. (1976). *Pupil Motivation and Values*. Dublin: Irish Association for Curriculum Development.

Raven, J. (1977). *Education, Values and Society: The Objectives of Education and the Nature and Development of Competence*. London, England, H.K. Lewis. Now available from the author at 30 great King St., Edinburgh EH3 6QH, Scotland..

Raven, J. (1980). *Parents, Teachers and Children: An Evaluation of an Educational Home Visiting Programme*. Edinburgh: Scottish Council for Research in

Education. Distributed in North America by the Ontario Institute for Studies in Education, Toronto.

Raven, J. (1984). *Competence in Modern Society: Its Identification, Development and Release*. Unionville, New York: Trillium Press.

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

Raven, J. (1988). Toward measures of high-level competencies: A re-examination of McClelland's distinction between needs and values. *Human Relations, 41*, 281-294.

Raven, J. (1989a). Parents, education and schooling. In C. Desforges (Ed.), *British Journal of Educational Psychology, Monograph Series No.4, Special Issue on Early Childhood Education* (47-67).

Raven, J. (1989b). The Raven Progressive Matrices: A review of national norming studies and ethnic and socio-economic variation within the United States. *Journal of Educational Measurement, 26*, 1-16.

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

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

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

Raven, J. and Dolphin, T. (1978). *The Consequences of Behaving: The Ability of Irish Organisations to Tap Know-How, Initiative, Leadership and Goodwill*. Edinburgh: The Competency Motivation Project.

Raven, J., Hannon, B., Handy, R., Benson, C. and Henry, E.A. (1975). *A Survey of Attitudes of Post Primary Teachers and Pupils, Volume 2: Pupils' Perceptions of Educational Objectives and their Reactions to School and School Subjects*. Dublin: Irish Association for Curriculum Development.

Raven, J., Johnstone, J. and Varley, T. (1985). *Opening the Primary Classroom*. Edinburgh: Scottish Council for Research in Education.

Raven, J., Raven, J. C., & Court, J. H. (1998, updated 2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General Overview*. San Antonio, TX: Harcourt Assessment.

Raven, J., Raven, J. C., & Court, J. H. (2000, updated 2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3: The Standard Progressive Matrices,Iincluding the Parallel and Plus Versions*. San Antonio, TX: Harcourt Assessment.

Raven, J., Ritchie, J. and Baxter, D. (1971). Factor analysis and cluster analysis: Their value and stability in social survey research. *Economic and Social Review*, 367-391.

Raven, J. and Sime, J. (1994). *Computerised Edinburgh Questionnaires*. Edinburgh, The Competency Motivation Project.

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

Raven, J. and Varley, T. (1984). Some classrooms and their effects: A study of the feasibility of measuring some of the broader outcomes of education. *Collected Original Resources in Education, 8(1)*, F4 G6.

Raven, J.C. (1939). The RECI series of perceptual tests: An experimental survey. *British Journal of Medical Psychology, XVIII*, Part 1, 16-34.

Ree, M.J., Earles, J.A. and Teachout, M.S. (1994). Predicting job performance: Not much more than ***g***. *Journal of Applied Psychology, 79*, 518-524.

Richardson, K. (1991). Reasoning with Raven – in and out of context. *British Journal of Educational Psychology, 61*, 129-138.

Roberts, E.B. (1968). A basic study of innovators: How to keep and capitalize on their talents. *Research Management, XI*, 249-266.

Rogers, E.M. (1962/83). *Diffusion of Innovations*. New York: Free Press.

Scottish Council for Research in Education (1933). *The Intelligence of Scottish Children*. London: University of London Press.

Scottish Council for Research in Education (1949). *The Trend of Scottish Intelligence*. London: University of London Press.

Scottish Council for Research in Education (1953). *Social Implications of the 1947 Scottish Mental Survey*. London: University of London Press.

