

## David Spendlove

Senior Lecturer,  
Liverpool John  
Moore's University

### Abstract

This paper identifies the current weaknesses and misuse of national assessments within design and technology in making sophisticated judgements about the success and failure of particular groups. Firstly, there are weaknesses in the accuracy of teachers' assessments. Secondly, there are weaknesses in ensuring consistency between schools. Thirdly, there are weaknesses in the interpretation and delivery of content prior to assessment and finally, there are weaknesses in matching pupil capability with assessment. This paper concludes that damaging judgements against particular groups of pupils based upon existing assessment practices cannot be justified.

The context of the debate about the increasing gap in performance between boys and girls (often referred to as the 'underachievement of boys') in design and technology is well documented (Spendlove, 2000, 2001). Within this context, teachers are often forced to search for quick solutions for raising attainment, as school examination results and performance tables continue to proliferate the marketization of education.

Many areas relating to the gender performance gap are, however, outside the teachers' controllable domain such as pupil biological and psychological differences in addition to changing social trends and the redefining of the male and female in society. Further to this is a collection of educational reforms that can be suggested to have individually and collectively conspired to create the differentials in performance that are now apparent between boys and girls.

The most significant factor, however, has been the development of national assessments over the last 15 years (National Curriculum and GCSE) and, more specifically, the nature and frequency of assessments.

Pupils in England are statutorily assessed at the ages of five, seven, 11, 14 and 16 (with the intention of additional testing at 12 and 13); more than any other country in the world. In design and technology pupils are formally assessed at seven, 11, 14, and 16. Ultimately, it is these statutory assessments that are defining the gender gap and the notional underachievement of boys. Quite simply, if the national assessments did not take place, the labelling of boys as failures from an early age would not occur. Given, however, that testing and assessments are to be taken regularly and their meaning is to be so great (as part of an increasing culture of accountability), then it is crucial that the quality and nature of

assessments are thorough, robust, accurate, unbiased and based upon design and technology capability.

Unfortunately, current assessment practices are far from robust and have become part of a divisive mechanism for the 'labelling of schools, teachers and pupils' through the publication of national and local 'league' tables. The tables present 'raw data' and are open to misinterpretation by parents and pupils (a small step in the right direction is the introduction of value added tables from 2001). Millions of pounds are spent each year on assessing pupils yet very little information is given to schools on what to do with this information. Therefore, an unfortunate outcome of the regular assessments has been an increasing frequency in labelling children as 'underachievers' from a younger age rather than providing an opportunity for early intervention and the development of compensatory education programmes.

... no educational assessment procedure is free from limitations, and it is now widely accepted that national assessment schemes have limited reliability (Wilmot, Wood and Murphy, 1996). Even within an area of the curriculum in a single year, any results produced will only give an approximation of the achievements of the pupils who are assessed. Assess them again with a similar set of assessments and the results will change. Assess them again on other aspects of that part of the curriculum and the results can be wildly different. Thus, assessment results in education are only ever at best approximations, giving a broad view of the achievements of students in relation to the things covered in their particular assessment. Such assessments are not precise and they are not at all robust, when it comes to using them to make sophisticated comparisons, or ascertain the answers to complex questions about their educational progress. (Murphy, 1996: 3)

To place an unjustifiably high value on national assessment results can lead to the demotivation of students and the setting of inappropriate targets. It can equally lead schools into teaching towards national tests and manipulating figures.

First, I am not sure that the tests are the right tests. The concept of levels is a very vague one; we need standardised tests in literacy and numeracy. Second, the tests have changed quite significantly over recent years so it is impossible to compare like with like. Third, a lot of individual tests are being administered in a creative way by some schools. (Woodhead, 1998: 1)

Within design and technology the assessment arrangements are clearly open to 'creative manipulation'. The end of key stage results that are based upon teacher assessments (TA) are not open to scrutiny, yet the outcomes are published nationally and used to make sophisticated judgements about the quality of provision.

