Routledge
Taylor & Francis Group

Education 3-13

International Journal of Primary, Elementary and Early Years Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rett20

‘Until now, nobody’s ever actually said straight out how do you feel about them’: children’s experiences of the impact of high-stakes testing through participatory methods

Joanna Wheldon & Tom Dobson

To cite this article: Joanna Wheldon & Tom Dobson (26 Oct 2024): ‘Until now, nobody’s ever actually said straight out how do you feel about them’: children’s experiences of the impact of high-stakes testing through participatory methods, Education 3-13, DOI: 10.1080/03004279.2024.2419471

To link to this article: https://doi.org/10.1080/03004279.2024.2419471

Published online: 26 Oct 2024.

EDUCATION 3-13
https://doi.org/10.1080/03004279.2024.2419471

'Until now, nobody's ever actually said straight out how do you feel about them': children's experiences of the impact of high-stakes testing through participatory methods

Joanna Wheldon and Tom Dobson

York St John University, York, UK

ABSTRACT

Research into the impact of high-stakes testing is often adult-centric. Using participatory methods with 10–11-year olds, this paper demonstrates how research with children provides a more nuanced understanding of the impacts of high-stakes testing on wellbeing, curriculum and pedagogy. This paper also demonstrates how children have an implicit understanding of the mechanisms of neoliberalism and how these mechanisms lead to competition with peers, which can exclude lower attaining children. We recommend further participatory research with children to inform wider policy and that schools develop children's emotional literacy to help them to manage peer relations around high-stakes testing.

ARTICLE HISTORY
Received 17 May 2024
Accepted 14 October 2024

KEYWORDS
High-stakes testing; wellbeing; curriculum and pedagogy; primary schools; neoliberalism; participatory methods

Introduction

A global report highlighted that 'one in seven 10–19-year olds experiences a mental disorder', with 'depression, anxiety and behavioural disorders' among the leading causes of illness and disability (World Health Organization (WHO), 2021). A recent systematic review reported high prevalence of anxiety, depression, sleep disorders and suicidal behaviour amongst young people (Hossain et al. 2022). Against this backdrop, the negative impact of 'high-stakes' testing on children's mental health is under scrutiny. Here we define high-stakes tests as tests to which professional and public consequences are attached for teachers and schools (Howell 2017). In England, where this research takes place, pre-COVID studies (e.g. Hutchings 2015) demonstrate how high-stakes testing negatively affected primary and secondary school students' wellbeing (Hutchings 2015). Given this and similar findings from recent research with children across Europe (Högberg and Horn 2022), the return to pre-COVID high-stakes testing in many countries could have a detrimental effect on the mental health of children.

The return to high-stakes testing could also continue pedagogical and curriculum narrowing (Berliner 2011), leading to didactic teaching, the implications of which are explored by the Organisation for Economic Development (OECD 2020). In England, pedagogical variation is restricted – a worrying finding given the link the OECD (2020) make between pedagogical variation and children's development of competencies required by employers. Pedagogical narrowing is shown to particularly affect disadvantaged children across the globe (OECD 2020). In England, those with lower attainment and special educational needs and disabilities (SEND) are most affected (Hutchings 2015), with SEND children removed from wider classes for English or maths, resulting in 'an apartheid system of schooling' (Berliner 2011, 292).

CONTACT Joanna Wheldon j.wheldon@yorksj.ac.uk

Conceptualising high-stakes testing as a mechanism of neoliberalism, this paper includes an analysis of research literature from the advent of increased high-stakes testing (2000 onwards) into the impacts of high-stakes testing. We focus on the extent to which this research is adult-centric with children's experiences often marginalised. Underpinned by a belief that the promotion of the voices of children in research is vital to understanding children's lived experiences (Punch 2002), we use participatory methods with children to explore how Year 6 children (10- and 11-year olds) in their final year of an English primary school perceive the impact of high-stakes testing. Our analysis of this data provides nuanced insights, which have implications for research, policy makers, school leaders and teachers.

Literature review

Neoliberalism and education

Neoliberalism positions educational institutions within a 'free market' economy, where resources are allocated to institutions delivering the best outcomes (Tett and Hamilton 2021, 1). Whilst definitions of neoliberalism vary and are contested, we are interested in how neoliberalism is operationalised in education and align with Ball's (2016) view that education reforms are driven by three interrelated mechanisms: market, management and performance. In England, 40% of primary schools and 80% of secondary schools operate as businesses through academisation (GOV.UK 2023), with competition and choice used as levers for improvement and governing bodies replaced by Boards of Directors and Chief Executive Officers. Ball (2016) describes this as 'reculturing', with less democratic forms of management and an education system driven by performance. Management is driven by performativity – 'a regime of accountability that employs judgements, comparisons and displays as means of control, attrition and change' (Ball 2013, 57). High-stakes testing is part of this mechanism and its global operation manifests in comparisons of student outcomes through the OECD's Programme for International Student Assessment (PISA). Successive British governments have argued that the raising of standards through improvement in PISA rankings is imperative to compete in a global economy (from Department of Education and Science (DES) 1985 to DfE 2016).

Since 1991, performativity in England has been measured through national curriculum testing, known as SATs – a high-stakes approach to testing with each school's results categorised, ranked and made public (Au 2007). Initially SATs were taken in primary and secondary schools, but in 2008 SATs in secondary schools were abolished following concerns about 'pervasive anxiety' in children (Curtis 2008). In primary schools, SATs were maintained, with the political discourse extolling their success in monitoring and raising standards (DfE 2016). However concerns have been continually raised about the perfunctory acceptance of SATs data by government ministers and the accuracy of such data as a reflection of improvement in standards (Linn 2000; Wyse, McCreery, and Torrance 2008).

Narrowing the curriculum, narrowing pedagogies: the impact on children's learning

Since 2000, research with adult stakeholders (parents, teachers, head teachers) has emphasised the negative impact of high-stakes tests on a broad and balanced curriculum. This criticism is not unique to England, but prevalent in the USA and Australia (Berliner 2011; Minarechová 2012) and non-Anglo spheric countries (Safa and Sheykholmoluki 2023), with a consensus that the focus on summative assessment has diverted schools' attention onto core subjects (English and maths) to the detriment of the broader curriculum (Berliner 2011; House of Commons Education Committee 2017; Madaus and Russell 2011). This has been intensified by the growth of test-based accountability systems intended to improve educational standards (Lane 2020).

