

Author(s) of article or chapter:	Murphy, P
TITLE of article or chapter:	Sources of inequity: understanding students' responses to assessment
Author(s) / Editor(s) of source publication:	
TITLE of source publication:	Assessment in education principles, policy and practice
Year, Journal Volume & Issue Number (if applicable):	1995 Vol. 2 (3)
Place of Publication and Publisher	Abingdon: Carfax
Pages (from - to): 249 - 269	ISBN / ISSN: 0969-594X

Copyright Notice

Staff and students of this University are reminded that copyright subsists in this extract and the work from which it was taken.

- access and download a copy;
- print out a copy;

This digital copy and any digital or printed copy supplied to or made by you under such terms of this Licence are for use in connection with this Course of Study. You may retain such copies after the end of the course, but strictly for your own personal use.

All copies (including electronic copies) shall include this Copyright Notice and shall be destroyed and/or deleted if and when required by the University.

Except as provided for by copyright law, no further copying, storage or distribution (including by e-mail) is permitted without the consent of the copyright holder.

The author (which term includes artists and other visual creators) has moral rights in the work and neither staff nor students may cause, or permit, the distortion, mutilation or other modification of the work, or any other derogatory treatment of it, which would be prejudicial to the honour or reputation of the author.

This document should not be downloaded or printed by anyone other than a student enrolled on the course designated below at the University of Bath, or the course tutor(s).

Course of study: ED50303 Assessment

Scanned by the University of Bath Library on 2011-10-04

Designated person authorising scanning: Tom Brumfit

Library Barcode of scanned copy: REQ-101314

Permission: Permission to copy this article/extract was granted through the Copyright Clearance Center

Assessment in Education, Vol. 2, No. 3, 1995

Sources of Inequity: understanding students' responses to assessment

PATRICIA MURPHY

School of Education, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

ABSTRACT This article considers some of the implications for assessment practice of current views about the nature of learning, and achievement. Evidence from national surveys is presented that demonstrates some of the different ways, 'tasks' and responses to them are understood by assessors and students. This is followed by a review of evidence about differential sub-group performance which is then examined to identify potential sources of inequity in assessment practice.

Introduction

Assessment remains at the forefront of educational reform. Like teaching, there is no single approach to assessment. Practice, i.e. the tasks and tests used and the treatment and interpretation of results, depends on underlying views of: learning, learners, achievement, and educational purpose and values. The present widespread commitment to educational assessment grew out of a concern about the mismatch between assessment practice and current understanding about the nature of learners and of the learning process. As Moss comments

Many of the arguments raging today about externally imposed versus classroom generated assessment or about multiple-choice versus performance assessments are warranted in terms of the consequences for instruction and learning and for equity. (Moss, 1992, p. 236)

Traditional assessment was based on a narrow definition of achievement which was assumed to be normally distributed in the population. Educational assessment, unlike traditional psychometric approaches, presumes that many achievements are attainable by all students but how and when they attain them varies from individual to individual. Hence, student potential is not viewed as predetermined and fixed but dynamic and changeable, responsive to adult and peer support and instruction. Educational assessment tasks had to reflect this and be seen to bring greater correspondence between assessment and learning. Tasks had therefore to: represent *authentic* examples of learning activities; be seen by students to be *relevant and purposeful*; engage students *actively* in *producing a response* and provide opportunities for them to apply their understanding as they would outside of school rather than

0969-594X/95/030249-22 © 1995 Journals Oxford Ltd

Supplied by The British Library - "The world's knowledge"

merely reacting to a set of givens as in traditional objective testing; and be *interactive* and multimode to enable the full range of student capabilities (including higher-order thinking and problem-solving skills) to be monitored and hence fostered.

Performance assessment is a sub-set of educational assessment. Assessment tasks that fit a performance-based model attempt to represent genuine learning activities so that they properly reflect the learning objectives specified in assessment criteria. This is seen to enhance both their face validity and to enable them to meet better the purpose of informing teaching and learning in a constructive way. Performance assessment is a very broad term open to numerous interpretations. Hence assessment tasks reflect the characteristics specified in varying degrees.

