

‘Dip, dip, sky blue, who’s it? NOT YOU’: children’s experiences of standardised testing: a socio-cultural analysis

Gerry Mac Ruairc*

University College Dublin

The recent decision by the Department of Education and Science in the Republic of Ireland to introduce the mandatory testing of children in Irish primary schools provides the broad context for this paper. This decision has particular implications for schools designated as disadvantaged. The main focus of this study is on identifying the strategies used by these children in completing standardised tests. The sample comprises 51, mixed gender, 6th class children from contrasting ideal type socio-economic urban communities. The findings reveal a considerable level of difficulty on the part of some children in designated disadvantaged schools, with most aspects of the test format and much of the testing process. A number of socio-linguistic and socio-cultural factors are identified that affect children’s attainment levels on these tests. Failure to take account of these issues will continue to present an overly crude picture of national attainment patterns and serve only to contribute to the stigmatisation and ghettoisation of children living in marginalised and disadvantaged communities. The paper argues for a broader discussion of the use standardised testing that will take account of the well-established patterns of socio-economic inequality.

Keywords: standardised testing; socio-economic inequality; class bias; test item response strategies; attainment; child perspective

Introduction

Tests should be labeled just like dangerous drugs: ‘Use with care!’ (Spolsky 1981, 20)

The enhanced profile of assessment in Irish primary school in recent years is a very welcome development in Irish educational discourse and practice. The work of the Irish National Council for Curriculum and Assessment (NCCA), among others, has been central in driving innovation in relation to this dimension of pedagogical practice in schools. The manner in which assessment is embedded in the structure of the 1999 Primary School Curriculum in Ireland, for example, has the potential to greatly increase the role assessment can play in enriching teaching and learning in schools. Further documentation from the Department of Education and Science, Republic of Ireland (DES) (2006) and the NCCA (2005, 2007) broadens the scope of assessment in recommending, very explicitly, the need to ensure a balance of practice in schools between assessment of learning and assessment for learning. The potential to provide quality information to students, parents and the system generally is considerable. Most of these recommended assessment strategies can support and assess the learning of all students irrespective of gender, race, ethnicity or class.

*Email: gerry.macruairc@ucd.ie

However, the use of standardised testing is a significant exception to this. The manner in which standardised tests are increasingly being used internationally for a variety of accountability and comparative purposes demands that this particular dimension of assessment is rigorously critiqued and debated at policy formation and implementation levels in order to limit the negative consequences of this form of assessment for particular groups of children in our schools.

Standardised testing in Irish primary schools – an overview of practice

In the Republic of Ireland, the practice of using standardised forms of assessment in primary schools both at school level and at system level has evolved over the past two decades. Two main testing instruments in English and mathematics, normed on the Irish primary school population, i.e. the Drumcondra Reading/Mathematics Tests and The Micra/Sigma-T tests are the main players in standardised assessment portfolio used in Irish primary schools. The profile of these tests has been enhanced over recent years by different policy developments with a reported usage of in the region of 95% (DES 2006). The primary school curriculum used in Irish schools recommended standardised testing as one of a variety of assessment methods (Government of Ireland 1999; NCCA 2005). The recent link between standardised assessment and the allocation of additional supports for special education provision in schools has added to the currency of testing throughout the system (DES 2000). Section 22 of the Government of Ireland Education Act (1998), while not mentioning standardised testing explicitly places a statutory obligation on schools to ‘regularly evaluate and periodically report the results of evaluation to the students and their parents’ (Section 22 (2b) Education Act 1998). The resolution of the debate surrounding mandatory testing in Irish primary schools has placed the standardised test instrument at the centre of the response to this statutory requirement by mandating schools not only to ‘administer standardised tests in English and mathematics to their pupils twice during their primary school years – at the end of first class or beginning of second class and at the end of fourth class or beginning of fifth class’ (NCCA 2007, 51) but to communicate these results to parents. In many ways, this outcome was inevitable given the international situation with respect to national standardised testing which has firmly embedded this practice within educational discourse and debate at all levels.

Testing and accountability

It is too early to assess the impact the use of mandatory testing will have on the development of assessment in Irish schools. The use that is made of test results at a school level and at a system level within the Department of Education and Science has yet to be fully determined. The now mandatory position allocated to testing aligns Irish policy more closely with international patterns and if international patterns are replicated here there are some negative consequences which accompany such policy developments. The arguably inevitable link between testing, once mandatory, and externally imposed accountability requirements is the source of considerable tension at school level in many countries. Where this practice is in place, policy-makers increasingly rely on testing to leverage accountability and instructional change (Shohamy, Donitsa-Schmidt, and Ferman 1996; Dixon 1999; Gillborn

and Youdell; 2000; Hill 2000; Shohamy 2001a,b; Gallagher 2003; Brown 2004; Anagnostopoulos 2005; Luxia 2005). The outcome of this often leads to 'far reaching and high stakes decisions and consequences about individuals and groups' (Shohamy 2001a, 373). The negative repercussions and 'washback effect' (Luxia 2005, 164) of a high stakes testing environment on the quality of education systems and the overall nature of the educational experiences for children can be sidelined in the pursuit of policy agenda based on accountability and escalation of standardised measurement of learning outcomes. These negative consequences include a 'teach to the test' culture in schools (Anagnostopoulos 2005; McNeil 2000; Lam and Bordignon 2001), the proliferation of inadequate pedagogy (McNeil 2000) and the avoidance of risk taking and innovative practice (Williams and Ryan 2000).