One study that also elicited children's perspectives alongside adult perspectives found a similar picture (Hutchings 2015). English primary school teachers reported an increase in time spent on core

subjects and children expressed a desire for more time on other subjects more useful for their futures (Hutchings 2015). One study from China that exclusively captured children's perspectives using focus group discussions, offers further insight (Cho and Chan 2020). Whilst the children in this study felt that certain subjects were more important for their future, they expressed frustration when 'major' (core) subject time spills over into 'non-major' (wider) curriculum subject time and recess. They saw this encroachment as detrimental to their mental health.

Research with adults suggests that curriculum narrowing creates educational inequalities for at-risk groups, including SEND and those with lower attainment (Ritt 2016). These groups are often withdrawn from classes and denied equal access to a breadth of curriculum compared to their peers (Berliner 2011). Research using the participatory methodology of longitudinal life history with lower attaining pupils provides key insight into the nature of this problem (Hargreaves, Quick, and Buchanan 2023). Teachers' focus on core subjects and performance blinded them to the 'passion' and 'knowledge' lower attaining children had for subjects like art, French, PE and computing – a blindness felt by the children themselves as increasingly 'constraining' (Hargreaves, Quick, and Buchanan 2023, 550).

Research with adults reveals that high-stakes testing has led to a concomitant narrowing of pedagogical breadth and creativity in favour of teaching to the test, repeated testing and instructional teaching (Au 2007; Polesel, Rice, and Dulfer 2014; Troman 2008). Research capturing both teachers' and children's experiences of pedagogy in a primary classroom illustrates how both parties understood the pedagogy of high-stakes testing as about children achieving the 'right' outcome (Hall et al. 2004) and how the development of skills, metacognition, problem-solving and collaboration were sidelined (Hutchings 2015). One study from Australia capturing children's perspectives through interviews found that children identified highly nuanced and specific changes that impacted their lived experience of school (Cranley et al. 2022). For example, changes to the physical layout of classrooms in preparation for tests, such as moving desks into rows, lessened enjoyment of lessons, causing anxiety for some and boredom for others, with learning 'less fun' due to increased use of teacher questioning for assessment purposes (Cranley et al. 2022, 1318).

Divided perspectives: impact of high-stakes testing on children's wellbeing

Most research into the impact of high-stakes testing on children's wellbeing is also adult-centric and focuses on negative impacts. This includes research which examines how high-stakes testing negatively affects children's confidence and motivation (Amrein and Berliner 2003; Black et al. 2002) and demonstrates the consequences of demotivation, disruptive behaviour, which has a detrimental effect on children's learning (Connor 2001). From the perspective of teachers in England, high-stakes testing impacts negatively upon children's mental health, with children experiencing 'physical sickness … tearfulness, interrupted sleep patterns … a loss of appetite, and over-revising' (Connor 2001, 106) – a narrative evident in interviews with primary school headteachers in England (Bradbury 2019). Teachers also felt their professional relationships with children shifted from a holistic approach to a more formal, objective-orientated approach in which the focus is attainment and results (Jeffrey 2002).

Some adult-centric research on the impacts of high-stakes testing on children's mental health takes a more nuanced view, highlighting how groups of children may be more positively or negatively affected. Webb (2006) and Beck et al. (2023) find that high-stakes testing motivates high attaining children towards performance goals, with those who fail demotivated. Tymms and Merrell (2007) and Jerrim (2021) conceptualise test-related stress, anxiety and self-esteem as bimodal: pressure becomes a challenge, leading to pride and achievement, or a threat, leading to anxiety. Similar research with adults highlights how high-stakes testing has the most negative impact on the self-esteem of lower attaining pupils (Beck et al. 2023; Webb 2006). Conversely, for motivated children, research in the USA (Clarke et al. 2003) found teachers reporting improved relationships due to an increase in quality of teaching due to high-stakes testing.

Research that includes children's and adults' perspectives provides further insight into the specific nature of these impacts. For example, children feel 'upset', 'guilty', 'ashamed', 'embarrassed', 'shaky' and 'doubtful' (McNess et al. 2001, 12), with widespread 'stress-related conditions' in children ranging from 'crying' to 'depression' and 'self-harm' (Hutchings 2015, 59). As well as providing a more nuanced understanding, research involving children's perspectives illuminates why they experience these feelings. In England, Ofqual's (the Office of Qualifications and Examinations Regulation) study concluded 'text anxiety' is caused by a 'disparity' between the child's belief of what they can achieve and expectations of performance (Howard 2020, 40). This disparity can be both caused by intra- and interpersonal factors as 'beliefs can be set by the students, but can also be influenced by parents, peers and teachers' (Howard 2020, 40). Longitudinal research by Connors et al. (2009, 9) captures 'a distinct temporal dimension' to SATs-related anxiety, occurring at different times for individual children and lasting for a variety of durations.

For high attaining children, child-centric research shows high-stakes testing can be motivating, resulting in 'the production of amenable pupil identities' (Hall et al. 2004, 806). Conversely, similar research identifies the negative impacts high-stakes testing has on disadvantaged student groups, including those from socially disadvantaged backgrounds, with Special Educational Needs and Disabilities (SEND) and whose attainment is low (Hall et al. 2004; Howard 2020; Hutchings 2015; McNess et al. 2001; Putwain et al. 2012). Accordingly, high-stakes testing has been shown to have the most negative impact on the self-esteem of lower attaining pupils (Howard 2020). Harlen and Deakin Crick (2002) found strong evidence that after the introduction of SATs in England, low attaining pupils had lower self-esteem than high attaining pupils. This is problematic as children form negative perceptions of their capabilities, which carry on into secondary schooling and beyond (Hutchings 2015).