With the exception of the Exemplification of Standards (SCAA, 1995) publication (a publication which aimed to help standardise levels; which is now out of date), there are no mechanisms for ensuring consistency of moderation between schools or educational authorities. This situation is unchallenged even though the Task Group on Assessment and Testing (TGAT) suggested that 'a complex process of group moderation through which teachers' assessments could be brought into line around a common standard was necessary'. (DES, 1988)

The statutory advice for determining a pupil's level of achievement is further limited as it uses the notion of 'best fit'. This is 'based on knowledge of how the pupil performs across a range of contexts, takes into account strengths and weaknesses of the pupil's performance and is checked against adjacent level descriptions to ensure that the level awarded is the closest match to the child's performance in each attainment target'. (QCA/DfEE, 1998)

The accuracy and approach used in teacher assessments (based upon 'best fit') is further open to scrutiny. Gipps (1998) found that teachers preferred the best-fit model of assessment compared to the previous statement of attainments for the first National Curriculum, as it provided a more manageable assessment system. However, the accuracy of the approach is variable.

The notion of 'best fit' is a consciously loose one. Because of this, teachers are taking a variety of approaches to making Teacher Assessment judgements. Some teachers will make quantitative judgements (to attain a level individuals must meet all the elements of a level description, 50%, or some other proportion); some will take a hurdle approach (individuals must be able to do x, y and z in order to reach Level 5); others will take an intuitive approach (this one feels like a good Level 4). Although not addressed in this study we know that some teachers will make ranking judgements (this individual is a clear Level 7, and this is a clear Level 6; less clear performances are then slotted in, in relation to these fixed points). Because of the lack of clarity of 'best fit', the differences in interpretation mean that, at times, there will have been a difference of

one level awarded to pupils and this is not acceptable in a 'high stakes' programme. (Gipps, 1998: 5)

The system is further devalued, as, in addition to there being little consistency in teacher assessment methods, there is no mechanism to ensure consistency with content. Although the Programmes of Study are prescribed, they are sufficiently flexible to encompass a vast range of activities. Teacher assessments can be focused upon any number of tasks, which ultimately have to be transposed through to a single level (the previous National Curriculum had a 3:2 ratio). The move from statements of attainment to level descriptions has been made because of the overload provided by the huge number of statements of attainment within the core subjects of the National Curriculum. However, the level descriptions are currently too global in their descriptions to be useful as assessment criteria, and if teachers are to use them for assessment purposes in anything more than a 'rough and intuitive' way, then they may need to break them down. Clear exemplification of levels is necessary in order to help classroom teachers make accurate assessments against level descriptions.

The most demoralising aspect of this series of inconsistencies is that pupils, predominantly boys, are being labelled as failures at an early age by an ultimately limited and weak assessment system.

The purposes to which national test results are now being put are profoundly de-professionalising, subverting good practice in assessment and consequently, in the process of teaching and learning. The ultimate losers are the pupils.

We must continue to question the educational value of these high cost, high stakes tests. Seven years on since the introduction of statutory assessment at Key Stage 3, where is the evidence that national tests are genuinely raising standards? The Qualifications and Curriculum Authority must as a matter of urgency review the way tests are both set and marked. (Smith, 1999: 1)

In situations where a combination of test and teacher assessments are made (maths, science and English) the test result is preferred, with teacher assessments being used as a backup. Hence, because of the combination of the boycott of tests in 1993 and the changing Orders for design and technology, the current assessment system is firmly based upon teacher assessment. Optional standard attainment tests were available for the 1997 assessments but have since been discontinued.

One of the overriding factors for not relying upon teacher assessments, which currently

underpins government reforms and provides comparative data, is the degree of objectivity within assessments. Assessments should be systematic, objective and based solely upon the subject criteria. However '...within this discourse it has been shown that subjective and erroneous evaluations of pupils' (sic) abilities that teachers make, often informally, can go on to produce a reality that reflects those original evaluations'. This resulted in 'the academically differentiated outcomes for children (which) lay in the working out of self fulfilling prophecies in the classroom.' (Hutchinson, 1994: 50)

Because of the nature of project work being used as the basis of assessments within design and technology, it is appreciated that teacher assessment is currently the best way of assessing pupils' work. The weakness of the system within design and technology is that there is too little emphasis on consistency, moderation and consensus. The mechanism could immediately be made more effective by insisting upon department to department, school to school and educational authority to educational authority, consensus moderation.