Pressure to provide what is perceived to be objective evidence of school performance emanates from a number of perspectives that focus on practice in schools. The concept of 'value added' which emerges from the school effectiveness perspective requires 'even more reliable and linked scales which can be accepted as objective' (Williams and Ryan 2000, 51). The international comparative dimension to measurement has also resulted in an enhanced role for standardised testing and an added legitimacy to the test itself and to the results of the testing process. The increasing focus on international evaluations such as Programme for International Student Assessment (PISA) which compares the effectiveness of national systems of education has resulted in producing a high stakes international league table that can be used to drive an agenda of reform. The comparative deliberations arising from positions on this list has led to 'policy anguish' in many countries, especially in countries that have 'under-performed' e.g. Germany (Stobart 2004, 244) and presents as evidence of the use that is made of this type of assessment. It is clear therefore that the pressure to measure the outcomes of schooling is increasingly becoming embedded in accepted practice. There is increasingly an inevitable sense that international and national educational policy and the practice that is guided and moulded by it will be influenced to a considerable degree by the discourse of measurement.

The contested terrain of testing

While originally considered to ensure equity and fairness, the role and function of standardised testing became a highly contested issue for debate in the US in the 1960's (Gallagher 2003). Specifically, critics of standardised testing voiced concern that women, limited English-proficiency students and racial minorities were penalised because of cultural differences (Gallagher 2003, 91). Since the early inception of standard-based reporting in the US, the three phases of testing identified in the literature (Haertel and Loricé 2004), have resulted in narrowing of the focus of inquiry in relation to the assessment of children's learning in school. The criterion-reference test (CRT) introduced by Glaser (Glaser 1963) and popular in the 1970s while not giving a prominent role to levels of performance, quickly became part of 'a form of measurement driven institutional planning in which the knowledge and skills to be learned were analyzed into narrow, carefully sequenced units' (Haertel and Loricé 2004, 83). This instructional approach that emanated from this testing era was based on behavioural psychology and was highly contested even at the height of its popularity (Haertel and Loricé 2004). The regime of testing which followed, i.e. minimum competency testing (MCT) (Brickwell 1978; Kohlfeld 1982;

Linn 2000) resulted in a high stakes testing culture for individual students. Most states in the US enacted legislation requiring students to demonstrate mastery of certain basic skills prior to graduation. The justification for this regime is more problematic than its predecessor. Three main difficulties are identified in relation to this genre of testing (Haertel and Lorie 2004). Because the stakes were high, procedural evidence of validity became more important than in the case of CRT. The outcomes were also loosely defined at policy level and as a result were open to local interpretation creating a potential for significant inequity. Finally, the MCT model of testing, with its typical multi-choice format, was actually less suitable for what it was trying to measure than the discredited CRT. A number of reviews of education policy and attainment in the US in the late 1980s early 1990s resulted in the replacement of the MCT with the current practice of standards-based testing. This present policy has resulted in creating a high stakes testing environment for students, schools and teachers with the result that the standards rhetoric has become deeply embedded in the discourse related to almost all aspects of schooling. These reviews and reforms in US education paralleled events elsewhere.

In the case of the UK, changes and developments can be traced back to the education reform act 1988 and the introduction of the national curriculum in the UK (Ball 2006). What has happened since 1988 has been a series of policy initiatives which have contributed to the process of shifting power away from teachers towards the centre (Gillborn and Youdell 2000; Ball 2008). The use of testing, the publication of league tables and more recently the concept of value added has been the central force in English education policy (Headington 2003) with the result that the British education system is now dominated by the 'rhetoric of 'standards' (Gillborn and Youdell 2000, 17) where all schools whatever the rhetoric must 'strive continually for more and more success; judged by traditional, biased and elitist criteria, where those who fail to measure up must look to themselves for the cause' (Gillborn and Youdell 2000, 18). The link between performance on standard assessment tests (SATs) at ages 7, 11 and 14 and the GCSE for 16 year olds and the heavy handed inspections carried out by the Office for Standards in Education (Ofsted) regime of inspections is particularly problematic. The discourse of 'failing schools', 'special measures', 'sink school' and a success criteria increasingly characterised by the phrase '5 A to C GCSEs' resonates with most teachers in England. It has been concluded that after almost 20 years of reforms what has emerged is an approach that is 'highly technicist, sociologically naïve, and in its consequences, deeply regressive' (Gillborn and Youdell 2000, 31). The most recent controversy, regarding the 2008 SAT tests results and specifically the identification of serious errors in the marking the 2008 English tests (Paton 2008; Shepherd 2008a,b), does little to contribute to a positive view with respect to the currency of these tests in English schools.

Bias in testing

Despite considerable scholarship in the area, going back to early work in relation to cultural bias in IQ tests, evidence continues to exist of what could be described as a 'denial of bias' among many policy-makers with respect to standardised tests. A consideration of whether a test is biased, while difficult to answer (Gipps and Murphy 1994, 18), points to the existence of a range of sources of bias. Broadly speaking test bias can occur in the design of test items, in the use that is made of the

test results and in the manner in which the test is administered. For the purposes of this paper, the first two sources of bias will be developed more fully.

Traditionally many examinations of bias in testing focused on test items and specifically on the extent to which gender and race impacted on test results. Test items were adjusted and manipulated in order to eliminate negative stereotypes, derogatory references to particular groups etc. (Goldstein 1986). Other attempts to identify bias in test items through statistical procedures using a form of item analysis referred to as differential item functioning (DIF) (Gipps and Murphy 1994; Tae-Il 2004; Abbot 2007) have identified differences between item bias and test bias. The case of item bias relates to questions which favour one group disproportionately whereas the latter refers to average test scores for various groups. Dealing with bias in the DIF perspective would result in loading a test with items biased against higher scoring groups. This particularly relates to attempts to control for linguistic bias, which is the subject of this paper. In this scenario deliberately using words unknown to one group, while it might be constructed to favour one socio-economic/race/gender group, will result in a biased test. Clearly, this does not provide an appropriate vehicle to ameliorate the situation.