A few studies focussed solely on lived experiences of children in relation to the impacts of high-stakes testing on their wellbeing, although only one (Cho and Chan 2020) took a more participatory approach. In terms of differences between student groups, whilst research with children reported children feeling positive and motivated when they achieved well in tests and demotivated when they did not, interviews with primary school children in England demonstrated that a causal relationship between motivation and failure is not simple as 'failure creates dislike, rationalised as lack of interest, which accounts for lack of effort' (Pollard et al. 2000, 145). From the same data, high attaining children feel increasingly confident and motivated; low attaining, vulnerable learners or those with SEND become disengaged with testing, reluctant to go to school during the test period (Pollard et al. 2000). This complexity is furthered in the research of Murphy et al. (2013) where interviews with children show how they can simultaneously hold seemingly contradictory positions, acknowledging the negative impacts upon their wellbeing whilst having positive attitudes towards science and its assessment. Cho and Chan's (2020) analysis of children's discussions explores how relations can be supportive and corrosive in terms of wellbeing, peers providing comfort and competition for children and families either placating or compounding stress through conflict.

Research questions

Our review demonstrates that research into the impacts of high-stakes testing on children is often adult-centric. Where researchers seek to capture children's voices using participatory methods – Cho and Chan's (2020) use of focus group discussion (2020) and Hargreaves, Quick, and Buchanan's (2023) use of life history – a more nuanced and sophisticated understanding of the impacts of high-stakes testing is afforded. This is because participatory research holds the potential to amplify children's voices, affording a deeper understanding of the impacts of high-stakes testing.

Accordingly, we have two research questions:

1. How do 10- and 11-year-old children experience the impacts of high-stakes testing?
2. How might the use of participatory methods with children deepen our understanding of how children experience high-stakes testing?

Methodology

A participatory approach

Wheldon sought to involve children as participants to understand their experiences and perceptions of high-stakes testing. Based on the idea that children 'hold deep knowledge about their lives and experiences' (Torre 2009, 111), participatory methods were used to place 'students at the centre of the process, viewing them as active participants rather than as subjects that will give us information' (Messiou 2012, 29). To illuminate the extent to which children were active participants, we draw upon Mayne, Howitt, and Rennie's (2018) Accountability Framework (AF). The AF consists of three interrelated domains: research valuing the 'rights' of children; the 'power' relations between researchers and children; and the 'culture' established by the researchers. In assessing this project design against the AF, our research accords with Article 12 of the United Nations Convention of the Rights of the Child (UN 1989). In terms of 'rights', the methods used mean the children are 'social actors' in the research rather than 'objects' or 'subjects'. The children fall short of being 'co-researchers', however, because the research questions were decided by Wheldon. In terms of 'power', this is research 'with' children rather than 'on' children. In line with the point above about co-researchers, the research falls short of being 'by' children. In terms of 'culture', the project is 'respectful' in promoting participation through inclusive methods.

Ethics

The research was conducted in line with the ethical principles set out by the British Educational Research Association (BERA 2024) and received institutional ethical clearance. This included: gate-keeper informed consent; parent and carer informed consent; and children's assent. In line with our participatory approach, assent was negotiated throughout the project through an 'interactive narrative approach' (Mayne, Howitt, and Rennie 2016, 8). When working with children, this approach improves the quality of informing as well as the quality of the participation of the children. This helped to create a respectful research culture and improved the quality of the research.

Context and participant selection

The research took place in a Roman Catholic primary school in North Yorkshire with below average numbers of SEND pupils and a slightly higher than average socio-economic intake. Wheldon was employed at the school, had taught the children previously, and was familiar to the setting and participants. At the time of data collection, 15 Year 6 children were in a mixed Year 5/6 class. Twelve met the criteria to take part in the research (i.e. would be sitting SATs). In consultation with the class teacher, it was decided that there were no individuals who were distinctively vulnerable to harm. Pseudonyms were assigned to ensure anonymity (see Table 1).

Table 1. Participant information.

Participant name	Gender	Prior academic attainment
Ben	Male	Middle attaining
Celia	Female	High attaining
Emmy	Female	Middle attaining
Harriet	Female	Middle attaining
Jake	Male	Middle attaining
Jasmine	Female	Middle attaining
Jasper	Male	Low attaining
Pippa	Female	Middle attaining
Sally	Female	High attaining
Stella	Female	High attaining
Vicky	Female	Low attaining – (SEND)
Zena	Female	High attaining

Data generation

Participants were randomly organised into two focus groups and data generation took place during two, 1-hour sessions – one before and one after SATs in May. Focus groups were used to explore how participants felt about a particular issue (Bell and Waters 2018; Bryman 2008). Focus groups were expedient with an interactive, dialogic approach, respecting children's experience and their contextualised understanding of the issue (Mayne, Howitt, and Rennie 2018). A range of participatory methods were used within the focus groups to include all children as outlined in Table 2.

The sessions were audio and video recorded and subsequently transcribed. The audio recordings of the focus group sessions contained ambiguities, overlapping conversations or inarticulate responses which could be clarified by seeing who had spoken, or by lip reading what was said. The video recordings were particularly useful for gathering 'non-verbal and verbal data' (McLafferty 2004, 191) and supported the writing of field notes after each session. Data in the form of images and words from the artefacts themselves were also analysed.

Body mapping: Children drew outlines of their bodies onto large pieces of paper and mapped their positive thoughts and feelings about SATs onto one half and their negative ones onto the other. Body mapping was used as the principal method of data collection in the first session as it suited the research question and supported our participatory design (O'Kane 2017) by offering participants a 'different way into' the research question (Gauntlett and Holzwarth 2006, 84). The creation of their body maps enabled children to participate individually and creatively: different parts of the body could be used as prompts to explore their individual feelings and experiences, acknowledging their 'multiple layers of understanding' (Mayne, Howitt, and Rennie 2018, 5). At the beginning of the second focus group session, the children were invited to revisit their body maps and re-engage with the research question and their previous responses.

SATs pressure chart: Children indicated sources of SATs pressure on a chart, placing stars against groups of people who were seen as a cause of pressure: 'Myself'; 'My peers'; 'Parents'; 'Brothers and sisters'; 'Teachers'. The children understood that the categories on the chart were taken from wider research findings into sources of pressure on children due to high-stakes testing, and that their responses would be situated within the wider context. This aligns with the AF by supporting children's rights and agency in research.

Timelines: To elicit data relating to the temporal nature of SATs, children created timelines of their 'journey', identifying significant events, both positive and negative, and adding information about what helped them. Minimal constraints were placed in order to maximise engagement and participation and to give agency and control to participants, allowing them to organise individual narratives. The artefacts were varied, rich in metaphor and individually creative. For example, some children chose to use emoticons – a language with which they were comfortable and familiar – to represent their varied feelings throughout the SATs journey.