Two particular areas of concern have been emphasised in the debate about the utility and feasibility of performance assessment. These are the lack of comparability across students and markers and the limited generalisability of performance across tasks. These concerns have emerged largely where performance assessment has been used for the purposes of certification (e.g. end of compulsory schooling examinations at 16 + in England, Wales and Northern Ireland) and accountability (e.g. National Curriculum assessment in England and Wales). Assessment experts offer a range of suggestions for addressing these concerns (see Gipps, 1994, for a discussion). Some experts, however, consider the focus on comparability and reliability to be inappropriate and argue the need to reconceptualise the notion of validity in relation to performance assessment (Moss, 1992). One of the major limitations on achieving this reconceptualisation is the absence of discussion about the models of learning and of knowledge that underpin performance assessment. Yet how knowledge is considered to develop in students, and how that knowledge is understood to be organised and accessed, fundamentally influences how, or even if, we consider comparability and generalisability can be achieved in assessment practice.

The current advocacy of performance assessment was an outcome of an emerging consensus about the inadequacy of earlier psychological theories of learning. However, no such consensus exists about the model of learning upon which performance assessment is predicated. Indeed, the models of learning underpinning various educational assessment initiatives are typically articulated in very general terms. A broadly constructivist perspective is common to most. What remains unclear is how this perspective is understood, or what influence it has on practice and on the treatment and interpretation of assessment results. A consequence of this is that an apparently common rhetoric can apply to widely differing practice. National assessment at ages 7, 11 and 14 in England is a good example of this, particularly if the original interpretation of performance assessment is compared with a more recent one (see Schools' Examination and Assessment Council [SEAC] 1991, 1993).

This article examines some of the implications for assessment of current views about the nature of learning and achievement. Of particular concern is the different way 'tasks' and responses are understood by assessors and students, and the implications of this for how outcomes are interpreted. Valid interpretation of assessment outcomes depends crucially on the assessor understanding the significant characteristics of learners and tasks and how these interact. As such, individual and group differences are of paramount concern to assessors. Yet scant attention has been paid in performance assessment development to the literature on differential performance. Indeed, recent research into exam performance at 16 + in England would suggest that those examiners involved in the development and moderation of exams are unaware of this evidence (Stobart *et al.*, 1992).

The relevant literature in this field is extensive. For these reasons the evidence drawn on is largely from national and international population surveys or established reviews of literature and research. Whenever it is possible to present only a summary position, references are given to allow readers to pursue issues in more depth. More extensive reference is made to the Assessment of Performance Unit (APU)'s findings (Foxman *et al.*, 1991) and those of the National Assessment of Educational Progress (NAEP); for this reason brief background methodological information about both is provided.

The article has three sections. In the first section, developments in understanding about the nature of learning and of achievement are considered in relation to assessment. Evidence from national population surveys (APU) in England, Wales and Northern Ireland are introduced to reinforce some of the messages emerging for assessment. This is followed by a focus on one emerging issue, that of differential performance between sub-groups, in particular males and females. In this section a summary of hypothesised sources of differential performance related to the sociocultural circumstances of students is presented, based on extensive reviews of research, in particular Gipps & Murphy (1994). The final section looks at evidence of sub-group performance from national and international surveys. These performance differences are then examined in the light of the various hypotheses proposed.

Learning and Assessment

Constructivism has been the dominant view of learning in education over the last few decades, particularly within the science and mathematics communities. Constructivism is essentially a theory of knowledge which involves conceptions of the learner, of knowledge and of the relationship between them. Key constructivist principles include the view that knowledge is not passively received by students but actively built up by them. It is also widely held by constructivists that personal knowledge, rather than informing us about the world, tells us about our experiences and how they are best organised.