The bias that occurs in the often uncontested manner in which test results are used is deeply problematic. At a macro level, the seemingly objective task of filling in bubbles (in the case of the Drumcondra test – little rectangles) with a no. 2 pencil takes on a different meaning in the context of national testing or a national assessment of reading attainment. At this level the results are passed through ‘a series of translations – from the individual to the aggregated publicised number’ (Graham and Neu 2004, 314). The outcome of this practice only serves to further stigmatise and pathologise the perceived failure and/or inadequacy of the majority of children in working-class communities. It is not surprising therefore that there is a considerable amount of international literature pointing to a more intensive stratification by class and race despite the fine sounding rhetorical language in which they are packaged (Gipps and Murphy 1994; Apple 2001).

It is notable that in the current climate this debate is largely absent. While there is some residual discussion in relation to bias in the discourse around testing it remains fair to say that the notion of a standardised test retains a dominant position in terms of its assumed objectivity therefore enjoying enormous trust and support on the part of the public and of institutions (Shohamy 2001b). There are a number of issues related to testing that need to be firmly articulated in all debate and discussion. At a broad level, the link between testing and the setting and maintenance of standards holds firm. The often uncontested nature of this discourse of standardisation needs to be examined rigorously. ‘What counts as standards, who should decide them, where they should come from, what their purposes should be in practice, how they are to be used. What counts as meeting them: These are the real issues’ (Apple 2000a, 725). Many believe that having standards and testing them will lead to higher achievement especially where it is currently problematic i.e. among disadvantaged students. Empirical evidence and analysis would not support this assertion (Delpit 2006; Lam and Bordignon 2001; McNeil 2000; Meier 2000; Shohamy 2001a). This negative impact is particularly notable in schools currently succeeding in connecting curricula and teaching to the realities of students’ cultures, backgrounds and economic conditions (McNeil 2000). At present, the focus of reform in many countries is driven by systems of National testing based on neo liberal notions of free markets, increased surveillance

and a neo conservative sense of the need to return to high standards, discipline and real knowledge. Testing, 'provides the mechanisms for comparative data that consumers need to make markets work as markets' (Apple 2002, 606). The policies that emerge from this rationale are often discussed as shared or common standards, yet what counts as common and who determines it is highly contested (Apple 2001).

At a micro level, what happens with test results in individual schools and the consequence of results on teacher and student classroom practice is largely overlooked (Anagnostopoulos 2005). Firstly, it is difficult to envisage how a positive outcome could be achieved from communicating test results to parents, where test results are persistently low. Secondly, it is necessary to examine the implications for the role test results play in instructional tasks in classroom and discourse at local school level. Tests possess inter-textual attributes (Bakhtin 1986) where the tests as texts are double voiced and where tests results are transformed into instructional texts and transform instructional discourse. In this way they facilitate the construction of particular meanings and subject position which constitute and are constituted by unequal power relations in school (Devine 2003). Where this inter-textuality has a positive impact on student learning i.e. where attainment levels are high do these tests function in a positive way and impact favourably on teacher expectations and the overall sense of success within the school? On the other hand, where results are predominately low, does this intertextual attribute result in negative consequences for the learning opportunities for students? This in turn can result in reduced teacher expectations and negative consequences for the belief in the efficacy of the school and schooling for children who are persistently failing on this type of test. There are a number of research findings from studies in the UK that indicate that poor levels of attainment and social class continue to be conflated at the level of school practice. Evidence from the dwindling number of studies probing current patterns of attainment in the UK reveal that a series of assumptions exist that present low ability and low attainment as almost a natural facet of the working-class cohort of students (Gillborn and Youdell 2000).

Test makers and test takers

The absence of debate in Ireland in relation to the impact of a policy of mandatory testing on children from working-class, marginalised communities is deeply regrettable. Although the judicious use of standardised testing in the overall suite of assessment tools recommended for use in schools is recommended there is no mention of the well-documented literature relating to standardised testing and children from working-class communities. This occurs despite the fact that a number of national reports in relation to reading attainment in Ireland have consistently pointed to the persistence of patterns of low attainment in disadvantaged schools (Cosgrove et al. 2000; Weir and Milis 2001; Eivers, Shiel, and Shortt 2004; DES 2005). While there has been a significant response at policy and provision level to the findings of these studies the root cause of low attainment is attributed solely at the level of the school and the individual child. Many of the initiatives and the policies that underpin them are often strongly positioned within a functionalist, meritocratic perspective which does not reflect the complexity and diversity of issues underpinning current attainment patterns. Problematising the test itself in an effort to examine the potential therein for linguistic or cultural bias has never been part of the discussion. Ireland is not

alone in this practice. The lack of discussion around the problematic use of Item Response Theory (IRT) (Bachman and Eignor 1997), model in PISA which seeks to order performance on a single dimension so that the test excludes untypical items which could validly represent cultural diversity (Goldstein 2004; Stobart 2004) represents a general lack of critique of the test instruments in educational discourse. 'More questions are asked about test takers than about the construction of tests and test makers' (Willie 2001, 472). More studies are needed on test makers, their personal characteristics and social experiences the selection of test items and how all of these interact with the various test takers (Willie 2001). Fundamental to this is the fact that standardised tests not only provide information, they also select the information to be provided (Graham and Neu 2004) and nowhere is this more discernable than in the case of the language used in standardised tests. It is thus that we are enabled to understand why these tests are harmful to marginalised/disadvantaged groups of students. The remainder of this paper will focus on the linguistic relativity of two of the main reading tests used in Irish primary schools. By examining the test answering strategies used by the children the findings will probe the asymmetrical patterns of attainment between working-class children and their middle-class peers. The linguistic relativity of the test instruments emerges as the single most stratifying element of the testing process and its outcomes.