Data analysis

Initially, a broad-brush approach to coding was taken using NVivo software, with emerging categories 'close' to the data (Charmaz 2006; Cohen, Manion, and Morrison 2018). The body maps and timelines proved a rich source of data, particularly for quieter children. Elements of Visual Discourse Analysis (VDA) (Albert 2009) were applied at various points during data analysis, taking a semiotic approach to the relationship between signs and symbols and their spoken or written

Table 2. Participatory methods used in focus groups.

Focus group session 1 (before SATs)	Focus group session 2 (after SATs)
Body mapping	Body maps revisited SATs pressure chart Timelines

language. This technique was particularly fruitful during the analysis of timeline transcripts in conjunction with related artefacts. The vectors that children drew carried meaning (Albert 2009) and many children chose to draw the line as a representation of the highs and lows in their narrative. Similarly, children drew sharks and whales lurking near the seabed during SATs. The bottom half of a picture often 'feels more threatened, heavier, sadder and constrained' (Albert 2009, 11). Detailed memos were written on each theme to document, reflect on and analyse the patterns that were emerging and 'build a foundation for theory development' (Saldana 2013, 252).

Due to the large amount of data and codes, it was decided to focus on those which were the most frequent, relevant to the research questions and to the literature. In this way, 'significant' categories were identified and aggregated into broad themes, with 'peripheral' categories put aside. Throughout the coding process, patterns were identified in relation to similarity, difference, frequency, sequence, correspondence, and causation (Saldana 2013, 6–7). Six broad themes emerged from the initial coding process:

1. Curriculum and pedagogy
2. Anxiety, stress and pressure
3. Motivation and equity
4. Self-esteem and identity
5. Relationships
6. Importance and purpose of SATs.

The final stages of analysis involved stepping back and revisiting the data sources to ensure that the coding and analytical process had comprehensively addressed the research questions and had been 'rigorously ethical' (Saldana 2013). This stage involved positioning the findings in relation to existing research and the wider theoretical frame of neoliberalism. This meant taking time to reflect on the data and their emergent patterns and themes, analysing them against the existing research field. Accordingly, the six original themes were reduced to four. The original theme relating to 'Self-esteem and identity' overlapped with the themes 'Relationships' and 'Anxiety stress and pressure' and we merged these themes. Similarly, 'The importance and purpose of SATs' was merged with 'Curriculum and pedagogy', and 'Motivation and equity' took on a more critical dimension by focussing on lower achieving pupils. Our final four themes are discussed below.

Discussion and findings

'Practice papers every day!' Curriculum and pedagogical narrowing due to high-stakes testing

A clear theme was the reductive effect of SATs preparation on the content and timetabling of the curriculum. Children described the suspension of usual lessons due to increased test practice. Pippa, for example, explained how, 'Before we used to do lessons and things like that … and we used to have fun. We used to do PE. But we never really did that when we had SATs it was literally every day, test test test'; Harriet testified, 'We haven't done an art lesson, I think, since … Christmas'. Some children felt that SATs preparation had begun as soon as they entered Year 6 whilst all children identified curriculum narrowing from the beginning of the second term. There was evidence that this increased sharply as SATs approached: Zena reported, '[after] February half term we start doing tests every day and get loads of homework'; Vicky commented, 'The week before SATs … we are [practising] like there is no tomorrow'; Sally summarised, 'Practice papers. Every day!' Evidence from timelines indicated there was a return to 'normal' after SATs week, although children described a post-SATs curriculum that was different again, dominated by activities there had not been time for such as cycle training, end of year performance and a residential. These findings mirror the international context of curricula distortion towards a narrow set of academic subjects deemed to be the most

important in high-stakes regimes (Berliner 2011; Cho and Chan 2020; Safa and Sheykholmoluki 2023) – a consequence of the neoliberal performance mechanism described by Ball (2016). The children were clear and unequivocal about the effect *on them* of the cessation of curriculum breadth during the weeks and months leading up to Year 6 SATs.

Data generated from the focus group added nuance to adult-centric research, which suggests an increase of teacher directed pedagogy in response to high-stakes testing (Au 2007; Berliner 2011). The children certainly perceived that SATs preparation impacted negatively on learning. Harriet expressed, 'I want to do more practical work but … we're too busy about SATs and we don't get to do much hands-on practical work and things'. Children reflected critically on their pedagogical experiences; revealing a distinct and sophisticated understanding. Zena clearly discerned the gap in learning potential between 'normal' and 'practice' lessons: 'It's changed the way we learn. Instead of lessons we do tests'. Ben's perspective adds further insight: 'Mostly in Year 6 you don't learn 'cause you've already done it all … you're kind of enforcing'. Children commented that over time they had become accustomed to this new 'rote' learning. As Stella pointed out, 'We've kind of got used to it now because we do it week in week out, but when it first happened it felt so different'. This dichotomy of compliance and criticality adds further insight to Hall et al.'s notion of children's 'amenable identities' (2004, 806).

Data from body maps revealed positive perceptions of aspects of SATs preparation on learning. Ben annotated his body map brain with, 'I think faster now' and Vicky claimed, 'It help [me] to concentrate'. Emmy commented on her increased knowledge, although her concept of 'knowledge' appeared limited to an increase in test scores: 'I know that my knowledge is getting bigger … from the beginning of autumn I've improved by quite a lot and every week I can see my scores getting higher and that's the thing that makes me happy'. Jake reported an increase in his knowledge whilst also recognising the distinctive dimension of this as 'test' knowledge: 'It's like, really helped my … knowledge. I don't think we'd have learned all these things that maybe aren't necessary but we have to learn in the SATs'. In keeping with Hall et al. (2004), pedagogy appeared to be largely limited to teacher-transmitted knowledge with few opportunities for creative learning or construction of meaning through varied pedagogies. Stella's response illustrates that even a lesson on 'problem-solving' was focused primarily on passing tests: 'Mr G has made a club for … problem solving … 'cause like problem solving's normally the two markers so to make us get the extra two marks'.