The function of cognition is adaptive and serves the organisation of the experiential world, not the discovery of ontological reality.... Thus we do not find truth but construct viable explanations of our experiences. (Wheatley, 1991, p. 10)

Taken to its extreme this means that people can never compare their understanding of their experiences with an independent objective reality. There is no reality outside an individual's mental construction, so there can be no certainty about our mental representations of the world. Von Glaserfeld applies this view to teaching in the following quote where he describes the teacher as

constructing a model of the child's notions and operations. Inevitably that model will be constructed not out of the child's conceptual elements but out of conceptual elements that are the [teacher's] own. It is in this context that the epistemological principle of *fit*, rather than *match* is of crucial importance. Just as cognitive organisms can never compare their conceptual organisations of experience with the structure of an independent, objective reality, so the...teacher can never compare the model he or she has constructed of a child's conceptualisations with what actually goes on in the child's head. In the one case as in the other, the best that can be achieved is a model that remains viable within the range of available experience. (von Glaserfeld, 1989, p. 186)

Critics of educational constructivism would take exception to this denial of access to the physical world instead arguing, as Ogborn does, citing (Rorty, 1991), that 'the required relation to the world of our beliefs about it is not provided by some impossible epistemological guarantee but by the fact that they are the beliefs of real live human beings engaged in causal interaction with the world' (Ogborn, 1995). There is, nevertheless, an important message for teaching and assessment that is consequent upon this view of knowledge. A teacher, rather than knowing what goes on in students' heads, constructs models of what she *believes* to be going on. Such models are based on a teacher's experience of, and beliefs about, students. In other words teachers *give* meaning to students' action and responses. The same is true of students, who similarly give meaning to teachers' actions and words and to the activities they face in learning and assessment situations. The meanings students give will clearly depend to an extent on their understanding of school and assessment practices.

Critiques of constructivism have increased in recent years, reflecting a growth in alternative perspectives on knowledge and learning. The belief that individual mental structures are the fundamental unit of cognition is challenged in these critiques. A constructivist model of individual knowledge construction is seen to lack 'the necessary social and communitarian dimension of cognition' (Matthews, 1993, p. 367). Research on situated cognition (Brown et al., 1989) and sociocultural approaches to the process of coming to know (Lave, 1988) are increasingly being applied by educationalists trying to understand how to promote effective teaching and learning. In constructivist theories thought is analysed in terms of conceptual processes located in the mind. Sociocultural theorists argue that such a view of thinking is inadequate because it fails to take account of the socially constituted nature of individuals. Individuals cannot be considered in isolation from their social and historical context. Furthermore, in this perspective individuals' engagement with activities has to take account of the context of the activity, i.e. the larger social, historical, political and economic influences that shape the activity. Wertsch (1991) describes a sociocultural approach in the following way.

The basic goal of a sociocultural approach to mind is to create an account of human mental processes that recognises the essential relationship between these processes and their cultural, historical and institutional settings. (p. 6)

Sociocultural theorists consider concepts to be socially determined and socially acquired. The acquisition of understanding from a sociocultural perspective is achieved by individuals coming to share in meaning through negotiation and discussion. Bruner considers the shared use of language as the key which unlocks others' minds to use. For Bruner culture is the

implicit, semi-connected knowledge of the world from which through negotiation people arrive at satisfactory ways of acting in given contexts. (Bruner, 1986, p. 65)

Both constructivist and sociocultural theories are currently exerting considerable influence in educational practice. For some researchers, both perspectives provide useful and necessary ways of looking at learning (Cobb, 1994). What is common to both perspectives is the view that the process of coming to know is constructive. This means that students are *actively* engaged in thinking and that the ideas and experiences they bring to situations matter. Whereas educational constructivists typically argue that students' interpretations and meanings are private and individual and, to an extent, unknowable, socioculturists consider that meaning derived in interactions is not exclusively a product of the person acting. Rather we should think of the individual acting in a setting engaged in *relational* activities with others (Lave, 1988). Therefore when students engage in school activities they do so with some shared understanding. However, how this shared understanding is used will reflect the students' understanding of, and involvement, in the activity (Rogoff, 1990). In stressing the significance of the activity and the context in which individual learning takes place, socioculturists argue that there is an intimate connection between knowing and doing. Human knowledge from this perspective is therefore situated in that the activity

is not separable or ancillary to learning and cognition. Nor is it neutral. Rather it is an integral part of what is learned. Situations might be said to co-produce knowledge through activity. (Brown *et al.*, 1989, p. 32)

