
The Tragic Illusion: Educational Testing

John Raven

Trillium Press

Unionville, New York

Toronto, Ontario

Oxford Psychologists Press

Oxford

England

ACKNOWLEDGEMENTS

I am deeply indebted to Jean Raven and Harold Berlak for their assistance in preparing this particular book. Sections of the material have previously been published in *Competence in Modern Society* (H.K.Lewis, 1984), the *British Journal of Educational Psychology, Monograph Series, No. 3* (edited by Black and Dockrell), and *Assessing Achievement: Toward the Development of a New Science of Educational Testing*, edited by Harold Berlak (SUNY Press, New York, 1990). Over the longer term, I am of course particularly indebted to my wife, but also to Lyle Spencer and David McClelland, and to Bram Oppenheim, Michael Fogarty, Bryan Dockrell, Betty Adams, and Stanley Nisbet. Funding has come from the Economic and Social Research Institute and the Scottish Education Department.

Copyright © 1991, J. Raven
All Rights Reserved.

Trillium Press

First Avenue
Unionville, NY 10988
(914) 783-2999
(914) 782-6359

Trillium Press

203 College St., Suite 200
Toronto, Ont M5T 1P9
(416) 351-0602
(416) 633-3010

Oxford Psychologists Press

311 Banbury Road
Oxford OX2 7JH
England
(0865) 510203

Printed in the United States of America by the Royal Fireworks Press
of Unionville, New York.

ISBN: 0-89824-523-0

CONTENTS

CHAPTER 1:	4
<i>Some Illustrations of the Limitations of Conventional Measurement Paradigms and their Implications.</i>	
CHAPTER 2:	15
<i>The Need for Basic Reform of Assessment.</i>	
CHAPTER 3:	47
<i>The Nature of Competence.</i>	
CHAPTER 4:	55
<i>A Formal Model of Competence, Motivation and Behavior and its Assessment.</i>	
CHAPTER 5:	64
<i>Implementing Generic Competence Assessments.</i>	
CHAPTER 6:	76
<i>The Question of Validity.</i>	
CHAPTER 7:	80
<i>Concluding Comments.</i>	
NOTES	89
REFERENCES	99

OVERVIEW

The first two chapters of this book discuss some limitations of the currently dominant measurement paradigm in education and psychology and the damaging effect it has on educational policy and practice. The third and fourth chapters outline a new model of competence, motivation, and behavior and a new framework for assessing their components. The fifth chapter describes the ways in which the assessment model developed in Chapter 4 can be operationalized. The sixth chapter presents evidence for the utility and validity of the new assessment procedures. The final chapter first explores the implications of what has been said earlier in the book for the way R&D is organized and then picks up a few of the specifically educational issues which have been mentioned in passing.

SUMMARY

In the first chapter, it is initially shown that conventional assessment procedures are unable to do justice to the most important outcomes of any educational process worth the name. In the first place, pupils' accomplishments are too diverse to be adequately documented by scores on any small number of "scales". Secondly, these procedures are unable to give pupils credit for having developed high level competencies — such as the ability to undertake activities which would merit the name "academic" (eg the ability to contribute in significant ways to genuine scientific, historical, archaeological or sociological studies) or the ability to do such things as communicate, take initiative, or lead.

One consequence of this is that pupils (and therefore teachers) are unable to get credit for possessing (or having fostered) high level competencies in the certification and placement process. This is important because it has the further consequence that schools tend to neglect the activities required to foster such competencies. Another is that it has precluded the development of the tools which teachers need to help them mount the kinds of educational program which are required to foster high level competencies.

The absence of appropriate assessment procedures has also resulted in some extremely misleading and damaging research and policy studies.

Since the most important outcomes of educational programs—both positive and negative—cannot be detected using conventional measures, studies of the kind advocated by the Joint Committee on Evaluation Standards fail to reveal either the harm which many schools and colleges do to students or the benefits of alternative approaches.

The author does not, however, argue against testing and assessment *as such*. On the contrary, he emphasizes that teachers do need to be able to identify students' interests, talents, and accomplishments, that administrators do need to be able to find out whether public money is being well spent, and that students do need to be able to identify what they have learned to do in the course of their education. To provide this information new kinds of measures are required.

The second chapter develops a number of other arguments demonstrating that it is vital to evolve a new assessment paradigm. Among these are that, without it, we will continue to try to administer the educational system using assessment procedures which lack both construct and predictive validity and which deflect schools from their main goals. We will remain unable help more teachers to mount effective educational programmes. And we will continue to be unable to prevent a disproportionate number of the wrong kind of people finding their way into influential positions in society.

In Chapter 3, the author shows that it is necessary to conceptualize high level competencies as *motivational dispositions*. This implies that they are *value-based* (rather than value-free) qualities. They have a series of interpenetrating, cumulative and substitutable, components. A two-stage (*not* a two-factor) measurement model is required to assess them: One must first find out what kinds of activity people "want" to undertake and then which components of competence they display whilst undertaking them.

These observations are formalized in Chapter 4. They lead, in the end, to the conclusion that we need to move away from any attempt to differentiate between people by reference to scores on a small number of internally-consistent "variables" and, instead, to adopt a *descriptive* framework (which has conceptually much in common with a chemical formula) in which we discriminate between people by reference to their dominant values, the competencies they display whilst carrying out activities they care about, the values the environments in which they are

placed trigger and release, and the concerns and competencies the tasks which they undertake lead them to develop.

Chapter 5 discusses the ways in which this model has been implemented in both programme evaluation and individual assessment. It is shown that, while externally-generated statements about people's values and competencies can provide useful information, better information is obtained by, in a sense, looking inside people's heads and identifying their motives, cognitions and feelings. This can be done using semi-structured interviews or procedures based on value-expectancy methodology.

Chapter 6 presents evidence for the validity of the information so obtained: The use of these methods has enabled us to document the effects of such things as both more and less effective forms of project-based education on outcomes which have previously been considered to be too intangible to capture ... and they have enabled us to show that the economic plight of America—and particularly Britain—(in comparison with, say, the Pacific rim countries) has much more centrally to do with values and the ability to sift and act on information than with the kinds of educational “failure” latched onto in reports like “A Nation at Risk”—and which policies like centralized prescription of curricula and National testing seek to remediate in such inappropriate and ineffective ways.

The last chapter argues that the barriers which have in the past prevented the requisite research and development being carried out merit urgent attention and that reform needs to begin in the universities. The universities need, in particular, to change both their teaching about the nature of science and the way understanding advances and the way in which they “educate” their students. However, if university staff *are* to transform their institutions in these ways, they themselves will need new forms of *political* competence. This observation reinforces one of the most important conclusions to emerge from the author's work on the nature of competence—namely that competence in modern society is *primarily* dependent on *political* competence.

Because what is said in the book has as many implications for educational practice as it does for assessment, the last few paragraphs of the last chapter catch a few of the loose ends that have been left dangling as the argument has progressed and refer readers to publications in which they can be followed up.

CHAPTER 1

SOME ILLUSTRATIONS OF THE LIMITATIONS OF CONTEMPORARY MEASUREMENT PARADIGMS AND THEIR IMPLICATIONS.

The Assessment of the Results of Good Educational Practice Defies Conventional Measurement.

In this section I will illustrate some of the limitations of current assessment paradigms from the work of one school class whose work we observed in the course of our research. We were asked to undertake the research because the body which is responsible for quality control in Scottish education—Her Majesty's Inspectorate of Schools—had found that, despite the reports and directives they had issued over the previous forty years, very little had changed^{1,1}. We were therefore asked—in the course of a project employing 1½ people for two years—to (a) find some examples of good practice, (b) portray the work of the teachers concerned in such a way that others could do likewise, (c) document the benefits—so that more parents, teachers, students and politicians could and would think it was worthwhile, (d) identify the barriers which had prevented change, and (e) suggest ways of overcoming the barriers. The results were published as *Opening the Primary Classroom*^{1,2}. The work described below comes from this study, although it could equally well have come from one of the High Schools whose work was studied in the course of the research which led to my *Education, Values and Society: The Objectives of Education and the Na-*

ture and Development of Competence^{1,3}. Unfortunately our notes on those schools were not stored in a way which enables me to use them here.

The educational activities undertaken by the students in the class whose work I will describe were unusual. But what is *most* significant about the work is not its rarity, but the fact that it represents one of only a very small number of classes where education was actually organized in the way recommended by numerous educators from the time of Parker and Dewey onwards—and in numerous government reports.

Most of the students' education took place in the course of interdisciplinary projects. These projects were very thorough-going. The students, as a group, carried out original investigations in the environment around the school. Their work inside their classrooms formed an integral part of these investigations. Within these overall projects, many students had personal projects, distinctive areas of specialization, and distinctive roles.

What was most striking about the approach was the teacher's distinctive concerns. She was not pre-occupied, as were most teachers, with coursework; with covering a syllabus. Instead, her attention focussed on the competencies which she hoped to help her students to develop through the activities they carried out. These competencies included reading, writing, spelling and counting. But they also included communicating, observing, finding the information which was needed to achieve goals (such information often having to be collected by observation or by talking to people rather than by reading books), inventing, persuading, and leading.

It is easiest to begin our discussion of the problems which this educational process poses for assessment by reviewing those aspects which are closest to the more widely discussed and assessed goals of education.

In the course of his environmentally-based project work, one student had become an expert on the distribution of different species of butterfly in the locality, their life cycles, and their relationship to their habitats. Another had become an expert on the history of a particular agricultural implement: he had related changes in the implement to a continuous—and apparently autonomous—series of improvements in the design itself and to changes in steel-making on the one hand, and patterns of agri-

culture on the other. Another student had become an expert on the relationship between improvements in that implement, the pattern of land use it demanded and facilitated, and changes in the social structure of the community. Yet another had become an expert on the current social structure of the area—who knew whom and what they talked about. Others had studied changes in the architecture and layout of the village and the occupations of its inhabitants.

The problems which these accomplishments pose for conventional measurement paradigms are almost insurmountable. The students' specialist knowledge simply would not show up on traditional attainment tests—indeed these students would get *low* scores on such tests because they would have devoted the time which other students would have spent mastering the knowledge required by the tests to these other activities. To do justice to the students knowledge, it would be necessary to administer a series of individualized tests which would tap each student's specialty.

However, these are the least of the problems which this work poses for the measurement paradigm which dominates education and psychology. More important than the unique store of specialist knowledge built up by the first student mentioned above was the fact that he had developed a selection of the *competencies* required to be a scientist. Among other things, he had learned to be sensitive to the cues which told him that he had an unresolved problem; he had developed the tendency to try to make glimmerings of insight on the fringe of consciousness explicit (indeed he would wake up at night in an effort to do this); he had learned to *invent* ways of making observations; he had learned to notice things which no one had noticed before; he had learned not only how to find information in journals, but also how to use what he did find to stimulate that kind of lateral thinking which is required to make use of the information that is obtained; he had learned to solicit and make use of the ideas of his fellow students and “ignorant” people in the community; he had learned to write to, telephone, and visit university lecturers who were interested in the same problem and he had spoken to them as equals; he had sharpened up his ideas by sparring with them; he had learned that he had a right to ask new questions and not merely answer other people's; he had learned that he could both ask and answer questions; he had learned to tolerate the frustrations which are involved in trying

to find better ways of thinking about things; and he had learned to invent ways of organizing and summarizing his data and communicating it to others—and not just in writing.

The competencies listed in the last paragraph are a sub-set of the competencies which cumulate to result in the effective pursuit of almost any goal which may be valued, and they can, to a degree, be substituted one for another^{1.4}. We have come upon them here in connection with discipline-based studies, but we could equally well have encountered them as a result of examining other activities which people might value and be motivated to undertake. But, pursuing the academic-discipline-oriented line of enquiry on which we have embarked, it is important now to note that the second student mentioned above had developed a different sub-set of these self-motivated preoccupations, sensitivities, thoughtways and perceptions in the course of undertaking an original historical study. The third had developed a similar—but by no means identical—selection of the competencies needed to be an excellent sociologist of one kind or another. And so on for the other students.

If our traditional assessment procedures are unable to cope with the problem of idiosyncratic, specialist, high-level, new, knowledge, they are even less able to document the growth of the subtle skills, motivated habits, thoughtways and pre-occupations which go to make up the repertoire of the competent scientist, historian, sociologist, photographer, reporter, cook, or mother.

But even this does not exhaust the problems which the educational process in which these students were engaged pose for assessment. The students had worked as a group. They had developed specialized roles in that group. In the process they had developed the competencies needed to function effectively in those roles. One student had become good at co-ordinating the activities of others. Another at putting others at ease and smoothing over interpersonal difficulties. Another as a negotiator. Another at presenting the results of other people's work to external visitors—a communicator rather than an original researcher. And so on. In the course of undertaking these activities all students learned to communicate, to invent, to make their own observations, to work with others, to lead and to follow.

These competencies defy conventional measurement. This is of the greatest importance. Without means of assessing these qualities, even students who have come through such educational programs cannot know that they are different from students who have come through other educational programs. Still less can they identify the ways in which they are different from them: they cannot know that they think differently, see things differently, have different priorities, tend to work differently with others or that they can do different things. Without means of assessing these qualities, teachers cannot build on the competencies which have been fostered in the course of one project, in the next one. Students cannot get credit for the talents they have developed when they come to scramble for a job or for entry to courses of further education. Teachers cannot get credit for having fostered these competencies in accountability exercises. At a societal level, we cannot prevent people who do not possess important high-level concerns and qualities like those mentioned being appointed to influential positions. But, perhaps worst of all, the absence of means of measuring these qualities limits our conception of what education *is*. In the current scheme of things, even the word "academic" fails to denote activities in which people observe, think, find better ways of thinking about things, make judgments, muster arguments, or even communicate important material effectively.

Before moving on it is worth noting that the fact that we have been able to make these observations shows that the measurement problem must, in principle, be soluble. What we have done in the course of making these observations is: (i) observed students as they were undertaking tasks they cared about, (ii) recorded the multiple and substitutable competencies they displayed whilst undertaking those tasks, (iii) adopted a *descriptive* approach when reporting our observations instead of trying to report them as "scores" on "variables", and (iv) distinguished between students in terms of the activities they cared about and the competencies they displayed whilst undertaking those tasks. The central message of the following chapters is that it is possible to build an alternative measurement paradigm based on precisely what we have done here.

Conventional Evaluation of Educational Programs Leads to Inappropriate Conclusions.

Having illustrated some of the problems which good educational practice poses for evaluation and assessment, I will now underline the need for an alternative measurement paradigm by citing some extremely misleading conclusions which have been drawn from evaluation studies which have been carried out within the dominant measurement paradigm.

Many studies have shown that Open Education depresses scores on conventional reading, writing and arithmetic tests. Stallings^{1.5} found the same thing when evaluating Headstart Follow Through. However, she also documented something else: Open Education *increased* the ability to perceive and think clearly, as measured by the *Raven Progressive Matrices*^{1.6}. Now, if one gets that kind of result on a test which is designed to be as little dependent as possible on educational experience, what would happen if one measured the outcomes which Open Educators hold most dear—leadership, the ability to work with others, the ability to communicate, and the ability to understand and influence society—and used measures which were *sensitive* to educational experience? The chances are that one would find substantial desirable effects of the program being evaluated.

In this context it is important to note that even the use of tests having higher ceilings in the areas of reading, writing and arithmetic might produce contradictory results. Traditional tests of reading do not measure such things as the ability quickly to discard books and articles which are irrelevant to one's purposes, the ability to use structure to find relevant information, or the ability to use what one does read as a stimulus to lateral thinking. Educational processes which promote the development of such high-level reading competence may well depress scores on tests which measure the willingness to read a passage one does not care about and remember answers which one could easily look up if one needed them. As Bullock^{1.7} and McClelland^{1.8} have observed, traditional tests of English do not measure the ability to communicate effectively ... and why would anyone who cared about such things as the use of innuendo, allusion, and context to influence a target audience

worry about irrelevant “rules of grammar”^{1.9}? (Cockcroft^{1.10} has made similar points about arithmetic).

It is all too easy to read the last paragraph, react by saying “Of course”, and yet fail to draw the most important conclusions. The dominant wisdom, as articulated in the Joint Committee’s *Standards for the Evaluation of Educational Programs, Projects and Materials*^{1.11}, asserts that evaluation studies should only report results which are obtained with reliable and valid tests. Unfortunately it is difficult to see how an *evaluation* can be regarded as valid if it fails to comment on the most important outcomes of an educational process—even if those outcomes cannot be measured with the available reliable and valid tests. Many of the most important outcomes of educational programs will, in fact, be of this kind because—like economic and social development—they will take many years to show up. Failure to discuss outcomes which cannot at present be measured with “reliable and valid tests”, therefore, deflects policy discussion away from the educational processes and outcomes which are most important and focuses it on those which are easiest to measure. The Joint Committee’s lack of concern with this issue, taken together with the fact that the Stanford Research Institute (where Stallings worked) lost their contract when they pointed out that, in order to do a competent job of evaluation, they needed to develop measures of other outcomes, suggests that there is a widespread failure to appreciate the implications of *not* finding ways of assessing a wider range of the outcomes of the educational process. One glaring consequence of this oversight is the current infatuation with low-level tests as indices of educational effectiveness. The fact that these fail to register any educational outcomes worth the name—such as those discussed in this chapter—somehow escapes notice. Notwithstanding the sentiments echoed in the introductory paragraphs of most reports, the attention of politicians, school reformers, administrators, teachers, parents and students is thereby focussed entirely on low-level outcomes which, as we will see later, are of very little importance to the individuals concerned or the societies in which they live. The result is that the reports and policies themselves have been correctly described by the former British Prime Minister Edward Heath as “a con trick”. There is no connection between the high-level objectives which are rightly recognized, the problems which are correctly identified, and the “quality control” measures which, it is

claimed, will lead to an amelioration of the problems and effective goal achievement. Yet the introduction of quality control measures which focus attention on outcomes other than those which it is hoped to achieve cannot fail to lead everyone in the wrong direction.

Our second illustration of misleading conclusions being drawn by researchers who have been too dependent on current measurement paradigms when evaluating educational activities comes from the “mastery learning” literature. If one asks what those students engaged in “mastery learning” programs who finish their work first do whilst they are waiting for the others to catch up, one learns that they do such things as help their slower classmates or go play football. It follows that, not only do these students learn that they are “smart” (able to complete routine tasks quickly): they also go on to develop additional competencies—such as those needed to tutor others or to play football. In other words, the variation between students’ competence in these other areas has actually been *increased*. Thus, far from mastery learning having equalized educational attainments, it has simply shifted the variance to unexamined domains. More seriously still, in an effort to get students to do things they are not good at, mastery learning programs—like mandatory testing of “basic skills” before students are allowed to move on to other tasks (such as undertaking projects) and most “compensatory” programs which require “special” students to spend all their time on “the basic skills”—exclude “slower” students from opportunities to exercise and develop their most important and distinctive abilities. In this way such procedures create a self-fulfilling situation in which it is only *possible* to distinguish between students in terms of a single “ability”. “Slow” students are excluded from opportunities to practice inventing, persuading, leading, communicating, and listening in the course of carrying out activities they care about. They are thereby deprived of the opportunity to demonstrate that, given a non-stultifying environment, they are, despite their “low ability”, actually good at doing some of these other things. This procedure *creates* a powerful general factor in “abilities”. Had the concepts of achievement and learning employed by those who have promoted and evaluated mastery learning been broader and multi-faceted, had they tried to foster multiple talents, and had the evaluators measured a wide range of educational outcomes, their conclusions would have been very different.

Mastery learning can also be used to illustrate another fundamental flaw in current test theory. Notwithstanding what has just been said about the evaluations of experimental mastery learning programs, the basic philosophical stance on which mastery learning is based is correct: if it is indeed important for certain students to master some skill, then they had better master it. If they have not done so, that is an indictment of the program. If it is not important for them to master the skill or material, then they should not have been on the course. Failure, therefore, means that either they had inadequate educational counselling and guidance or that they have been incompetently taught. Either way, a half-learned skill is an indictment of the students' education, not the student. The implications for current measurement models are devastating—because most of them are built around the assumption that a “normal” curve in educational outcomes is not only to be expected, but is actually desirable. Whatever may be the case for natural abilities, that assumption cannot be justified for *educated* abilities.

The third illustration of seriously erroneous conclusions being drawn from the use of traditional measures in evaluation studies comes from the work of Coleman^{1,12} and the International Association for the Evaluation of Educational Achievement^{1,13}, which were widely interpreted to mean that “schools make no difference”. Both sets of studies used traditional tests to measure educational achievement. Both then set out to identify the teacher variables which “made a difference”. They then discovered that, once the effects of home background and “ability” had been statistically removed, “schools made no difference”. This absurd conclusion derives from the fact that the tests used were explicitly designed to measure content which was common to all educational programs. Specialist knowledge introduced by particular teachers was unlikely to show up. In other words, the *most likely* differential effects of teachers could not show up.

When we ourselves set out to measure idiosyncratic learning—to document *what* students had learned whilst working with particular teachers (rather than whether they had learned something which some investigator thought they ought to have learned) and included undesired and undesirable as well as desired and desirable learning—we found dramatic variation from one teacher to another in students' self-images, priorities, values, and patterns of competence. Students in different classes had

learned quite different things, many of them being not only unintended but also dysfunctional. And, in the course of a single program of project-based educational activity, what any one student had learned was quite different from what other students had learned. The students had chosen to attend to different things—that is, their *environments* had been different.

Conclusion.

I have now shown that traditional attainment tests are unable to do justice to the outcomes of any educational process worth the name, that they are unsuitable for identifying students' talents, that they do not help teachers to administer educational programs, that reliance on tests in attempts to improve the quality of education directs teachers' attention away from the very goals which are emphasized when identifying the shortcomings of the educational system and toward low-level goals, that they cramp our concepts of education in general and academic education in particular, and that, used routinely in evaluation studies, they lead to conclusions which are, at best, misleading, and, in practice, often not in the long-term interests of either the students concerned or society itself.

It remains to add that teachers *do* need to be able to identify students' interests and talents in order to be able to implement effective educational programs. They *do* need to be able to monitor progress toward their goals so that they can take corrective action when necessary. Students *do* need to be able to identify what they have learned and how they are different from others who have been enrolled in other types of program. They *do* need to be able to get credit for their accomplishments. Teachers and administrators *do* need to be able to identify the relative merits of different types of educational program. They *do* need to be able to find out which aspects of those programs are working well and which are not, and how to improve them^{1,14}.

Clearly, a new measurement paradigm is urgently required.

But the examples I have given do more than offer criticism. They also suggest a basis on which it might be possible to build an alternative "measurement" paradigm: they show that, instead of using "reliable and valid" measures of a few outcomes of an educational process, it would be possible to sample the entire domain of possibly relevant out-

comes^{1.15}. This suggests that an alternative “measurement” paradigm might be grounded in ideographic description—as in Chemistry—instead of in variables—as in Physics. We will build on this insight in the following chapters. However, because we have as yet neither encountered all the damaging effects of the current measurement paradigm, nor evidence of the basis on which an alternative might be built, I will now review some more of this evidence.

CHAPTER 2

THE NEED FOR BASIC REFORM OF ASSESSMENT.

In this chapter I will begin to develop seven reasons why it is vital for those interested in education to find ways of assessing such qualities as initiative and the ability to understand and influence society — qualities which have eluded psychometricians since the dawn of psychology. The seven reasons which will be discussed are (somewhat telescoped): that what is assessed controls what gets attention in schools and what gets discussed in policy debates, that teachers and students need appropriate tools to monitor progress toward their goals and achieve them effectively, that current academic assessments lack construct and predictive validity and do not permit us to exclude incompetent people who have inappropriate concerns from influential positions in society, and that current assessment procedures lack objectivity.

1. What is assessed in the certification and placement process controls what happens in schools, and currently drives any education worth the name out of schools.

In research which has now been replicated in England^{2.1}, Scotland^{2.2}, Northern Ireland^{2.3}, Eire^{2.4}, Belgium^{2.5} and the United States^{2.6} students, their teachers, their parents, ex-students and employers have been asked to say how important is each of a number of potential objectives of education. A selection of the results, obtained from male Irish adolescents, is shown in Chart I.

Chart I
Importance of Objectives: Boys.
 Percentages of boys rating each objective "Very Important."