Pupils' experience of anxiety and stress: a complex phenomenon

A complex picture emerged of children's individual experience of anxiety, adding to the argument that test anxiety is a multidimensional construct (Jerrim 2021; Putwain et al. 2012). Anxiety about SATs was a dominant theme across the data generated in focus groups with all children reporting some degree of anxiety, feeling 'nervous', 'stressed', 'scared' or 'worried'. However, whilst some children expressed high levels of anxiety, others appeared unaffected: Jake for example, asserted, 'I wanna pass them but I'm not really bothered … it hasn't really changed the way I act'. Children reported physical symptoms: butterflies were a common theme on the children's body maps, as were chewed nails and sleeves. They also described sweating, dizziness, tiredness, lack of focus, feeling sick and upset, being restless, having a 'quicker pace' and fidgeting more. Data collected indicated there was a temporal dimension to anxiety; on timelines, many children drew a dip in the line or supplemented their written comments with sad or straight-faced emoticons to represent anxiety or pressure as SATs approached and in SATs week itself. In line with research by Connors et al. (2009), there was significant variation amongst the children as to when they began to experience anxious feelings. Sally noted, 'Brief worrying on SATs [at the start of Year 6] as Mr G had already started talking about them'. Stella's experience was, 'Just kind of the whole build up kind of makes you scared'. For Harriet, the process started in Year 5 when she began 'thinking about SATs next year'. Observation data revealed a distinct difference in the children's general demeanour in the focus

groups taking place after SATs; they were more relaxed, livelier and there was a palpable sense of 'release'. This was also represented on timelines with an upward trajectory, steep spikes pointing upwards, positive emoticons and by using different colours for 'before' and 'after' SATs. After SATs, Jake described, 'Chilling out and waiting for our results (happy face)' and Emmy described the difference as, 'I feel like I've changed like a hundred percent … I think it's just the fact that actually SATs are over now'.

According to Wigfield and Eccles (1989, 161), *emotionality*, i.e. 'the more physiological/affective part of anxiety', is less harmful than *worry,* i.e. 'the cognitive component of anxiety'. Similarly, in our research there were fewer than half as many references to 'stress' and 'worry' across the data than references to 'feeling nervous', suggesting children's feelings of anxiety may have been more aligned with emotionality rather than worry. This could be interpreted as a 'healthy' or 'normal' response to SATs rather than a chronic, damaging cognitive reaction. However, other children did experience more pernicious effects of worry and stress, such as those described above. This is in tension with large scale, adult-centric research (Jerrim 2021), which suggests there is scant evidence to support a direct correlation between high-stakes tests and children's 'socio-emotional outcomes'.

Furthermore, there was evidence of a bimodal conceptualisation of the impact of high-stakes tests, with qualitative data offering nuanced insight into the complexities of how individual children experience test-related anxiety. Several children described simultaneous feelings of 'nerves' and 'excitement' and for some, experiencing 'butterflies' was negative and positive. The data also offered further nuances into how individual children's innate beliefs about themselves might affect their experience of test anxiety. Jasmine, who was very anxious about SATs, appeared to demonstrate a fixed view of her ability, 'It can boost confidence if your scores get better, but then if it goes down you could lose your confidence and think well I'm going to fail my SATs so I might as well not try'; whereas Jasper was more optimistic about improvements, '1 week until the SATs (smiley face on timeline) not really that nervous because I been doing better'.

'It can make you depressed if your friends get higher'; high-stakes testing shaping relationships with peers

A key reason identified by most of the children for high-stakes testing impacting negatively upon their mental health and wellbeing was the pressure they felt from peers and how competition resulting from high-stakes testing shaped peer relationships. In terms of pressure, this was evident in the chart, with the majority of children indicating they felt pressure at some point during the build-up to SATs, 'my peers' accounting for 50% of the overall causes ('my peers' = 46 stars; 'myself' = 18 stars; 'siblings' = 14 stars; 'teachers' = 8 stars; 'parents' = 7 stars). This emphasis on peers causing pressure is different from research involving adults, which tends to identify teachers and parents as the primary sources (Hall et al. 2004; Jeffrey 2002). Whilst anxiety has been attributed to relational causes in adult-centric research, including parents, teachers and peers (Howard 2020), how peer relationships are shaped by high-stakes testing has not been explored in detail.

Children in this study talked about pressure from their peers, predominantly in the form of competitive culture. Zena spoke of how she would compare her results with peers and how this would make her feel when the comparison was negative: 'It made a difference what everyone else got to how you thought you did. I think there was a bit of pressure … you don't want to feel that you're the odd one out if you do horribly and the rest do amazing'. For Emmy, the competition with peers had an affective dimension: 'When I get a good score my friends get even better and it can sometimes make me upset … Sometimes, like, I'm not as good as them so I can't be up there'. This affective dimension that was more strongly felt by Jasmine: '[It] can make you upset or depressed if your friends get higher'. The most extreme response came from Harriet, for whom the negative effects of competition were so great she changed her friendship group: 'Sometimes … at break times I

try to play with new people that … are maybe younger than me so that I have someone else that … doesn't know about SATs yet so that they won't keep on bringing it up and things'.

These negative effects can be seen as symptomatic of neoliberalism's mechanism of performance (Ball 2016) – 'a regime of accountability that employs judgements [and] comparisons' (Ball 2013, 57). What is unusual is that the negative impacts of neoliberalism on education typically focus on teachers rather than children (Ball 2013). This insight into the ways in which high-stakes testing as a mechanism of performance impacts upon children indicates the reach of neoliberalism and its worrying impact on children's mental health, wellbeing and relationships.

In line with the discussion of theme 2 above, the effects of competition presented as bimodal (Putwain et al. 2012), with some children feeling the positive benefits of competition as motivation for improvement. Jake claimed, 'I wanted to beat my friends but then I always wanted them to, like, get a good score'. Bimodal effects of competition with peers occurred within individual children at different points in time. Emmy, who had spoken of being 'upset' because of gaining a lower score than her friends, later felt, 'That bit of competitiveness gives you adrenalin … because you want to do well. So you try and work harder'. She felt that when her results compared favourably with her friends, this had a positive effect on her self-esteem: 'Can I just say Celia and Zena are very clever and when you beat them you feel good'.