Methodology

This examination of standardised testing in a sample of primary schools in the Republic of Ireland was part of a larger, in-depth study of children's experiences both in school and outside school. The communities from which the sample is drawn represent ideal-type socio-economic groupings based on a combination of Small Area Population Sample (SAPS) census data provided by the Irish Central Statistics Office and individual school records. The schools in the disadvantaged areas are categorised as DEIS band one schools under Department of Education and Science classification. Friendship-based, focus group discussions were held over one school year with a total of 53 12-year-old (6th class) children from two working-class schools and one middle-class school. The friendship focus group was selected as the main methodology for the study because as a model it has the potential to outweigh the presence of an observer and assist in approaching the idea of recording everyday language and the normal experience it describes. The discussions with the children on their experience of standardised testing took place near the end of the research process. By this time, the children were very familiar with the researcher and the manner in which the research was being conducted. In order to ensure that the testing experiences remained fresh in the children's experience, the school visits were organised for dates immediately following the completion of the standardised test in the respective schools. The method used with each group was as follows: The discussion began with a general discussion of the children's experience of the test. This was followed by a more detailed critique of the test completed by the sample of children. Participants were then asked to complete a selected section of both the test already completed and the alternative test (i.e. Micra-T level 4¹ if they used the Drumcondra Reading Test Level 4² or the opposite where applicable). The manner in which the different groups of children approached the test was observed and a discussion of the strategies used in completing the test constituted the final part of the focus group discussion.

During all stages of the research, the author was aware of the need to ensure that the highest standards of integrity were maintained. University College Dublin provide comprehensive guidelines, outlining a code of good practice, in order to promote a culture of research practice which will ensure integrity and rigour in all research activity (University College Dublin 2005). All of these principles of good ethical practice were central to the manner in which this study was conducted. Because the research was carried out with children, the author was particularly mindful of the guidelines and policy framework in the University for working with vulnerable groups. The rationale for working with children is clearly established by this policy. In particular the use of the friendship focus group as research was found to be particularly child friendly. It provided the children with the security of their friends at all times thereby ensuring that they always felt safe. The research was carried out in the children's own school and in a room in the school that was familiar to the children. The author was also careful to create an atmosphere where children felt relaxed and to ensure that any question, targeted at probing an issue, was carefully measured. Children were free to 'opt out' of any discussion they were not comfortable with, however, this did not occur at any time in the process. The research was designed in such a way so as to ensure that the conducting of the research did not impact on these children's time in school.

The phonological differences between children of the various social class groups are not the chief focus of this paper. However, the interviews are scripted to reflect the differences in the children's language varieties. A simple non-phonetic style of scripting is used in order to best represent the children's pronunciation (e.g. 'dis' for this, 'sayn' for saying). Where included, the language variety used by the researcher will be transcribed in a similar fashion.

Findings

Standardised test attainment patterns

The results of the standardised reading tests in English of the participating 6th class groups were obtained from each of the schools. Table 1 provides details of the mean scores of these results. The findings of this data reveal a marked difference between the achievement of pupils in the middle-class school and their peers from the schools in the working-class areas. These findings are similar to those cited in other research by the author (Mac Ruairc 2004) and also correspond to international findings of student attainment based on race and class in the US (Cooper 2000; Booker Baker 2005) and the UK (Mamon 2004). The attainment patterns of the children in the schools in the working-class areas are broadly in line with the overall attainment

Table 1. Mean scores of standardised reading tests.

Name of School	Number of Children Tested	Mean Score
Cully Hill School in working class community	24	33.9%ile
Heather Rise School in working class community	22	16.6%ile
Ard Aisling Middle Class School	31	73.5%ile

patterns outlined in a range of reports and studies including Eivers, Shiel, and Shortt. (2004) and DES (2005).

When these scores are ranked (see Figure 1) the differential attainment levels of individual children on the basis of their socio-economic status are decidedly marked.

While the difference in terms of the children’s attainment levels between the working-class schools is a significant issue, it should not detract from the bleak picture evident in the attainment gap between the socio-economic groups. These patterns cannot be ignored. It is vital to conduct a full examination that will reveal the extent of the factors that contribute to the persistence of this degree of difference. This paper explores the strategies used by the different groups of children in completing test items in an effort to identify the impact this aspect of the testing process has on attainment.

Strategies used to complete standardised test items

The analysis that follows provides an insight into the main test item response strategies used by the children in this study. While there were a number of commonalities across the social class groupings with respect to the choice of strategy there were significant differences between the children from the different socio-economic groups in terms of the outcome of the selected strategies. The test item is included, where applicable, prior to the children’s commentary in order to provide each quotation with the appropriate context.

Strategy one: means the same as . . . is the correct word for . . .

It is the accurate use of this strategy that will ensure that children complete the different test items correctly. The test score would be, therefore, an accurate measure of their competency in English reading, vocabulary and comprehension. The

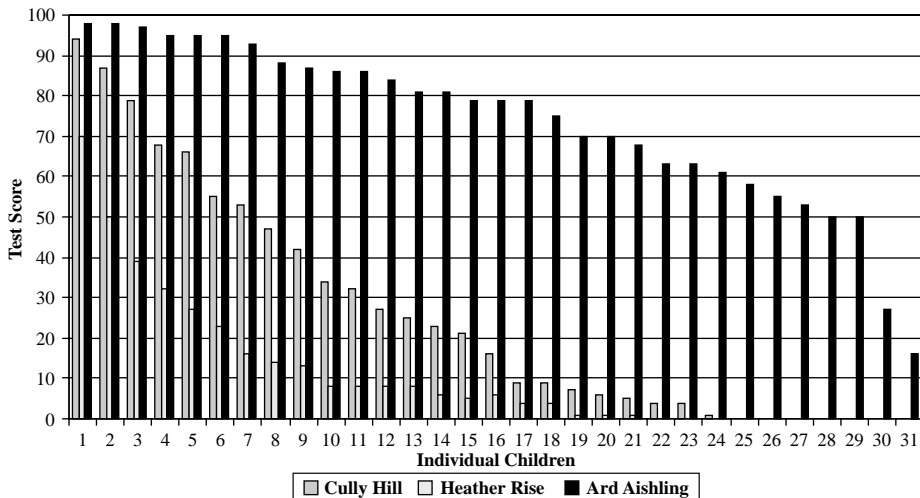


Figure 1. Comparison between percentile scores (ranked): Cully Hill, Heather Rise and Ard Aisling.

findings in this study reveal that middle-class children used this frequently in answering the test. Its use was regularly associated with a high level of confidence and certainty of outcome for this group of children. They could read the test item and all of the distractors and were very clear in relation to the answer. This degree of certainty in terms of the choice of distractor and the discourse that framed this choice indicates a considerable continuity between the language of the test and the language with which the children themselves have direct experience. As a result of this there was a sense of confidence among the group around the completion of the test items with children regularly supporting each other in their choice of answer.