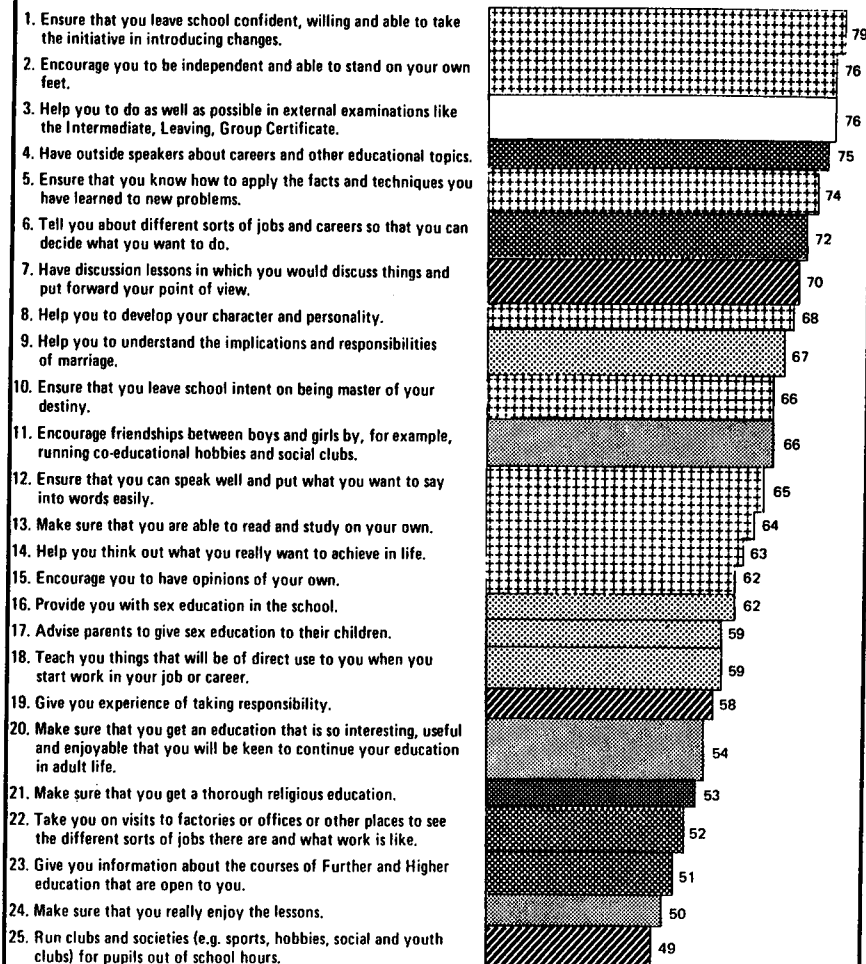
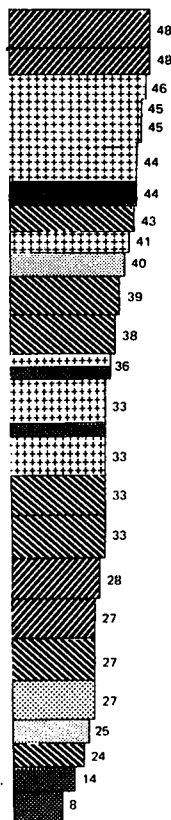


Chart I (Contd)

26. Educational visits in connection with your subjects – such as to see chemical plants, museums and theatres.
27. Give you a say in how the school is run.
28. Make sure you are confident and at ease when dealing with people.
29. Ensure that you can express yourself clearly in writing.
30. Help you to develop a considerate attitude towards other people.
31. Make sure you go out into the world determined to make Ireland a better place in which to live.
32. Teach you about what is right and wrong.
33. Provide facilities for pupils to do their homework at school.
34. Help you to get on with other people.
35. Encourage you to have a good time.
36. Help you to take an interest in and to understand what is going on in the world now.
37. Introduce you to new subjects, e.g. philosophy, sociology, archaeology etc.
38. Encourage you to have a sense of duty towards the community.
39. Make sure that you leave school aware of the prolonged struggle for Irish freedom and determined to uphold the ideals which inspired it.
40. Ensure that you feel confident and at ease when dealing with figures and numbers.
41. Enable you to develop an interest in subjects other than those studied for examinations.
42. Teach you about a wide range of cultures and philosophies so that your own can be seen to be only one of many.
43. Have Project work, that is work in which you have to make something or do some investigation and write it up.
44. Make sure you have opportunities to give short lectures and talks to the rest of your class.
45. Ensure that you are aware of aspects of school subjects which you do not have to know for the examinations.
46. Teach you about bringing up children, home repairs, decorating and so on.
47. Take you on holidays in this country or abroad.
48. Run courses for adults as well as young people.
49. Have rules about the clothes and hairstyles you may wear in school.
50. Have rules about the sort of things you may do outside of school hours



For technical reasons the base varies from item to item, but the base for most figures is 5–600. See technical report (Raven, 1975) for exact numbers.

KEY TO SHADING IN CHARTS I–IV



Self-Initiated Competencies and Qualities of Character



Guidance



Information which can be used in everyday jobs and lives



Academic knowledge content



Prescriptive Oughts



Activity Methods, mostly directed towards developing Self-initiated Competences



Emotional Feelings at the time

This Chart is of interest for a number of reasons, of which only three will be discussed here.

The first of these is that, in the boys' opinion, the most important goals of education include developing the confidence and initiative required to introduce change, independence, the ability to apply knowledge to new problems, and to develop their characters and personalities. These data not only challenge the widely-held belief that schools should not be concerned with social change: they specifically call into question Goodlad's conclusion that there is little "demand" for schools to pursue personal development goals^{2.7}. It would seem that schools' customers *do* want them to foster the qualities which many educators have long argued that they should foster.

Secondly, the students thought that career guidance was very important. When we asked them to say whether schools should do more to achieve each of these goals, it emerged that they thought that career guidance was the most neglected of the goals they considered most important. Flanagan^{2.8} and Bachman^{2.9} in their longitudinal studies have both found that these opinions are largely correct. An enormous amount of time is lost, and a great deal of distress is caused, as people flounder around in the job market until they find a niche which at last taps their (previously unidentified) personal interests and talents. Furthermore, work is typically the first occasion on which people have an opportunity to identify *and develop* their true potentials. Students (like most researchers, including Flanagan himself^{2.10}) are, however, wrong to conceptualize the problem as being one of *career* guidance, because any occupational group encompasses a wide range of people who have very different motives and talents and who do very different things. Thus, some psychologists run companies, some do research, some teach school, some edit journals, some are politicians, and some are secretaries. Calvin Taylor has shown that there are 12 types of outstanding research scientist and 25 types of outstanding physician^{2.11}. All are different types of people and they have very different talents and areas of specialist knowledge. What the data suggest is, therefore, that what students really need is some means of identifying, developing, and finding ways of utilizing and being rewarded for their own particular talents.

Chart II
Importance of Objectives: Girls.
 Percentage of girls rating each objective "Very Important."

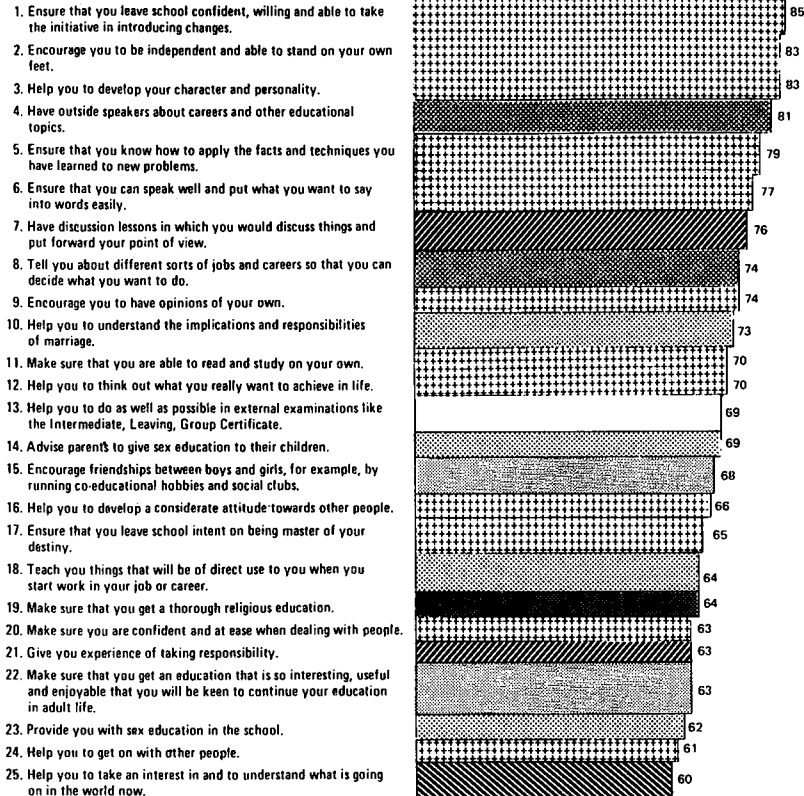
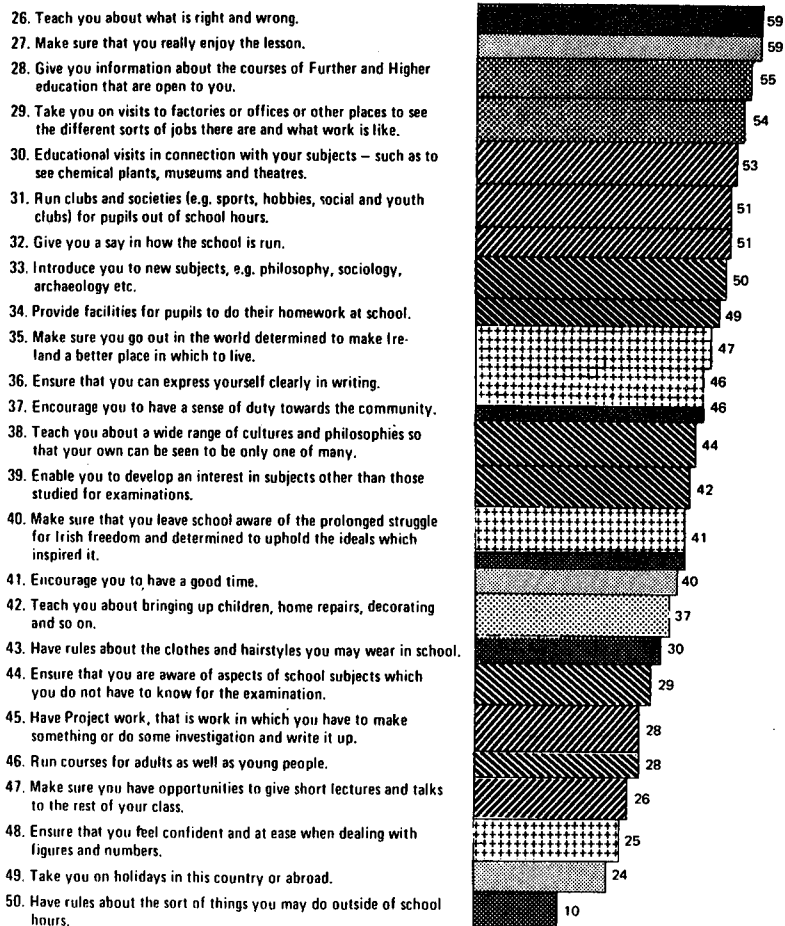


Chart II (Contd)

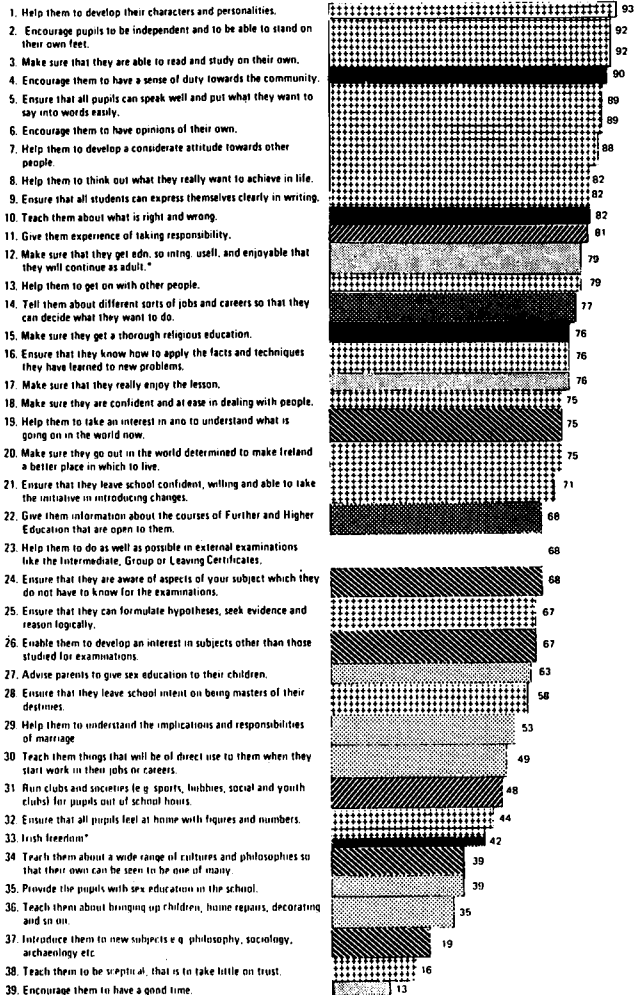


For bases see footnote to Chart IV.

Chart III

The Importance of Educational Objectives.

Percentage of teachers saying each objective "Very Important" for "more academic" pupils.



*These two items have been abbreviated from the items on the questionnaire which read:

"Make sure that they get an education that is so interesting, useful and enjoyable that they will be keen to continue their formal education in adult life"

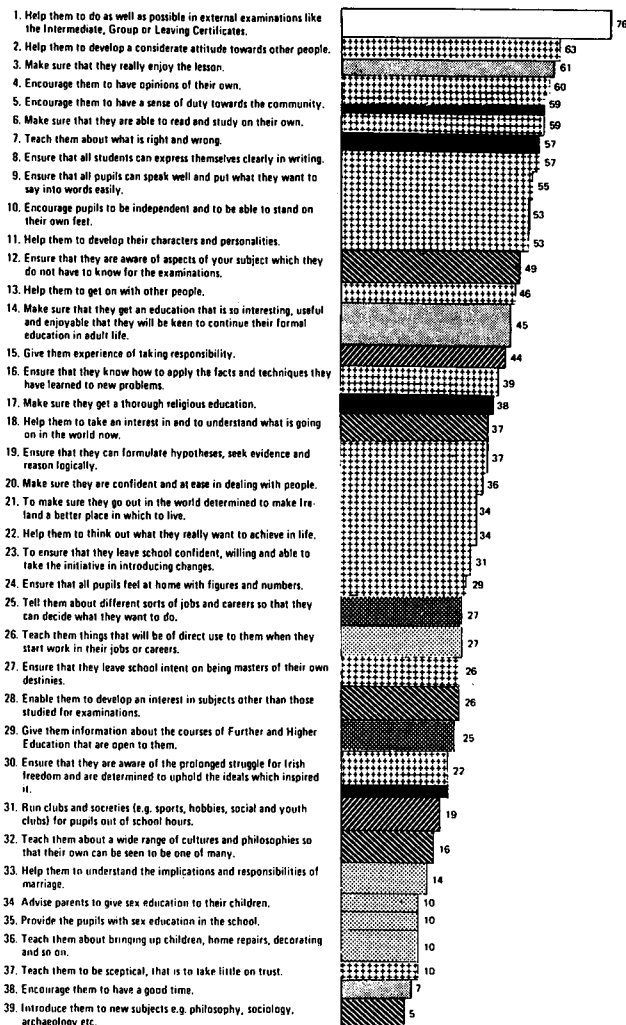
"Ensure that they are aware of the prolonged struggle for Irish freedom and are determined to uphold the ideals which inspired it."

Weighted base 1 = 100% All teachers rating objectives for "more academic" pupils: 612.

Chart IV

Objectives Receiving Most Attention.

Percentage of teachers (other than heads) who Tried Very Hard to Achieve Each Objective in Their Own Lessons With the "More Academic" Pupils.

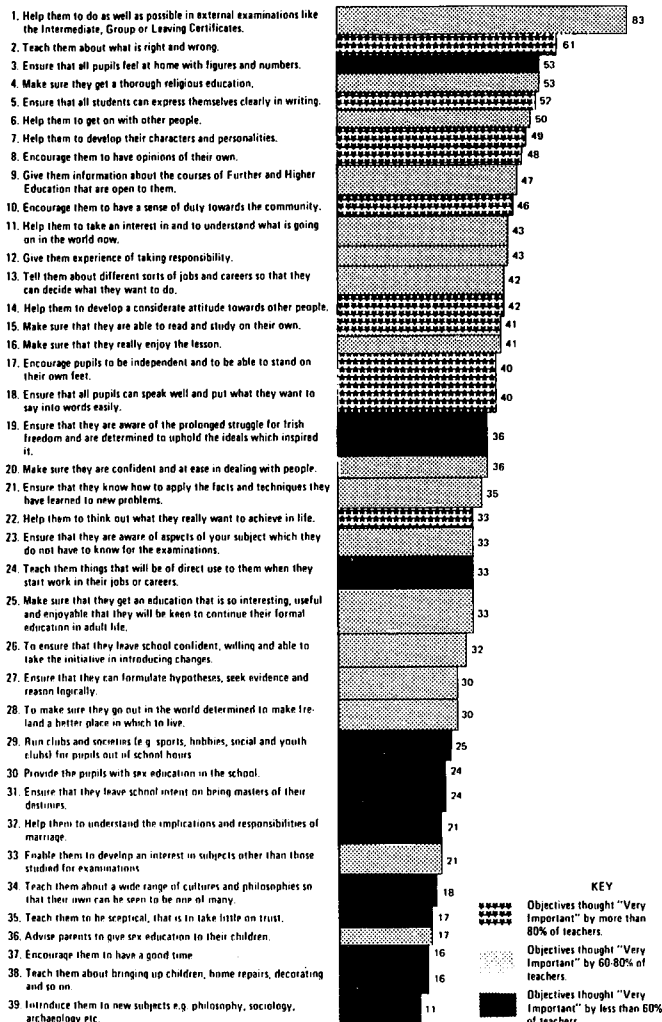


Weighted base (= 100%) all teachers other than heads rating objectives for "more academic" pupils: 528.

Chart V

Success with which Educational Objectives are Attained.

Percentage of teachers saying Education "Very Successful" or "Moderately Successful" in Achieving each Objective with the "More Academic" Pupils.



Weighted base (100%)

All teachers answering for the "more academic" pupils. 617

The third thing to be noted from the Chart is that, while boys put getting through examinations in second place, they put learning more about academic subjects than is required for tests and studying non-examined academic subjects near the bottom of their priorities (actually in positions 45 and 41 respectively). This discrepancy, taken together with other information—such as the annoyance they felt if they were given a lesson which was informative and enjoyable but did not relate to the syllabus—suggests that what they mean when they say they want an academic education is a high grade. The content is unimportant. Goodlad's^{2.12} data amply support this conclusion.

All of these results are of the greatest importance: male adolescents *do* want change, they do want schools to pursue personal effectiveness goals, they do want schools to help them to identify, develop and get formal recognition for their own personal talents, and they place no great emphasis on the *content* of what is commonly regarded as an academic education. What they mean when they say they want schools to focus on academic goals is that they want certificates which will buy entry to jobs, and these certificates can, at present, only be obtained through “academic” studies (which, in practice, do not merit such a designation).

The obvious next question is whether male students' views are shared by females, parents, teachers and employers. The girls' data are given in Chart II, and it is clear that the answer to the question just raised is a resounding “yes”.

Data are not available for these pupils' parents but data from parents are available from an earlier survey in Great Britain^{2.13}, contemporary work in the US^{2.14} and later work in the US^{2.15}. Comparative tables are available in *Education, Values and Society: the Objectives of Education and the Nature and Development of Competence*^{2.16}. Suffice it to say that, while there are important differences, the parents' data is similar to the pupils' data.

There are, however, significant differences between the data obtained from pupils and parents on the one hand and teachers and principals on the other. Teachers' educational priorities are shown in Chart III. The first thing to note is that, while these have a more prescriptive ring about them than the students, teachers still think that the character development goals of education are very important.

Given this broad measure of agreement, the next question becomes: "Why are these goals neglected?". There are many reasons for this which it would not be appropriate to go into here^{2.17}. However what I will now do is present some data which suggest that, if these personal development goals are to be achieved, it will be necessary, among other things, to assess progress toward them as part of the certification and placement process.

As can be seen from Chart III, teachers place getting students through examinations in position 23. However, Chart IV shows that, according to the teachers (and we will shortly see that they are correct) examination-oriented activities dominate the school day. And Chart V shows that these are the *only* objectives which significantly more than half the teachers feel they attain even moderately well.

It would seem to follow from both the students' and the teachers' data that it is what is assessed in the formal certification and placement process which controls what happens in schools. This observation has been confirmed by others^{2.18}. If, therefore, we want to influence what happens in schools, we must influence what is assessed in the certification and placement process.

But do we want schools to focus on different goals? We have seen that students, parents and teachers think so (provided, of course, that the change does not jeopardize pupils' chances of getting a good job). But are they right?

In a series of studies (including an extensive review of the literature) we have confirmed that their opinions *are* largely correct. The qualities which they want schools to foster are those which distinguish more from less effective machine operatives^{2.19}, bus drivers^{2.20}, laborers^{2.21}, salesmen^{2.22}, farmers^{2.23}, teachers^{2.24}, managers^{2.25}, businessmen^{2.26}, doctors^{2.27}, social workers^{2.28}, engineers^{2.29}, scientists^{2.30}, officers in the armed forces^{2.31}, diplomats^{2.32} and politicians^{2.33}.

Not only are their priorities for education correct: they are also right when they say that these goals are neglected and poorly attained. The many studies which confirm their opinions have included classroom observation studies, surveys of educational practice, and evaluations of educational outcomes. They have been carried out in both the UK and US^{2.34}.

There are many reasons for the discrepancy between precept and practice which is now so apparent^{2.35}. However, only two of them are of direct concern to us here. (Anyone interested in finding ways of enabling schools to reach their main goals should, however, examine the other reasons for the discrepancy which are spelled out in *The Most Important Problem in Education is to come to Terms with Values*, which will be published by Trillium Press.)

One of the reasons why schools rarely foster qualities like initiative, the ability to work with others, and the ability to understand and influence society, is that people only learn to do these things by practicing them in the course of undertaking activities they care about. Yet there are no tools—ie assessment procedures—to help teachers to identify individual students' interests, priorities, and patterns of competence, to help them to invent appropriate developmental experiences for each student, and to help them to monitor their reactions. Without such tools, most teachers cannot manage, simultaneously, a large number (30-odd, in the case of the typical school class) of individualized, competency-oriented, educational programs.

Another important reason why these goals are neglected is that there is no way in which students can get credit in the certification and placement process for having developed these competencies. And, since teachers' reputations depend on their students' success, this means that teachers cannot get credit for having fostered them. Many teachers are acutely aware of the dilemma this poses, pointing out that they would be jeopardizing their students' life chances if they spent time fostering these qualities and thereby deprived them of opportunities to secure high grades on attainment tests. Parents, too, are often also acutely aware of the problem^{2.36}.

To make the point again in a way that moves us forward: what happens in schools is not determined by the wishes of parents, teachers, Inspectors, Ministers of Education or anyone else. It is determined by our knowledge of developmental processes, by the tools available to teachers, and, most importantly, by what is assessed in the certification and placement process which controls entry to further education and jobs.

It is vital to appreciate the full implications of our observations (1) that schools are failing to achieve their main *educational* goals, and, (2)

that what they do is primarily controlled by what is assessed in the certification and placement process. What these observations are telling us is that the main function of schools is not to educate but, as Jencks suggested, to legitimize the rationing of privilege^{2.37}. In other words, it is not assessment *per se* which controls what goes on in schools, but the use to which assessments are put—namely to allocate position and status.