As well as competition being divisive and negatively felt and holding the potential to be motivating and positively felt, some children indicated that high-stakes testing brought them closer, providing a shared sense of identity as they experienced the tests together. For Sally this meant, 'You're going through the SATs together. It's not like you're doing them by yourself'. Vicky used the personal pronoun 'we' to indicate a collective endeavour: 'SATs are over. We survived it. We got it [done]'. This complex picture of relationships, with friends being damaged and strengthened by high-stakes testing, is also identified in Cho and Chan's (2020) focus group with children, where children identified how their friends provided comfort and competition throughout the test period.

An apartheid system of schooling? The damaging effects of high-stakes testing upon lower achieving children

Whilst theme 3 adds to the under-researched area of the impact of high-stakes testing on peer relationships in schools, there is already strong evidence from more adult-centric research and research involving children that high-stakes testing negatively impacts upon lower attaining children (Beck et al. 2023; Harlen and Deakin Crick 2002; Howard 2020; Hutchings 2015; Webb 2006). Despite there only being two lower achieving children in our study, our analysis evidences a complex picture in terms of the effects of high-stakes testing on these children. In relation to anxiety, no clear pattern was evident between achievement and anxiety, although most (but not all) children who said they felt excited by the tests were higher attainers. Visual analysis of the children's timelines also revealed an arbitrary picture. Sally, who was one of the highest attaining children but one of the most anxious, drew a timeline with a steep downward trajectory during SATs, plunging to the bottom of the sea where a killer whale and a shark with sharp teeth in an open mouth lurked, indicating threat and oppression. Jasper, the lowest attaining boy in the group, did 'get butterflies sometimes in my stomach 'cause I feel nervous' but he also drew a straight arrow timeline with only one reference to feeling nervous during the SATs build-up, accompanied by a straight-face emoticon. Vicky, the lowest attaining girl, said, 'Sometimes I hear good things about SATs but sometimes bad and that makes me worried'.

Closer analysis offers an explanation for Jasper and Vicky's perspectives. It has long been argued that any attempt to identify patterns between test anxiety and children's ability needs to consider children's inner beliefs about whether they feel their ability is fixed or mutable (Wigfield and Eccles 1989). Some higher achieving children, like Jasmine, were very anxious about SATs and appeared to demonstrate a fixed view of their ability. Vicky, on the other hand, spoke of personal growth, 'something about the good is … I can see me getting better and better each day'. Similarly,

Jasper's timeline was a narrative of growth: '1 week until the SATs (smiley face) not really that nervous because I been doing better'. In line with having a growth mindset, Jasper reflected upon his resilience when he found the tests difficult: 'Whenever I know I've tried my hardest I always feel really proud of myself that … I didn't give up on questions'.

From a social perspective, observations at the start and end of the focus group suggested that Vicky, the lowest attaining girl and the only SEND child, separated herself from the rest of the class in lessons by working at her own table. Despite attesting to feeling part of a collective in thinking about a shared experience of the tests, an argument could be made that disparity in attainment between herself and peers fostered social separation in Vicky, which in turn led her to distance herself physically from the group when they were discussing the impact of the tests. This is reflected in Vicky drawing a perfect 'poop' emoticon on her timeline, accompanied by the unambiguous statement: 'I think that sats are rubbish'! Vicky's embodied experienced of high-stakes testing supports the research which suggests high-stakes testing promotes an 'apartheid system of schooling' (Berliner 2011, 292).

Conclusion

Our participatory approach to research into children's experiences of high-stakes testing offers new insights in comparison to previous, adult-centric research. Analysis of the data generated shows how the impact of high-stakes testing on curriculum, pedagogy and wellbeing is experienced and understood by children. In a profound way, the children in this study understood the neoliberal mechanisms of management and performance (Ball 2016): how these mechanisms directly narrowed the curriculum at specific moments in time; how this affected pedagogy, with more transmissive approaches taken; and how this affected their own mental health, especially in relation to the temporal nature of the anxiety they experienced. The effects of neoliberalism in relation to the third mechanism, the free market, are often conceptualised in terms of competition between school and teachers (Ball 2013). Our research identifies a potentially more worrying development – how children become competitive with peers and how this shapes and damages relationships. Whilst our research project is small in terms of sample and short in terms of timeframe, children holding fixed mindsets appear to be vulnerable to competition with their peers in the context of high-stakes testing. There is also some evidence to suggest that lower attaining children, including those with SEND, are most vulnerable to this dynamic, as they become socially excluded.

These new insights gained by involving children in research into high-stakes testing using participatory methods lead us to making three key recommendations. First, in the context of a mental health crisis (WHO 2021), as well as a need for pedagogical variation to develop children's skills and competencies (OECD 2020), more participatory research into the impacts of high-stakes testing with children will be invaluable. Second, policy makers taking heed of this research with children should consider reform of both curricular and testing to foster children's development of skills and competencies and positive wellbeing. And third, as long as high-stakes testing remains in primary schools, school leaders and teachers should draw upon their children's implicit understanding of neoliberal mechanisms to engage in open and explicit discussions about high-stakes testing. By doing so, children's emotional literacy will be developed in such a way as to enable them to navigate the negative effects of competition with their peers, confront any tendencies towards fixed mindsets, and mitigate the potential social isolation of lower attaining children.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Joanna Wheldon http://orcid.org/0009-0004-0365-3418
Tom Dobson http://orcid.org/0000-0001-5354-9150

References

Albert, P. 2009. "Reading Students' Visual Texts Created In English Language Arts Classrooms." *Language Arts Journal of Michigan* 25 (1): 6–16. https://scholarworks.gvsu.edu/lajm/vol25/iss1/3/.

Amrein, A. L., and D. C. Berliner. 2003. "The Effects of High-Stakes Testing on Student Motivation and Learning." *Educational Leadership* 60 (5): 32–38. http://www.wou.edu/~girodm/611/testing_and_motivation.pdf.

Au, W. 2007. "High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis." *Educational Researcher* 36 (5): 258–267. https://doi.org/10.3102/0013189X07306523.

Ball, S. J. 2013. *The Education Debate (2nd Ed.)*. Bristol: The Policy Press.

Ball, S. J. 2016. "Neoliberal Education? Confronting the Slouching Beast." *Policy Futures in Education* 14 (8): 1046–1059. https://doi.org/10.1177/1478210316664259.

