

“THE BIG TEST”: A SCHOOL COMMUNITY EXPERIENCES STANDARDIZED MATHEMATICS ASSESSMENT

Fiona Walls

James Cook University, Australia

Standardized testing in primary school mathematics is widely used as a monitoring device in education systems world-wide. This paper uses Foucault's notion of assessment as a technique of power, to examine how members of one Australian school community experienced the annual compulsory standardized Year 5 Aspects of Numeracy Test. The paper examines the ways in which the “numerate” child is inscribed within the “normalizing” apparatus of standardized assessment.

INTRODUCTION

Standardized testing of literacy and numeracy skills of primary school students has become a common method of assessment in Australia. It is widely accepted that such tests are objective and reliable, providing essential information for the provision of quality education. Wiliam, Bartholomew & Reay (2004) challenge such claims:

“For the most of last century, educational assessment derived its principal research paradigm from psychology...the creation of tests and other forms of assessment has been regarded as an essentially technical and objective undertaking although there has, during the last quarter-century, been an increasing acceptance that educational assessments have social consequences –people change what they do because of assessments.” (p. 43)

In examining UK secondary school students' experiences of the SAT (Scholastic Aptitude Test), Wiliam et al demonstrate a powerful relationship between assessment, learning and pupil identity, arguing that test performance plays a significant part in students' “becoming” as mathematical subjects and in determining the shape of students' learning trajectories.

This paper reports on research that also investigated relationships between standardized assessment and pupil identity. The research gathered the accounts of children, parents, teachers and managers of one primary school community who reported their experiences of the 2005 Queensland Year 5 Aspects of Numeracy Test.

The specific purposes of the Years 3, 5 and 7 tests are “to collect data from the population of Years 3, 5 and 7 students for reporting to parents/carers and schools for systematic reporting,” and, “to accommodate the assessment of students against national benchmark standards” (Queensland Studies Authority, 2005a). All eligible children in Queensland are required to take part, with exemptions granted only in exceptional cases. The test is administered within classrooms in the manner of an examination. Teachers guide the children through the two-hour test following a prescribed script and adhering to strict time allocations. The test is of the pencil-and-paper type with over half of the forty-two questions structured in a multiple-choice format, and the remainder requiring digits to be written in answer boxes.

Walls

Although its authors claim the test assesses children's thinking, reasoning and working mathematically (The State of Queensland (QSA), 2005a), the test does not ask children to explain or justify their answers. This method of assessment assumes correct answers match "correct" thinking, which in turn indicates sound mathematical understanding. Mathematics education researchers who have investigated such tests (e.g. Ellerton and Clements, 1997) claim a question/answer/understanding mismatch rate of as high as 30% and argue that the tests are unacceptably unreliable. This method of assessment continues to be used by Education Queensland as the primary source of statewide data about children's achievement in primary mathematics.

Children's test results were reported as dots on horizontal scales, pinpointing overall achievement as well as achievement by the separate mathematics strands of Number, Measurement & Data, and Space, as shown in Figure 1 below. The dots positioned the child's test score relative to the state average (solid line), the "middle" 50% of scores in the state (shaded box), and Australian national benchmarks (dotted line). On the example below, the hypothetical child has scored above the state average for both measurement and number, below the state average for Space, but within the middle 50%, giving an overall score slightly better than the middle 50%.

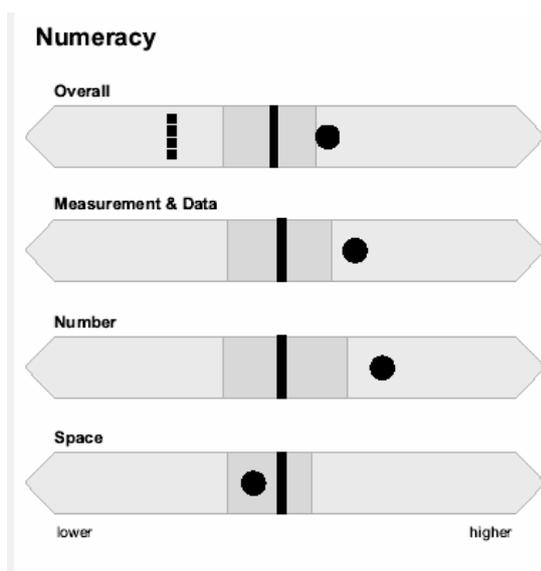


Figure 1: Sample of Numeracy Test results (The State of Queensland (QSA), 2005b)