Much current assessment practice continues to assume that meaning is inherent in the words used to communicate assessment tasks. There is the further belief that problems can be given ready-made to students. Taking some of the key ideas about learning discussed, it is possible to identify some fundamental problems for assessment as it is typically practised. For instance, if students are 'active meaning-makers who continually give contextually based meaning to...others' words and actions' (Cobb, 1988, p. 88) then more attention has to be paid to students' interpretations of assessment tasks. Furthermore it is the students who have to be considered as 'the best judge of what they find problematic' (Cobb *et al.*, 1991, p. 157).

These two challenges, if accepted to any degree, have serious consequences for assessment. In the first case, account has to be taken of the different contextually based meanings that students might use to make sense of an assessment situation. Students' inadequate performance cannot be explained without an understanding of

these. Nor can it be assumed that inadequate performance reflects a lack of achievement. In the second case, purpose cannot be assumed to be shared between assessor and student or student and student. Yet the purpose of a task defines what knowledge students consider appropriate to draw on and, ultimately, what task they tackle. Variation of purpose between students and groups of students will limit the interpretability of assessment outcomes (Murphy, 1993).

The problem of task definition has consequences for the construct validity of assessment. One aspect of construct validity is concerned with the evidential basis of test interpretation, i.e. the justice of the interpretations made of test or item scores, the construct being that which it is assumed the test or item measures. Messick (1989), as part of his description of which evidence to explore when researching the validity of assessments, suggests

directly probe the ways in which individuals cope with the items or tasks, in an effort to illuminate the processes underlying item response and task performance...investigate uniformities and differences in these test processes and structures over time across groups and settings. (Messick, 1989, p. 6)

The probing of the processes underlying task performance is precisely what is demanded in a model of assessment that assumes that it is the student who gives meaning to assessment tasks and that students' meanings are determined by their sociocultural experiences. Lerman's (1993) description of the consequences of such a view for teaching can be usefully applied to assessment

since any pupil will be situated in many contexts, depending on his or her socio-cultural experience, the teacher [or the assessor] may not always be able to predict what will be called up by the activities [tasks] s/he offers. Consequently teachers [and assessors] need to find ways of enabling pupils to find, create and negotiate their meanings. (Lerman, 1993, p. 8)

When such explorations have been conducted, they have revealed the highly task-specific nature of students' performance. There are various ways of explaining this specificity. One explanation derives from the view that thinking and learning are situated. Such a view of students' knowledge challenges the validity of generalising about achievement, or indeed lack of achievement from a small number of assessment instances. Good assessment practice has to recognise the tentative nature of judgements made about students' achievements.

The APU surveys found evidence that relates to this. The APU conducted a series of annual surveys in England, Wales and Northern Ireland to monitor the performance of populations of 11, 13 and 15-year-old pupils in maths, language, science and design and technology from 1978 to 1989. Particular reference is made in the article to the APU science results (Department of Education and Science [DES], 1988a, b, 1989). The APU science project was based on the domain-sampling approach. In this approach a large pool of questions is generated for a particular 'ability' or 'skill' against a criterion referenced definition of that skill. The pool of questions generated is assumed to act like a universe of all such possible questions that might have been developed, hence any random sample of questions from the pool is representative of the pool and of the 'ability' or 'skill' being assessed. The questions allocated to a particular domain have to be agreed to 'fit' the definition and the domains defined have to be judged to have educational value and relevance. The problem with this approach is defining an agreed and implementable domain (see Johnson, 1989, for discussion of the survey methodology and question validation procedures).

f

F

1

The science assessment framework analysed science performance into five broad categories of achievement. Between 12,000 and 16,000 students drawn from between 500 and 1000 schools were tested at each age in every year. Over 400 science questions were administered each time. The tasks used involved an extensive range of methods of presentation, operation and response. The data collected ranged from generalised population scores on test categories to detailed diagnoses of students' errors on individual tasks. In addition, questionnaires were used to collect data on issues thought to influence science performance, such as out-of-school activities and interests.