Since this conclusion is as unwelcome as it is unexpected, it is important to cite some further evidence in support of it. For fifteen years, a series of committees in England deliberated about the reform of examinations. These committees never arrived at firm recommendations. Then the Minister for Education established a new committee and insisted that it make recommendations within six months. The resulting report^{2.38} is of great interest because of what is said between the lines. When I reviewed it I called it “a damning commentary on our society”. It first notes that there is plenty of educational evidence which shows that young people have a wide variety of potential talents and that schools are capable of fostering them. It then examines the qualities which people need at work and in society. It again notes the need for a wide range of talents and areas of specialization. It then notes that, if schools are to foster the wide range of talents which are available—and which are so badly needed by society—it will be necessary to have a wide range of courses directed toward different goals, covering different content, and making use of different educational processes. It next points out that, in order to testify to the achievement of all these outcomes, it will be necessary to create a wide variety of examining Boards offering a wide range of syllabi, tests, levels, and modes of assessment. The different modes will make it possible for the assessments to be geared to programs with different objectives, and allow a wide range of talents to be observed and recorded. So far, so good. All highly laudable, comprehensible and rational. But then the committee does an extraordinary—and at first sight inexplicable—thing. In a single sentence embedded in a summarizing paragraph it says that “the results (of this plethora of syllabi, programs, modes, and levels of assessment) will be expressed on a single scale of seven points in a subject area”. This recommendation in practice negates everything which has been said previously. What could have led the committee to make it? The only

suggestion I can make is that there are hidden sociological forces at work. It would seem that there is a sociological need to have a clear and unarguable criterion to legitimize the allocation of social and occupational position and status and that this sociological force is so powerful that it over-rides all rational and explicit educational and occupational considerations.

Although the evidence cited above shows that social forces mediated by the assessment process lead teachers to behave in ways which do not accord with their own priorities, most of the teachers we have interviewed refused to acknowledge the inevitability of the controlling power of assessment. Most of them simply wanted assessment to go away: 84% said that employers and universities should have their own selection procedures and leave schools free to get on with education^{2.39}. This is sociologically naive. Our own evidence—and common experience—is that they will end up teaching to the tests—and to whatever goals are assessed—regardless of who does the assessing.

There are good reasons why teachers are reluctant to come to come to terms with these sociological forces. Many people become teachers because they want to help young people to grow, because they want to communicate, and because they want to do a worthwhile job in the community. The idea that their *main* job is to allocate students' life chances is an anathema to them. Yet, if they were to entertain such an unthinkable thought, they might (with the aid of researchers and others) be able to find ways of harnessing these sociological forces in such a way that they would push them in the direction in which everyone wants them to go, instead of away from their goals. We would, after all, merely be doing what Watt did when he harnessed nasty, dangerous, scalding, wasteful steam to do useful work for mankind. My suggestion is that this could be done by inserting into the certification and placement process, measures of the very qualities which most people want the educational system to foster. We would then have an educational steam engine, still fuelled by sociological forces, but under the control of educators rather than in control of them^{2.40}.

The suggestion that the activities of teachers and pupils be controlled by including measures of important qualities like initiative and the ability to work with others in the assessment process, of course, creates serious dilemmas which need to be made explicit and thought through. Many

of these will be taken up later. But before doing so, it may be useful to remind the reader that this was not the only reason for emphasizing the need to find ways of assessing these competencies. Without better means of assessing them, most teachers will find that it is too difficult to implement multiple-talent educational programs^{2.41}, and they themselves will be unable to get credit for such things as finding better ways of meeting their students' needs in staff appraisal systems.

2. The assessment paradigm controls what shows up in policy evaluation, and thus what gets discussed in policy debate.

In the last section, I showed that the narrowness of the range of qualities which can be assessed using the currently dominant measurement paradigm in the certification and placement process has seriously constricting effects on what happens within schools. In this section, I will show that the restriction which our current measurement paradigm places on what is assessed in evaluation studies debases our concept of education and restricts the range of issues which get discussed when educational policy is being formulated.

To start with, some of the observations which were made in the last chapter may be summarized. There, I showed that educational "evaluations" can be misleading, if not immoral (in the sense that they harm individuals in the short term and society in the long term), if they are not broadly-based. I further suggested that the Joint Committee's injunction to use only tests which are reliable and valid for evaluation purposes tends to trap evaluators into such studies. This occurs because there are no reliable and valid measures of many of the most important outcomes—both positive and negative—of educational processes. I hinted that, without radical change in our measurement paradigm, it would not be possible to develop such measures—because the most important outcomes of education include the development of initiative, beliefs about how organizations should be run, and the competencies required to achieve one's goals effectively. In this context, I suggested that, if measures of such outcomes were included in evaluation studies, the bottom line would often be that, while the program being evaluated did improve students' scores on traditional tests, it simultaneously had effects which were—as authors like Goodman, Reimer, and Illich have repeatedly suggested—profoundly damaging^{2.42}.

A more subtle effect of using tests which have been constructed to satisfy the traditional criterion that “good” tests will have “normal” (ie Gaussian, chance) distributions, is that they are rarely suited to the task of demonstrating that the educational system as a whole, or any subsection of it, is *not* achieving important goals. One cannot get a “normal” distribution across the entire population when only a few teachers are fostering the talents one is concerned with. This is one reason why it has so rarely been demonstrated that schools, in general, stunt the development of qualities like initiative and the ability to work with others.

A corollary of this is that few of the tests which are currently available are suited to the task of demonstrating that the educational system—or any given teacher—is (or is not) producing a wide variety of different types of excellent student. To do this one would need to stand current concepts of discrimination on their side and demand tests which identify what each student is good at, rather than whether he or she is good at performing specific tasks selected by the teacher or evaluator.

There is another way in which the use of tests which incorporate the assumption that scores should be “normally” distributed leads evaluators to overlook important educational issues. The adoption of such tests as appropriate measures of *outcomes* encourages researchers to adopt multivariate designs in which it is taken as axiomatic that the *inputs* will, like the outcomes, vary *in strength* from place to place. A more plausible assumption would seem to be that a few teachers will have created classrooms in which relevant educational activities occur and others will not. One would therefore expect to find the desired (or undesired) outcomes in some places and not in others. Thinking in terms of input “variables” (which are likely to be monotonically related to outcome “variables”), instead of about the nature of the educational *processes* involved and their probable effects, hardly leads researchers to study what is going on in such a way as to be able to identify *which aspects* of the process are important and which are not—although this is often exactly what the administrator who commissioned the research really wanted to know. The adoption of a statistical (viz high-status) standard multivariate design seems somehow to absolve the evaluator from the responsibility of having to spell out the supposed chain of causal linkages between the “inputs” (eg money) and the outcomes to which those inputs are supposed to be related—such as improved test scores. For this reason, evaluation

studies rarely tell one much about how to improve educational processes. They give one still less insight into how to get from what *is* to what *might be*—for the constellation of activities which would yield the desired result may be very rare in the world as it exists: that is to say, in the world as it can be studied using a multivariate design. This does not mean either that the effects of the components of an effective program of education could not be studied at different sites, or that it would be difficult to create more appropriate educational programs.

One particularly important consequence of the widespread failure to examine educational processes (itself deriving from a pre-occupation with “variables”) and the institutional and sociological contexts in which they are embedded, is that researchers have often overlooked the fact that the failure of a particular innovation to achieve a desired outcome is often due, not so much to the inappropriateness of the input itself, as to the negation of the effects of that input by other aspects of the system. For example, one regularly comes across evaluations of curriculum-development projects in which a number of teachers have been trained in “new methods”. These teachers have then been dispersed around a number of schools. No steps have been taken to ensure that parents, other teachers, or students understand the distinctive goals of the project. Crucial equipment and books are often absent and there is often no support to help the teachers to implement the innovation wholeheartedly. The students spend much of their time with other teachers whose teaching may, at best, neither support nor reinforce the new methods, but who will often deliberately change their teaching with a view to undermining the “new-fangled” ideas of the project teacher. Believing that they are acting in the best interests of the students, these other teachers may increase the formal content of their teaching to counter the “damage” done by the innovator. Projects which have been able, simultaneously, to change both the content of teaching and the tests used to assess students and pass final judgment on them are conspicuous by their absence. For this reason, other teachers, parents, and students sense that the life chances of the “experimental guinea pigs” are being jeopardized, and intensify their efforts to help them reach traditional goals. Unaware of all this, the project evaluators administer their tests and add the project to the long list of “failed attempts at curriculum reform”.

One does not really want an evaluation to tell us what actually is the effect of some change, weakly implemented, in the absence of a supportive context, on outcomes of little individual or social importance—outcomes which are, in any case, only tangentially related to the goals of the innovation. One wants to know what the effects—both positive and negative—of some change in the educational process would be on important long-term outcomes (like the economic and social development of the society concerned) if the changes were properly implemented in a supportive context. And one wants to know which changes—both educational and contextual—are crucial to obtaining desired outcomes and avoiding undesired ones^{2.43}.

A still more serious variant of the sequence whereby a focus on outcomes leads researchers and evaluators to fail to study the relevant processes, is that it encourages researchers and evaluators to dismiss as unimportant educational activities which do not affect the very narrow range of outcomes with which evaluators have traditionally been preoccupied (ie those which can be measured with “reliable and valid tests”). This is particularly likely to occur if the educational processes in question are themselves regarded as intangible and unmeasurable. Examples of these include most of the educational processes which are designed to lead—and which from time to time do lead—to the transformation of students. As Jackson^{2.44} has emphasized, this includes most activities which are intended to lead to spiritual or moral conversions and those which are designed to release previously unsuspected capacities, talents, and energies. Jackson, Havighurst and Taba^{2.45}, McClelland^{2.46} and Dolphin and Raven^{2.47} have noted that such transformative experiences tend to occur as a result of such things as exposure to people who are good at doing things one wants to do oneself or people one would like to be like. (Such role models may be personal, or embedded in stories or parables). Other transformative experiences include coping with stressful and demanding situations^{2.48}. Some of those we interviewed in the course of the environmental studies project mentioned earlier claimed that relevant experiences included encounters with something greater than oneself whilst undertaking demanding activities one cares deeply about. Alternatively, they may be associated with the sense of awe and wonder which sometimes comes from contemplating the complexities of organisms or symbiotic relationships. Foshay’s^{2.49} “spiri-

tual”—or self-transcendent—dimension—which, he argues, should form part of all discipline-based studies—should, perhaps, be included here.

Another set of processes which the adoption of a variable-based analytic framework has led researchers to neglect are those associated with people's tendency to actively select and construct the environments in which they live and work.

There are many who claim that such processes are not amenable to scientific study. This has led some to assert that both education itself and educational evaluation are arts in which science, or at least positivism, has little place. Both positions reveal a profound misunderstanding of the nature of science^{2,50}. The history of science is an account of a process of making the intangible visible: Ampere used a magnetic needle to make electricity visible; cloud chambers are used to make the paths of neutrons visible.

The adoption of “reliable and valid” tests as the only legitimate basis for scientific evaluation both arises from, and reinforces, a view of science in which pride of place is not given to the evolution of new understandings, new ways of thinking, and new ways of making the intangible visible, and this mistaken concept of science leads administrators to be unwilling to fund—never mind commission—work which is designed to develop an understanding of educational (including transformative) processes or to find ways of measuring “intangible” inputs to, and outcomes from, the educational process.

I may summarize what I have said in this section by saying that the use of traditional tests, judged to be excellent against such criteria as discriminating power, internal consistency, reliability and validity has led to the adoption of inappropriate evaluation designs, to misleading research conclusions, to socially and ethically unjustifiable policy recommendations, policies and practices and to the perpetuation of inappropriate views of both the research process and science itself. It would seem that, instead of judging the goodness of tests against these criteria, we need to evaluate them against such criteria as whether they enable people to identify, develop, and get recognition for their talents, tend to yield new insights into the educational process, offer important insights into the benefits and disbenefits of particular educational programs, help us to identify what is working and what is not working—and why—and

how to improve the educational process, or enable teachers to identify, foster, and credit talents of each and every student in their class.

3. Current assessments do not recognize students' talents, thereby damaging most students' development and life chances and depriving society of their most important potential contributions.

Factor-analytic studies conducted by psychologists since the turn of the century have, on the whole, reinforced the tendency of teachers to think in terms of "ability" rather than "abilities" (or multiple talents). There is a general feeling in education that, by and large and in the main, the tests which are currently available distinguish between "able" and "less able" students. Anger that current tests not only fail to recognize important talents, but actively stunt their growth, is not widespread. In the paragraphs which follow, I hope to show that this complacency is not justified.

In the course of our research^{2.51} we have encountered a few teachers who have been able to help all of their students to identify and develop important talents. What was most distinctive about their work was that they persisted in looking for students' strengths even when those students were conspicuously unable to do some of the things which other students accomplished effortlessly. They sought—and obtained—the help of their students in this process. But, to make this process effective, they had to supply the students, as a group, with the concepts they needed to think about multiple talents—and they had to create developmental environments in which those talents could be displayed, developed, and exercised for the benefit of the group.

In chapter 1 I gave some examples of the kinds of talent and competence which they were able to identify and foster: the ability to persuade, the ability to make acute observations about bio-physical processes, people or society, the ability to make good judgments, and the ability to defuse tension in a group. What emerged was that it was much more appropriate to discriminate between students in terms of what they were good at than in terms of their "level of ability"^{2.52}.

Taylor and his colleagues^{2.53} have shown the same thing more systematically, portraying the results in terms of "totem poles". In a similar

vein, Smith^{2.54} found that the simple expedient of requiring teachers to identify what each student was good at, how he or she had contributed to the school as a community, and how the school was going to help the student, with the parents' assistance, to develop his or her particular talents during the next term, resulted in a virtual revolution in the school precisely because these procedures led the staff to discover, not only that all students were able (albeit in different ways), but that those abilities could be fostered and recorded. Burgess and Adams^{2.55} have set up a more general system with similar aims. In our own work^{2.56} we have identified such a long list of important talents that it is impossible to believe that any one person could develop more than a few of them: different people must possess different talents.

Further evidence that the range of tests traditionally used in schools measure—and focus attention on—only a tiny fraction of the talents which schools could be identifying, fostering, and recording comes from studies of what happens when people get to work. Bachman and his colleagues^{2.57} not only found that 80% of the young adults they interviewed said that they had been able to identify and develop their talents at work (compared with 13% at school): they also found that the experience of work led to the most significant developments in competence and changes in personality that had occurred in these young people's entire lifetimes. We^{2.58} found that one of the things which young people liked about work was that they were able to do things *which they were good at* and not the uni-dimensional boring things—which they were not good at anyway—which they were obliged to do at school.

The conclusion is inescapable: human talent is extremely varied and diverse and its varieties bear little relationship to the types of talent required to do well at school. We therefore do an injustice to most students by failing to help them to identify and develop their talents. As Flanagan^{2.59} has shown, the cost—in terms of personal suffering and loss to society—is enormous. Teachers are not only not helped to implement multiple-talent programs by the tests which are currently available: they are actively prevented from doing so by the need to goad pupils to obtain high scores on tests which will buy entry to courses of further or higher education and good jobs. If we are to tackle these problems, if we are to help schools and individual teachers to pursue their *educational* goals, if we are to enable schools to foster our children's

talents and recognize those talents once they have been developed, we need a very different measurement model.

4. *Construct Validity: Most tests measure neither "academic" nor "intellectual" ability.*

Her Majesty's Inspectors of Schools in Scotland (whose work I mentioned above)^{2.60}, Goodlad^{2.61} and Raven *et al*^{2.62} have shown that most school days are filled with boring, non-cumulative, routine activities which rarely involve analysis, evaluation, judgment, critical thinking, reconciling different points of view, communicating, or developing new insights into historical, literary, scientific or social issues, let alone identifying or solving new problems. There is little sensitive, respectful facilitation of the development of students' particular talents. In Goodlad's words, "Teachers did not respond to students because students rarely initiated anything". There is little opportunity for students to practice doing such things as thinking, planning, inventing, reassuring, leading, working with others, or developing their own understanding of how society works and taking the steps needed to influence it. There is even little writing of continuous prose, let alone prose which is revised and re-written over an extended period of time in order to convey something which is important to the originator to an audience which he or she wishes to influence, or to a recipient who needs the information. Language teaching mostly involves students underlining words in sentences and learning teacher-generated "rules" of grammar. Arithmetic mainly involves applying mechanical rules without understanding. These, of course, are the operations measured and rewarded by the tests currently in use: the effective teachers whose work was mentioned above (and which is portrayed in more detail in *Opening the Primary Classroom*^{2.63}) pointed out that such things as the ability to use context and structure to locate desired information, the ability to discard most of the material which crosses one's desk because it does not relate to one's purposes, the ability to use material one reads—even if it is not directly relevant to one's purposes—to stimulate lateral thinking about problems one is trying to tackle, and the ability to use innuendo, layout, and gestures to convey a message, does not show up on traditional tests. Nor does possession of the strategies which will be required to perform arithmetic correctly once mechanical associations (eg $7 \times 9 = 63$) have been lost—

and, make no mistake about it, the IEA's research^{2.64} demonstrates that such associations *are* rapidly forgotten once they are not practiced daily. Conversely, the HMIs^{2.65} and Cockcroft's work^{2.66} shows that many students cannot even use their knowledge that $7 \times 9 = 63$ to calculate in a shop what 7 objects, each priced at 9c, will cost. Thus, not only do most of the available tests not measure the ability to undertake any form of academic or intellectual (or even practical) activity worth the name—ie they lack construct validity—but they also trap teachers and students into activities which mean that it is inappropriate either to describe most schools as academic or vocational institutions or to describe the activities which take place in them as intellectual or practical.

This conclusion that most tests lack construct validity in the sense that they fail to measure academic, intellectual, or practical competence, is reinforced by analyses which show that they measure only temporary mastery of small and arbitrary samples of low-level information drawn from the vast domains of knowledge which exist in each discipline^{2.67}. 50% of the information which students have memorized has been forgotten after one year and 80% after two years^{2.68}. The low-level, non-specialist knowledge which is required to do well on the tests is out-of-date when it is taught, is non-cumulative, and is unlikely, even if remembered, to be of value to those concerned in the future. When people require knowledge, they usually need up-to-date specialist knowledge—which frequently did not even exist when they were at school. While half of the middle-class adults we interviewed in one of our studies^{2.69} said that their education (ie their exam grades) had helped them to get a job, only 13%—even of this group—said it had helped them to develop useful skills. Working-class informants were, understandably, even less positive. *Their* “education” did not even enable them to *get* a good job. When one gets beyond the “3Rs”, the tests currently available have even less construct validity. There is no sense in which temporary knowledge of a smattering of out-of-date scientific facts can be said to be a valid index of scientific knowledge—for such a label gives the impression that the knowledge assessed is in some sense a representative sample of the whole domain of scientific knowledge. Still less can a score on such a test be described as an index of the ability to think scientifically, the ability to keep up to date in a specialist field, the ability to find information relevant to problems one encounters, or the ability

to make observations which aid in solving those problems. Actually the problem is even worse than that because, as Taylor and his colleagues^{2.70} have shown, there are at least 12 different types of outstanding *research* scientist (never mind scientists in general). These all possess different concerns and patterns of competence. Furthermore, *none* of them is the sort of person who gets high grades on traditional achievement tests. It is therefore misleading to describe the tests which are commonly used as measures of "scientific knowledge"—still less as measures of "scientific ability".

This is a convenient point at which to draw attention to one more implication of the data which have been presented. We have seen that vocational competence demands high-level competencies, including intellectual and academic competencies. We have also seen that any form of intellectual or academic activity worth the name also demands many of these competencies. Thus, contrary to common assertion, there is no tension between the goals of general education and vocational education. The achievement of both demands high-level, professional, educational activity which is conspicuous by its absence in schools. The real tension is between educational goals, personal development needs, and vocational needs on the one hand, and the sociological need to have a clear and unarguable criterion for legitimizing the rationing of privilege on the other. It is this tension which resulted in the previously-mentioned recommendations of the Waddell Committee—"that the results (of a plethora of attainment measures based on different syllabi, levels, and modes of assessment) should be expressed on a single scale of seven points in a subject area"—finding itself in competition with an explosion of "profiling" systems^{2.71} which, it is (vainly) hoped, will make it possible for schools to give students credit for their talents.

Although I have said enough to discredit the whole testing enterprise as it is currently organized, I cannot resist rubbing salt in the wound. Many traditional academic attainment tests lack both reliability and the ability to discriminate: Spencer found that 60% of the variance in non-multiple-choice grades arises from variance between examiners; only 40% is due to variance between the students^{2.72}. The raw score difference between an A and a D grade is typically only 8 raw score points. Most of this small difference is attributable to differences in presentation, not differences in knowledge of the subject matter. Wolf^{2.73} has shown

(1) that “mastery” of particular operations depends very much on how the questions are asked or tasks are set (thus, for example, the proportion of students who are “able to undertake simple divisions” is dramatically different if the questions are posed in the form of $\frac{112}{7}$ rather than in the form $112 \div 7 = ?$ and the dependence of the answer to the question of whether the “criterion” has been reached on the way in which the task is presented becomes much worse if the criterion consists of being able to undertake any kind of task remotely resembling real-life), (2) examiners commonly do not agree even on whether a particular performance should be assigned to the top rather than the bottom half of the distribution, let alone in their more detailed ratings, (3) individual examiners typically allot very different marks to the same performance on different occasions, and (4) examinees perform the same task very differently on different occasions. Her conclusion is that the constructs to be indexed have very little generalizability and that re-test reliabilities are extremely low even when the measures have high internal consistency. (Indices of internal consistency are, of course, typically presented as acceptable surrogates for re-test reliabilities.) These tests therefore not only do not measure any form of academic or intellectual ability worth the name: they are, even as indices of differences between students in temporary knowledge of content or skill, unreliable and lacking in construct validity. Their continued use to assess students—and, through such assessments, according to the current world-wide zeitgeist, educational institutions, courses and teachers, thereby (according to the argument) providing choice and accountability (thus, in turn, nominally stimulating the improvement of educational provision)—cannot possibly be justified. Nor can it lead to the selection of the right people for important positions in society.

5. *Current assessments lack predictive validity.*

If the construct validity of academic assessments has rarely been carefully investigated, the same cannot be said of their predictive validity, as Ingenkemp's^{2,74} review of the extensive literature makes clear. However, not only are Ingenkemp's conclusions none too positive, the selection of appropriate criteria against which to validate achievement tests is problematical, to say the least.

Although employers regularly use scores on attainment tests to select employees, the Hunters^{2.75} meta-analysis of some 90 studies shows that their predictive validity even to success in *training* is only about .2. This is significantly, although not dramatically, less than the predictive validity of intelligence tests. This is, however, not the end of the story, because Berg's^{2.76} work shows that the correlation between earlier and later "educational achievements"—even in the same subject—drops to zero if the subsequent courses are taken some time after leaving school, and especially if those courses are really required to improve job performance.

Hunter and Hunter minimize the inability of educational attainment tests to predict more than 1% of the variance in actual occupational *performance* (over 425 studies) by saying that the criterion measures which are available are unreliable and that, in any case, too many other factors come into operation. Under such circumstances, however, what justification can there be for basing job selection on educational credentials? Such an admission supports Jencks^{2.77} claim that they are merely a charade to legitimate the rationing of privilege.

The work of Taylor^{2.78} and Bray, Campbell and Grant^{2.79} has revealed something else of importance: in those studies in which educational attainments do appear to predict occupational performance, the explanation is that those with higher credentials were, *on appointment*, assigned to fast tracks (where they were given preferential treatment) and to higher status and better-paid jobs. Once this difference has been corrected for, the apparent predictive validity of educational grades disappears.

In point of fact, a large number of studies summarized by McClelland^{2.80}, Berg^{2.81}, Collins^{2.82}, Dore^{2.83}, Jencks^{2.84}, Holland and Richards^{2.85}, Hoyt^{2.86} and the author^{2.87} show that, although *level* of education is markedly related to life success, grades obtained at any terminal level have no predictive validity. Bachman and his colleagues^{2.88} have further shown that, if students do not already possess the characteristics of those who stay on at school, staying on does not improve career prospects. All in all, therefore, the evidence points to the conclusion that not only do test scores not predict life success to any socially significant extent, the educational system itself, in general, adds little to people's competence^{2.89}.

On reflection, how *could* traditional tests reliably predict occupational success? Any one occupational group contains within it a wide variety of people who do very different things. As Berg^{2.90} has emphasized, occupational groups are to be understood, not as psychological, but as sociological, phenomena. Their function is to regulate competition. By maintaining “qualified” entrants in short supply their incumbents are able to achieve scarcity value. Berg’s data show that entry qualifications are raised when the number of “qualified” entrants is about to increase, and Folger and Nam^{2.91} have shown that this occurs despite there being no change in the tasks undertaken, or the skills exercised, by the group concerned.