I believe that teacher assessment is bound to be richer, more varied and more comprehensive – in short, more valid – than any kind of externally set task or test ... above all I would argue for the power of consensus moderation across schools in enhancing teachers understanding of the curriculum, in widening their horizons of teaching approaches, teaching materials and assessment approaches, and in bringing them to a common shared understanding of levels or grades. Consensus moderation is a very powerful form of professional development. (Nuttall, 1995: 239)

Ensuring consistency, however, is only one feature of the assessment process. Ensuring the removal of gender bias in assessments and the use of gender analysis is another matter.

The stereotypical labelling and the use of assessment as a means of disciplining pupils, in particular boys, provides a further contributory factor in boys' underachievement. 'Some teachers view a low grade as a short, sharp spur to better effort, but it is not always perceived that way by pupils whose self-confidence is more brittle than we think... there is a danger that some boys will regard them as a sign of 'street cred' with their peers'. (Bleach, 1998: 45)

School practices, including assessment and reporting, curriculum design, language, and teaching and learning processes, are not gender neutral. They shape, challenge, or

endorse particular beliefs about gender. Assessment and reporting processes intersect with gender in a number of ways including students' beliefs, experiences and understandings of appropriate gender relations, the gendered 'nature' of knowledge – particular ways of learning and bodies of knowledge are credited as being as important as teachers' beliefs, values and practices concerning gender and assessment.

There is nothing wrong with assessment being human and subjective. It only becomes problematic when particular kinds of knowledge and particular kinds of assessment devices are seen as being more important, 'objective', valid, or reliable than others.

Consideration of gender perspectives in assessment and reporting cannot be a 'bolt-on' exercise, but should be integral to existing assessment and reporting processes so that teachers can acknowledge their students' different (gendered) starting points whilst also recognising their students' preferences for different learning styles and processes.

When assessment results are analysed by gender, statements are made concerning how well, or poorly, girls are doing in particular learning areas compared to how well, or poorly, boys are doing. This method can only have validity if neutral assessments along with consistent learning and socialising experiences are apparent. This is clearly unlikely. However, the contexts for assessment activities can be analysed to ensure a consistent range of balanced activities rather than bias to either group.

The APU surveys have demonstrated that consistent differences between the performance of boys and girls are due to sub-effects related to the items and/or their administration. The overall difference depends on the combination of effects within the item and the test instrument overall. Thus, it would be possible to engineer group differences or remove them by changing the assessment item...

Altering the characteristics of items can effect (sic) group differences by advantaging one group but in so doing another group would be disadvantaged. This follows because up to a point what is being measured are learnt gendered behaviours and or expectations rather than achievements. The findings, therefore, support the hypothesis concerning the influence of psychosocial variables of performance both in terms of how boys and girls come to view themselves and become viewed by others and how subjects become constructed and achievements within them defined. (Gipps, 1994: 147)

Within design and technology activities at Key Stages 2 and 3, the contexts for assessment opportunities through design and make assignments (DMA) and focused practical tasks (FPT) are derived by subject teachers. The selection of these activities and anticipated learning and gendered behaviours is crucial in avoiding bias and stereotyped assessments. Studies have consistently shown certain activities to favour one gender group or another.

The application of mathematics and the assessment of results in technological projects are two instances of reflective tasks that play to feminine strengths, thereby enabling them to compete with boys in subjects in which the latter had formerly held an advantage. (OFSTED/EOC, 1996: 17)

At GCSE, the nature of pupil projects are developed through negotiation with a teacher or through adoption of examination board set projects. The balance of projects can clearly be open to a gender bias to either boys or girls. Kimbell suggests that 'one is led to the somewhat sinister conclusion that it would be possible – given an understanding of these effects – to design activities deliberately to favour any particular nominated group'. (1996: 96) The APU study found that with reflective activities within design and technology, boys were more able to get to grips with these aspects of capability when they are practically engaged in developing solutions. Whilst with aesthetics and people contexts girls showed more understanding than boys in almost all tests. With energy systems, 'the complete dominance of boys in this conceptual area is the most straightforward of all our findings... boys demonstrate more understanding than girls in every instance'. (APU, 1991: 221)

About communication the APU concluded 'the difference between girls' communication and boys' communication is strongly contrasted and the girls outperform boys in all areas'. (1991: 219)

The difficulty for teachers is in developing projects which are gender neutral (environmental orientated projects were consistently found to be gender neutral), or to recognise bias and act upon it. This may mean ensuring balance over a range of projects and ensuring that awareness exists of likely weaknesses for either sex, with compensatory programmes being developed to overcome these weaknesses. Adding to the complexity of the issue is that findings are not consistent with all ability ranges with low ability girls' performance being found to be 'particularly fragile'. The APU remains the only thorough

analysis of design and technology by assessment and gender and clearly illustrates a way forward through further research in this area to offset the current gender imbalance.