Middle-class responses – Drumcondra test

They accumulated money all their lives.

- A. *wasted*
- B. *needed*
- C. *craved*
- D. *hoarded*

Shane: They accumulated money all their lives. [Reads the list of distractors] Wasted... needed... craved... hoarded... Hoarded.

Researcher: Why?

Shane: Because it's right. I know it is

She frequently visits the museum

- A. *seldom*
- B. *often*
- C. *sometimes*
- D. *never*

Devlin: I went through them all and seldom means ... ah ... like hardly ever. Often means frequently, sometimes is sometimes so it was obvious ... often

The overall degree of confidence among the group of children that characterised the middle-class practice was notably absent from the children in the working class schools. They regularly struggled with reading the test items and seldom used 'the means the same as' strategy. When used it was confined to a small number of students or to an occasional test item and accompanied by a lack of certainty and confidence with respect to the selected answer.

Working-class responses – Drumcondra test

He spoke reluctantly about his experiences.

- A. *incessantly*
- B. *rationally*

- C. *wearily*
- D. *unwillingly*

Max: I put unwillingly 'cause I think that reluctantly means that.

They were unable to subdue him.

- A. *quieten*
- B. *rescue*
- C. *assist*
- D. *replace*

Keith: I wrote quieten cause subdue means to calm him down.

The meeting ended abruptly.

- A. *amiably*
- B. *suddenly*
- C. *peacefully*
- D. *unexpectedly*

Andy: I picked unexpectedly ... I didn't know which to pick ... I think it means quickly. I'm not really sure.

Strategy two: guessing

Guessing was a feature of almost all of the children's completion strategies at some stage during the test. Children from the different socio-economic groups used the guess strategies differently with significantly varied degrees of success with respect to choosing the correct answer. The strategies used to assist or lead to a guessed answer can be grouped into two broad typologies – the quick pick or 50/50 guess.

The quick pick

This involved a simple guess and usually happened after failure to read the item. In some cases where reading was successful but where comprehension was not clear the children scanned ahead briefly and then choose (picked) a word. The children regularly refer to this type of activity as picking – '*I picked b*'. This straightforward guess was used as a strategy 45 times in total. The breakdown between school types is as follows: Heather Rise (working class) accounted for 62% of the quick pick guesses, Cullyhill accounted for 31% (working class) and Ard Aisling (middle class) accounted for 17% of this type of guessing. There was a clear preference among working-class students for this strategy. In many cases the children's difficulty in reading the test items left them with little option but to guess.

Working-class responses – Drumcondra Test

Researcher: What did you do when you didn't understand the word?

Jake: Guess – Eeeny meany miney mo.

- Pat: It's the first one that comes to your head. You don't make a fuss about it.
Sinead: I guessed all of them. I did... all of them.
Alma: I guessed all of them.
Jake: I picked C.
Steve: I didn't know what the word means – I guessed D.

Working-class responses – Micra-T

- Lauren: I read everything and yeh have to fill in all these.
Researcher: And did you understand it?
Rebecca: No, I wrote in anything

The use of this straightforward guess by the middle-class children was infrequent and confined to times when other options were deemed not to be viable.

Middle-class responses – General

- Stephanie: If you really get stuck then you have to guess.
Eire: They say on the instructions make your best guess.

The 50/50 guess

This strategy involves the use of a more considered guess. The child typically reports narrowing down the options prior to guessing the final option. Sometimes this simply involved picking the only word on the list understood by the child. Children from both social class groups used this strategy but with a greater degree of success when deployed by the middle-class children. These children were often more sure of the number of options that were not suitable. As a result, even though they were still guessing the answer, they were more likely to be correct. In the case of the working-class children the outcome was little different from the quick pick type.

Working-class response – Drumcondra test

The supply of food was adequate.

- A. Scarce
- B. Constant
- C. Guaranteed
- D. Sufficient

- Ali: Scarce.
R: Why?
Ali: Cause I didn't know what any of the other words meant.

Ann answered the question courteously.

- A. Correctly
- B. Immediately
- C. Politely
- D. Cleverly

Frank: She answered the question cur . . . curtesly like. I thought that it was good like correctly . . . I guessed that the word meant correctly.

Middle-class response – Drumcondra test

The supply of food was adequate.

- A. *scarce*
- B. *constant*
- C. *guaranteed*
- D. *sufficient*

Steve: The supply of food was adequate . . . I picked D ‘sufficient’. It was kind of a guess . . . like you know what it means . . . but you don’t know exactly what it means . . . if you know what I mean.

The President denounced the agreement.

- A. *ignored*
- B. *ratified*
- C. *welcomed*
- D. *condemned*

Sean: Ratified because I knew what all the other were. Ignore means you’re kinda putting off the agreement, welcomed means to welcome, condemned – I don’t think it would be. So I chose ratified I didn’t know what it meant so I thought it might be it.

Other strategies used by the working-class children included picking a word that ‘looked like’ the target word in some way, picking a word that started with the same letter or picking a word that was similar ‘in mood’ to the target word – ‘it wouldn’t be advised – that sounds like a threat’ (Kieran). In most cases these strategies did not result in the correct answer.