Beck, K. C., H. L. Røhr, B. Reme, and M. Flatø. 2024. "Distressing Testing: A Propensity Score Analysis of High-Stakes Exam Failure and Mental Health." *Child Development* 95 (1): 242–260. https://doi.org/10.1111/cdev.13985.

Bell, J., and S. Waters. 2018. *Doing Your Research Project: A Guide for First-Time Researchers. 7th Ed.* London: Open University Press.

BERA. 2024. "Ethical Guidelines for Educational Research – 5th Edition." https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-fifth-edition-2024.

Berliner, D. 2011. "Rational Responses to High Stakes Testing: The Case of Curriculum Narrowing and the Harm That Follows." *Cambridge Journal of Education* 41 (3): 287–302. https://doi.org/10.1080/0305764X.2011.607151.

Black, P., P. Broadfoot, R. Daugherty, J. Gardner, W. Harlen, M. James, G. Stobart, and D. Wiliam. 2002. *Testing, Motivation and Learning*. Cambridge: Nuffield Foundation and University of Cambridge.

Bradbury, A. 2019. *Pressure, Anxiety And Collateral Damage: The Headteachers' Verdict*. London: UCL Press.

Bryman, A. 2008. *Social Research Methods (3rd Ed.)*. Oxford: Oxford University Press.

Charmaz, K. 2006. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. London: SAGE Publications.

Cho, E., and T. Chan. 2020. "Children's Wellbeing In A High-Stakes Testing Environment: The Case of Hong Kong." *Children and Youth Services Review* 109:104694. https://doi.org/10.1016/j.childyouth.2019.104694.

Clarke, M., A. Shore, K. Rhoades, L. Abrams, J. Miao, and J. Li. 2003. *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from Interviews with Educators in Low-, Medium- and High-Stakes States*. Boston: National Board on Educational Testing and Public Policy.

Cohen, L., L. Manion, and K. Morrison. 2018. *Research Methods In Education (8th Ed.)*. London: Routledge.

Connor, M. J. 2001. "Pupil Stress and Standard Assessment Tasks." *Emotional and Behavioural Difficulties* 6 (2): 103–111. https://doi.org/10.1080/13632750100507660.

Connors, L., D. Putwain, K. Woods, and L. Nicholson. 2009. "Causes and Consequences of Test Anxiety in Key Stage 2 Pupils: The Mediational Role of Emotional Resilience." In *British Educational Research Association Conference: Examination Anxiety in Primary, Secondary and Sixth Form Students*, 2–5 September 2009, Manchester.

Cranley, L., C. Robinson, G. Hine, and D. O'Connor. 2022. "The Desks Have Changed; It Must Be Naplan Time': How Naplan Affects Teaching and Learning of Mathematics." *Issues in Educational Research* 32 (4): 1306–1320.

Curtis, P. 2008. Sats for 14-Year-Olds Are Scrapped. *The Guardian*. https://www.theguardian.com/education/2008/oct/14/sats-scrapped.

DES. 1985. "Better Schools: a Summary." http://www.educationengland.org.uk/documents/des/betterschools.html.

DfE. 2016. *Educational Excellence Everywhere*. London: The Stationery Office.

Gauntlett, D., and P. Holzwarth. 2006. "Creative And Visual Methods For Exploring Identities." *Visual Studies* 21 (1): 82–91. https://doi.org/10.1080/14725860600613261.

GOV.UK. 2023. "Schools, Pupils and Their Characteristics – Academic Year 2022/23." https://explore-education-statistics.service.gov.uk.

Hall, K., J. Collins, S. Benjamin, M. Nind, and K. Sheehy. 2004. "Saturated Models of Pupildom: Assessment and Inclusion/Exclusion." *British Educational Research Journal* 30 (6): 801–817. https://doi.org/10.1080/0141192042000279512.

Hargreaves, H., L. Quick, and D. Buchanan. 2023. "National Curriculum and Assessment in England and the Continuing Narrowed Experiences of Lower-Attainers In Primary Schools." *Journal of Curriculum Studies* 55 (5): 545–561. https://doi.org/10.1080/00220272.2023.2253455.

Harlen, W., and R. Deakin Crick. 2002. *A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation For Learning*. London: EPPI-Centre, Institute of Education, University of London. https://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=Pbyl1CdsDJU%3D&tabid=108&mid=1003.

Högberg, Björn, and Daniel Horn. 2022. "National High-Stakes Testing, Gender, and School Stress in Europe: A Difference-In-Differences Analysis." *European Sociological Review* 38 (6): 975–987. https://doi.org/10.1093/esr/jcac009.

Hossain, M. Mahbub, Fazilatun Nesa, Jyoti Das, Roaa Aggad, Samia Tasnim, Mohan Bairwa, Ping Ma, and Gilbert Ramirez. 2022. "Global Burden of Mental Health Problems Among Children And Adolescents During Covid-19 Pandemic: An Umbrella Review." *Psychiatry Research* 317:114814. https://doi.org/10.1016/j.psychres.2022.114814.

House of Commons Education Committee. 2017. "Primary Assessment: Eleventh Report of Session 2016–17." https://committees.parliament.uk/committee/203/education-committee/publications/reports-responses/?SearchTerm=assessment&DateFrom=&DateTo=&SessionId.

Howard, E. 2020. *A Review of the Literature Concerning Anxiety for Educational Assessments*. Coventry: Ofqual. https://assets.publishing.service.gov.uk/media/5e45825340f0b677bf6eb3ea/A_review_of_the_literature_concerning_anxiety_for_educational_assessment.pdf.

Howell, Angelique. 2017. "'Because Then You Could Never Ever Get A Job!': Children's Constructions of Naplan as High-Stakes." *Journal of Education Policy* 32 (5): 564–587. https://doi.org/10.1080/02680939.2017.1305451.

Hutchings, M. 2015. *Exam Factories: The Impact of Accountability Measures on Children and Young People*. London: National Union of Teachers.

Jeffrey, B. 2002. "Performativity and Primary Teacher Relations." *Journal of Education Policy* 17 (5): 531–546. https://doi.org/10.1080/02680930210158302.