It would not be appropriate to conclude this discussion without examining some findings which appear to conflict with our conclusion that traditional measures of academic achievement have little predictive validity. Hope^{2.92} has shown that 60% of social mobility, both upward and downward, in both the US and Scotland, can be explained (in the statistical sense) by IQ. Given that IQ tests measure the common factor running through measures of academic achievement, how can one reconcile these results with those cited above? Part of the explanation follows from the previously noted fact that *level* of education does make a difference. In fact, our work^{2.93} shows that children who leave the educational system at different ages differ in many ways which go way beyond IQ. They differ in their values, the competencies they want to exercise, the sorts of things they want to do with their lives, the satisfactions they want from their jobs, what they want from education, and their priorities in life. The differences between the values and IQs of different socio-economic groups are not, as many believe, mainly a product of background. Nor are they, as Kohn^{2.94} believes, a product of occupational experience. They are a product of some poorly-understood process of anticipatory socialization, in which young people acquire the values and behaviors which are characteristic of the groups they will later enter. As Kinsey^{2.95} demonstrated in relation to sexual behavior, this is true even of beliefs, attitudes and values they have never discussed and behavior they have never had—and never will have—an opportunity to observe. (One implication of this is that we need to develop much more respect for people who competently undertake activities which are very different from those which we ourselves value.) Another part of

the explanation is that, as we have seen, educational credentials, despite their lack of predictive validity, are the only socially acceptable means of controlling entry to occupations. Hope devotes a whole chapter to the dilemmas which this poses. He notes that senior public servants may be highly intelligent, but still possess neither the inclination nor the qualities which are required to act in a creative and inventive way in the public interest. It is on finding ways of broadening the selection process to include these other qualities that we must focus. We may, at present, not only be paying through the nose for an educational system which provides only dysfunctional custodial care and an acceptable way of allocating privilege, but also operating a system which allocates jobs on the basis of criteria which are, at best, irrelevant and, more probably, socially dysfunctional.

6. Current assessments select the wrong people for important jobs.

In the previous section I suggested that traditional tests of academic achievement have little value for one of the purposes for which they are most commonly used (namely for occupational selection) but wound up discussing Hope's concern that we may actually be promoting the wrong people—possibly people who are over-concerned with their own advancement or people whose only ability is facility with words—into influential positions in society. Hope is not the only person who has been concerned about this^{2.96}. In the course of our own work^{2.97}, we encountered many managers who had vandalized their workplaces—destroying the developmental potential of their sections and their subordinates, not to mention depriving many of them of their livelihoods—in order to advance their own careers by demonstrating that they could run their sections efficiently by “reducing costs”. Tomlinson and Tenhouten^{2.98} have shown how a few children at school exploit the system by ingratiating themselves with their teachers and by preparing and rehearsing “original” examination answers. Interestingly, however, while most children know that they could or “should” do these things in order to “get on”, most refrain from doing so because it seems to them to be somehow immoral^{2.99}. Our research among adults suggests that a disproportionate number of those who do it turn up in influential positions in society. As Hope says, selectors should recruit those who

possess the qualities required to do the job well. Promotion should not even be regarded as a reward for having performed well in some other job, let alone as a reward for having performed a personal service for someone in authority. To implement this procedure it would be necessary to have a set of assessment procedures which it would be difficult to introduce not only because they would be very different to those in current use, but also because they would particularly threaten the more personal-power and advancement-oriented people in society^{2,100}.

7. The assessment procedures currently in use lack objectivity.

Finally, the most commonly encountered objection to any change in assessment procedures is that the alternatives are “subjective”. However, current assessment procedures are anything but objective in any non-trivial sense. How objective is a picture of a student or employee which ignores the most important concerns and priorities of that person and the talents which he or she is able to exercise in pursuit of them? How “objective” is an evaluation of an educational program which ignores its most damaging effects? Yet such highly selective descriptions are precisely what we are most commonly offered—with the justification for their one-sidedness being the supposed objectivity of the information which is provided.

The fact is that value-based decisions are built into *all* evaluations but have, in most cases, been obscured by a facade of pseudo-science. What is assessed and reported is *primarily* dependent on the preoccupations, values and purposes of those who construct and select the tests that are used, and not on the qualities of the person or programs being assessed or evaluated. Value-free assessment, value-free evaluation, and, indeed, value-free education, are self-contradictory concepts; impossibilities; things which cannot be; oxymorons. Not only is value-free assessment an impossibility: the assertion that assessment should be value-free is itself a value-based statement asserting a value-based criterion for the assessment of assessment procedures which has the effect of concealing, even if it is not intentionally designed to conceal, the values of the person making it. The sooner this fact is grasped *and built into our concepts of assessment, evaluation, and education*, the sooner we will be able to make progress.

A more important limitation of current concepts of objectivity is, however, the fact that the qualities which someone will display are vitally dependent on the situation in which he or she is placed, and on the opportunities which he or she has had to develop the qualities which are being assessed. It is difficult to demonstrate acceptable forms of initiative in a Latin class. The selection of a Latin class as an appropriate context in which to rate "initiative" amounts to a statement about the *kind* of initiative which is valued by the assessor. Consequently, not only are assessments of initiative made in Latin classes unlikely to have much predictive validity to other situations: they are unlikely to reflect the capacity of those being rated to display initiative in other situations. This does not invalidate the concept of initiative; it simply shows that its display and recognition is dependent on the values of both the person being assessed and the assessor. Thus, insistence on standardized test items and situations, while perhaps conducive to replicability in behavior, are not after all conducive to the objective assessment of competence. Even if the candidate values the task set or the benefits which might be obtained by performing it "well", the situation may preclude the levels of persistence, persuasion, and planning which are crucial to any successful initiative. Likewise, differences between candidates may tell us more about the opportunities they have had in the past to practice taking initiative in those kinds of situations, than they tell us about what those same candidates are objectively capable of doing in those situations—or habitually do in other situations which are of concern to them. Thus, differences between candidates in typical assessment center tasks—such as persuading unwilling colleagues to build bridges across rivers—usually tell us more about whether they have had relevant experiences than they do about what the candidates would do if they had relevant experience, or what they do when trying to undertake activities which they care about and of which they have experience. They may have neither a predilection nor the capacity to persuade others to build a bridge which someone else has told them to get built, but they maybe excellent at inventing new materials for building bridges and persuading others to reveal relevant information. Failure to record this other fact would seriously challenge the assessment center's claim to objectivity.

The observations made in the last paragraph mean that failure to document (1) the context in which observations are made and the behaviors

which that situation tends to elicit, fails to elicit, and suppresses, (2) the opportunities which the person being assessed has had in the past to develop relevant or alternative competencies, (3) what the candidate can do if he or she is able to work for an extended period of time at a task which he or she cares about, and (4) what kinds of situation would engage the values, motives and talents of the candidate, is therefore a serious indictment of the objectivity of any assessment process. Yet few of the current procedures which claim to be objective include any of this information. The typical disclaimer—that the test score reports only what the student did on a specific test at a particular time—shifts the responsibility for any injustices which may result from the assessment from the assessor to the user. But it does not make the process any more objective. What is more, since the user has neither the information and expertise required to realize what has been left out of the assessment process and its interpretational context nor information on the social consequences of the assessments that are made, whereas the assessor's professional association has opportunities to do all of these things, abdication of responsibility for publicizing the limitations and potentially damaging consequences of the assessments and evaluations that are made is highly unethical.

Conclusion.

In this chapter I have suggested that there are seven good reasons why it is vital to develop an alternative measurement paradigm in education: because they play such an important role in allocating position and status, the assessments which are currently made deflect schools from their goals, help to focus debate about educational policy on trivial issues, condemn people who possess important talents which schools could help them to foster to lives of degradation and humiliation, lack objectivity, construct validity and predictive validity, and neither enable us to select the right people into, nor exclude the wrong people from, influential positions in society. Nor do they help teachers to diagnose students' learning difficulties and prescribe remedial action, help teachers to administer individualized, competency-oriented, educational programs, or help politicians or administrators to run more cost-effective educational systems—let alone help them to run more educational systems. They

trap evaluators into an inappropriate concept of evaluation, and educational researchers in general into an inappropriate concept of science. They trap psychologists into unethical positions. And, because the long term personal and social consequences of their continued use are dysfunctional, they can only be regarded as immoral.

In the next chapters, I will suggest a possible way forward. Here I would like to conclude by underlining one of the last sentences of the last paragraph. The fact that the issues which I have discussed have so rarely been raised suggests that the model of science with which social scientists have been working has not been one in which pride of place has been accorded to asking new questions, re-formulating existing questions, and surfacing and debating conceptual issues with a view to advancing understanding. It has been one in which priority has been given to blind empiricism. That, in itself, would seem to be sufficient grounds for asking whether the educational system is achieving its main goals, for paying more attention to alternative goals, and for asking how the main goals of education can be achieved more effectively.

CHAPTER 3

THE NATURE OF COMPETENCE.

In the previous chapters, I have shown that new forms of assessment are required so that:

1. Teachers can administer (manage) individualized competency-oriented educational programs. If high-level competencies are to be fostered, students must be able to practice them in the course of undertaking activities they care about. If they are to be able to do this, teachers must be able to identify each student's concerns, interests and patterns of competence and monitor his or her reactions to his or her experiences.
2. Students can identify their own distinctive talents, monitor their progress as they develop them, and get credit for their accomplishments. New forms of assessment are also required to enable people to get credit for the talents they have developed at work and in the community – for this is where people obtain their most important developmental experience and information. Only in this way will it be possible to break the stranglehold which educational institutions currently have on job-entry qualifications.
3. Teachers can get credit for their accomplishments in accountability exercises and evaluation studies. They, like other public servants, need to be able to get credit for doing such things as paying attention to their clients' needs and inventing better ways of meeting them.
4. Evaluators can design studies which will enable administrators to find out how to improve educational programs and policies.
5. It becomes possible to implement a more effective manpower policy based on more sensible guidance, placement and development

procedures, and selection policies which are better at getting the right people into, and excluding the wrong people from, important positions.

In this chapter I will suggest a basis on which an alternative measurement paradigm might be built. Later I will describe the ways in which this paradigm has already been operationalized and ways in which its implementation could be improved.

The place to begin to build an alternative *measurement* paradigm must be with the question: "What is competence?". This has been discussed at some length in my *Competence in Modern Society: Its Identification, Development and Release*^{3.1}. The discussion which follows will be limited to issues which are essential as a basis on which to build an alternative assessment paradigm, and will ignore other issues having to do with the release and development of competence.

We can read almost anywhere that initiative is a quality which it is vital for educators to foster. It is seen as essential to both competitive capitalism and successful socialism. What *is* initiative?

To take a successful initiative, people have to be self-motivated. Self-starting people must be persistent and devote a great deal of time, thought and effort to the activity. They need to initiate innovative action, monitor the effects of that action, and learn from those effects more about the problem they are trying to tackle, the social, political and environmental context in which it is situated, and what is effective and ineffective about the strategies they are using. To succeed, they must anticipate obstacles in the future and invent ways of circumventing or overcoming them. They will need to build up their own, unique, set of *specialist* knowledge. They will have to get help from others. More often than not, it will be necessary to establish coalitions with others to gain control over social and political forces which would otherwise deflect them from their goals.

Perhaps the crucial point to be emphasized in attempting to clarify the nature of competence is that no-one is going to do any of these things unless they care about the activity they are undertaking. Values are, therefore, central^{3.2}. In practice it turns out that that which is valued may be a particular *outcome* (such as stopping a factory polluting a river) or it may be a particular *style of behavior* (such as finding better ways of doing things or getting people to work together effectively).

What has been said has major implications for psychological and educational measurement. It means that one must know an individual's values, preoccupations, or intentions before one attempts to assess his or her abilities. Important abilities demand time, energy and effort. As a result, people will only display them when they are undertaking activities which are important to them. It does not make sense to attempt to assess abilities except in relation to valued goals.

The above analysis of the nature of initiative also implies that it does not make sense to attempt to assess separately the cognitive, affective, and conative^{3,3} components of an activity. One cannot meaningfully assess "the ability to develop better ways of thinking about things" independently of the pleasure which the person concerned derives from doing so, and his or her determination to make glimmering insights explicit. *These affective and conative components are an integral part of what we mean by the ability to cognize*^{3,4}. Not only do the three components interpenetrate: if the behavior in question—the initiative—is to be successful, these components must be in balance. Both determination exercised in the absence of understanding, and the converse, are unlikely to make for success.

These observations are in sharp conflict with many traditional canons of psychometry. I have argued that one cannot assess abilities independently of values. This means that it is essential to adopt a two-stage approach when assessing competence. We must first find out which types of behavior someone values, and then, and *only* then, assess his or her ability to bring to bear a wide variety of potentially important cognitive, affective and conative behaviors to undertake the activity effectively.

It is important to emphasize that the widely held view that one can use one set of scales to assess values and another, independently, to assess knowledge, skills, abilities or competencies, simply does not make sense. The latter will only be developed and displayed when the situation in which the individual finds him or herself triggers or releases the former. Furthermore, since people often cannot tell one what their distinctive preoccupations and concerns are (since they do not know what other people's are) one of the best ways of finding out what people care about is to ask: "In the course of pursuing what kinds of activity does this person display multiple and high level competencies?"

Our example—initiative—also highlights another way in which the assumptions on which the dominant measurement paradigms in psychology and education are based fail to engage with important aspects of competence. Conventional psychometric theory places great stress on internal consistency or factorial purity. Scores derived from tests composed of items which do not correlate with each other are said to be meaningless. Yet it would seem from our example that this assertion is incorrect. People's initiatives are more likely to be successful the *more* independent and different things they do in the course of pursuing their goals. For example, they are more likely to be successful if they re-conceptualize the problem, obtain the help of others, persist over a long period of time, and so on. Yet their inclination and ability to do any one of these things in pursuit of their goals is unlikely to be closely related to their inclination and ability to do others. Furthermore, if they do any one of them particularly well it will, to some extent, compensate for their failure to do others.

It follows from the observations made in the last paragraph that, if we are to assess such qualities as initiative, instead of trying to develop measurement tools which are as internally consistent as possible, we need to try to develop *indices* made up of items which are as little correlated with each other as possible^{3.5}. This is actually not so heretical as at first sight it appears, because it is standard practice to make use of multiple regression equations which involve summing over maximally independent variables to obtain the best prediction.

The insights we have developed so far may be summarized as follows: if we are to find ways of assessing important human traits we will need to abandon our desire to develop value-free, internally consistent measures. Instead, we will need to develop value-based, maximally-internally-heterogeneous *indices* which do justice to the psychological complexity of these qualities.

Cognitions of Institutional Structures.

This is an appropriate point to introduce one more disturbing insight which has emerged in the course of our work: value-based cognitions of social processes are central to competent behavior and need to be documented in any meaningful assessment of competence.

Behavior is very much determined by such things as people's beliefs about how things *should* be done, who should relate to whom, and about what. It is very much influenced by their perceptions of roles—by what they think it is appropriate for someone in their position to do, by what they think other people expect them to do, by how they think other people will react to their behavior. It is determined by their understanding of what is meant by terms like “management”, “participation”, “majority decision-taking”, “managerial responsibility”, “wealth”, and “democracy”.

The disturbing conclusion is that, if we are to assess competence in any meaningful way, it will be necessary to assess these, essentially political, beliefs.

Because this conclusion raises the spectre of social control and teachers brainwashing children, it is necessary to reinforce it by saying that we initially came to this conclusion from exactly the opposite starting point: when we compared more with less competent farmers, teachers, bus drivers, blacksmiths, managers and military officers we found that, in each case, it was the nature of the political behavior (with a small p) which the more effective people engaged in which was most important. Put the other way round, the most important source of incompetent occupational behavior in modern society is the inability and unwillingness to do something about the wider social, institutional, and political constraints arising from outside one's job—because it is these which overwhelmingly determine what one *can* do within it.

The need to describe the situation in which an individual finds himself as an integral part of the assessment.

Although the way in which people define the situation in which they find themselves has a marked effect on their behavior, that context has other direct and indirect effects. It influences their behavior directly through the constraints which it places on what they can do, and it influences it indirectly through the concepts, understandings and competencies which people are able to practice and develop.

It therefore emerges that, if one wishes to assess competence, it is necessary to assess both the perceived and the actual institutional context in which it occurs. As we saw earlier, it is either meaningless or wildly

prejudicial to say that people lack the ability to do something which they have never had the opportunity to practice doing. That is why "back to basics" re-enforces a "single factor" model of ability. The only way out of the dilemma is to make assessment of the context part of the assessment of the individual.

Identification of Values and Cognitions.

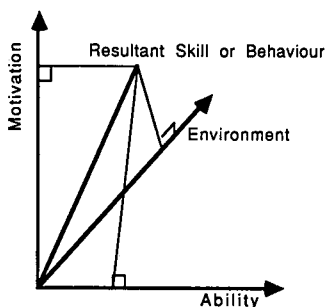
Although satisfactory measures of competence must be value-based and include the wider social and civic perceptions and understandings just mentioned, one unfortunately cannot discover these simply by asking people to identify the behaviors they value and their beliefs about how society works and their role in it. Because they do not know much about the values, preoccupations and thoughtways of others, they cannot perceive, still less identify, the ways in which they themselves are distinctive. That is why it is impossible for students who have come through one type of educational program to tell one how the issues on which they will in future tend to focus, and their ways of approaching those problems, differ from those of others who have come through other programs.

Not only are people unable to perceive and identify their own distinctive values and beliefs: some of the most important value-based social cognitions are shared by all members of a cultural group. For this reason, even an observer from that group who was able to see into the heads of his peers would be unlikely to notice and report them.

Recapitulation and Re-statement.

In the course of these remarks I have introduced some ideas which my colleagues and I have taken many years to stumble upon and make explicit—and which contrast sharply with many traditional assumptions in psychology and education. For this reason, many people have found it helpful for me to re-present the same ideas in different way. I will now do this, making use of a 3-dimensional diagram proposed by Ron Johnson and shown in Fig. 1.

Johnson argues that behavior is a resultant of three sets of variables: skills and abilities, motivation, and the situation in which people find

Figure 1:Johnson's Model

themselves. For our purposes we can substitute “components of competence” for “skills and abilities” and “values” for “motivation”. So far so good. But I have also argued that:

- (1) Components of competence will only be developed and displayed whilst those concerned are undertaking tasks they care about. They cannot be abstracted in the way suggested by the diagram and assessed independently of motivation. Motivation is *an integral part* of competence.
- (2) Effective performance—the resultant—is much more dependent on the number of independent and substitutable competencies which are brought to bear in a wide variety of situations in order to reach a goal, than it is on the level of competence or ability displayed in relation to any one of them in a particular situation. It is the total number of competencies which individuals display in many situations over a long period of time in order to reach their valued goals that we need to assess, not their level of ability in relation to any one of them. Any *overall* index of a person’s “ability” or “motivation” is virtually meaningless.
- (3) The situation in which an individual is placed influences the values which are aroused and the competencies which are practiced and developed directly—quite apart from its influence on the behavior which emerges at the end—when a person with a particular pattern of motivation and abilities is placed in a particular situation. Not only do environments have the power to transform people, people actively select themselves into, and attend and respond to different features of, particular environments. Johnson’s diagram does not recognize this. It gives the impression that a change in some feature

of the environment will lead to an increase (or decrease) in the quality or frequency of a particular behavior, with the motivation and ability of the actor remaining the same.

Despite these limitations, Johnson's diagram is useful because it emphasizes (i) that it is important to assess all three sets of variables, (ii) that behavior is a product of all three sets of variables, (iii) that the components of competence can only be assessed in relation to a task the individual cares about, (iv) that behavior is influenced by people's perceptions of the situation in which they find themselves, their understandings of the way the organization works, and the reactions they expect from others, and (v) that people will only display the levels of competence of which they are capable if they define the situation in which they are placed as one which will enable them to undertake activities they care about.

Above all, the diagram emphasizes that the competence with which people perform tasks they are given—the resultant—cannot, on its own, be treated as a meaningful index of their current competence to perform that task, let alone an index of the competencies they possess.

The diagram can also be used to illustrate the fact that other people's ratings of observed behavior—the resultant—are even less valid indices of the ratee's competence than is the behavior itself: for what raters perceive depends on their own values and priorities, what they take to be the demands of the task and situation, and their subjective ability to manage the ratee—who has values, priorities and talents which may well differ from their own. Many teachers (and managers) lack confidence in their own ability to manage independent, thoughtful, questioning students. This makes them unwilling to create situations in which such qualities could be developed and displayed. And it has a marked effect on the interpretation they place on such behavior when it occurs.

Having said that, it is important to note that it is only the (already "contaminated") "resultant" behavior, further contaminated by their own values and abilities, which any observer can see with the unaided eye. The only way round this difficulty involves, on the one hand, getting inside the ratee's head, and, on the other, making the values, priorities, assumptions and competencies of the rater as explicit as possible.

CHAPTER 4

A FORMAL MODEL OF COMPETENCE, MOTIVATION, AND BEHAVIOR AND ITS ASSESSMENT.

We return now to the task of elaborating our model of competence and the way in which its components are to be assessed. We have seen that it is inappropriate to try to assess the self-motivated competencies which make for effective behavior except in relation to activities which the person concerned cares about. We have also seen that there are many components of competence, that many of them are relatively independent of each other, and that these competencies are cumulative and substitutable.

This way of thinking about competence may be made more concrete by reference to Grid 1.

On it, some of the types of behavior which people value have been listed across the top. These behaviors have been grouped into the three clusters (Achievement, Affiliation, Power) identified by McClelland in 1958 and confirmed empirically in our own previous work^{4.1}. Down the side are listed a number of components of behavior which, if present, are likely to result in the activity being successful. These components of competence include cognitive activities like making plans and thinking about obstacles to goal achievement, affective activities like enjoying the activity or wishing that a necessary but distasteful task was completed, and conative activities like will, determination and persistence. However, also listed is a number of other factors which contribute to successful performance—like having the support of others and believing that one's behavior is consistent with both one's own and others' views of what it is appropriate for someone in one's position to do^{4.2}.

Achievement

Affiliation

Power

Doing things which have not been done before.

Inventing things.

Doing things more efficiently than they have been done before.

Finding better ways of thinking about things.

Providing support and facilitation for someone concerned with achievement.

Ensuring that a group works together without conflict.

Establishing warm, convivial relationships with others.

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.

Ensuring effective compliance with
one's demands.

Cognitive

Thinking (by opening one's mind to experience, dreaming and using other sub-conscious process) 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: eg "I know there will be difficulties, but I know from my previous experience that I can find ways round them."

Personal normative beliefs: eg "I would have to be more devious and manipulative than I would like to be to do that."

Social normative beliefs: eg "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.

Conative

Putting in extra effort to reduce the likelihood of failure.

Summoning up energy, determination and will-power.

Persisting over a long period, alternately 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 routinised, 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.

[illegible]

The importance of separating these value and efficacy components in assessment can be re-emphasized by taking another example. An individual who values success at football may show a great deal of initiative in relation to football, be very sensitive to feedback from the environment, seek the help of others to improve performance, monitor and continuously improve his or her style, seek out new techniques and ideas, be sensitive to minor cues which suggest ways to improve, be sensitive to the approval or disapproval of his or her peers, have the willpower to persist in the face of difficulty, and be able and willing to persuade local politicians to provide a pitch or field. Nevertheless, if the ability of this same person to engage in these complex, cognitive, affective, social, and conative activities is assessed in relation to performance at mathematics—a goal which, for the sake of argument, we may assume this individual does not value—then one might erroneously conclude that he or she is unable (and not just unmotivated) to engage in the activities which have been mentioned. Teachers, psychologists, and managers have, in the past, too frequently been guilty of drawing such erroneous conclusions.

Attention should be drawn to the fact that, while this model is readily comprehended as a model designed to help us to understand and assess motivation—the styles of behavior someone values and his or her ability to pursue those goals effectively—it is, in reality, a model of competence.

Descriptive Statements and Profiles.