THEORETICAL FRAMEWORK

This paper investigates how such standardized assessment might operate in the production of the child as "numerate". It draws on Foucault's theories of power, knowledge, and the discursively produced subject (Foucault, 1994) that researchers have found useful in investigating issues of pupil identity in mathematics education (e.g. Walkerdine, 1998; Walls, 2004, Cotton, 2004). The paper uses Foucault's suggestion that measurement and classification operate as apparatuses of recognition and normalization. Foucault (1970) looked to historical sources to understand how human systems of classification over time and across differing cultural settings have been concerned with "noticing", "describing", "defining" and "ordering". He argued that classification systems structure our world view, and that classification of the "self" has become a pervasive managerial technique within institutions such as

schools, whose purpose it is to *correct* or *train*. In measuring, classifying and ordering our “selves” according to matrices of normalization such as levels, stages, benchmarks and standards, we make it possible to delimit, proscribe, admit passage, or debar. Foucault (1977) described the test or examination as the mechanism central to the process of classification of “selves”:

“The examination combines the techniques of an observing hierarchy and those of a normalizing judgment...It establishes over individuals a visibility through which one differentiates them and judges them. That is why, in all the mechanisms of discipline, the examination is highly ritualized. In it are combined the ceremony of power and the form of experiment, the deployment of force and the establishment of truth. At the heart of procedures of discipline, it manifests the subjection of those who are perceived as objects and the objectification of those who are subjected.” (pp. 184-185)

Foucault regarded schools as disciplinary institutions in which, “The perpetual penalty that traverses all points and supervises every instant...compares, differentiates, hierarchizes, homogenizes, excludes – in short, *normalizes*” (Foucault, 1977, p. 183). In this view, schooling acts as a discursive domain productive of “ability” such as the “numerate child”, created and maintained through a framework of assessment practices of which the standardized test is an essential component.

METHOD

This research sought to reveal the complexities of the social dimensions surrounding the classifying and normalizing apparatus of one measure of numeracy, the standardized Queensland Year 5 Aspects of Numeracy Test. The research was conducted within a large suburban Queensland school. Perspectives were gathered from members of the school community, including children, parents, teachers and school managers. The school community was investigated as a site where, at the capillared extremities (Foucault, 1997) of a centralized education system and within the contingent arenas of school and family, children might be *produced* as “numerate subjects” through the classifying mechanisms of the test.

The research was based on the *critical collaboration* model described by Atweh, (2004) and the *critical analytic* approach described by Freebody (2003), whose key concepts are the provision of multiple ‘voices’, the location of the research process within authentic accounts, and an emphasis on the *process* of research activity as much as the *outcomes*. The research involved gathering participants’ experiences of the test by means of informal conversation immediately after the test had been administered, and again just after the results had been received. It was conducted in four phases: (a) orientation in which the researcher spoke with school managers, reviewed the test preparation materials and observed one class sitting the test; (b) post-test conversations with teachers, children and parents which were audio-recorded and transcribed; (c) transcript review in which transcripts were returned to participants for comment or amendment; (d) post-results conversations with teachers, parents and children. Participants’ conversations were reviewed for statements that might provide evidence of Foucault’s contention that examinations operate to *normalize*: that is, to *compare*, *differentiate*, *hierarchize*, *homogenize*, and *exclude*.

RESULTS

The children spoke of the test as an extraordinary event in their schooling within which they became separated, measurable, and individual performers.

Ana: We couldn't sit near to each other; we couldn't open our desks, or get out of our chairs, or anything.

Ari: At the beginning I was struggling, but I had some water and I could do it better.

They spoke of the test as a highly charged emotional experience which they approached with excitement and apprehension. Excitement was linked to the possibility of rewards gained from "doing well" while anxieties were linked to the possibility of punishments for doing badly such as losing face in front of friends, parents or siblings, loss of privileges at home, or being kept back.