In APU tests, where the science assessment tasks used corresponded to the situations in which students had acquired the knowledge being assessed, their performance was at the expected level, i.e. 15-year-olds outperformed 11 and 13-year-olds. However, if the tasks required students to apply that knowledge to new situations outside school, then 15-year-olds reverted to their everyday understanding and performed at a similar level to 11-year-olds (Donnelly, 1988). Those who accept that students' performance is task-specific recommend that more tasks are used and the time-scale for testing extended in order to enhance the validity and generalisability of measures (see Linn, 1993). The APU evidence clearly showed that for domain-referenced performance assessment, the more tasks sampled from the domain the less error associated with the estimated population score. Of course, the number of tasks required depended on the variability within the domain (see Johnson, 1989, for a discussion). Unfortunately, such a recommendation involves costs, particularly if performance assessments are administered by teachers in classroom situations, as in England. These costs are not just monetary but relate to the effects on teaching morale when assessment is seen to dominate teaching and learning. These effects have been found to defeat, in many cases, the major aim of assessment reform, which is to establish assessment as a tool for teachers and learners.

Another explanation for the lack of generalisability of performance outcomes suggests that the tasks used are inappropriate as they fail to share the *same characteristics*. The notion that comparable tasks can be developed remains an underlying belief of many concerned with assessment. One assumption implicit in this belief is that test developers can be aware in advance of what constitutes a demand, or an ancillary ability, in a task. This, certainly in the relatively novel area of performance assessment and given current understandings about the situated nature of personal knowledge, is highly unlikely. There is also the implicit assumption that unidimensionality is achievable and desirable in tasks, i.e. that only one construct or factor is measured. The APU science assessment framework categories

such as *observation* were understood not to be unique and mutually exclusive. It was accepted that certain skills defined in one category were preconditions for others. Thus questions were known to be multidimensional. Dimensionality is used here to describe the factors or constructs assumed to be being measured by a particular assessment instrument for the population or sub-group for whom the scores are to be interpreted. Dimensionality is therefore an *outcome* of the interaction of the individual student, sub-group or population with the assessment task.

A characteristic of the APU survey results was the variability in individual performance across a test category or a construct defined within it. This variability in performance was found to be due to sub-effects related to the tasks or their administration and students' interaction with them. The research demonstrated that what is demanded in tasks, albeit labelled, for example, as science and judged by external validation to 'fit' within one criterion element of that subject, actually required the understanding of numerous demands which spanned other aspects of science achievement and other curriculum subjects (DES, 1989). An example of this from APU findings (DES, 1988b) showed that on a relatively tightly defined construct, i.e. to read information from tables or graphs, the performance of a representative sample of 15-year-olds was high, about 90% across a range of tasks. However, if any of the following demands were included in the general requirement to read information, e.g. scale deduction, interpolation or manipulating decimal fractions or non-integer values on scales, performance was depressed by as much as 50%. Nor were these demands found to exert a consistent influence across all students.