In principle, Grid 1 can be used to identify the behaviors which people value and the components of competence they tend to display in pursuit of them. For any one person, an assessor could, after having made relevant observations, enter checks in the appropriate cells under the behaviors the person values. By adding up the checks in any one column, the assessor can obtain an index of how likely it is that the person concerned will undertake that kind of behavior effectively. By summing the scores obtained in adjacent columns under each of the overall headings, scores which indicate the probability that a person will reach achievement, affiliation and power goals can be obtained. This yields a profile which is directly comparable with those published by McClelland, and which he (in the present context, misleadingly) refers to as profiles of motivation.

It is important to note, however, that because, as has been indicated, the Grid should be considerably extended, the procedure would become cumbersome if it were applied wholeheartedly. A way round this problem will be suggested shortly.

Heterogeneous Indices or Internally Consistent Factor Scores?

Not only must values be assessed as an integral part of the assessment of competence, the components of competence we have identified cannot be meaningfully analyzed or identified in factorial or dimensional terms. The scores obtained by summing down the columns in Grid 1 are, quite obviously, not “uni-dimensional”. Indeed, the more independent and heterogeneous the competencies which are composited the better—provided, of course, each relates to goal achievement. At this point many readers will (as a result of their training in the dominant internal-consistency, factor-analytic, measurement paradigm) be thinking: “Such scores are not meaningful!” It is therefore important to note that, while the factor analyst’s claim that such heterogeneity shows that the scores which are obtained are not uni-dimensional is correct, the assumed corollary—that they are not meaningful—does not necessarily follow. No one would argue that multiple regression coefficients are meaningless simply because they are derived from summing over a large number of independent variables.

Overall Indices vs Detailed Descriptive Statements.

In practice, an account of the types of behavior which a person values and the competencies they display in the course of carrying out those activities provides much more useful information than a single total score. Such a description is radically different from a profile of scores across a series of factorially independent dimensions. The assumptions behind a factorial profile are that behavior is best to be described and understood in terms of people’s relative scores on a small number of dimensions. The assumption behind the model developed here is that behavior is best to be understood by identifying people’s values, compulsions, perceptions and expectations, and the components of competence they tend to display spontaneously in pursuit of their valued goals.

“Atomic” vs “Variable” Models

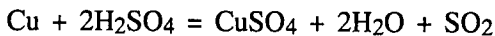
The difference between factorial profiles and descriptive *statements* can be illustrated by using examples from physics and chemistry. Physicists have shown that the behavior of a projectile is best described by some such equation as:

$$s = ut + \frac{1}{2}at^2$$

(the distance travelled at a particular time is determined by the initial velocity multiplied by the time elapsed plus half the acceleration multiplied by the square of the elapsed time).

The factor analysts' model is analogous. For example, it may assert that the degree of leadership which will be displayed is a function of the person's scores on a number of other variables, such as extroversion and intelligence.

Unlike physicists, chemists have found a quite different type of equation to be most useful in their work. They argue that substances and the environments in which they are placed are best described by listing the elements of which they are composed and the relationship between these elements. The descriptors (elements) are drawn from a large set known to all chemists. The elements which are not present do not need to be listed. The behavior of the substance in a particular environment is then described by equations which make it possible to describe transformations as well as monotonic combination:



(Copper plus sulphuric acid yields copper sulphate, water and sulphur dioxide.)

It is being argued here that human beings might best be described and understood by adopting a model which has more in common with that used by chemists than that used by physicists. Such a framework would enable us to indicate people's values and the components of competence which they show a spontaneous tendency to display in pursuit of them, together with relevant and significant features of their environments, without restricting us to the small number of variables which characterize factor-analytic models^{4,3}.

We may now attempt to push our chemical analogy a little further. Following this model we might find ourselves writing a *summary* description of an individual and the environment in which he or she lives and works. This might take the following form (the symbols which are used are examples only, and should in no way be taken to suggest that we have developed even a preliminary version of a more complete table of "human elements"):

Ach₄Pow₃;Auth₄PartCit₂;NuP₄HostP₃;DP(T)₁

Such a statement might be interpreted to mean that the individual concerned showed a spontaneous tendency to display four components of competence in pursuit of achievement goals, and three in pursuit of power goals. Four items contributing to the set dealing with authoritarian perceptions of society, and only two of the set dealing with participatory citizenship, were endorsed. Four aspects of the environment were supportive of the individual's goals; the manager modelled achievement behavior but did not delegate, encourage participation, nor create developmental tasks for his subordinates. There was "hostile press" from other people in the individual's environment. Concern with efficiency and effective leadership were scorned. The task which the individual was set had little developmental potential; it was a routine task which prevented the person concerned from developing perceptions and expectations appropriate to innovation.

If the equation were written in some way which permitted of movement, one would conclude that the individual would be likely to become frustrated and lose motivation to engage in achievement and leadership behaviors.

In fact, of course, such summary statements could be filled out in a great deal more detail, and very usefully too. One could identify exactly what type of achievement or power behavior the individual thought it was important to engage in; one could identify exactly which competencies were brought to bear in pursuit of each; one could identify the particular perceptions and expectations which encouraged and prevented the person concerned from engaging in such behavior; one could say more about the role models to whom he or she was exposed by managers, colleagues and subordinates; and one could say more about the tasks set and their probable effects on the person's future development and motivation.

The next point to be made is that such *statements* about people can be extended to include statements about their environments—combined statements which enable us to describe (or, more correctly, model) the *transformational* processes which occur in homes, schools and workplaces.

Equally importantly, however, they also enable us to give due weight to the major role which the environment, by engaging (or otherwise) people's values, plays in facilitating (or suppressing) both the initial development of high level competencies and their subsequent exercise and display.

It is of considerable theoretical and practical significance to note that by, in this way, in a sense, going beyond the juxtaposition of statements about individuals and their environments to actually *incorporate* statements about the environment *into* statements about individuals, one can handle the problems which have been highlighted by Greeno^{4,4} and his colleagues as a result of noting the situational specificity of behavior and observing that “the same” psychological activities (eg “problem solving” behavior) often take different forms in different contexts. Greeno *et al* claim that their findings show that such terms have no *generalized* meaning. But this does not necessarily follow from their data. The argument developed in this chapter implies both that the competencies which people will display depend on whether the situation in which they find themselves taps their values and that the same psychological qualities can look very different in different environments and in the context of different environments and different values and supporting competencies. To pursue our chemical analogy: elements look and behave very differently when combined with varying other elements. Copper, for example, looks and behaves very differently when it is in combination with only oxygen compared with when it is in combination with sulphur as well ie CuO_2 looks and behaves very differently from CuSO_4 . This does not mean that copper has changed its nature. If the analogy holds, there is no reason to suppose that “problem solving behavior” will have the same surface characteristics when it occurs in combination with differing clusters of other competencies and is displayed by people who are pursuing different types of valued activity in different environments. Thus the “hunch-based experimental interactions with the environment” which form such an important component in some forms of “problem solving” behavior may take a very different form when they are engaged

in by someone who is good at, and cares about, putting people at ease and when engaged in by someone who is good at, and cares about, making an original contribution to abstract scientific theory^{4.5}.

By insisting that statements about the individual are accompanied by statements about both the situation in which the observations were made and the relevant previous experience of the individual—and thus prior opportunities to learn—we can also overcome the serious challenges to faith in the possibility of “objective” psychological and educational assessment which were posed in the last chapter. There we saw that: (1) values are triggered, and competencies thus released or suppressed, by the situations in which people find themselves; (2) people can only have developed high level competencies if the situations in which they have previously found themselves have tapped their values (or, put the other way round and in a more concrete form, they may be perfectly capable of learning to do something they are currently unable and unwilling to do if they are placed in a situation which engages their values); and (3) people may be able to unleash high level competencies which they currently do not display if they come to value the task they are being expected to undertake. Our position is, therefore, that *a description of the situations which have in the past tapped people's motives, the competencies they displayed in those situations* (and would therefore probably transfer to any new task they might now come to value), *and whether the situation in which they currently find themselves* (and are now perhaps being observed and assessed) *taps their values must form an integral part of any meaningful assessment of their competence.*

For the sake of clarity, I will now briefly recapitulate the argument of this chapter. We first noted that people's areas of competence can be identified by putting check marks in the cells of a two-dimensional grid which has valued behaviors across the top and components of competence down the side. We then noted that internally-heterogeneous summary scores (analogous to multiple regression correlation co-efficients) which can be obtained by summing the check-marks in adjacent columns of the Grid are conceptually identical to McClelland's “motivation” scores. However, we also noted that the original (check-mark based) “description” of the behaviors the individual valued and the competencies displayed whilst undertaking those activities was much more revealing than the summary scores. We further noted that, since the

printed Grid was only an illustrative sample drawn from a much larger theoretically-definable grid (with the result that putting check marks on the grid would become cumbersome if pursued wholeheartedly), we could achieve the desired effect by writing “chemist-style” *descriptive statements* about people. These identify the behaviors they “value” and the competencies they display whilst undertaking those activities. We then noted that this very same procedure would enable us to describe the relevant (and only the relevant) features of the environments in which people live and work—and those in which they had previously lived and worked. We finally noted that this would enable us to both model the transformational processes which have proved so intractable in developmental psychology and education and handle the problems which the situational specificity of behavior pose for conventional (trait-based) concepts of “ability”.

One final observation may be made about our research and the nature of the future scenario which would stem from its adoption. The crucial—almost idiosyncratic—feature of what we have been doing has been that we have been mapping and sampling relevant *domains* of competence—including their motivational basis and their cognitive, affective, and conative components. At the present stage in the development of our science, this has been no routine, handle-cranking, concurrent-validity-driven, activity. Quite the opposite: it is only possible to carry it out effectively after one has developed a thorough understanding of the area one is dealing with. To pursue such work one needs, not so much a new *methodology*, as a climate which emphasizes that scientists should devote a considerable amount of time to what is, after all, the crucial phase of any scientific enquiry worth the name—namely to developing concepts and understanding. However, as the framework for thinking about and mapping the domains of competence becomes clearer, the task of assessing people will become more like carrying out a chemical analysis than “measuring” their height with a ruler or taking their temperature with a thermometer after the manner of a physicist.

CHAPTER 5

IMPLEMENTING GENERIC COMPETENCE ASSESSMENTS.

Assessment performs many functions. Teachers need it to monitor the effects of their actions. Students need it to find out how well they are doing and improve their performance. Administrators need it to study the effects of individual teachers, schools, and groups of schools. Summative assessments are required at the point of interface between schools and society, so that students can get recognition for the competencies they have developed and thus get an opportunity to use them—and develop them further—in the course of employment.

The preoccupation with traditional tests and the criteria established to assess test “quality” has not only resulted in invalid, unreliable and dysfunctional tests: it has also resulted in the failure to develop diagnostic and prescriptive tools more suited to such purposes as evaluating and improving educational programs on the one hand, and to diagnosing and remedying students’ learning difficulties (eg in reading) and offering individualized programs of competence-based education geared to each student’s interests, values and talents, on the other. To either assist with reading difficulties or implement competency-oriented education, teachers need to be able to obtain information on the motives and potential interests of each student, invent a possibly developmental experience for that student (ie one which harnesses the student’s motives, builds on the competencies which have already been developed, and addresses the problems the student has in pursuing his or her own goals), monitor the student’s reactions to that experience—and especially his or her specific difficulties—and take corrective action when necessary.

In addition, school authenticating and governing Boards need to be able to assess particular policies and programs of study: they need to be able to document the distinctive features of the programs and demonstrate that they have distinctive consequences for those who pass through them. They also need to be able to find out whether individual teachers are identifying and developing at least some of the talents of each of the students in their classes. Educational officials, administrators, teachers, students and employees need to be able to undertake stock-taking exercises to look at the human resources available. To do these things they must, both individually and collectively, assess the quality of the developmental environments and experiences which are available, and their probable consequences. For these purposes there is a need for a set of "mirrors" which enable people, individually or collectively, to take stock of what is happening so that they can, if appropriate, decide to change it.

In this chapter I will summarize what we have been able to do, using the model developed above, to fill some of these gaps. In the course of so doing, the methodologies we have employed to operationalize the model will be illustrated. However, if the reader is not to be too disappointed with what is to follow, it is important for him or her to approach the material with realistic expectations. Virtually all the work on which this discussion has been based has been carried out in "spare time", on an unfunded basis, as private skirmishes on the edges of a series of unrelated and non-cumulative projects which were commissioned for reasons having little to do with the central theme of this and subsequent chapters. Given the unquestioning acceptance of the dominant paradigm by those who control funds and review research proposals, and given the desire for quick returns and immediate answers among those who commission research and evaluation studies, it has proved impossible to obtain funds for research which would have addressed these issues directly. In setting appropriate expectations it is also important to say that, precisely because there has been no continuity in funding or projects, there has been no continuity in staffing either. No sooner have those concerned been socialized into (earlier versions) of the way of thinking presented here than—complete with their hard-won insights and expertise—they have had to move on.

In this context, the progress which has been made looks less insignificant. It has proved possible to use the measurement model outlined above without difficulty in program evaluation. It was used in both our evaluation of the Lothian Region Educational Home Visiting project (which was a Levenstein-like program of adult education designed to "emphasize the unique and irreplaceable role of the mother in promoting the development of her children")^{5.1}, and in our evaluation of the links established between primary schools and agencies of non-formal education such as zoos and museums^{5.2}. In both cases, it enabled us to show that, contrary to the received wisdom, adults (whether parents or teachers) had, for better or worse, dramatic effects both on children's and adults' values and on their competence to undertake valued activities effectively. It has also been employed without difficulty when assessing what might loosely be called national and organizational climates and patterns of competence associated with economic and social development and decline^{5.3}.

We have had more difficulty in using it for *individual* assessment purposes. However, even here, one set of procedures (Behavioral Event Interviewing and Records) provides relevant and useful information in an elegant and cost-effective way, and other procedures (based on value-expectancy methodology) have been shown to have considerable potential. In this chapter, methods based on externally generated *statements* will be reviewed first, followed by Behavioral Event Interview Methods, and then methods based on Value-Expectancy-Instrumentality theory.

Statements.

There are two essential pre-requisites to obtaining meaningful external assessments of competence. First: assessors should be thoroughly familiar with the conceptual framework summarized above and developed more fully in *Competence in Modern Society*^{5.4}. Second: they, like good mothers^{5.5} and managers^{5.6}, should both have gone out of their way to pay attention to what their students and subordinates say and do (and to the meanings of their gestures and innuendoes) and thereafter have created situations in which students or subordinates can enthusiastically pursue activities which they care about, growing in confidence and competence in the process. If they have done these things, teachers, lecturers and managers will, if they are good observers, find it relatively easy to

put checks in the cells of an extended version of Grid 1 to indicate which activities their students or subordinates value and the competencies they display spontaneously whilst pursuing them. An alternative is for assessors simply to list, after the manner of a chemist (or doctor, when writing a prescription), the behaviors which those being assessed value and the competencies they display while pursuing those valued activities. The lists of values and components of competence published in *Competence in Modern Society* may be used as aides-memoires for this purpose. If this approach is adopted, teachers and managers can also usefully describe the situations in which students and subordinates have worked, using the framework presented for describing classroom and organizational climates (in terms of the motives they tend to arouse and the behaviors they tend to encourage) presented in *Education, Values and Society* and *Competence in Modern Society*.

It is important to note that, whereas most external assessments of people take the form of *ratings*, what one gets by following the procedures described above is a series of *statements* about people and the environments in which they have been observed.

It will be readily apparent that this procedure requires teachers, lecturers and managers first to become thoroughly familiar with the ideas summarized above (a task no more difficult than that required of every student who aspires to be a chemist), and then to devote a considerable amount of time to the process of (a) studying students' or subordinates' interests and talents and (b) creating situations in which those talents can be expressed. (If teachers or managers have failed to create appropriate individualized developmental environments, or failed to make their observations in such environments, any statements made about, or ratings made of, high-level competencies will be meaningless.) Because of the time required, the use of rating systems—such as are often found in staff appraisal systems—is not a feasible, or at least a sensible, proposition in many settings. On the other hand, familiarity with (indeed, day-to-day use of) the framework is crucial to the development, release and effective deployment of human resources. It is therefore essential that teachers, lecturers and managers develop the habit of thinking more carefully about their students' and subordinates' talents and how best they can be developed and deployed. This objective might best be achieved, however, not by pitching them directly into assessing these qualities, but

by encouraging them to use the results of the more student-based assessment procedures and climate surveys to be described below.

In the past we have experimented with, indeed advocated the use of, Behaviorally Anchored Rating Scales^{5.7}. In essence, this procedure requires raters to agree on, for example, precisely what level of initiative is indicated by a specific behavior of a particular ratee. At first, the approach appeared to be very promising. However, we encountered serious difficulties when trying to implement the necessary procedures. The reason for this took some time to emerge. Although it was obvious from our earliest trials that behavior which one teacher would describe as "initiative" would be described by another as "the student trying to ingratiate himself with his teacher", it was not until we had recognized the centrality of values in the assessment of competence that we were able to appreciate that this problem could not be resolved without first finding out what the *student's* values were—and then respecting those values, *whatever* they were. That done, we could begin to get some agreement about what was meant by such qualities as "initiative" *in relation to the student's own priorities*. But, even then, if one wished to assess his or her competence, one had to develop behaviorally anchored scales for all the competencies listed in *Competence in Modern Society* in relation to all possible goals. The task became even more cumbersome than putting checks in an extended version of Grid 1. We backed off.

Behavioral Event Interviews.

Behavioral Event Interviews^{5.8}—or their development as collections of Personal Reports on Critical Incidents (or Records of Behavioral Events)—require teachers and lecturers to share more of the responsibility for assessment with students. Students are asked to think of—or keep records of—times when things went particularly well and particularly badly for them; they are asked to report both events which they were particularly pleased about, and events that led them to feel frustrated and uncomfortable. They are asked to record what happened, what led up to the situation, and what the outcome was. They are asked to say what they were trying to do or accomplish. (In this connection care has to be taken to reassure them that it is both appropriate and important to record "unacceptable" goals—such as passing the time as

pleasantly as possible in warm friendly conversation—because workplaces and society need people who value such behavior and do it well.) They are asked to describe their thoughts and feelings while they were engaged in the activity. And they are asked to say what others did, what they did, and how others reacted.

These records are then scored by the teacher or lecturer, or by an external agency, using a variant of Grid 1. The student's or subordinates' values and the competencies displayed when pursuing them are very apparent to anyone familiar with the conceptual framework developed above. The basic interview or record sheets remain available should those being assessed wish to challenge the overall statements which are made about their values and pattern of competence. If the interviewing and scoring are carried out jointly by student and teacher—and possibly the students' peers—a wealth of information is available to guide future placement and development. The methodology is elegant and, provided all concerned are prepared to take personal development seriously, it has the potential progressively to initiate both staff and students into ways of thinking about human resources and their development and utilization which are essential to the future development both of the educational system and society.

Variants of this methodology have been developed independently by Stansbury^{5.9} and by Burgess and Adams^{5.10}. Their work is important for two reasons. On the one hand it indicates that it is feasible to envisage that such assessments might be much more widely employed in schools. On the other, it alerts prospective users to the amount of time which is required if students are to be offered the guidance and counselling which is required as a basis for effective competency-oriented education, itself a pre-requisite to meaningful assessments of multiple talents.

The Assessment of Competence Using Value-Expectancy Methods.

Value-expectancy methodology is designed to get inside people's heads, assess the (re-interpreted) three dimensions in Johnson's diagram, and compute the resultant(s). The methodology enables us to assess people's values, their perceptions of relevant features of their environment, what they expect the effects of their actions to be, and how much

importance they attach to each of the consequences which are anticipated. The consequences which are examined include those arising from the individual's own competence (or the lack of it) and consequences which follow from other people's reactions to that behavior. If appropriate, people's confidence in their ability to deal with the reactions they expect from others are also documented. The methodology enables the assessor to combine these bits of information together in order to calculate the strength of the resulting disposition to undertake different kinds of task effectively in particular kinds of situation.

It is easiest to introduce the theoretical basis of value-expectancy-instrumentality methodology by reference to the work of Fishbein. In the late 60s, Fishbein^{5,11} stimulated a paradigm shift in the then quiescent area of "attitude" measurement by emphasizing, and finding an elegant way of handling, something which everyone had always known—but which had not been taken into account in the theories or practice of attitude measurement current at the time (and which is still neglected in the measurement of personality and abilities). This is that behavior—such as buying biscuits or using contraceptives—is primarily determined by multiple beliefs and feelings which come into play in particular situations, rather than by single underlying "attitude" or personality variables^{5,12}.

Fishbein made two fundamental contributions to our ability to think about, and handle, these issues. First, he focussed attention on something which has been repeatedly emphasized in this book, namely that it is the respondent's attitude toward, or value for, *the behavior in question*—and not his value for the object of the behavior—which it is important to assess. One should study the respondent's attitude toward *using* those contraceptives—rather than his attitude toward the contraceptives. Second, he found a means of tying together three well-established, empirically-based, theoretical viewpoints about behavior determination in psychology and sociology.

The first of these traditions holds that people will be inclined to engage in an activity if they are relatively certain that the activity will lead to satisfactions which they value. The second holds that they will be more likely to do something if they feel that the behavior is consistent with their self-images—with their view of the sort of person they want to be. The third viewpoint is that people will be more likely to engage in a behavior the more certain they are that other people expect them to do

so, and the more dependent they are on a favorable reaction from those other people.

There is considerable evidence^{5,13} to support each of these viewpoints taken individually. The predictive validity of measures based on any one of them is typically of the order of .4. The beauty of Fishbein's work was that, for the first time, it enabled us to assess each set of variables more systematically and then tie the three sets of variables together. The method of combining and weighting the component parts is itself supported by a considerable body of empirical research. The effect of these developments is that predictive validities of .8 to .9 are not uncommon.

Before moving on, attention may be drawn to the way in which Fishbein's model parallels that developed above in connection with Grid 1. There, we argued that the capacity to undertake a valued activity effectively was multiply determined and that it was dependent on bringing to bear a number of relatively independent—but substitutable—competencies, each having cognitive, affective, and conative components. It was argued that effective behavior depends on having an appropriate self-image, on perceiving oneself as having the support of relevant reference groups, and on having an appropriate institutional framework in which to work (ie on *shared* beliefs about priorities, relationships, and ways of doing things).

In non-technical language, what the Fishbein version of value-expectancy-instrumentality theory does is ask people what they think the consequences would be if they were to engage in any particular behavior and then weight those consequences with the importance attached to each. Three domains of possible consequences are systematically studied. These may loosely be called *personal* consequences, *self-image* consequences, and *reference groups' reactions*.

The *personal* consequences which are studied include such things as "It would take up a great deal of time which I would prefer to devote to other things"; and "I would have a lot less money for other things".

The *self-image* (or, more correctly, personal normative belief) consequences include such things as "No self-respecting person would do this"; "It is my duty to do this"; and "I would be working for the long-term good of mankind if I did this".

The *reference group* consequences include "My grandmother would object to my doing this"; "My workmates would encourage me to do this"; and "God will punish me if I do this".

Each of these perceived consequences has to be weighted by the importance attached to (or motivation to comply with) them: What my grandmother thinks won't have much influence on my behavior if I don't *care* what she thinks.

So, to apply the model fully, we first have to find out how *certain* the people we are assessing are that, if they engaged in the behavior, each consequence would follow—and then how important each of those consequences is to them. We then multiply the certainty ratings by the probability ratings and sum the resulting products.

To use value-expectancy-instrumentality theory to index the likelihood that people will display selected competencies in the course of undertaking tasks they care about we first identify tasks which they have a "felt need" to carry out by asking them to complete a *Quality of Life* questionnaire (see Fig. 2a).

On this Questionnaire they are first asked to indicate how *important* various features of the environment are to them, and how important they think it is to be able to do various things at work. Thereafter they are asked to say how *satisfied* they are with each of these same features of the environment, and with their opportunity to do each of the things they have said they would like to do. Their responses are then examined in order to identify an item which they have rated both important and unsatisfactory.

The *Consequences* Questionnaire is then used to explore their perceptions of the consequences of trying to do something about this unsatisfactory state of affairs. What do they think would happen if they tried to persuade other people to do something about it? What would happen if they tried to do something about it themselves?

The consequences which are studied cover the domains identified in Fishbein's model: they include such things as conflict with other values, whether doing it would enable them to be the sort of person they want to be, and their perceptions of how their reference groups would react.