In the case of the cloze test format used in the Micra-T, there are several examples from the answers given by the children in the working-class school which indicates that the correct methodology was used to complete the item – the children were aware of the need to read the full sentence, choose a word to complete the sense of the sentence, and re-read the sentence to check the answer. However, this did not always result in a successful choice of answer. In the example quoted below, Elliot read this sentence aloud and neither he nor the others in the group saw any problem with the outcome of his selection (cloze selections inserted by in bold). This pattern of answering was very commonly used by the children in the schools in the working-class areas in response to the cloze items in the Micra-T.

Working-class responses – Micra-T

Little _____ they suspect the terrible disaster that lay _____.

- Elliot: Little *people* they suspect the terrible disaster that lay *passenger*.
R: Would anyone change any of the words he has used there.
Barry: No, It's grand.

Strategy three: checking the word in context

There were a number of different approaches taken by the children in seeking a context for their answer in their own linguistic repertoire. These approaches can be grouped under two broad headings and in each case the socio-economic divide in terms of linguistic difference is clearly notable.

It makes sense . . . it sounds right

When children were unsure of their answer, they often read all the distractors/options into the sentence to check for sense. In the case of the middle-class children, this was often simple a checking strategy to confirm the suitability of their answer. For the working-class children, this was an unsuccessful strategy. In many cases the language in the test did not resonate, in a meaningful way, with the children's linguistic experience in the sentence. As a result of this most of the distractors were viewed as making sense. There were numerous examples of incorrect answers arising from this strategy in the data.

Working-class response – Drumcondra test

She is a very competent person.

- A. *Lazy*
- B. *Quiet*
- C. *Proud*
- D. *Able*

Frank: It's what suits it. Yeh go . . . she is a very lazy person . . . she is a very quiet person . . . she is a very proud person . . . she is a very able person. And yeh see which one sounds best and the one that sounds best is proud.

The search for context

In some cases the children tried to draw on their own experience of the use of the word in order to try to figure out which option to choose. This was a tried and tested part of the children's test response repertoire and with different degree of success for different socio-economic groups.

The action we took was appropriate.

- A. *suitable*
- B. *considerate*
- C. *profitable*
- D. *necessary*

Rachel: But appropriate means like that like my ma would say you need to wear appropriate clothes so it means suitable.

However, it was in the use of this strategy that the degree of linguistic discontinuity between the test and the linguistic repertoire of the children in the schools in the working-class areas became most evident. The level of experience among the working-class children of the type of words used in the test was very limited. Their views in relation to the language used in the test clearly indicate the gap that exists between the children and the linguistic register of the test instrument.

Peter: They're very hard to understand.

Sandra: They're big weird yokes... I never really heard any of them before.

Emma: Hello have you stumbled? [hypercorrect voice/mocking tone]

Sandra: You wouldn't hear them unless you were going in for one' a them meetings and ye'd have to wear a little suit an' all like an' high heels.

Emma: Ye'd have them in a real posh school.

It is not surprising therefore that seeking a familiar context for the language in the test was regularly unsuccessful. In some cases the children's use of the strategy revealed fundamental misunderstandings about the meaning of word as a direct result of the particular context in which the child had or thought they had experienced the use of the word.

Working-class responses – Drumcondra

They were unable to subdue him.

A. *quieten*

B. *rescue*

C. *assist*

D. *replace*

Declan: They were unable to subdue him ...

James: I put rescue cause I thought subdue meant to ... like ... get him back.

Pat: I thought he was stuck down the hole and you get him back.

James: To subdue him back to your gang, yeah that's it.

Matt: I wrote quieten cause subdue means to clam him down.

Frank: I put replace ... cause in work if some fella left and you couldn't get another fella for the job then you couldn't replace him.

Are dogs domestic animals?

A. *foreign*

B. *friendly*

C. *tame*

D. *sensible*

Shelly: I picked friendly because domestic means clean from domestos the bleach and clean ... well ... that's good and so is friendly.

Ali: I picked friendly cause they are ... dogs are friendly.

In the case cited below the boys completely ruled out the correct word and found the idea that this choice could be appropriate as source of amusement.

You have made one obvious mistake.

- A. *evident*
- B. *careless*
- C. *silly*
- D. *slight*

Jack: I put silly.

Fred: I put careless.

Kevin: I think careless, silly and sight all make sense with it cause you can make a slight mistake

R: What about evident?

Fred: Evident . . .(laughing) no that wouldn't make sense that's legal stuff . . . evidence.

There was a marked contrast between the experiences of the children in the middle-class school who were notably more at ease with the language in the test. This search for context strategy resonated with their own linguistic habitus. The responses from this group of children ranged from a significant level of familiarity to a 'gut feeling' based on a view that they had heard the word used before. This provided a degree of familiarity that often was enough to successfully guide their choice of answer.

Middle-class response: Drumcondra Test

Are dogs domestic animals?

- A. *foreign*
- B. *friendly*
- C. *tame*
- D. *sensible*

Derek: Domestic means to do with the house to do with where you're living. It's like a domestic appliance . . . so tame is supposed to be the correct answer. It's really stupid none of them are the correct answer but the one you'll get the marks for is tame.

Ann answered the question courteously.

- A. *correctly*
- B. *immediately*
- C. *politely*
- D. *cleverly*

Researcher: Why did you choose politely?

Jackie: Courteous kinda means nice kinda you know . . . polite.

Eva: I don't know why I knew the meaning. I've heard it like loads of times before.

Eire: You read it in books.
Stephanie: You ... kinda ... hear it around the place.

The confrontation lasted three hours.

- A. *argument*
- B. *ceremony*
- C. *lecture*
- D. *conference*

Emma: I don't know ... I thought ... I have heard the word confrontation before but I can't remember. I thought it meant something about ... like ... arguing. You know confronting something like you know. I thought it meant arguing. I wasn't too sure.

