

Article

“Which Child Left Behind”: Historical Issues Regarding Equity in Science Assessment

Peter J. Fensham ^{1,*} and J. Joy Cumming ²

¹ Monash University and Queensland University of Technology, Kelvin Grove campus, Victoria Park Rd, Kelvin Grove, QLD, 4059, Australia

² Griffith Institute for Educational Research, Mount Gravatt campus, Griffith University, 176 Messines Ridge Rd, Mt Gravatt, QLD 4122, Australia; E-Mail: J.Cumming@griffith.edu.au

* Author to whom correspondence should be addressed; E-Mail: p.fensham@qut.edu.au; Tel. +61-7-3138 3338, Fax. +61-7-3138-3985.

Received: 22 April 2013; in revised form: 24 July 2013 / Accepted: 24 July 2013 /

Published: 7 August 2013

Abstract: Assessment of learning plays a dominant role in formal education in the forms of determining features of curriculum that are emphasized, pedagogic methods that teachers use with their students, and parents’ and employers’ understanding of how well students have performed. A common perception is that fair assessment applies the same mode of assessment and content focus for all students—the approach of assessments in international comparative studies of science achievement. This article examines research evidence demonstrating that the act of assessment is not neutral—different forms of assessment advantage or disadvantage groups of students on the basis of family backgrounds, gender, race, or disability. Assessment that implicitly or explicitly captures the social capital of the child serves to consolidate, not address, educational equity. The article provides an overview of ways that science curriculum focus and assessment can introduce bias in the identification of student achievement. It examines the effect of changes to curriculum and assessment approaches in science, and relationships between assessment of science and the cultural context of the student. Recommendations are provided for science–assessment research to address bias for different groups of students.

Keywords: science assessment; equity; social capital; gender; culture; language; disability

1. Introduction

At the biennial Symposium of the International Organization for Science and Technology Education (IOSTE) in Tunisia in November 2012, a Swedish group presented a symposium that explored the issue of “*Which Child Left Behind?*”, an unexpected twist of the aspirational US rhetoric in the Bush years about the internationally-known US *No Child Left Behind* Act (NCLB) of 2001 [1]. The question suggests that whenever we revise the emphasis of the curriculum for science, encourage new pedagogies, and make use of different assessment practices, some groups of students will be disadvantaged and some advantaged—a rather sobering thought for educational innovators.

Using reported accountability as the driver, one intention of NCLB in US is to ensure focus on effective educational provision for all students regardless of racial, ethnic or social class background, gender or disability. The underlying philosophy is that all children can learn, and teachers should have expectations that all children can achieve high standards [2].

Educational reforms, especially those that rely on social indicators such as NCLB’s student achievement reporting, can have unintended outcomes [3], and considerable criticism has resulted as to the negative effects of NCLB on the learning of all students [4]. While the reforms of NCLB are not central to this discussion, they highlight that reforms in education in general, and in assessment more specifically, are not neutral acts in education. Change has consequence.

Research in the field of assessment has noted for some time both positive and negative effects of assessment in driving student learning [5–8]. Educational assessment signifies to the learner the knowledge that is valued. Positive and negative effects occur regardless of whether assessment is occurring at a systemic level, such as by external examination or participation in international testing programs such as Trends in International Mathematics and Science Study (TIMSS) or the Programme for International Student Assessment (PISA), at a school level, or within classrooms.

Educational assessment research today focuses not only on valid assessment of student achievement but also on assessment that can be used to improve student learning, and assessment for formative purposes [9]. Research on effective assessment to improve learning emphasizes three critical points: First, students need to understand educational goals underpinning educational assessment and to be able to self-evaluate their progress towards these; second, the mode or form through which student knowledge is assessed will affect what students learn and how well they are able to demonstrate their knowledge; and third, but not least, within classrooms and school environments, interactions among teachers and students are critical in effective assessment. Educational assessment at all levels is socially-situated practice [10,11]. As noted assessment expert, Patricia Broadfoot, has observed: “Assessment is not an exact scientific process—the involvement of human beings in every aspect of its design, execution and use makes it irrevocably a social project” [12]. The nature of the knowledge itself has been identified as socially-constructed [13,14].