The process may be illustrated by taking an example: supposing we are interested in exploring the consequences which students expect to

Fig.2a The Assessment of the Components of Competence
 An Illustration from the Edinburgh Questionnaires.
 (Note: This is a schematic representation only; it does not bear a direct relationship to the questionnaires)

(a)The Process

Importance

How important is it to you to:

Hi Lo

1. Work in a clean environment. ☐ ☐

2. Be able to do new things which have not been done before. ☐ ☐

Satisfaction

How satisfied are you with:

Hi Lo

1. The cleanliness of your work environment. ☐ ☐

2. Your opportunity to do new things which have not been done before. ☐ ☐

Consequences

If you have said that it is very important to you to work in a clean environment and that you are dissatisfied with the current state of affairs: what would happen if you tried to get something done about it? How likely is it that each of the following would happen?

Very Unlikely
Likely

Personal Reactions

I would enjoy trying to get something done about this. ☐ ☐

I would not know where to begin. ☐ ☐

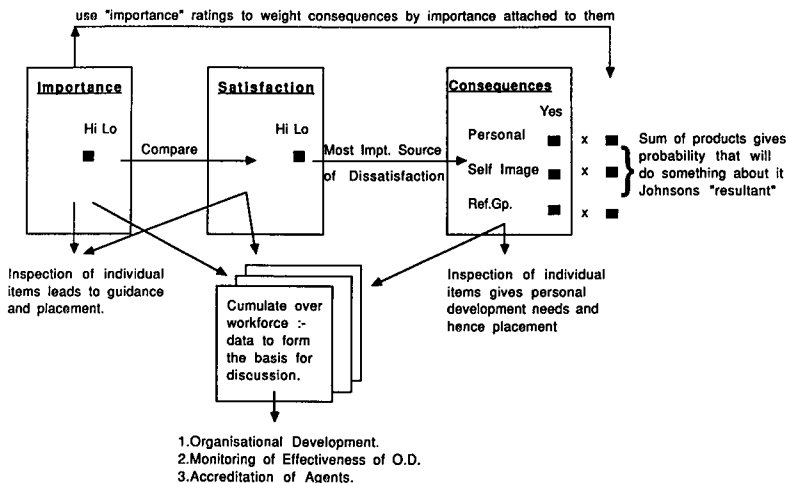
Self Image

I would have to be devious and manipulative. ☐ ☐

Reference Groups' Reactions

My boss would promote me. ☐ ☐

Fig.2b Flow Chart (oversimplified and schematic)



follow from trying to persuade their fellows to behave more responsibly. The students would first be asked what they thought the *personal* consequences would be. They often think that trying to do this would make them uncomfortable and unhappy, leave less time for other activities they value, and demand abilities which they feel they do not possess. In the absence of these abilities, any attempt on their part to persuade other people to behave responsibly would demand a great deal of effort, lead others to think that they were getting above themselves, and the whole thing would be a disaster. They would look, in their own eyes and in the eyes of others, very foolish indeed.

After they have been asked what they think the general consequences would be, they are asked what sort of person would do these things and whether they would like to be that kind of person. They sometimes feel that the sort of person who would try to persuade his fellows to behave more responsibly would be a rather pious, priggish, killjoy, and that, to be successful, they would have to be devious and manipulative. They may not themselves wish to be any of these things.

Finally, they are asked how others would react: would their friends support or reject them, would their teachers condemn them because they would have exposed their behavior as self-interested, rather than concerned with the good of all; would they, like Socrates, be deprived of career opportunities because they had identified themselves as the sort of person who takes moral issues seriously?

If one cumulated these results one would have a clear assessment of the strength of the student's disinclination to engage in the activity! (See Fig 2b).

But, by going through the process we have described, one obtains a great deal more useful information than this single index. In the case just described, one would have learned a great deal which would be of value in helping one to devise an individualized, generic-competency-oriented, developmental program to help the student concerned, if he or she so wished, to resolve value conflicts and thus release energy into chosen tasks, and to practice and develop competencies required to reach valued goals. The student could, for example, be brought to pay more attention to the probable long-term social consequences of not behaving in a socially responsible way. He or she might be encouraged to meet

other people who *had* behaved in a responsible way and not been punished or forced to behave in ways which were incompatible with being the sort of person they want to be. As a result of getting to know them, the student might learn how to persuade other people more effectively without having to be obnoxious. He or she could be helped to practice the skills required to obtain the co-operation of others.

Classroom Climate Measures.

Not only would the information obtained by using the above procedure be of value in making it possible to design an *individual* program of development for this particular student. The data collected from all students in a class would be of value in enabling the teacher (or an external accrediting agency) to assess the quality of the teacher's overall program of placement and development and his or her ability to release the know-how, goodwill, and enthusiasm of all students—and thereafter to improve his or her performance in both these respects.

Accreditation of Institutions Combined with Observer Judgments.

This is the basis for the final suggestion to be made in this chapter. I have already indicated that this methodology can be used to document the effects of educational programs and to highlight deficiencies in them. I have argued that such assessments would enable us to place the validation of courses on a sound basis. Let us now back up one step. It is a relatively straightforward matter to determine the presence or absence of classroom processes which are likely to lead to the identification and development of the talents of each student^{5.14}. Having demonstrated that some teachers had created developmental environments for their students, one could infer that they must have had opportunities to *observe* their students exercising high level competencies in the course of undertaking tasks they cared about. Under these circumstances, any statements they made about students' values and areas of competence would have a good chance of being meaningful^{5.15}.

THE QUESTION OF VALIDITY.

The validity of value-expectancy-instrumentality measures like those described has been established in a number of studies. However this has not been done by computing conventional correlation coefficients but rather, as Messik^{6.1} has recently advocated, by establishing a network of connections between the "measures" and their causes and consequences.

The first study in which the theoretical model outlined here was fully operationalized was in an evaluation of an Educational Home Visiting Scheme^{6.2}. The Scheme sought to "underline the unique and irreplaceable role of the mother in promoting the educational development of her children". Home Visitors, who were all trained teachers, visited the homes of 2-3 year old children for an hour a week for about 9 months. By working with the children in the mothers' presence, they sought to portray effective mothering behavior in such a way as to lead the mothers to do likewise. The evaluation showed that the Home Visitors had a dramatic effect on the mothers' beliefs and expectations, but very little effect on their behavior. The only behavior which changed significantly was that the mothers became more likely to hand their children over to professional careers. This was, of course, exactly the opposite to what was intended. We were able to show that this occurred because, although the mothers now believed that it was both important and effective to do such things as talk to their children, and had come to believe that intelligence was more readily influenced than they had previously thought, neither the environmental constraints on their behavior nor their basic values had changed. As far as environmental constraints were concerned, they, for example, still lacked the time needed to do the things they had always believed they should do, and now recognized to be

even more important than they had imagined. As far as their values were concerned, they would still have preferred their children to be dependent on, rather than independent of, them. Thus, despite the fact that they now believed even more strongly than before that the behaviors which the Home Visitors modelled were both important and efficacious, they were still prevented from doing them by environmental constraints and value conflicts. They resolved this dilemma by handing their children over to professionals—for whose competence they had developed a great respect. They ended up feeling even more guilty than before about not doing things they already knew they should be doing.

It emerged that, if the program was to be effective, the Home Visitors would, among other things, have had to set out to *influence* the mothers' values. They would also have had to help them, as a central objective of the project, to develop the competencies they needed to get more control over their own lives—and especially to influence public provision. This, naturally, posed serious dilemmas for the Home Visitors. Our interest here is, of course, methodological rather than substantive. The point is that the value-expectancy measures we developed proved to be sensitive to the effects of the educational program, enabled us to identify what worked and what did not and the reasons why, helped us to understand counter-intuitive effects of the intervention, and enabled us to identify the (often unexpected) remedial actions which were necessary.

In another project^{6.3} we used the methodology to study the effects of different types of educational program on elementary school students. We found that, contrary to common assertion, different teachers had dramatically different effects on students' concerns, priorities in education, priorities in life, behavior, and patterns of competence. Thus, some teachers led their pupils to feel that it was important to select tasks which were socially important and to obtain the cooperation of others to carry out those tasks effectively. The students learned how to tackle such tasks and how to win others' cooperation. That they had learned to do these things could be demonstrated by examining the consequences they anticipated: one did not need to observe their behavior. They understood how the local democracy and bureaucracy worked, how to identify leverage points within it, and how to influence it. They knew the strengths of their fellow students. Students in other classes did not think it was important to do these things, did not think it was important to learn how

to do them, and (rightly) anticipated disastrous consequences should they try to do them. One of the most striking results of the project was the discovery that, contrary to common assertion, what teachers did reflected (even if it did not match) their priorities^{6.4}, and the patterns of educational activity they created were in turn reflected in their students' values and patterns of competence. Once again, therefore, the methodology enabled us to document teachers' concerns and patterns of competence—and it also enabled us to develop measures of program outcomes which were sensitive to the effects of the educational programs students were offered. These measures in turn enabled us to pinpoint strengths and deficiencies in the programs.

The objective of the work reported in *Competence in Modern Society* was to develop a set of tools (The Edinburgh Questionnaires) which would be useful in staff guidance placement and development and in organizational development^{6.5}. The work showed, somewhat unexpectedly, that the "British Disease" stemmed from a lack of interest in doing such things as finding better ways of doing things, finding new things to do, finding better ways of thinking about things, working at a task which would in the long run benefit the whole organization or society, or getting people to work together effectively. Surprisingly, the negative consequences which were anticipated should they decide to do any of these things were not a significant deterrent to undertaking these activities. They simply did not think it was important to do them. The methodology worked: the problem was other than that which it was assumed to have been. *The Edinburgh Questionnaires* have since been used in Samoa, Tonga, Japan, China, Hong Kong, Singapore, the Philippines, Canada and the US^{6.6}. Whereas some people had reacted to the Scottish data by saying "Of course, could it be otherwise?", the cross-cultural data make it transparently obvious that things not only *could* be otherwise but *are* dramatically otherwise (in Japan and Singapore) and have the consequences with which we are all too familiar.

Taking these results together with those obtained with the Taylor-Nelson Monitor^{6.7} one obtains an acutely disturbing picture: the British, with the Dutch, are most likely to support the "new" values: conservation, re-cycling, quality of life defined in other than "economic" terms, community support networks, humanitarian values ... but they are unwilling to do the things they would need to do to translate those values into

effect. Given the predictable paralysis, it is not surprising that those who espouse the "old" values are able to impose their will on the majority, creating a deeply divided society in the process. Once again, then, use of the methodology described in this book has enabled us both to understand how the societies concerned got into the economic situation in which they find themselves and to document what would need to be done if they are to create the new kind of society they want. The methodology has also enabled us to generate data which enable people, as individuals or as groups, to take a look at their beliefs, attitudes, priorities and expectations in a kind of a mirror and ask themselves whether they like the look of what they see and, in particular, what they think the probable consequences of those beliefs, understandings and perceptions will be. (If they do not like what they see, they can go on to ask what they could do about it.)

Taken together, these studies suggest that the methodology does have considerable validity in that it enables us to understand, predict and influence behavior. It has enabled people to get more control over their own lives and organizations. Additional, convincing, evidence of the validity of value-expectancy methodology, interpreted and applied more narrowly, will be found in Feather^{6,8}.

CONCLUDING COMMENTS.

How is the Necessary R&D to be Organized?

In this book I have shown that new forms of assessment are required, at both school and system level, to administer educational programs. I have shown that new forms of assessment are required to credit the outcomes of the educational process and to enable people to identify, develop, get recognition for, and utilize their talents at work and in society. I have shown that new forms of assessment are required for research purposes—and not least for studying the social and economic consequences (at individual, group, and societal levels) of educational programs which foster different preoccupations and talents.

I have also shown that it is possible both to re-conceptualize the nature of talent and its assessment and, without further basic research, to operationalize an alternative measurement paradigm in such a way as to generate useful information. Implementing these procedures more widely in program evaluation, and developing the administrative procedures required to make them less cumbersome for individual assessment, would cost but a small fraction of what is currently spent on testing and evaluation^{7.1}. Finally, I have shown that, if there is a need to caution against unrealistic expectations in this area, there is also a need for wider recognition of the limitations of the tests currently available, and the damaging personal and social consequences which stem from their use.

It remains to add a few observations about the barriers which have inhibited development of this work in the past, and to reflect on the steps which are needed if more progress is to be made in the future.

At first sight, it is extraordinary that the vast sums which are spent on educational and personnel evaluation and testing should have contributed so little to the necessary developments. By and large, we are still working with models of ability and assessment which were around at the turn of the century. In referring to the vast sums which have been spent, I mean to include not just the huge federal contracts which have been awarded for the evaluation of Headstart and similar experimental programs and for the identification of the variables (sic) which make for school effectiveness. I also have in mind (i) current expenditures on the seemingly innumerable bureaux of research and accountability located in almost every school system in the country; (ii) the budgets of organizations, like ETS, which develop and market tests, whether for the diagnosis of special educational needs or for assessing scholastic aptitude or achievement; (iii) the vast pool of school psychologists and others who administer, score and interpret such tests; (iv) the student- and teacher- time which is devoted to assessment over and beyond that needed to diagnose learning difficulties and prescribe appropriate developmental experiences (the latter being conspicuous by its absence^{7,2}); and (v) military expenditures on personnel assessment.

Given the scale of this testing and evaluation enterprise, it seems remarkable that the developments which I have summarized should have had to be based on the work of perhaps half a dozen people—such as David McClelland and Calvin Taylor—who have all made their contributions only as a result of extraordinary personal commitment to research and to society, and without significant public funding.

The institutions which might have been expected to devote substantial time and money to relevant R&D include the sponsors of educational research and evaluation (such as NIE), test agencies, the military, consortia of bureaux of research and accountability, and the universities. However, I will argue that it is on the reasons why the universities—the educational institutions—have not been offering more of a lead that we should focus our attention.

The lack of sponsorship from *official* agencies is not too remarkable. The conventional wisdom is that one cannot expect research customers to commission research to solve problems they themselves have not noticed or to find ways of achieving goals which have not crossed their minds—or even goals which have crossed their minds but which they

do not know how to reach. Nor can consortia of bureaux of research be expected to press their case effectively. The classical entrepreneurial problem is that people are not generally aware of their needs until someone puts a solution in front of them.

What is remarkable is that the researchers who have had most day-to-day contact with testing, evaluation, the problems of education, and the psychometric and educational research literature, have neither emphasized the need for these developments nor gone out of their way to contribute to the new thinking which is required. What is remarkable is that there is not a single reference, either to the need for more broadly-based assessment procedures or to the need for a new measurement paradigm, in the papers presented to ETS's invitational conference on "The Redesign of Testing for the 21st Century"^{7.3}. There is not, for example, in that collection of papers (or in the Special Issue of "Applied Psychology" dealing with Computerized Psychological Testing edited by Eyde) a single reference to the possibility of using computers to implement psychometric models which could not be implemented in paper and pencil format. There is also no reference to the problems which I have sought to address here in the Joint Committee's *Standards for the Evaluation of Educational Programs, Projects and Materials*^{7.4}; in the *Handbook of Research on Teaching*^{7.5}; the 1982 AERA *Encyclopaedia of Educational Research*^{7.6}; or in the ten-volume *International Encyclopaedia of Education*^{7.7}.

Why have the test agencies not promoted the necessary research and development? They have, after all, assumed the mantle of the classical entrepreneur. In capitalist theory, their *raison d'être* is to identify previously un verbalized needs, invent ways of meeting those needs, and place the appropriate product on the market. Unfortunately, as McClelland^{7.8} and others^{7.9} have shown, the classical theory is wrong: the necessary activities are too long-term and too risky: the original inventor rarely profits. (Perhaps the Japanese could teach us how to handle this problem.) Perhaps as serious is the problem that, as with so many of the tools and procedures we so badly need in modern society, the necessary products will not be hard and marketable. Then again, despite the theory, commercial organizations are not noted for their enthusiasm to introduce new products when these will upset their existing market. I suspect, however, that the real problem is that the attention of those who control the flow of development funds in these organizations has not

been drawn to the unmet needs and the possibility of satisfying them because they do not have the "parallel organization" structure described by Kanter^{7.10}. The time needed to notice the needs, question basic assumptions, and undertake the necessary development work has not been available—and, indeed, it is an extraordinary fact that the time required to do these very things is often deleted from contracts for evaluation activity—witness SRI and Headstart Follow Through.

These reflections suggest that the central problem rests with the universities. They need to think more carefully about the ways in which scientific jobs are defined, the responsibilities of scientists, the competencies they need, and the qualities required by those who become researchers. It has not been ignorance of established scientific knowledge which (as the writings of those like Lerner^{7.11}, who emphasize routine testing, imply) has led to the failure to advance understanding and capitalize upon the discoveries which are made, but a lack of scientific competence as exemplified in the first chapter of this book. We need to rethink the training we offer prospective researchers^{7.12}. We need to involve our prospective scientists—and others—in more project-based educational programs of the kind described at the beginning of the first chapter and by Winter, McClelland and Stewart^{7.13}. We need to disseminate more appropriate understandings of the research process so that research sponsors—and the Joint Committee—cease to accept inappropriate beliefs about the nature of science and the research process and become, as a result, more willing to fund the adventurous research which is required to advance understanding and which will lead to the development of the tools and procedures which are needed.

And so we come a full circle—for we are once again discussing the nature of (occupational) competence, the role which educational institutions could and should be playing in fostering it, the institutional arrangements which our educational organizations lead people to accept as appropriate, and the types of people whom our assessment procedures lead us to appoint to influential positions in society. In seeking to identify what has prevented a significant proportion of the funds devoted to testing being applied to development work in this area, we find ourselves discussing why the universities have been unable to undertake the necessary research, change their own offerings, and foster in students the

necessary adaptability, conceptual ability and beliefs about the research process, society, and its operation.

The problem is not simply the usual one of scientists discarding outworn theories, paradigms of measurement, science and teaching. The reasons why the universities have been unable to foster scientific competence, adaptability, and willingness to draw attention to research needs, include the now familiar fact that they do not know how to foster these qualities, and the fact that students cannot get credit for possessing them or professors for fostering them. But they also include some of the other processes we have touched on in this book. As Schon^{7.14} has emphasized, the universities have developed highly-fragmented "disciplinary" structures. These fragmented specialist-knowledge-based "disciplinary" structures serve to minimize competition both within the universities and among the "professionals" they "train". The sociological functions performed by these divisions are, therefore, of considerable importance. However, the effect is that students are forced to "master" (temporarily) vast quantities of information rather than develop the generic competencies which are common to the advancement of a number of disciplines—and to non-disciplinary professional competence as well. Still less are they generally encouraged to study the workings of society and take responsibility for doing something about the problems they notice. The effects of these oversights only show up when those concerned start working for test and evaluation agencies, school systems, and government departments^{7.15}. University staff themselves have the greatest difficulty in branching out into new areas: they have at least to behave as if they accepted the conventional wisdom on how science advances (hypothetico-deductive method) in order to secure the research contracts on which their careers and the prestige of their departments depend, and they have to teach the received wisdom to get their courses accepted by professional organizations (the main sociological functions of which are to protect the interests of their incumbents). University staff are not encouraged to reflect on, or spell out, more appropriate institutional arrangements because the framework in which they work is itself so dysfunctional and so difficult to influence because of the sociological functions it performs.

In coming to these conclusions we are, once more, underlining the point that competence in modern society has centrally to do with the ability to understand and influence sociological processes. The effect

of our observations in this chapter is therefore to project everything which has been said in this book onto a giant TV screen in our own back yard. Unfortunately, while this may make us as uncomfortable as did those pictures of our young men in Vietnam or the effects of our economic and political policies on Ethiopian peasants, we may feel just as powerless (incompetent) to do anything about the problem. It would appear, therefore, that educational and scientific competence demands civic competence in every bit as great a measure as does the ability to intervene in the global political economy to protect the biosphere.

It is tempting to revert to a discussion of the professional and civic activities which are needed if we are to do anything about this depressing situation, the nature of the civic understandings and competencies we need, and the steps which we personally could take to influence the future. However, this is not the place to do so. Relevant discussions will be found in the publications listed in a footnote^{7,16}. Here it is sufficient to say that, to run modern society effectively, we need nothing less than new concepts of wealth, democracy, citizenship, the public service and the role of the public servant. Dissatisfaction with current arrangements is endemic: the most important source of dissatisfaction with life in modern society derives from people's relationships with politicians and bureaucrats. Perhaps the most important single activity which is required if the necessary civic activities are to be released is for someone to write a book entitled *The New Wealth of Nations*. This would re-orient thinking about the modern politico-economy as effectively—and appropriately—as did Smith's *Wealth of Nations* two centuries ago. It would, in particular, articulate things which people already "know" and legitimize things which are already happening. It would specifically legitimize the socially participative activities which are required to produce modern wealth.

Implications for the Process of Education.

In this book I have spelt out the implications for assessment of our research into the nature of competence. However, I began the first chapter by describing some outstanding educational activities. I must now set what we have since learned about the nature of competence and its assessment in an educational context. Historically, we began the work

which has been summarized here with a study of the goals of education and the perceived barriers to achieving them^{7.17}. In the course of this study (and its replications) we discovered (a) that the wider goals of general education—ie its goals beyond basic numeracy and literacy—were extremely unclear; (b) that the methods to be used to reach these wider goals were still less clear; (c) that there were major non-obvious, and previously unidentified, barriers to reaching these goals; and (d) that the absence of appropriate assessment procedures played a major role in deflecting schools from their goals. Accordingly, I sought to initiate—and in a less than perfect way in due course succeeded in undertaking—further work in areas (a), (b), and (d). By and large, it was the simultaneous pursuit of research in these three areas which led to the model of competence and its assessment summarized here. However, the work on assessment proved to be unexpectedly important. It led all concerned to become clearer about the nature of the qualities which were to be fostered, and thus clearer about how they were to be fostered. Some readers may be particularly interested in the insights we have developed into the educational process whilst we were engaged in this work. Unfortunately, a summary of these insights would occupy at least another book. Fortunately, summary articles and books are already available^{7.18}. Trillium Press is also publishing a book (*The Most Important Problem in Education is to Come to Terms with Values*) which identifies the serious and unsuspected barriers to the dissemination of effective educational programs which stem from the value-laden nature of competence—and the steps which are required to manage the educational system in such a way as to overcome them.

Despite the availability of these discussions of effective education, it may nevertheless be useful to highlight, very briefly, a number of conclusions which have already been mentioned in passing in this book. One is that any meaningful form of competency-oriented education must be grounded in the student's values. Teachers must therefore be able to identify, and be willing to respect (rather than feel they have to change) individual students' values. They have not always been noted for their ability or willingness to do either of these things. Another is that it is necessary to foster different competencies in different students and foster the same competencies in different students in different ways. This does not fit easily with the current focus on standardized, age-determined, national curricula. Al-

though many reports^{7.19} talk of individualization at the same time as emphasizing the need to prescribe curricula, prescriptive curricula debase the meaning of the word “individualization”. It is, indeed, impossible to discern *what* Goodlad has in mind when he uses the term^{7.20}.

More basically, as Dore^{7.21} has also argued, the kind of individualization which is needed is fundamentally at odds with our preoccupation with equality in public education—for equality is most commonly understood to mean that everyone will do the same thing, rather than have equal access to a wide variety of alternatives which have demonstrably different consequences for the individual concerned and for the society in which he or she lives. Nor can it be reconciled with the goal of developing teacher-proof materials. Other conclusions follow from our discovery that social and civic competence is central to competent behavior in modern society. This finding raises a number of important dilemmas for those who care about public education.

Private schools have always recognized the importance of value-laden and political education, and have set about inculcating social values and political beliefs without the slightest qualms. Those who did not accept what a particular school taught could, of course, always go elsewhere. This observation suggests that, if public schools are to provide effective education, it will be necessary for them to offer students many more options embodying different value positions, and encourage them to choose between them. If the public schools are to do this, teachers’ and administrators’ first priority (apart from commissioning the necessary R&D) will have to be to engage the general population in programs of adult education which will lead to a general reconsideration of beliefs about how society works and should work—and, in particular, beliefs about equality in public provision.