Chloe: I'm feeling a bit excited. Because they (teachers) said I'd probably do well.

Perry: I was so excited to do the test, I thought about my Mum and Dad and my family, if I do well, what would they say and everything, if I got a good mark...All I'm worried about is getting some money taken off me or something if I get a bad mark...I get stuff taken off me or lose money or get grounded.

Chris: I had butterflies. I didn't even want to do the test...I was like, "Oh my God, please be good, please be easy"...I don't want to get my results...My mum and dad says that if I don't get over 10 or 11, I have to do Year 5 again.

Sally: I felt sad that I wouldn't do really well.

Children talked about the purpose of the test. Their statements show that they believed the test to be an inevitable and essential part of their schooling; without its *differentiating* and *comparing* judgements, they would be unable to learn, progress, or make it in the world beyond primary school.

Ana: I think to see what level we're at, you know

Kim: If you get all the test wrong or you don't do a test, then you won't learn for next year. It's like a cycle...And if you didn't get a good result out of it, that means that we didn't learn much in Grade 5 so we should stay back

Tony: My mum said, "Don't worry 'cause it's not going to affect your life... (Dad said) "You probably won't get kept down."

Chris: It actually does affect your life, if you don't go really good, it affects your career.

Perry: Some people won't get jobs they want because they won't know as much.

Zarn: The test is important. [If you missed the test] people wouldn't be able to find out how they went...so the teachers won't know how much they know, what to do in case they need to be kept down or need to do more maths.

These statements contain evidence of the children's understanding of the test as part of a process of *differentiating*, *including* and *excluding*, but also of *homogenising*; the need for those who fail to "stay back" or "do more maths" in order to close gaps and meet needs; ideally, within the same curriculum everyone reaches the same standard.

The children explained their reading of the results (refer Fig 1). While some struggled to make sense of the graphic representation of their performance, their statements show that they understood the test to be *comparing*, *differentiating* and *ranking*:

- Adele: I think I did pretty good...I saw that and I thought, “Wow!”...I’m off the graph.
- Tony: I got all mine in the higher place...they’re like the highest grade.
- Bart: I got over the...(points to the line indicating the state average).
- Kim: It means what rating you got. And that’s like the middle line (pointing). And, I don’t know what that darker square is, I think it’s not low, but “before low”. [I feel] Bad...pretty sad...I thought I would have done better...’cause it was pretty easy and I knew all the maths.
- Perry: Well I’m very good at Space, and Number I’m OK at. Measurement & Data I’m not that good at it. Overall, yeah, I’m good... I thought I’d be better at Number than Space. But now I’ve found out that I’m better at Space than Number.
- Ari: (Pointing to each of his dots in turn) Well that’s like, “*all right*”, and that’s, “*good*,” and that’s “*all right*,” that’s my guess...’cause at my old school I didn’t get this, I’m used to getting numbers, marks...
- Britta: I wanted to be a bit higher.

Reading the results was a social process. Interactions with family, peers and friends helped children to gain a sense of the *worth* of their results. Comparisons proved affirming for some, but distressing for others:

- Ana: My friends said they’re good [results]...I’ve got two [dots] low and that’s in the middle, and I don’t know about that one (indicates ‘overall’ category).
- Sally: She (identical twin sister) got higher than me. Simon (younger brother) got high in everything... He can pick it up straight away and for me and Ingrid we have to keep on working for it...I didn’t want Simon to look at mine because I didn’t want him to see he was better than me...Yeah, I’ve never been good at maths.
- Bart: I don’t think Martin’s (classmate) maths [result] is going to be very high. He’s not good at maths. I’m good at maths.

Teachers’ statements revealed not only concern about the *comparisons* created by the test, but also the pressures they faced in preparing children for the test, revealing how their teaching behaviour changed to optimise children’s performances.

- Allie: There are percentages of these tests where some of it is baseline and then it gradually gets harder to let those children who are very bright, shine. I just find that very frustrating for the children who aren’t the shiners.
- Dan: You do to a point have to teach to the test, just to get them used to the process of doing one...it’s a stress to make sure that you’ve covered everything.
- Col: We talk about pressure on the kids but also on the teachers as well... we’ve got to prepare for this because you’re not giving the kids the opportunity to show their best...to me it rearranges my teaching format, you’ve got to do bits of everything to ensure that they get a range.