There are also enormous incompatibility issues between design and technology capability and assessment principles. Atkinson (2001) has considered how high order thinking such as creativity, problem solving and analytical thinking (some of the key features of capability) impacts upon pupils GCSE design and technology performance. Ultimately it was found that evidence of these capabilities was not required. In fact being highly creative could be a hindrance.

Their frustration throughout their project (pupils) was strongly in evidence. Some became the class 'fool'; some spent their time using their creative ability helping others solve their design problems; some found excuses for missing lessons; some became 'quiet wanderers' around the classroom avoiding their teacher, who in turn seemed to try avoid noticing them. (Atkinson, 2000: 276)

The lack of reward and incentive to apply capability and the use of long inflexible coursework (Creswell, 1990) which fails to reflect capability are conspiring against pupils. Examination boards and teachers are misleading pupils about their ability in design and technology.

First GCSE examination boards need to encourage overtly the use of flexible, appropriate design and problem solving strategies through modified schemes of assessment. At present teacher assessors tend to reward 'thin' evidence that is well presented rather than rewarding the use of higher order skills, in particular creative thinking and appropriate design processes ... Teachers are unlikely to change their current 'successful' teaching strategies at the public examination stage of a pupil's education unless pupils are explicitly rewarded with higher marks for such activities. (Ibid)

From this short paper it can be seen that national assessment systems are not sufficiently standardised or robust for them to be used in a sophisticated or comparative way. To do so is dangerous with the casualties (predominantly and currently certain groups of boys) being unfairly labelled as failures from an early age.

**References**

- APU (1991) *The Assessment of Performance in Design and Technology*, London: HMSO
- Atkinson, S. (2000) 'Does the Need for High Levels of Performance Curtail the Development of Creativity in Design and Technology Project Work?', *International Journal of Design and Technology Education*, September 2000
- Bleach, K. (1998) *Raising Boys' Achievement in Schools*, London: Trentham Books
- Creswell (1998) in Arnot, M. and Weiner, G. *Recent Research on Gender and Educational Performance*, London: Open University
- Department of Education and Science (1989) *Task Group on Assessment and Testing – A Report*, London: DES
- Equal Opportunities Commission (1996) *The Gender Divide*, London: HMSO
- Gipps, C. and Murphy, P. (1994) *A Fair Test: Assessment, Achievement and Equity*, Buckinghamshire: Open University Press
- Gipps, C., Clarke, S. and McCallum, B. (1998) *The Role of Teachers in National Assessment in England*, Leeds: AERA Conference symposium
- Hutchinson, D. and Schagen, I. (1994) *How Reliable is National Curriculum Assessment?*, Berkshire: NFER
- Murphy, R. (1996) *Drawing Outrageous Conclusions from National Assessment Results*, Lancaster: BERA Conference, 12-15 September
- Murphy, R. and Broadfoot, P. (1995) *Effective Assessment and the Improvement of Education – A Tribute to Desmond Nutall*, London: Falmer Press
- QCA/DFEE (1998) *Teacher Assessment*, London: HMSO, 1, 12
- SCAA (1996) *Exemplification of Standards Design and Technology: Key Stage 3*, London: HMSO
- Smith, P. (1999) 'Union Attacks 'Pointless' Tests', *Times Educational Supplement*, 1 January
- Spendlove, D. (2001) in Barnes, B., Morley, J. and Sayers, S. *Issues in the Teaching of Design and Technology*, London: Routledge
- Spendlove, D. (2000) in Kimbell, R. (Ed) *Design and Technology International Millennium Conference*, Wellesbourne: DATA
- Spendlove, D. (2000) 'A Better Look at Boys', *Times Educational Supplement Curriculum Special*, Autumn
- Woodhead, C. (1998) 'Tests are Unreliable Says Chief Inspector', *Times Educational Supplement* 23, 12