Besides this, it will be necessary for all concerned—whether as providers or as students—to reconsider what is *meant* by political education. Instead of inculcating political, social and civic beliefs, teachers will need to foster in students the motivation and the competencies required to examine the workings of society and take action on the basis of those observations. As Robinson’s^{7.21} work has shown, however, and as the experience of Headstart^{7.22} has confirmed, even this threatens established economic and political interests to such an extent that they mount highly-organized campaigns to sabotage it.

The last of our findings to which attention may be drawn here is that any move toward effective education—ie toward competency-oriented education—involves a major change in the teachers' role. It implies a shift from a concept of teaching as telling to a concept of teaching as facilitating growth. It implies a shift from being the center of attention and the source of wisdom, to being a back-seat manager of growth and development. This not only conflicts with satisfactions which teachers often go into teaching to obtain—such as to communicate and exert authority—it also means that managerial, not didactic, competence is what is required to do the job effectively ... and managerial competence above all involves being able to influence the wider social forces which, otherwise, so much limit what one can do in one's job. In conclusion, then, the introduction of effective education requires more than a change in assessment. It involves unexpected changes in beliefs about the nature of public provision and the way it is managed, unexpected changes in the teachers' role, and new patterns of competence among teachers. Above all, it threatens a wide range of established interests in fundamental ways. Paradoxically, the long term interests of almost all of these groups would be best served by espousing it. Disturbingly, human beings seem to have great difficulty acting in their own long term interests.

NOTES AND REFERENCES

- 1.1. Walberg (1984) has documented a similar lack of change in the United States.
- 1.2. Raven, Johnstone and Varley (1985). It is of more than chance significance that these terms of reference are reflected in this book.
- 1.3. Raven (1977). Copies of this are still available by mail order direct from the publisher, Oxford Psychologists Press, 311 Banbury Road, Oxford OX27JH.
- 1.4. Raven (1984)
- 1.5. Stallings and Kaskowitz (1974)
- 1.6. The correlations were equal in magnitude but opposite in sign.
- 1.7. Bullock Report (1975)
- 1.8. McClelland and Dailey (1973)
- 1.9. See discussion below and Raven (1984).
- 1.10. Cockcroft Report (1982)
- 1.11. Joint Committee for Educational Evaluation (1981); Raven (1984)
- 1.12. Coleman et al (1966)
- 1.13. See, eg, Peaker (1975); Walker (1976); Thorndike (1973); Comber and Keeves (1973); Carroll (1975)
- 1.14. The three evaluation studies cited above show that it is not appropriate to use tests which are designed to discriminate between students in research and evaluation studies whose objective is to investigate the differential effects of different types of educational program. In papers published by the World Bank, Schwarz and I (Schwarz and Raven, both in Searle, 1985) have shown that their limitations are even more serious if the objective is to find ways of improving educational programmes. Teachers need a whole range of tools to help them to administer individualized, competency-oriented, educational programs: they need to be able to identify students' interests and talents and invent a personalized program which will enable the student to practice a wide variety of competencies, and they need to be able to monitor students' reactions to those programs. In addition, teachers need to be able to diagnose learning difficulties in such a way as to get help in prescribing remedial action. Traditional tests are not suited to any of these purposes. It is for this reason that, although system-wide testing directs pupils', parents', and teachers' attention to the goals that are assessed, Harlen (1984) found that no one—neither pupils, teachers, nor administrators—actually uses the scores. The whole process is cumbersome and ineffective.
- 1.15. One implication of this is that the researcher's first task is to develop an adequate theoretical understanding of the educational process that is to be studied, rather than to focus on "research design" and "methodology".

- 2.1. Morton-Williams et al (1968); Morton-Williams, Raven and Ritchie,(1971)
- 2.2. MacBeath et al (1981); CES (1977)
- 2.3. Bill et al (1974)
- 2.4. Raven et al (1975, 1975); Raven (1976, 1977)
- 2.5. De Landsheere (1977)
- 2.6. Flanagan and Russ-Eft (1975); Johnston and Bachman (1976)
- 2.7. Goodlad (1983) at first declares that "parents want it all". He then insists on his respondents answering a forced choice question which requires them to choose whether schools are mainly about academic, social, personal development, and vocational goals. Still, a large proportion of his respondents said that non-academic goals were the most important. Despite this he, in the later chapters of his book, still manages to assert that most parents, teachers and pupils think that the main goals of education are to achieve academic goals—despite the fact that his own work shows that what happens in schools does not merit such a description. As I have shown elsewhere (Raven 1986), Goodlad's conclusions are contradicted by the work of Johnston and Bachman (1976) and others.
- 2.8. Flanagan (1978)
- 2.9. Bachman et al (1978)
- 2.10. Flanagan (1958, 1976); Flanagan et al (1975); Schools Council (1971-82, 1973); Closs et al (1985); Closs (1986)
- 2.11. Taylor, Smith and Ghesilin (1963); Price et al (1971)
- 2.12. Goodlad (1983)
- 2.13. Morton-Williams et al (1968)
- 2.14. Johnston and Bachman (1976)
- 2.15. Goodlad (1983)
- 2.16. Raven (1977)
- 2.17. Raven (1977, 1980, 1990)
- 2.18. See, eg, Frederiksen (1984)
- 2.19. Flanagan and Burns (1955); ITRU (1979)
- 2.20. Van Beinum (1965)
- 2.21. Sykes (1969)
- 2.22. McClelland (1961)
- 2.23. Rogers (1962)
- 2.24. Schneider, Klemp and Kastendeik (1981); Huff et al (1982); Raven et al (1985)
- 2.25. Klemp, Munger and Spencer (1977); Jaques (1976)
- 2.26. McClelland (1961); Schwartz (1987)

-
- 2.27. Price et al (1971)
 - 2.28. McClelland and Dailey (1974)
 - 2.29. Beuret and Webb (1983)
 - 2.30. Taylor et al (1963); McClelland (1962)
 - 2.31. Klemp et al (1977)
 - 2.32. McClelland (1973)
 - 2.33. Raven (1984)
 - 2.34. Raven (1977); Raven et al (1985); CES (1977); HMIs (1980); Johnston et al (1976); Goodlad (1983); Nash (1973); ORACLE; Bennett (1976); Rutter et al (1979)
 - 2.35. Raven (1977, 1977, 1980, 1984, 1987)
 - 2.36. Raven (1977, 1980, 1983)
 - 2.37. It is important to distinguish the argument put forward here—which is that the assessments are made in schools determine position and social mobility—from the argument, which our data do not support, that the current assessment system is a facade which only students from higher-status backgrounds understand, and which therefore enables the higher-status groups to recruit their own, and only their own, children.
 - 2.38. Waddell (1978)
 - 2.39. Raven (1977)
 - 2.40. I am not arguing that it is essential to continue to organize society on a hierarchical model. However, it is important to note that the effective operation of an egalitarian society in which everyone would be able to develop, utilize, and get recognition for his or her talents would be dependent on the development of exactly the same tools as we are concerned with in this chapter—for they are crucial to the implementation of any effective manpower guidance, placement and development policy. It would also be dependent on the development of the tools needed to administer and evaluate diversity in public provision (see Raven, 1987, 1989). The problems of the so-called communist economies stemmed, not from failure to adopt the mechanisms of the marketplace, but from failure to develop these alternative administrative tools.
 - 2.41. Fraley (1981) has shown that this is one of the main reasons why otherwise well-resourced attempts to implement Core Curriculum and Dewey-style progressive education have failed.
 - 2.42. The implications for the design of evaluation studies are that failure to get a rough fix on an important benefit or disbenefit of an educational program, failure to index its failure to achieve one of its main goals, or failure to identify a variable which contributes in some important way to its success or failure, is a more important oversight than failure to assess those inputs and outcomes using measures whose reliability and validity has already been established. The result is that,

whereas accuracy is possibly a hallmark of a good academic study, the hallmark of a good evaluation study is its comprehensiveness. For a further discussion see Raven (1984) "Some Limitations of the Standards" and Raven (1985) "Some Questions and Lessons for the Bank".

- 2.43. These issues have been discussed more fully both by Schwarz and Raven in Searle (1985).
- 2.44. Jackson (1986)
- 2.45. Havighurst and Taba (1949)
- 2.46. McClelland (1962, 1982, 1982); Winter, McClelland and Stewart (1981).
- 2.47. Raven et al (1985); Raven and Dolphin (1978); Raven (1984)
- 2.48. Winter et al (1981); Raven and Dolphin (1978); Lempert (1986)
- 2.49. Foshay (1987)
- 2.50. Eisner (1985), for example, while importantly legitimizing an approach to evaluation which enables evaluators to comment on some of the wider aspects of education with which I have been concerned in this chapter, has done both science and education a disservice by accepting and reinforcing widely-held, but mistaken, beliefs about the nature of science. Instead of speaking about the art of educational evaluation, he should have used his observations to challenge widely-held concepts of science.
- 2.51. Raven (1977); Raven et al (1985)
- 2.52. There are two problems with the factor-analytic studies which have reinforced the concept of 'ability'. The first is that the range of abilities sampled is very narrow: they measure only so-called "academic" abilities, entirely neglecting qualities like leadership, decision-taking ability, or the ability to put others at ease. When these are added, many of the correlations fall to 0.2—which means that 96% of the variance on one variable cannot be predicted from the general factor. Secondly, they have overlooked the fact that most important abilities only develop with practice in appropriate environments—ie they have failed to note that the context in which the measures were taken did not foster a wide range of abilities.
- 2.53. I.e. Taylor (1985)
- 2.54. Smith, L. See numerous issues of IDEAS, which he edited, and MacIntosh and Smith (1974). See also James (1968).
- 2.55. Burgess and Adams (1986)
- 2.56. Raven (1977, 1977, 1984)
- 2.57. Bachman et al (1978)
- 2.58. Morton-Williams et al (1968); Raven (1977)
- 2.59. Flanagan (1978)
- 2.60. HMIs (1980)

-
- 2.61. Goodlad (1983)
 - 2.62. Raven et al (1985)
 - 2.63. Raven et al (1985)
 - 2.64. See, eg, Walker (1976); Carroll (1975).
 - 2.65. HMIs (1978)
 - 2.66. Cockcroft Report (1982)
 - 2.67. Spencer, E (1979, 1983)
 - 2.68. Goodlad (1983)
 - 2.69. Raven (1980)
 - 2.70. Taylor et al (1963)
 - 2.71. See Broadfoot, (Ed.) (1986) for an overview of these and Burgess and Adams (1986) for details on the best of these schemes.
 - 2.72. Spencer, E (1979, 1983). See also the review of relevant studies by Ingenkamp (1977).
 - 2.73. Wolf (1987)
 - 2.74. Ingenkamp (1977)
 - 2.75. Hunter and Hunter (1984). The Hunters' main point is that intelligence tests are the only tests which account for a substantial proportion of the variance in occupational performance. Across thousands of studies of a wide range of occupational groups they account for about 25% of the variance in performance. (The figure drops to 9% in longitudinal studies). No other measures—of interests, biography, handwriting or anything else (including assessment center-based ratings) account for more than 1% of the variance. A good Vocabulary test which takes only 5 minutes to administer will do as well as longer, more complicated, "Intelligence" tests (Raven, Court and Raven, 1988).
 - 2.76. Berg (1973); Freedman and Berg (1978)
 - 2.77. Jencks et al (1973)
 - 2.78. Taylor et al (1963); Taylor (1971, 1971)
 - 2.79. Bray, Campbell and Grant (1974)
 - 2.80. McClelland et al (1958); McClelland (1973)
 - 2.81. Berg (1973); Freedman and Berg (1978)
 - 2.82. Collins (1979)
 - 2.83. Dore (1976)
 - 2.84. Jencks et al (1973)
 - 2.85. Holland and Richards (1965)
 - 2.86. Hoyt (1965)

- 2.87. Raven (1977)
- 2.88. Bachman et al (1978)
- 2.89. It is important to note that this does not mean that some schools do not make a difference or that more could not do so—for both Winter, McClelland and Stewart (1981) at university level, and my colleagues and I at elementary and high school level (Raven 1977, Raven et al 1985), have shown that some teachers do indeed make a dramatic difference to the development of general competence.
- 2.90. Berg (1973)
- 2.91. Folger and Nam (1964); O'Toole (1975)
- 2.92. Hope (1985)
- 2.93. Raven (1977)
- 2.94. Kohn (1969)
- 2.95. Kinsey (1948)
- 2.96. Many of those who joined the Progressive Education movement did so because they noticed that very few of the "more academic" pupils who are selected and promoted by the educational system are the right sort of people to hold high office. Nor are they the best potential managers—or even teachers. Nuttgens' (1988) experience, based on a lifetime in education, is that, through the educational system, we promote "those who are least able to do anything" into influential positions in society. Bernstein (1975) argued that a disinterested observer would conclude that the goals of elementary education had been deliberately obfuscated so that the system would promote those who are best able to work out what those in authority want to hear and most willing to do it. M.F.D. Young (1971) and others have likewise argued that educational knowledge has been (deliberately?) organized and framed in such a way that students have to work out what the authorities will count as knowledge and then supply that particularly arid form of "knowledge" in order to secure advancement. What these observations amount to is a statement that, from this point of view, not only are the manifest goals of education mere window dressing, the fraudulent institutions of our society (including the educational system) require fraudulent people who will, on the one hand, perpetuate the myths of our society and engage in the kinds of fraudulent activity which are required to purchase their own advancement in that system and, on the other, decline to question or seek to influence the workings of society. This thesis is examined much more fully in the authors' companion volume *The Most Important Problem in Education is to Come to Terms with Values* (Trillium Press, 1990). Here it is sufficient to note that while the educational system's tendency to both breed and select such people may in the past have been functional for both the individuals concerned and the societies in which they live, that is, given the economic and ecological crises which confront us, unlikely to be true in the future. To tackle the pressing financial, social, environmental, economic and ecological problems which confront us, we must ensure that the leaders and

managers of our society (in both the public and private sectors) have the concerns and the abilities that are required to act on information in the long term interest of society.

- 2.97. Raven and Dolphin (1978).
- 2.98. Tomlinson and Tenhouten (1976)
- 2.99. Miller and Parlett (1974)
- 2.100. Winter (1973); McClelland (1964, 1975); McClelland et al (1972); McClelland and Burnham (1976)
- 3.1. Raven (1984)
- 3.2. The term "value" is not quite right—because the behaviors in question often seem to be rather compulsive. People engage in them "despite themselves". This is difficult to reconcile with the term "valued activity", which conjures up an image of a 'freely chosen' activity. Yet people do usually agree that these activities are important to them, and it is in this sense that they can truly be said to value them. McClelland has tried to avoid the difficulty by using the term 'need'. Unfortunately, this has led him to claim that his measures are not measures of values. This is not only untrue, it has also, as I have shown (Raven 1988), caused endless confusion and unnecessary argument.
- 3.3. The conative components are those concerned with determination, persistence, and will. In the American literature (other than that associated with McClelland) these components have either been ignored or inappropriately subsumed under 'affective' (although Snow, influenced by Heckhausen, himself influenced by McClelland, is currently bent on re-instating them). Yet a person can very much enjoy doing something without being determined to see it through—and he or she can hate doing something, but still be determined to do it.
- 3.4. This does not mean that it is not useful to think about behavior in terms of these categories: only that, in practice, attempts to assess the components separately are mistaken.
- 3.5. While it may be thought that the viewpoint developed here might be reconciled with traditional factor-analytic theory by focussing on qualities like "the ability to make one's own observations", a little reflection shows that this is not the case. Our argument is precisely that such qualities cannot be assessed independently of valued goals. They have no generalized meaning. Therefore, they cannot be assessed by factorially-pure scales.
- 4.1. Raven, Molloy and Corcoran (1972); Raven (1977)
- 4.2. These components of competence are spelt out in more detail in Raven (1977, 1984).
- 4.3. It is not, in fact, difficult to reconcile some such model with the facts to which factor-analysts point as a justification for their model. They point out that most human traits are correlated with each other. They go on to argue that it is un-

necessary to retain a large number of independent dimensions, or categories. However, many of the correlations are of the order of 0.2 and most are of the order of 0.3 to 0.5. Even the latter leave some 75% of the variance on one trait 'unexplained' by the variance on the other. There is, therefore, a good chance that someone who is not good at one thing will be good at another. Even factor-analysts point out that this is because the second ability has probably caught the interests of the person concerned and, therefore, been practiced and developed. While the factor-analyst's model does, in fact, provide for such possibilities (by including provision for specific factors) these are generally neglected in practice. If we were forced to state our case in factor-analytic terms, we would therefore find ourselves arguing that the important things to record about an individual are his specifics, not his generalities.

- 4.4. Greeno (1989); Brown et al (1989)
- 4.5. It is failure to unscramble values from the competencies which are required for problem solving which is the fundamental flaw in Sternberg's (1985, 1986) triarchic theory of intelligence. Academic, Social and Practical "intelligence" involve the application of the competencies which are required for problem solving to different valued goals. They are not different "types" of intelligence. Furthermore, "problem solving" is a difficult and demanding activity which involves persistence over time and in the face of obstacles. It is also crucially dependent on such things as sensitivity to the slight feelings on the fringe of consciousness which tell one that one has a problem or the germ of a solution (ie on affective activities) and on the initiation—on the basis of hunches— "experimental interactions" with the environment which rely on the use of feelings to "monitor" their effectiveness and learn from them. It therefore emerges that what appears to be the archetype of a "pure cognitive ability" is none other than a complex motivational disposition, or competence, of the kind we have been centrally concerned with in this book.
- 5.1. Raven (1980); McCail (1981)
- 5.2. Raven et al (1985)
- 5.3. Raven (1984); Graham and Raven (1987)
- 5.4. Raven (1984)
- 5.5. Raven (1980)
- 5.6. Klemp et al (1977)
- 5.7. Smith and Kendall (1963); for a summary of the procedure as we have used it, see Raven (1977).
- 5.8. McClelland (1978); Spencer, L.M. (1983)
- 5.9. Stansbury (1976, 1980)
- 5.10. Burgess and Adams (1986)

-
- 5.11. Fishbein (1967); Fishbein and Ajzen (1975). However, see also Vroom(1964); Porter and Lawler (1968); Feather (1982); Mitchell (1982).
- 5.12. This is the explanation of the still widely-encountered statement that "there is little relationship between attitudes and behavior". It is true that there is little relationship between behavior in a particular situation and scores on a single, factorially-pure, attitude or personality scale. But there is a very close relationship between behavior and "attitudes" (or behavior tendencies) indexed by identifying and summing the perceptions, beliefs and feelings which come into play in the particular situation using the techniques under discussion here.
- 5.13. This evidence is reviewed in Raven and Dolphin (1978).
- 5.14. Raven (1977); Walberg (1979-85); Howard (1982)
- 5.15. This would place the procedure advocated by the Scottish Examination Board (1985) for Social and Vocational Education on a firm basis. For this syllabus, the SEB does not require teachers to assess individual pupils' social and vocational competence. It insists only that they certify that the course was likely to lead to these outcomes. This is something we are in a position to validate. From there it would be but a short step to trust the teachers' judgment about individuals.
- 6.1. Messick (1989)
- 6.2. McCail (1981); Raven (1980). For earlier applications of the partially-developed model see Raven et al (1972).
- 6.3. Raven and Varley (1984)
- 6.4. I am talking here at a fairly gross level. Teachers varied a great deal from one to another in their educational objectives. In relation to this variation between teachers, the slippage between one teacher's objectives and what he or she did looks less stark than it does when researchers focus on such things as the discrepancy between teachers' reporting that they have conducted an 'open-ended' discussion lesson, and external observers' ratings of the 'openness' of that discussion. What was striking in our study was how few teachers thought it was important to have open-ended discussions. Those teachers who thought it was important to do so got on with it, albeit imperfectly. Not surprisingly, the rest conducted no such discussions.
- 6.5. Raven (1982, 1984)
- 6.6. Graham and Raven (1987); Graham, Raven and Smith (1987)
- 6.7. Large (1986)
- 6.8. Feather (1982). It is, perhaps, useful to mention that, although the evidence of validity cited here has not been expressed in the form of correlation coefficients, it provides exactly the information which those coefficients seek to provide: namely, evidence that scores based on verbal behavior relate to other aspects of behavior and vary with experimental manipulation (cf Messick, 1989).

- 7.1. Peters (1987), and Roberts (1968) and Flanagan (1958, 1976) have underlined the costs to society of not implementing more effective procedures for ensuring that people's talents are identified and developed.
- 7.2. Goodlad (1983)
- 7.3. ETS (1985); See also Eyde (1987)
- 7.4. Joint Committee (1981)
- 7.5. Travers (1973); Wittrock (1986)
- 7.6. AERA (1982)
- 7.7. Husen and Postlethwaite (1985)
- 7.8. McClelland (1961)
- 7.9. Roberts (1968); Oeser and Emery (1958)
- 7.10. Kanter (1985)
- 7.11. Lerner (1987)
- 7.12. I have discussed these issues at some length in Raven (1975, 1983, 1984, 1985, 1987).
- 7.13. Winter et al (1981)
- 7.14. Schon (1987)
- 7.15. Cf Schon (1987); Klemp et al (1977); Winter et al (1981); Hope (1985); Raven (1984); Raven et al (1985).
- 7.16. De Bono (1980); Emery et al (1974); Kanter (1985); Morgan (1986); Toffler (1981); Raven (1980, 1981, 1982, 1983, 1984, 1986, 1987, 1988)
- 7.17. Morton-Williams et al (1968)
- 7.18. Raven (1977, 1984, 1985, 1986, 1987, 1990, 1990, 1990)
- 7.19. Eg Goodlad (1983); Carnegie Reports.
- 7.20. See Raven (1986)
- 7.21. Dore (1976)
- 7.22. Robinson (1983)
- 7.23. Bronfenbrenner and several others have told me that Headstart began in the Office of Economic Opportunity. It was initially a program of adult education designed to foster the competencies required to promote community development. It proved to be so easy to unleash competent (political) behavior among adults that the situation had to be defused by changing the program goal from one which was attainable but disconcerting to one which was laudable but unattainable. A convenient goal of this sort consisted of raising children's IQ's. This was one reason why those who commissioned program evaluations were not interested in identifying programs which actually promoted the growth of competence.

REFERENCES.

- American Educational Research Association. (1982). *Encyclopedia of Educational Research*. 5th Edition. New York: AERA.
- Bachman, J.G., O'Malley, P.M. and Johnston, J. (1978). *Adolescence to Adulthood: Change and Stability in the Lives of Young Men*. Ann Arbor, Michigan: The Institute for Social Research.
- Bennett, N. (1976). *Teaching Styles and Pupil Progress*. London: Open Books.
- Berg, I. (1973). *Education and Jobs: The Great Training Robbery*. London: Penguin Books.
- Beuret, G. and Webb, A. (1983). *Goals of Engineering Education*. London: CNA.
- Bill, J.M., Trew, C.J. and Wilson, J.A. (1974). *Early Leaving in Northern Ireland*. Belfast: Northern Ireland Council for Educational Research.
- Boyer, E.L. (1983). *High School: A Report on Secondary Education in America*. The Carnegie Foundation for the Advancement of Teaching, New York: Harper and Row.
- Bray, D.W., Campbell, R.J. and Grant D.L. (1974). *Formative Years in Business*. New York: Wiley.
- Broadfoot, P. (Ed.) (1986). *Profiles and Records of Achievement*. Eastbourne, England: Holt, Rinehart and Winston.
- Brown, J.S., Collins, A. and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, Jan-Feb, 32-42.
- Bullock Report. (1975). *A Language for Life*. London: HMSO. Burgess, T. and Adams, E. (1986). *Records of Achievement at 16*. Windsor: NFER-Nelson.
- The Carnegie Report. See Boyer (1983), Feistritzer (1983), Perrone (1985).
- Carroll, J.B. (1975). *The Teaching of French as a Foreign Language in Eight Countries*. Stockholm: Almqvist and Wiksell; New York: John Wiley, The Halsted Press.
- Centre for Educational Sociology, University of Edinburgh. (1977). *Collaborative Research Dictionary*.
- Closs, S.J. (1986). Current and Future Developments of the JIIG-CAL System. *British Journal of Guidance Counselling*, Vol 14.1, January.
- Closs, S.J., MacLean, P.R. and Walker, M.V. (1985). *An Evaluation of the JIIG-CAL System*. London: Hodder and Stoughton.
- Cockcroft, W.H. (Chairman) (1982). *Mathematics Counts. Report of the Committee of Enquiry into the Teaching of Mathematics in Schools*. London: HMSO.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Hood, A.M., Weinfeld, L.D. and York, R.L. (1966). *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office.
- Collins, R. (1979). *The Credential Society*. New York: Academic Press.