Sigel, I.E. (1986). Early social experience and the development of representational competence. In W. Fowler (Ed.), *Early Experience and the Development of Competence. New Directions for Child Development, No.32*. San Francisco: Jossey-Bass.

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.

Stansbury, D. (1980). The record of personal experience. In T. Burgess and E. Adams, *Outcomes of Education*. Basingstoke: MacMillan Education.

Sternberg, R.J., Wagner, R.K. and Okagaki, L. (1986). Practical intelligence: The nature and role of tacit knowledge in work and school. In R.J. Sternberg (Ed.), *Practical Intelligence*. New York: Cambridge University Press.

Stufflebeam Joint Committee on Standards for Educational Evaluation (1981). *Standards for Evaluations of Educational Programs, Projects and Materials*. New York: McGraw Hill.

Styles, I. (1999). The study of intelligence – the interplay between theory and measurement. In M. Anderson (Ed.), *The Development of Intelligence*. Sussex, England: Psychology Press.

Taylor, C.W. (1971). *All of Our Children are Educationally Underprivileged*. Salt Lake City: Department of Psychology, University of Utah.

Taylor, C.W. (1976). *Talent Ignition Guide*. Salt Lake City: University of Utah and Bellvista Public School.

Taylor, C.W. and Barron, F. (Eds.). (1963). *Scientific Creativity: Its Recognition and Development*. New York: Wiley.

Taylor, C.W., Smith, W.R. and Ghiselin, B. (1963). The creative and other contributions of one sample of research scientists. In C.W. Taylor and F. Barron (Eds.), *Scientific Creativity*. New York: Wiley.

Tellegen, A., Lykken, D.T., Bouchard, T.J., Wilcox, K.J., Segal, N.L. and Rich, S. (1988). Personality similarity in twins reared apart and together. *Journal of Personality and Social Psychology, 54(6)*, 1031-1039.

Tharp, R.G., Jordan, C., Speidel, G.E., Au, K.H.P., Klein, T.W., Calkins, R.P., Sloat, K.C.M. and Gallimore, R. (1984). Product and process in applied developmental research: Education and the children of a minority. In M.E. Lamb, A.L. Brown and B. Rogoff (Eds.), *Advances in Developmental Psychology, Volume III* (91-144). Hillsdale, NJ: Lawrence Erlbaum.

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.

Trevarthen, C. (1992). The self born in inter-subjectivity: The psychology of infant communicating. In U. Neiser (Ed.), *Ecological and Interpersonal Knowledge of Self*. New York: Cambridge University Press.

Van Beinum, H. (1965). *The Morale of the Dublin Busman*. London: Tavistock Institute of Human Relations.

Vernon, P.A. (1989). *Speed of Information-Processing and Intelligence*. Norwood, NJ: Ablex.

Vernon, P.A. (1991). Studying intelligence the hard way. *Intelligence, 15*, 389-395.

Vernon, P.A. (1993). Intelligence and neural efficiency. In D.K. Detterman (Ed.), *Current Topics in Human Intelligence, Volume 3*. Norwood, NJ: Ablex.

Vernon, P.E. and Parry, J.B. (1949). *Personnel Selection in the British Forces*. London: University of London Press.

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

Vodegel-Matzen, L.B.L., van der Molen, M.W. and Dudink, A.C.M. (1994b). Error analysis of Raven test performance. *Personality and Individual Differences, 16(3)*, 433-445.

Waller, N.G., Kojetin, B.A., Bouchard, T.J., Lykken, D.T. and Tellegen, A. (1989). Genetic and environmental influences on religious interests, attitudes and values: A study of twins reared apart and together. *Psychological Science, 1(2)*, March, 138-142.

Vygotsky, L.S. (1981). The genesis of higher mental function. In J.V. Wertsch (Ed.), *The Concept of Activity in Society Psychology*. Annank, NH: Sharpe.

Winter, D.G. (1973). *The Power Motive*. New York: Free Press.

Wright, B.D. and Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement, 29*, 23-48.