Conclusion

This study was a small scale, in-depth examination of the experiences children from different socio-economic groups had with respect to standardised testing. The sample was small and as a result it is not possible to generalise the findings. However, clear patterns are identifiable in the data set and a number of issues emerge which require consideration in the context of the type of testing culture that will develop in Irish schools going forward. The commonalities in the type of test response strategies used between the contrasting socio-economic groups reveal that children are generally equipped with a similar repertoire of skills and strategies with which to approach the test situation. This is most likely to be the result of prior experience of testing over their years in primary school and direct guidance from teachers with respect to completing test items. However, when the use of the different strategies is interrogated, significant differences between the socio-economic groups in terms of the efficacy of each strategy emerge. The findings strongly reveal that at the core of this issue is the marked difference in the linguistic habitus (Bourdieu 1991) and its associated linguistic capital between the two groups of students. The level of resonance between the language used in test items and the middle-class linguistic repertoire positions the middle-class student in a notably privileged position in terms of the potential for higher levels of attainment on the test. This high level of linguistic resonance permeates the whole test experience for the middle-class child, with the result that each of the strategies used, including simple guessing, have a great likelihood of success. The testing experience for the working-class child on the other hand, is characterised by varying degrees of struggle with the linguistic challenges that the tests presents. In many cases, for this group of children, the experience of the standardised test is predominately negative. It provides a context for the further erosion of many children's self-efficacy in meeting the demands of schools, and serves as a clear reminder of the inferior position of the working-class linguistic repertoire on the linguistic hierarchy that prevails in education.

The bias inherent in standardised tests exists at a fundamental level in favour of middle class students. Other literature in relation to bias (Graham and Neu 2004; Gipps and Murphy 1994; Tae-Il, 2004; Abbot, 2007; Goldstein 1986; Shohamy 2001a; Anagnostopoulos 2005) provides a framework for examining bias that does

not adequately take account of, or articulate clearly, the deep level at which this bias operates. In choosing to test children on the dominant linguistic code, the test and the testing process negate a whole way of being, a complete and well-established way of making sense of the world by failing to recognise the language variety used by specific groups to construct meaning. The meaning that is communicated to children in marginalised communities in relation to the lack of prestige and stigmatised nature of their linguistic capital is overtly articulated in the practice of testing. The data reported here clearly exposes the children's struggle to make sense of the language of the test. The attempts by the students to locate a resonance for the language register of the test in their linguistic repertoire is not acceptable practice in a school system that purports to enable the child 'to realise his or her potential as a unique individual' (Government of Ireland 1999, 7). This surely must involve validating the linguistic and cultural lifeworld of each individual child.

What is required is a fundamental recognition of the impact that linguistic differences have on the processes used in completing standardised tests and the attainment results that are derived from them. What needs to ensue is a long-overdue rigorous questioning both of the current and future role of standardised testing in Irish primary schools by focusing on the impact the policy has on those who benefit least from the current *status quo*. The findings of this study provide clear evidence that we do not exist on a level linguistic playing field and that this factor has direct causal links to the asymmetrical patterns in attainment on standardised testing of reading. It is vital therefore that the current policy drive that enhances the currency of standardised testing in the system is prevented from following the well established trend where policy initiatives 'when actually put in place, tend to benefit those who already have advantages' (Apple 2000b, 312).

Notes

1. The Micra-T is a group administered standardised reading test. The test contains five passages, all of a cloze procedure format. The tests were re-developed for the school year 2003–04 and are considered by the test developers to contain material suitable for different ability groups. The version of the test used in this study, was standardised on a nationally representative sample of more than 10,000 Irish pupils during the 2002–03 school year.
2. The Drumcondra Primary Reading Test (DPRT) is a group administered, standardised test of silent reading used in the Republic of Ireland. The test assesses two aspects of reading: reading vocabulary and reading comprehension. The vocabulary sub-test only was examined and explored as part of this study. This sub-test consists of 40 questions each containing a target word embedded in a short sentence and four distractors. The pupil must determine which of the four distractors is closest in meaning to the target word. The edition used in schools i.e. the 1996 edition was used in this study.

References

- Anagnostopoulos, D. 2005. Testing, tests and classroom texts. *Journal of Curriculum Studies* 37, no. 1: 35–63.
- Abbot, M.L. 2007. A confirmatory approach to differential item functioning on an ESL reading assessment. *Language Testing* 2, no. 1: 7–36.
- Apple, M. 2000a. Educational politics from below. *Educational Policy* 15, no. 5: 724–9.
- . 2000b. *Official knowledge*. 2nd ed. New York: Routledge.
- . 2001. Will standards save public education? *Educational Policy* 14, no. 2: 311–8.
- . 2002. Does education have independent power? Bernstein and the question of relative autonomy. *British Journal of Sociology of Education* 23, no. 4: 607–16.