The multidimensionality of tasks is not an assessment artefact but rather an expression of the nature of learning activities. Hence the requirement in performance assessment to capture authentic learning activities cannot be met by assessment tasks whose parameters are tightly controlled. Indeed, advocates of performance assessment typically recommend the reverse, i.e. the use of a range of tasks covering different modes of presentation, response and operation to assess a particular aspect of student achievement. Such tasks will, however, assess different achievements. For example, as part of the APU surveys, 13-year-old students' strategies for carrying out practical investigations in science were assessed using two methods, direct observation and judgement of written records. A comparison of the results showed that students' level of performance was depressed if the written records rather than observations were used as the source of evidence (DES, 1989). The requirement to report their strategy and findings imposed further demands on students to 'know' what was relevant in their actions and to communicate this. In the assessment by direct observation the assessor made the judgements of relevance on students' behalfs and the onus on communication was absent. Both methods of assessment provided useful information about the nature of students' investigative skills but what they represented were *different* insights. If it is only possible to use a limited range of tasks and methods assessors need to be certain what insights the tasks provide and, importantly, which insights they do not.

A further potential source of variability in performance outcomes is students' different experiences. This variability has direct consequences for equity in perform-

ance assessment. Student achievement in an educational model of assessment is not viewed as pre-determined, but as changeable. An expected source of variability is therefore students' access to, and experience of, teaching in school. What is often overlooked as problematic for the interpretation and generalisability of performance assessment is the potential for differential out-of-school experiences. When students try to make sense of assessment situations they do so using the totality of their understanding of the world, of school and of assessment and testing. The sources of knowledge and experience that students draw on are what Lerman (1993) refers to when he talks of students being 'situated in many contexts'. The consequences for assessors of this are two-fold. First they need to consider the *potential* social and cultural bases of students' experiences when setting tasks which, as Lerman advocates, enable them to 'find, create and negotiate their meanings'. Second, they need to establish the sociocultural patterns in the meanings students derive in order to interpret their responses to assessment tasks.

3

١

2

I

ł

£

1

l

The APU science results were based on a model that assumed tasks were multidimensional and that students' experiences within and without of school would vary. Consequently, background information concerning students' interests, attitudes and experience outside as well as inside school was collected. Examination of the survey results at each age revealed consistent differences in the performance of girls and boys. These differences revealed that sub-effects in performance results arise in part from the affective responses of students. The impossibility of separating the cognitive from the affective in students' responses has to be considered if interpretations of assessment outcomes are to be valid and therefore just.

The complexity of performance assessment outcomes cannot be underestimated. The complexity arises in part from the nature of the tasks and in part from the students themselves. Both aspects of this complexity need to be understood by assessors. In the next section a brief description of sex-role development research is provided to indicate some possible sources of gender differences in performance. This is followed by a review of sub-group differences in performance established by national and international surveys of students' performance in schools. The intention is to suggest the hypotheses that appear to have most relevance to assessment. These hypotheses illuminate the different ways that boys and girls may be 'situated' and thus provide a basis for probing the contribution of students' affective responses to task performance.

Sources of Differential Performance: some hypotheses and trends

Numerous social and psychological factors have been cited in hypotheses about the sources of gender differences in performance. Psychosocial explanations have proved powerful because of the evidence from many studies that similarities in males' and females' performance far outweigh any differences observed for the majority of the population (Halpern, 1992). Furthermore, many of the major reviews of studies into cognitive sex differences have revealed empirical trends in the size and extent of

differences which provides support for psychosocial explanations (see, for example, Hyde *et al.*, 1988, 1990). The literature in the area is vast but it is possible to provide a picture of the significant factors in the socialisation of males and females that emerge from the research. The brief description offered comes from Wilder & Powell's (1989) review, but for more detailed discussion see Halpern (1992) and Gipps & Murphy (1994).

The picture of socialisation processes offered by Wilder & Powell (1989) highlights the different ways parents respond to boys and girls and encourage them to interact with the world and with people. Parents' expectations differ for boys and girls. These different expectations are reflected in the activities and toys they provide for them and in their reactions to them. Boys and girls continue to engage in different hobbies and pastimes from an early age and demonstrate diverging interests that increase with age. A consequence of these differences is that children develop different ways of responding to the world and making sense of it, ways which influence how they learn and what they learn. These different treatments also influence children's views of what constitutes appropriate behaviours for them and what others' expectations of them are.