Walls

Comparison, differentiation, hierarchization and *exclusion* were recurrent themes in the parents' accounts. Test anxiety and feelings of failure were poignantly described. Their accounts of exclusion show how *homogeneity* emerged as a desired state – if testing *excludes*, then corrective action is needed to ensure that no child is left behind.

- Ursula: When she (daughter Sally) does these tests, she gets really, *really* stressed...She ends up in tears, it's not worth it because it highlights her ability, her weakness, which is maths...We got her results back yesterday. Sally was pretty upset...Crying. I tried to explain to her that it didn't mean as much as her general report card because the teacher knows what she's been doing through the year. But when she saw that her dots were all below in everything, she was very upset...her brother's in Grade 3 and he was off the scale which makes them feel inferior...it causes friction between the siblings, and embarrassment, like yesterday I was doing Ingrid's (Sally's identical twin sister), I look at their reports individually, and Ingrid was sitting there and telling Sally to go away [saying] "I don't want you to look at mine". But what happens where they've got an overall rating that's below the benchmark? There's got to be a purpose to this testing, to put kids through the stress and then just leave them there. There should be extra tutoring or something, they should be trying to get them up there.
- Marlene: Britta (daughter) stresses about the test to begin with...and then for her to be able to compare it with other kids and to see she's not doing so well is really upsetting to her...they started talking about it in Term 2 and she started building up from there. "It's that big one again, Mum...it's all of Queensland." She'd looked at [the test results] at school and seen other kids', and she said, "Mum, mine was lower than my friends and that was really bad." That was quite a shock for her to open it in front of other kids, who've gone, "Look where your dot is!"...She just sees it as, "I'm not even in the middle"...I don't really want to know where my child is as a dot in Queensland...My child's not a dot. The comparison I don't think is fair. I just don't think it's something that the kids need to go through.
- Carol: We're very proud of (daughter) Adele's results. She's got a high level for everything, but Leo (son) sat the Year 3 Test and we're very disappointed with his results...if that's the benchmark, he's back here, he's behind the eight ball.
- Tina: She (daughter Zarn) did really, really well on her results... this is probably surprising (points to lower Number result) and that she'd be off the graph of what they'd expect (points to Measurement & Data)...They need to have a standard that they can reach... I think that's what is important about the Numeracy testing is they actually find out where they're sitting with the state.
- Olivia: Chris (son) thinks he's no good at maths. His result's just above the middle line, so he's actually in the higher side of the average...His dad said, "Mate, you can't tell me you're no good at maths...It makes a lot of difference with Chris's results to my oldest son; he was well below the state average in everything.
- Marie: He (son Andrew) thought he did better in Numeracy than Literacy, he said he thought the Numeracy was easy, but it was the other way around. So I wonder, are these results really Andrew, or was it just him panicking on the day? He's a little sad that Charlie (younger brother) may be a lot quicker than him at maths. Charlie did really well (on the Year 3 Test). I haven't shown him Charlie's

results. I don't want them to compare, but it's how to talk to him about it...I can't pinpoint what went wrong so I can't help him.

The school manager's views were summed up in the following statement.

Principal: It's about schools, getting data about trends to improve performance and it's about systems, so it's very political. At the coal face when you've got these little Year 3s doing this test and getting anxious about it, or we had one Year 5 and he was away in Year 3 when they did test and he was very anxious because he'd never done one of these before, and going through the whole trauma and then knowing it's not actually about helping your child, it's a game, they have to jump through the hoop, but it's not going to benefit them; it's hard to justify as an educator...we need to be honest about the purpose of the test, what the audience for the results are, and the audience is not really the parents, it's not the students, its hardly for teachers...It's for administrators and the school.

CONCLUSION

The conversations of the participants in the research reveal that the Years 3, 5 & 7 Aspects of Literacy and Numeracy Tests loomed as a major event in school and home life in its perceived authority to tell the truth, that is, to objectively measure and rank each child. In this perception, school management and teacher behaviour were modified, pupil identity reworked, and relationships within families adjusted.