Jerrim, J. 2021. "National Tests and the Wellbeing of Primary School Pupils: New Evidence from the UK." *Assessment in Education: Principles, Policy & Practice* 28 (5–6): 507–544. https://doi.org/10.1080/0969594X.2021.1929829.

Lane, Suzanne. 2020. "Test-Based Accountability Systems: The Importance of Paying Attention to Consequences." *ETS Research Report Series* 2020 (1): 1–22. https://doi.org/10.1002/ets2.12283.

Linn, R. L. 2000. "Assessments and Accountability." *Educational Researcher* 29 (2): 4–16. https://doi.org/10.3102/0013189X029002004.

Madaus, George, and Michael Russell. 2010. "Paradoxes of High-Stakes Testing." *Journal of Education* 190 (1–2): 21–30. https://doi.org/10.1177/0022057410190001-205.

Mayne, F., C. Howitt, and L. Rennie. 2016. "Meaningful Informed Consent with Young Children: Looking Forward Through an Interactive Narrative Approach." *Early Child Development and Care* 186 (5): 673–687. https://doi.org/10.1080/03004430.2015.1051975.

Mayne, F., C. Howitt, and L. Rennie. 2018. "Rights, Power and Agency in Early Childhood Research Design: Developing A Rights-Based Research Ethics and Participation Planning Framework." *Australasian Journal of Early Childhood* 43 (3): 4–14. https://doi.org/10.23965/AJEC.43.3.01.

McLafferty, I. 2004. "Focus Group Interviews as a Data Collecting Strategy." *Journal of Advanced Nursing* 48 (2): 187–194. https://doi.org/10.1111/j.1365-2648.2004.03186.x.

McNess, E., P. Triggs, P. Broadfoot, M. Osborn, and A. Pollard. 2001. "The Changing Nature of Assessment in English Primary Classrooms: Findings from the Pace Project 1989–1997." *Education 3-13* 29 (3): 9–16. https://doi.org/10.1080/03004270185200301.

Messiou, K. 2012. *Confronting Marginalisation In Education: A Framework for Promoting Inclusion*. Oxon: Routledge.

Minarechová, Michaela. 2012. "Negative Impacts of High-Stakes Testing." *Journal of Pedagogy/Pedagogický casopis* 3 (1): 82–100. https://doi.org/10.2478/v10159-012-0004-x.

Murphy, C., L. Lundy, L. Emerson, and K. Kerr. 2013. "Children's Perceptions of Primary Science Assessment in England and Wales." *British Educational Research Journal* 39 (3): 585–606. https://doi.org/10.1080/01411926.2012.674921.

OECD. 2020. *PISA 2018 Results (Volume 6): Are Students Ready to Thrive in an Interconnected World?* Paris: PISA OECD Publishing.

O'Kane, C. 2017. "Participatory Research On Kinship Care In East Africa." In *Research with Children: Perspectives and Practices*, edited by P. Christensen, and A. James, 180–202. Oxon: Routledge.

Polesel, J., S. Rice, and N. Dulfer. 2014. "The Impact of High-Stakes Testing on Curriculum and Pedagogy: A Teacher Perspective from Australia." *Journal of Education Policy* 29 (5): 640–657. https://doi.org/10.1080/02680939.2013.865082.

Pollard, A., P. Triggs, P. Broadfoot, E. McNess, and M. Osborn. 2000. *What Pupils Say: Changing Policy and Practice in Primary Education*. London: Continuum.

Punch, S. 2002. "Research with Children: The Same or Different from Research with Adults?" *Childhood (Copenhagen, Denmark)* 9 (3): 321–341.

Putwain, D. W., L. Connors, K. Woods, and L. J. Nicholson. 2012. "Stress and Anxiety Surrounding Forthcoming Standard Assessment Tests in English Schoolchildren." *Pastoral Care in Education* 30 (4): 289–302. https://doi.org/10.1080/02643944.2012.688063.

Ritt, M. 2016. The Impact of High-Stakes Testing on the Learning Environment. Retrieved from Sophia, the St. Catherine University repository website: https://sophia.stkate.edu/msw_papers/658.

Safa, M. A., and H. Sheykholmoluki. 2023. "The Impacts of a Nationwide High-Stakes Test from High School Teachers and Principles' Perspectives: A Qualitative Study." *International Journal of Language Testing* 13 (1): 104–132. https://languagetestingasia.springeropen.com/articles/10.1186s40468-023-00254-0.

Saldana, J. 2013. *The Coding Manual for Qualitative Researchers (2nd Ed.)*. London: SAGE Publications Ltd.
Tett, L., and M. Hamilton. 2021. *Resisting Neoliberalism In Education – Local, National And Transnational Perspectives*. London: Palgrave.
Torre, M. E. 2009. "Participatory Action Research and Critical Race Theory: Fueling Spaces for Nos-Otras to Research." *The Urban Review* 41 (1): 106–120. https://doi.org/10.1007/s11256-008-0097-7.
Troman, G. 2008. "Primary Teacher Identity, Commitment and Career in Performative School Cultures." *British Educational Research Journal* 34 (5): 619–633. https://doi.org/10.1080/01411920802223925.
Tymms, P., and C. Merrell. 2007. *Standards and Quality in English Primary Schools Over Time: The National Evidence (Primary Review Research Survey 4/1)*. Cambridge: The University of Cambridge Faculty of Education.
UN. 1989. *Adoption of a Convention on the Rights of the Child*. Geneva: United Nations. https://downloads.unicef.org.uk/wp-content/uploads/2010/05/UNCRC_united_nations_convention_on_the_rights_of_the_child.pdf.
Webb, R. 2006. "Teachers' Perspectives on Teaching and Learning in a Performativity Culture." In *Changing Teaching and Learning in the Primary School*, edited by R. Webb, 33–46. Maidenhead: Open University Press.
WHO. 2021. "Mental Health of Adolescents." https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health.
Wigfield, A., and J. S. Eccles. 1989. "Test Anxiety in Elementary and Secondary School Students." *Educational Psychologist* 24 (2): 159–183. https://doi.org/10.1207/s15326985ep2402_3.
Wyse, D., E. McCreery, and H. Torrance. 2008. *The Trajectory and Impact of National Reform: Curriculum and Assessment in English Primary Schools (Primary Review Research Survey 3/2)*. Cambridge: The University of Cambridge Faculty of Education.