- Bachman, L.F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- . 2000. Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing* 17: 1–42.
- Bachman, L.F. and D.R. Eignor. 1997. Recent advances in quantitative test analysis in *Encyclopedia of language and education* 7, eds. C. Clapham and D. Corson, 227–42.
- Bakhtin, M.M. 1986. *Speech genres and other late essays*. Austin: University of Texas Press.
- Ball, S.J. 2006. *Education policy and social class*. London: Routledge.
- . 2008. *The education debate*. Bristol: Policy Press.
- Booker Baker, P. 2005. The impact of cultural biases on African American students' education: a review of research literature regarding race based schooling. *Education and urban society* 37: 243–56.
- Bourdieu, P. 1991. *Language and symbolic power*. Cambridge: Polity Press.
- Brickwell, H.M. 1978. *Minimum competency testing: 1978*. Colorado: Education Commission of the States.
- Brown, G.T.L. 2004. Teachers' conception of assessment: implications for policy and professional development. *Assessment in Education* 11, no. 3: 301–18.
- Cooper, R. 2000. Urban school reform from a student of color perspective. *Urban Education* 34, no. 5: 597–622.
- Cosgrove, J., T. Kellaghan, P. Forde, and M. Morgan. 2000. *The 1998 national assessment of English reading*. Dublin: Educational Research Centre.
- Department of Education and Science. 1999. *Primary school curriculum* Dublin: Government Publications Office.
- . 2000. *Learning support guidelines*. Dublin: Government Publications Office.
- . 2005. *Literacy and numeracy in disadvantaged schools: Challenges for teachers and learners*. Dublin: Government Publications Office.
- . 2006. Supporting assessment in primary schools. Circular letter 0138/2006.
- Delpit, L. 2006. *Other people's children: Cultural conflict in the classroom*. New York: The New Press.
- Devine, D. 2000. Constructions of childhood in school: power, policy and practice in Irish Education. *International Studies in Sociology of Education* 10, no. 1: 23–41.
- . 2003. *Children, power and schooling: How childhood is structured in the primary school*. London: Trentham Books.
- Dixon, H. 1999. *The effect of policy on practice: An analysis of teachers' perceptions of school based assessment practice*. Albany: New Zealand Massey University.
- Eivers, E., G. Shiel, and F. Shortt. 2004. *Literacy in disadvantaged primary schools*. Dublin: Educational Research Centre.
- Gallagher, C.J. 2003. Reconciling a tradition of testing with a new learning paradigm. *Educational Psychology Review* 15, no. 1: 83–99.
- Gallas, K. 2001. 'Look, Karen, I'm running like Jello': Imagination as a question, a topic, a tool for literacy research and learning. *Research in the Teaching of English* 35: 457–92.
- Gipps, C., and P. Murphy. 1994. *A fair test?*. Philadelphia: OUP.
- Gillborn, D., and D. Youdell. 2000. *Rationing education: Policy, practice, reform and equity*. Buckingham: Open University Press.
- Glaser, R. 1963. Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist* 18: 519–21.
- Goldstein, H. 1986. Gender bias and test norms in educational selection. *Research Intelligence* 23: 2–4.
- . 2004. Pisa study check up. *Assessment in Education* 11, no. 3: 324–48.
- Government of Ireland. 1999. *Primary School Curriculum – Introduction*. Dublin: Government Publications Office.
- Graham, C., and D. Neu. 2004. Standardized testing and the construction of governable persons. *Journal of Curriculum Studies* 36, no. 3: 295–319.
- Grant, S.G. 2001. An uncertain lever: exploring the influences of state level testing in New York State on teaching social studies. *Teachers College Record* 103, no. 3: 398–426.
- Haertel, E.H., and W.A. Lorie. 2004. Validating standards-based test score interpretations. *Measurement* 2, no. 2: 61–103.

- Headington, R. 2003. *Monitoring, assessment, recording, reporting and accountability: Meeting the standards*. London: Fulton.
- Hill, M.F. 2000. *Remapping the assessment landscape: Primary teachers reconstructing assessment in self managing schools*. Hamilton, New Zealand: University of Waikato.
- Kohlfeld, B.C. 1982. An analysis of opinions regarding minimum testing competency. *Education* 5, no. 1: 54–7.
- Lam, T.C.M., and C. Bordignon. 2001. An examination of English teachers opinions about the Ontario grade 9 reading and writing test. *Interchange* 32, no. 2: 131–45.
- Linn, R.L. 2000. Assessment and accountability. *Educational researcher* 29, no. 2: 4–16.
- Luxia, Q. 2005. Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing* 22: 142–73.
- Mac Ruairc, G. 2004. Schools, social class and children's perception of language variation. In *Primary voices*, eds. J. Deegan, D. Devine & A. Lodge. Dublin: IPA.
- Mamon, S. 2004. Mapping the attainment of black children in Britain. *Race Class* 46: 78–91.
- McNeil, L. 2000. *Contradictions of reform*. New York: Routledge.
- Meier, D. 2000. *Will standards save public education?*. Boston: Beacon Press.
- National Council for Curriculum and Assessment (NCCA). 2005. *Assessment in the primary school curriculum: Guidelines for school*. Dublin: (NCCA). http://www.ncca.ie/uploaded-files/primary/ass_pr_schs_05.pdf
- . 2007. *Assessment in the primary school curriculum: Guidelines for school*. NCCA: Dublin.
- Paton, G. 2008. School SATS results hit by delay. *Daily Telegraph*, July 4.
- Shepherd, J. 2008a. Sack SATS markers Tories demand. *Education Guardian*, July 17.
- . 2008b. New SATS blow as US firm is relieved of remarking. *Education Guardian*, July 29.
- Shohamy, E. 2001a. Democratic assessment as an alternative. *Language Testing* 18: 373–91.
- . 2001b. *The power of tests: A critical perspective on the uses of language tests*. London: Longman.
- Shohamy, E., S. Donitsa-Schmidt, and I. Ferman. 1996. Test impact revisited: Washback effect over time. *Language Testing* 13: 299–317.
- Spolsky, B. 1981. Some ethical questions about language testing. In *Practice and problems in language testing*, ed. C. Klein-Braley and D.K. Stevenson. Frankfurt: Peter Lang.
- Stobart, G. 2004. Developing and improving assessment instruments. *Assessment in Education* 11, no. 3: 243–5.
- Tae-il, P. 2004. DIF for examinees with different academic backgrounds. *Language Testing* 21, no. 1: 53–73.
- Weir, S., and L. Milis. 2001. The relationship between the achievement of 6th class pupils from disadvantages backgrounds and their attitudes to school. *The Irish Journal of Education* 32: 63–83.
- Williams, J., and J. Ryan. 2000. National testing and the improvement of classrooms teaching: can they coexist? *British Educational Research Journal* 26, no. 1: 49–73.
- Willie, C.V. 2001. The contextual effects of socioeconomic status on student achievement test scores by race. *Urban Education* 36, no. 4: 461–78.

Copyright of Irish Educational Studies is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.