Boys and girls also experience schooling differently. For example, the interactions between teachers and boys and teachers and girls have been found to vary in frequency, duration and content. Teachers' judgements of girls' and boys' achievements and needs have also been found to vary in stereotypical ways, as do their expectations of them. Consequently, boys and girls develop different perceptions of their abilities and relationships with academic disciplines. Children's judgements of their own competences and potential, and those of other children, correspond closely with those of their teachers. From this picture of socialisation processes it is possible to identify factors that might influence assessment performance.

If students' self images are socially and culturally determined, their reactions to assessment tasks may vary depending on the correspondence between their self image and the stereotyping of the task. Thus if a task reflects a typical male activity, girls may feel disinclined to attempt it irrespective of whether they possess the actual achievement being assessed. A more general effect related to students' differing self-images concerns their attitudes to different academic disciplines. Both teachers and students' share beliefs about which domains are appropriate for girls and which are appropriate for boys. This can result in students valuing particular subjects differently depending on whether they are a girl or a boy. Negative attitudes to subjects can have a two-fold effect. First, they can lead to demotivation which can limit students' engagement with learning opportunities eventually leading to underachievement. Second, they can influence whether students feel able to engage with particular assessment tasks which, in turn, can affect the quality of their engagement. A further predicted outcome of students' differential images and attitudes is that they develop different expectations of success with regard to particular achievements. These different expectations can have a marked influence on students' confidence in their achievements. A belief in one's ability or lack of ability, whether well founded or not, will influence the quality of students' interaction with assessment tasks. The hypothesised relationship between levels of confidence and measured

performance continues to be the subject of much research into assessment performance (see Halpern, 1992).

The different *out of school experiences* of males and females, it is argued, provide them with different opportunities to learn. Hence, girls and boys come to school with different achievements. These differences reflect differences in opportunity to learn rather than ability, but often this goes unrecognised by teachers in their judgements of children. As a consequence of their different social and cultural experiences, children develop particular views of what is relevant. Hence, presented with the same situation they focus on different aspects of it. This is particularly significant in assessments, where typically it is assumed that the perception of what is relevant in a task is both shared by students and by assessors. Further, it is argued that because teachers, like parents and students, develop sociocultural expectations of groups of individuals, the differences students bring to school are compounded in school. It is hypothesised that *different approaches to the teaching of sub-groups* encourage differential skill development. There is evidence that classroom dynamics influence achievement but the effects are not simple or well understood.

A further hypothesis to explain differences in performance is that the actual *structure of domains is biased* against certain groups, the bias being evidenced in the learning styles preferred, the modes and styles of expression valued, and the achievements that are assessed (Kelly, 1978). This bias, it is argued, becomes part of a domain because of the over-representation of members of a particular sub-group within it. Hence maths and science are denoted as 'masculine'. It is predicted that while other groups' responses to assessment tasks may have merit in terms of specific domain achievement, this merit is not recognised by assessors because of their domain-specific expectations.

The effects of psychosocial variables on students' engagement in learning and assessment tasks can lead to real differences in achievement between sub-groups. However, it is essential to distinguish between lack of achievement as an outcome of teaching and lack of achievement arising from a lack of *opportunity to learn*. It is also evident that the effects of psychosocial factors in assessment situations can mask students' actual achievements. This can arise because features of tasks function as barriers to certain students, preventing them from accessing the task or leading them to derive alternative tasks and alternative solutions not recognised or valued by assessors.

Trends in Performance: sub group effects

The National Assessment of Educational Progress (NAEP) in the USA has assessed national samples of 9-year-olds, 13-year-olds and 17-year-olds since 1969, and from 1983 began sampling students by grade as well as age. A wide range of subject areas is tested, and a range of background variables related to schools collected. NAEP, unlike the APU Science project, used item response modelling (IRM) to estimate the average proficiency for the nation and the various sub-groups of interest. IRM makes a number of assumptions about assessment tasks. First, that a task measures a single construct or a constant combination of two or more for the whole popu-