- Comber, L.C. and Keeves, J.P. (1973). *Science Education in Nineteen Countries*. Stockholm: Almqvist and Wiksell; New York: John Wiley, The Halsted Press.
- De Bono. E. (1980). *Future Positive*. Mddx, England: Pelican Books.
- De Landsheere, V. (1977). On Defining Educational Objectives. *Evaluation in Education*, 1, No.2, 73-190. Oxford: Pergamon Press.
- Dore, R. (1976). *The Diploma Disease*. London: Allen and Unwin.
- Eisner, E.W. (1985). *The Art of Educational Evaluation*. Lewes: The Falmer Press.
- Emery, F. et al. (1974). *Futures We're In*. Centre for Continuing Education, Australian National University.
- ETS. (1985). *The Redesign of Testing for the 21st Century*. Princeton, N.J.: ETS.
- Eyde, L.E. (Ed.) (1987). *Computerised Psychological Testing*. Special Issue: *Applied Psychology*, 36, nos 3&4.
- Feather, N.T. (Ed.) (1982). *Expectations and Actions: Expectancy-Value Models in Psychology*. Hillside, N.J.: Erlbaum.
- Feather, N.T. (1982). Human Values and the Prediction of Action: An Expectance-Valence Analysis. in Feather, N.T. (Ed.), *Expectations and Actions: Expectancy-Value Models in Psychology*. Hillside, N.J.: Erlbaum.
- Feistritzer, E. (1983). *The Condition of Teaching*. New York: Carnegie Foundation for the Advancement of Teaching.
- Fishbein, M. (Ed.) (1967). *Readings in Attitude Theory and Measurement*. New York: Wiley.
- Fishbein, M. and Ajzen, I. (1975). *Belief, Attitude, Intention and Behaviour*. Reading Mass.: Addison Wesley.
- Flanagan, J.C. (1958). *Personal and Social Performance Record: Teachers Guide and Materials*. Palo Alto, California: American Institute for Research.
- Flanagan, J.C. (1976). *Planning Life and Career Goals: A Cluster of Materials and Manuals*. Monterey, Calif.: CTB/McGraw-Hill.
- Flanagan, J.C. (1978). *Perspectives on Improving Education from a Study of 10,000 30-year-olds*. New York: Praeger Publishers.
- Flanagan, J.C. and Burns, R.K. (1955). The Employee Performance Record. *Harvard Business Review*, 33, 95-102.
- Flanagan, J.C. and Russ-Eft, D. (1975). *An Empirical Study to Aid in Formulating Educational Goals*. American Institutes for Research: Palo-Alto, California.
- Flanagan, J.C. et al. (1975). An Individual Instructional System: PLAN. in Talmage, H. (Ed.), *Systems of Individualised Education*. Berkeley, California: McCutchan Publishing Corp.
- Folger, J.K. and Nam, C.B. (1964). Trends in education in relation to the occupational structure. *Social. of Education*. Fall, 19-34.
- Foshay, A.W. (1987). The Curriculum Matrix. *The Education Forum*. Vol 51, No.4, Summer, 341-353.
- Fraleigh, A. (1981). *Schooling and Innovation: The Rhetoric and the Reality*. New York: Tyler Gibson.

- Frederiksen, N. (1984). The Real Test Bias: Influences of Testing on Teaching and Learning. *American Psychologist*, 39, 193-202.
- Freedman, M. and Berg, I. (1978). Investment in Education: Public Policies and Private Risks. *New Directions in Education and Work*, 2, 19-34.
- Galton, M. and Simon, B. (1980). *Progress and Performance in the Primary Classroom*. London: Routledge and Kegan Paul.
- Galton, M., Simon, B. and Croll, P. (1980). *Inside the Primary Classroom*. London: Routledge and Kegan Paul.
- Goodlad, J. (1983). *A Place Called School*. New York: McGraw-Hill.
- Graham, M.A., Raven, J. and Smith, P.C. (1987). Identification of High Level Competence: Cross-Cultural Analysis between British, American, Asian and Polynesian Labourers. To be published in *Organisation Forum*.
- Graham, M.A. and Raven, J. (1987). International Shifts in the Workplace—Are We Becoming an “Old West” in the Next Century? Provo: BYU Dept. Organizational Behavior.
- Greeno, J.G. (1989). A Perspective on Thinking. *Amer. Psychologist*, 44, 134-141.
- Harlen, W. (1984). The impact of APU science work at LEA and school level. *J. Curr. Studies*, 16, 89-94.
- Havighurst, H.J. and Taba, H. (1949). *Adolescent Character and Personality*. New York: Wiley.
- HMI. (1978). *Primary Education in England: A Survey by H. M. Inspectors of Schools*. Dept. Education and Science, London: HMSO.
- HMI. (Scotland) (1980). *Learning and Teaching in Primary 4 and Primary 7*. Edinburgh: HMSO.
- Holland, J.L. and Richards, J.M. (1965). Academic and Non-Academic Accomplishments. *J. Ed. Psychology*, 56, 165-175.
- Hope, K. (1985). *As Others See Us: Schooling and Social Mobility in Scotland and the United States*. New York: Cambridge University Press.
- Howard, E. (1980). *Some Ideas on Improving School Climate*, Colorado: Department of Education.
- Howard, E. (1982). *Involving Students in School Climate Improvement. New Designs for Youth Development*. Tucson: Associations for Youth Development Inc.
- Howard, E. (1982). *Successful Practices for Making the Curriculum More Flexible*. Denver: Colorado Department of Education.
- Howard, E. (1982). *Instrument to Assess the Educational Quality of Your School*. Denver: Colorado Department of Education.
- Hoyt, D.P. (1965). *The Relationship Between College Grades and Adult Achievement*. Iowa: American College Testing Program, Res. Report No. 7.
- Huff, S., Lake, D. and Schaalman, M.L. (1982). *Principal Differences: Excellence in School Leadership and Management*. Boston: McBer & Co.
- Hunter, J.E. & R.F. (1984). Validity and Utility of Alternative Predictors of Job Performance. *Psychol. Bull*, 96, 72-98.

- Husen, T. and Postlethwaite, N. (Eds.) (1985). *International Encyclopaedia of Education*. London: Pergamon.
- Igenkamp, K. (1977). *Educational Assessment*. Windsor, Berkshire: NFER.
- ITRU (1979). *The A-Z Study: Differences between Improvers and Non-Improvers among Young Unskilled Workers*. Cambridge: The Industrial Training Research Unit.
- Jackson, P.W. (1986). *The Practice of Teaching*. New York: Teachers College Press.
- James, C. (1968). *Young Lives at Stake*. London: Collins.
- Jaques, E. (1976). *A General Theory of Bureaucracy*. London: Heinemann.
- Jencks, C., Smith, M., Acland, H., Bane, M.J., Cohen, D., Gintis, H., Heyns, B., and Michelson, S. (1972/3). *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books; London, England: Penguin Books.
- Johnston, L.D. and Bachman, J.G. (1976). *Educational Institutions*. in Adams, J.F. (Ed.), *Understanding Adolescence*, 3rd Edition, p.290-315. Boston: Allyn and Bacon.
- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for Evaluations of Educational Programs, Projects and Materials*. New York: McGraw-Hill Book Co.
- Kanter, R.M. (1985). *The Change Masters: Corporate Entrepreneurs at Work*. Hemel Hempstead: Unwin Paperbacks.
- Kinsey, A.C. (1948). *Sexual Behaviour 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). *Class and Conformity: A Study in Values*. Illinois: Dorsey Press.
- Large, P. (1986). Article summarising work conducted by the Taylor-Nelson Monitor. in *The Guardian*, 7 Feb. 1986.
- Lempert, W. (1986). *Sozialisation und Persönlichkeitsbildung in beruflichen Schulen, dargestellt am Beispiel der Entwicklung moralischer Orientierung*. *Die berufsbildende Schule*, 38, 148-160.
- Lerner, B. (1987). *A National Census of Educational Quality – What Is Needed?* NASSP Bulletin, March, 42-60.
- MacBeath, J., Mearns, D., Thomson, B. and How, S. (1981). *Social Education: The Scottish Approach*. Glasgow: Jordanhill College of Education.
- McCail, G. (1981). *Mother Start. An Account of an Educational Home Visiting Scheme for Pre-School Children*. Edinburgh: Scottish Council for Research in Education.
- McClelland, D.C. (1961). *The Achieving Society*. New York: Van Nostrand.
- McClelland, D.C. (1962). *On the Psychodynamics of Creative Physical Scientists*. in Gruber, H.E. (Ed.), *Contemporary Approaches to Creative Thinking*. New York: Atherton.
- McClelland, D.C. (1964). *The Roots of Consciousness*. Princeton, N.J.: Van Nostrand.
- McClelland, D.C. (1973). *Testing for Competence rather than for "Intelligence"*. *Amer. Psychologist*, 28, 1-14.
- McClelland, D.C. (1975). *Power: The Inner Experience*. New York: Irvington.

- McClelland, D.C. (1978). *Guide to Behavioural Event Interviewing*. Boston: McBer.
- McClelland, D.C. (1982). *Education for Values*. New York: Irvington.
- McClelland, D.C. (1982). What Behavioural Scientists Have Learned About How Children Acquire Values. in McClelland, D.C. (Ed.), *The Development of Social Maturity*. New York: Irvington Press.
- McClelland, D.C., Baldwin, A.L., Bronfenbrenner, U. and Strodbeck, F.L.(1958). *Talent and Society*. Princeton, N.J.: Van Nostrand.
- McClelland, D.C. and Burnham, D.H. (1976). Power is the great motivation. *Harvard Business Review*, 54, No.2.
- 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.
- McClelland, D.C. et al. (1972). *The Drinking Man*. New York: Free Press.
- MacIntosh, V. and Smith, L. (1974). *Towards A Freer Curriculum*. London: University of London Press.
- Messick, S. (1989). Meaning and values in test validation: The Science and ethics of assessment. *Educational Researcher*, 18, No.2, 5-11.
- Miller, C.M.L. and Parlett, M. (1974). *Up to the Mark: A Study of the Examination Game*. London: Society for Research in Higher Education.
- Mitchell, T.R. (1982). "Expectancy-Value Models in Organisational Psychology", p293-312. in Feather, N.T. (Ed.), *Expectations and Actions: Expectancy-Value Models in Psychology*. Hillside, N.J.: Erlbaum.
- Morgan, G.A. (1986). *Images of Organisation*. Beverly Hills: Sage.
- Morton-Williams, R., Finch, S., Poll, C., Raven, J. and Ritchie, J. (1968). *Young School Leavers*. London: HMSO.
- Morton-Williams, R., Raven, J. and Ritchie, J. (1971). *Sixth Form Teachers and Pupils*. London: Schools Council/Books for Schools.
- Nash, R. (1973). *Classrooms Observed*. London: Routledge and Kegan Paul.
- National Task Force For Economic Growth. (1983). *Action for Excellence*. Denver: Education Commission of the States.
- Nuttgens, P. (1988). *What Should We Teach and How Should We Teach It?* Aldershot: Wildwood House.
- Oeser, O.A. and Emery, F.E. (1958). *Information, Decision and Action*. University of Melbourne.
- ORACLE. See Galton & Simon (1980), Galton, Simon & Croll (1980), Simon & Willcocks (1981).
- Oppenheim, A.N. and Torney, J. (1974). *The Measurement of Children's Civic Attitudes*. Stockholm: IEA; New York: John Wiley, The Halsted Press.
- O'Toole, J. (1975). *Human Resources Development and Competency-Based Education*. Syracuse University Research Corp.: Educational Policy Research Center.

- Passow, A.H., Noah, H.J., Eckstein, M.A. and Mallea, J.R. (1976). *An Empirical Study of Twenty-One Educational Systems*. Stockholm: Almqvist and Wiksell.
- Peaker, G.F. (1975). *An Empirical Study of Education in Twenty One Countries*. Stockholm: Almqvist and Wiksell; New York: Wiley.
- Perrone, V. (1985). *Portraits of High Schools*. New York: Carnegie Foundation for the Advancement of Teaching.
- Peters, D.H. and Roberts, E.B. (1967). *Unutilized Ideas in University Laboratories: A Preliminary Study*. Mass.: Alfred P. Sloane School of Management, Institute of Technology.
- Peters, R. (1987). *Practical Intelligence*. New York: Harper and Row.
- Porter, L.W. and Lawler, E.E. (1968). *Managerial Attitudes and Performance*. Homewood Illinois: The Dorsey Press.
- Price, P.B., Taylor, C.W., Nelson, D.E. et al. (1971). *Measurement and Predictors of Physician Performance: Two Decades of Intermittently Sustained Research*. Salt Lake City: University of Utah, Dept. of Psychology.
- Raven, J. (1975). *Social Research in Modern Society: I: The Role of Social Research; II: The Institutional Structures and Management Styles Required to Execute Policy-Relevant Social Research*. *Administration*, 23, 225-246 and 247-268.
- 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*. Oxford: Oxford Psychologists Press.
- Raven, J. (1977). *On the Components of Competence and their Development in Education*. *Teachers' College Record*, 78, 457-475.
- Raven, J. (1977). *School Rejection and its Amelioration*. *Educational Research*, 20, 3-9.
- Raven, J. (1980). *The Most Important Problem in Education is to Come to Terms with Values*. *Oxford Review of Education*, 7, 253-72.
- Raven, J. (1980). *Parents, Teachers and Children*. Edinburgh: The Scottish Council for Research in Education.
- Raven, J. (1981). *Social Research and Development in the Doldrums*. *New Universities Quarterly*, 36, 72-86.
- Raven, J. (1982). *The Edinburgh Questionnaires: A Cluster of Questionnaires for use in Organisational Development and in Staff Guidance, Placement and Development*. Oxford: Oxford Psychologists Press.
- Raven, J. (1982). *Public Policy in a Changed Society*. *Higher Education Review*, 14, 80-89.
- Raven, J. (1983). *The Relationship Between Educational Institutions and Society with Particular Reference to the Role of Assessment*. *International Review of Applied Psychology*, 42, 249-274.
- Raven, J. (1983). *Towards new Concepts and Institutions in Modern Society*. *Universities Quarterly*, 37, 100-118.

- Raven, J. (1984). *Competence in Modern Society: Its Identification, Development and Release*. Oxford: Oxford Psychologists Press.
- Raven, J. (1984). Quality of Life, the Development of Competence, and Higher Education. *Higher Education*, 13, 393-404.
- Raven, J. (1984). The Role of the Psychologist in Formulating, Administering and Evaluating Policies Associated with Economic and Social Development in Western Society. *J.Econ.Psychol.*, 5, 1-16.
- Raven, J. (1984). Some Barriers to Educational Innovation from Outside the School System. *Teachers College Record*, 85, 431-443.
- Raven, J. (1984). Some Limitations of the Standards. *Evaluation and Program Planning*, 7, 363-370.
- Raven, J. (1984). What is Language Teaching? *New Horizons*, 25, 82-90. Also published in *Teaching English*, 18, No.1, 23-27.
- Raven, J. (1985). The Institutional Framework Required for, and Process of, Educational Evaluation: Some Lessons from Three Case Studies. in Searle, B. (Ed.), *Evaluation in World Bank Education Projects: Lessons from Three Case Studies*. Washington, DC: The World Bank, Education and Training Dept. Report EDT5 141-170.
- Raven, J. (1986). A Nation Really at Risk: A Review of Goodlad's "A Place Called School". *Higher Education Review*, 18, 65-79.
- Raven, J. (1986). *Fostering Competence*. in Burgess, T. (Ed.), *Education for Capability*. London: NFER-Nelson.
- Raven, J. (1986). Review Article: The Future of Work. *Journal of Occupational Psychology*, 84, 343-348.
- Raven, J. (1987). Choice in a Modern Economy: New Concepts of Democracy and Bureaucracy. in Maital, S. (Ed.), *Applied Behavioural Economics*. Brighton: Wheatsheaf Books.
- Raven, J. (1987). The Crisis in Education. *The New Era*, 68, 38-44.
- 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. (1989). Questionable assumptions in test construction. *Bulletin of the International Test Commission*, Nos.28 & 29, 67-95.
- Raven, J. (1989). Democracy, bureaucracy and the psychologist. *The Psychologist*, Vol.2, No.11, November, 458-466.
- Raven, J. (1989). Equity in diversity: The problems posed by values—and their resolution. pp59-101 in F. Macleod (Ed.), *Families and Schools: Issues in Accountability and Parent Power*. Brighton, England: Falmer Press.
- Raven, J. (1990). *Giving Effect to the New Values*. New York: Trillium Press.
- Raven, J. (1990). *The Most Important Problem in Education is to Come to Terms with Values*. New York: Trillium Press.
- Raven, J. (1990). *Quality Control in Public Education*. New York: Trillium Press.
- Raven, J. (1990). The Barriers to Achieving the Wider Goals of General Education. *British Educational Research Journal*, No. 16, 273-296.

- 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. and Dolphin, T. (1978). *Toward Value-Expectancy Measures of Human Resources*. Edinburgh: 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 1: Teachers' Perceptions of Educational Objectives and Examinations*. Dublin: Irish Association for Curriculum Development.
- 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: The Scottish Council for Research in Education. Distributed in the United States by Trillium Press.
- Raven, J. Molloy, E. and Corcoran, R. (1972). Toward a Questionnaire Measure of Achievement Motivation. *Human Relations*, 25, 469-492.
- 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, No.1, F4 G6.
- Raven, J.C., Court, J.H., and Raven, J. (1988). *A Manual for Raven's Progressive Matrices and Vocabulary Tests*. London, England: H.K.Lewis; San Antonio, Texas: The Psychological Corporation.
- Roberts, E.B. (1968). A Basic Study of Innovators: How to Keep and Capitalize on their Talents. *Research Management*, XI, 249-266.
- Robinson, D.W. (1983). *Patriotism and Economic Control: the Censure of Harold Rugg*. Rutgers University, New Jersey: D.Ed. Dissertation. 1984: Ann Arbor, Michigan: University Microfilms International.
- Rogers, E.M. (1962). *Diffusions of Innovations*. Glencoe: Free Press.
- Rutter, M., Maughan, B., Mortimer, P. and Ousten, J. (1979). *Fifteen Thousand Hours: Secondary Schools and their Effects on Children*. London: Open Books.
- Schneider, C., Klemp, G.O. and Kastendiek, S. (1981). *The Balancing Act: Competencies of Effective Teachers and Mentors in Degree Programs for Adults*. Boston: McBer & Co.
- Schon, D. (1987). *Educating the Reflective Practitioner*. San Francisco: Jossey-Bass.
- Schools Council. (1971-1982). *Work Series*. Harlow, England: Longman.
- Schools Council. (1973). *Environmental Studies 5-13: The Use of Historical Resources*. Schools Council Working Paper 48. London: Evans/Methuen Educational.
- Schwartz, H.H. (1987). Perceptions, Judgment and Motivation in Manufacturing Entrepreneurs. *J.Econ. Behav. and Organisation*. (in press).
- Schwarz, P.A. (1985). Evaluation of the Pakistan Primary Education Project; A Methodological Case Study. in Searle, B. (Ed.), *Evaluation in World Bank Education*

- Projects; Lessons from Three Case Studies. Washington D.C.: The World Bank, Education and Training Dept, Report No. EDT5.
- Scottish Examination Board. (1985). Scottish Certificate of Education: Social and Vocational Skills on the Standard Grade. Revised Scheme for Moderation of Internal Syllabuses and Assessment Arrangements.
- Searle, B. (Ed.) (1985). Evaluation in World Bank Education Projects; Lessons from Three Case Studies. Washington D.C.: The World Bank, Education and Training Dept, Report No. EDT5.
- Simon, B. and Willcocks, J. (Eds.) (1981). Research and Practice in the Primary Classroom. London: Routledge and Kegan Paul.
- Smith, L.A. (1966). The Raising of the School Leaving Age: Pilot Course Report. London: Goldsmiths College.
- Smith, P.C. and Kendall, L.M. (1963). Retranslation of Expectations. *J.App.Psychol.*, 41, 149-155.
- Spencer, E. (1979). Folio Assessments or External Examinations? Edinburgh: Scottish Secondary Schools Examinations Board.
- Spencer, E. (1983). Writing Matters Across the Curriculum. Edinburgh: The Scottish Council for Research in Education.
- Spencer, L.M. (1983). Soft Skill Competencies. Edinburgh: The Scottish Council for Research in Education.
- Stallings, J. and Kaskowitz, D. (1974). Follow Through Classroom Observation Evaluation 1972-1973. Menlo Park, California: Stanford Research Institute. Report URU-7370.
- Stansbury, D. (1976). Record of Personal Experience, Qualities and Qualifications. (plus tutor's handbook.) South Brent: RPE Publications.
- Stansbury, D. (1980). The Record of Personal Experience. in Burgess, T. and Adams, E. Outcomes of Education. Basingstoke: MacMillan Education.
- Sternberg, R.J. (1985). Beyond IQ: A Triarchic Theory of Human Intelligence. New York: Cambridge University Press.
- Sternberg, R.J. (Ed.). (1986). Practical Intelligence. New York: Cambridge University Press.
- Sykes, A.J.M. (1969). Navvies: Their Work Attitudes. *Sociology*, 3, 21 and 157.
- Taylor, C.W. (1971). All of Our Children are Educationally Underprivileged. Dept. Psychology: University of Utah.
- Taylor, C.W. (1971). All Our Students Are Educationally Deprived. Utah: University of Utah Press.
- Taylor, C.W. (1985). Cultivating Multiple Creative Talents in Students. *Journal for the Educationally Gifted*, Vol.VIII, No.3, 187-198.
- Taylor, C.W. and Barron, F. (Eds.) (1963). Scientific Creativity. 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 Taylor, C.W. and Barron, F. (Eds.), *Scientific Creativity: Its Recognition and Development*. New York: Wiley.

- Thorndike, R.L. (1973). *Reading Comprehension in Fifteen Countries*. Stockholm: Almqvist and Wiksell; New York: John Wiley, The Halsted Press.
- Toffler, A. (1981). *The Third Wave*. New York: Bantam Books.
- Tomlinson, T.M. and Tenhouten, D. (1976). *Awareness, Achievement Strategies and Ascribed Status of Elites*. Washington, D.C.: National Institute of Education. Unpublished Report.
- Travers, R.M.W. (Ed.) (1973). *Second Handbook of Research on Teaching*. Chicago: Rand McNally.
- Van Beinum, H. (1965). *The Morale of the Dublin Busman*. London: Tavistock Institute of Human Relations.
- Vroom, V.H. (1964). *Work and Motivation*. New York: John Wiley.
- Waddell, J. (1978). (Chairman). *School Examinations*. London: HMSO. Walberg, H.J. (Ed.) (1979). *Educational Environments and their Effects*. Berkeley, Ca.: McCutchan.
- Walberg, H.J. (1984). *Improving the Productivity of America's Schools*. *Educational Leadership*, 41, No.8, 19-30.
- Walberg, H.J. (1985). *Classroom Psychological Environment*. in Husen, T. and Postlethwaite, N. *International Encyclopaedia of Education*. London: Pergamon.
- Walberg, H.J. and Haertel, D. (1980). *Validity and Use of Educational Environmental Assessments*. *Studies in Educational Evaluation*, 6, 225-238.
- Walker, D.A. (1976). *The IEA Six Subject Survey: An Empirical Study of Education in Twenty One Countries*. New York: John Wiley.
- Wilson, J. et al. (1973). *Early Leaving in Northern Ireland*. Belfast: Northern Ireland Council for Educational Research.
- Winter, D.G. (1973). *The Power Motive*. New York: Free Press.
- Winter, D.G., McClelland, D.C. and Stewart, A.J. (1981). *A New Case for the Liberal Arts*. San Francisco: Jossey Bass.
- Wittrock, M.C. (Ed.) /AERA. (1986). *Handbook of Research on Teaching: Third Edition*. New York: AERA.
- Wolf, A. (1987). *Work Based Learning: Trainee Assessment by Supervisors*. Bradford, England: MSC Sales, ISCO.
- Young, M.F.D. (Ed.) (1971). *Knowledge and Control. New Directions for the Sociology of Education*. London: Collier-Macmillan.