In mathematics education in Australia, "normalizing" classification through standardized testing begins in early primary school, measures children according to state or national standards and benchmarks, and is linked to the politically significant international studies. In this way, the mathematical achievement of every child in Australia can be documented and tracked. Through such a regime, the "numerate child" becomes a subjected entity made visible for teachers, parents, and policy makers. Within the establishment of a numerate norm, the "innumerate child" and the "exceptional child" are also brought to life. For the subjects who are thus objectified, beliefs about mathematics, cognition, affect, and identity become entwined, as Britta, whose test results were not what she had expected, so clearly articulates:

Britta: I don't like maths, it's not fun, but I do my best...sometimes it's confusing... I thought maybe, if I done [the test] again, I probably could have done it better than I did, could have answered the questions.

With little choice but to accept the judgement of the test and it's identification of her as less-than-acceptably-numerate, Britta still tries to resist its condemnation by retaining the hope that she "could have done it better". Inscribed by the "normalizing" processes of obligatory state testing, Britta's enjoyment of mathematics and her faith in her mathematical abilities have been eroded.

The research showed that irrespective of how "well" the children "did", the test contributed significantly to their *becoming* as mathematical subjects. Children, parents, teachers and school administrators all grappled with the test's licence to create a normalizing "truth" (Foucault, 1970). In light of the limited information the test elicited about children's mathematical thinking, and in view of the questionable

validity of the test, the inappropriate and disproportionate capacity of the test to compare, differentiate, rank, homogenise, exclude and “normalize” must be rigorously challenged. Further community-based research of this nature is necessary.

References

- Atweh, B. (2004). Towards a model of social justice in mathematics education and its application to critique of international collaborations. In I. Putt, R. Faragher & M. McLean (Eds.) *Proceedings of the 29th Annual Conference of Mathematics Education Research Group of Australasia*. (47-54) James Cook University, Townsville: 2004.
- Cotton, T. (2004). What can I say? What can I do?: Archaeology, narrative and assessment. In M Walshaw (Ed.), *Mathematics education within the postmodern*. Greenwich CT: Information Age Publishing.
- Foucault, M. (1997). *Society must be defended: Lectures at the Collège de France, 1975-76*. Translation, 2003. London: Penguin Books.
- Foucault, M. (1994). Power. J. Faubion (Ed.), *Essential works of Foucault 1954-1984, Vol 3*. London: Penguin.
- Foucault, M. (1977). *Discipline and punish: The birth of the prison*. London: Allan Lane.
- Foucault, M. (1970). *The order of things: An archaeology of the human sciences*. New York: Random House Inc.
- Foucault, M. (1969). *The archaeology of knowledge*. Routledge Classics Edition, (2002). London: Routledge.
- Freebody, P. (2003). *Qualitative research in education: Interaction and practice*. London: Sage.
- Ellerton, N. & Clements, K. (1997). Pencil-and-paper tests under the microscope. In A. Begg (Ed.) *People in mathematics education. Proceedings of the 20th Annual Conference Mathematics Education Research Group of Australasia*. Rotorua: (155-162)
- The State of Queensland (QSA), (2005a). *2005 Queensland Years 3, 5 and 7 Tests in Aspects of Literacy and Numeracy: 2005 Test preparation handbook*. Brisbane: The State of Queensland (Queensland Studies Authority).
- The State of Queensland (QSA), (2005b). *Queensland Years 3, 5 and 7 Tests in Aspects of Literacy and Numeracy: 2005 Test reporting handbook*. Brisbane: The State of Queensland (Queensland Studies Authority).
- Walls, F. (2004). The “mathematically-able child” in primary mathematics education: A discursive approach. In I. Putt, R. Faragher & M. McLean (Eds.) *Proceedings of the 29th Annual Conference of Mathematics Education Research Group of Australasia*. Townsville: 2004.
- Walkerdine, V. (1998). *Counting girls out: Girls and mathematics*. New Edition. London: Falmer Press
- William, D., Bartholomew, H. & Reay, D. (2004). Assessment, learning and identity. In P. Valero & R. Zevenbergen (Eds.), *Researching the socio-political dimensions of mathematics education: Issues of power in theory and methodology*. (43-61) Dordrecht: Kluwer Academic Publishers.