When the complexity of assessment practice is considered, its potential to have inequitable impact on students is clear [10,15]. Research shows that changes in curriculum can lead to changes in assessment directions. A major curriculum reform in science education has been the move to a focus on deeper scientific learning and to modes of assessment that in themselves create different performance expectations, and outcomes, for students. In this article, we provide a historical overview of relationships between students’ learning in and engagement with science, changes in curriculum,

and demonstration of achievement through different forms of assessment. We explore the issue of cultural or social capital, that is, the experiences that students have in diverse family backgrounds, including race (cultural experience/knowledge and language), socio-economic status (SES) and gender. Disability, the focus of most recent assessment-equity research, can also be viewed holistically from a cultural or social capital framework. Cultural difference underpins language and linguistic differences, and experiential learning of students. These considerations can affect students' science learning and demonstration of their scientific understanding, according to the identified focus of scientific learning and assessment mode.

It may be that what follows provides answers to the embarrassing question posed at IOSTE.

2. Social Capital and Science Education

A large study of Australian doctoral students in Chemistry in the 1970s led to several unexpected findings. The levels of educational qualification of both their fathers and mothers were less than those of the whole cohort of students studying first year Science. Seventy-three per cent of fathers and 84% of mothers had not matriculated from high school compared with 54% and 67%, respectively, for the first year students. On a 1–7 point scale (1 highest), the SES status of fathers of the doctoral students clustered between four and six compared with one and four for the first year students [16].

The great majority of these doctoral students were male. They had continued directly through school into undergraduate studies, and then into doctoral work with little or no study of subjects outside of science and mathematics in their final school and undergraduate years. These data paint a picture of lower SES boys, first generation university students, who did well in mathematics in primary school and then, encouraged by their teachers, continued in this success line into secondary school and hence into the physical sciences at university. Unlike their higher SES peers, at the end of schooling, they did not consider other professional paths, like medicine, dentistry or engineering, that their expertise in science and mathematics could also have enabled.

The mathematics and sciences these doctoral students studied at school were the then recently revised curricula that stemmed from what is now known as the “ALPHABET” era due to the short hand titles, such as Biological Sciences Curriculum Study (BSCS) Biology, Chem Study Chemistry and Physical Science Study Committee (PSSC) Physics, for these large scale and lavishly funded projects developed on both sides of the Atlantic. These revisions brought the school curricula in mathematics and the sciences into line with the way these subjects were then being taught at university level. The content for learning became more conceptual and more mathematical at the expense of the descriptive detail and more historical and applied emphases in earlier curricula [17]. These new curricula coincided with a major shift in the mode of assessment from the shorter and longer descriptive essays commonly required previously in the sciences to short answer and multiple choice items focused on concepts. In retrospect, these changes in content emphasis meant that the abstractness of these subjects made them more immune to socio-cultural capital compared with subjects in the humanities, the arts and social sciences, and along with the low demand for written work, allowed working class boys, with encouragement, to succeed in them.

This research from nearly four decades ago raises interesting issues about the more recent curricular changes being introduced into school science. Since the later 1980s a gradual change in the curricular

rhetoric about school science has been to make the links between science and society more explicit. Over this 25 year period these intentions have been supported initially by innovative curriculum materials generated by the Science/Technology/Society movement [18] and more recently by the use of context-based teaching in science classrooms, an approach that begins with a real world context involving science and technology [19].

The two large international assessment projects for science learning that began in the 1990s, Trends in International Mathematics and Science Studies (TIMSS) and Programme for International Student Assessment (PISA), have to varying degrees continued to encourage these links between science and society. The PISA project particularly emphasizes the links by contextualizing test items in socio-scientific situations [20]. Both projects have introduced open-ended items requiring short and longer prose answers and minimized mathematical content in their science items. While issues are raised later regarding broader cultural and linguistic biases of both testing programs, both projects have found that in many countries there is a quite steep SES gradient for performance [20,21]. The SES status is based on several variables related to family cultural background. When compounded with greater parental experience of education, such a background creates socio-cultural capital that may mean some forms of assessing science knowledge and performance advantage the students with “richer” socio-cultural capital compared with those who are “poorer”.

Changes to school science curriculum and assessment could therefore be adding to socio-economic inequity in the learning of school science, compounded by newer modes of science assessment. The mainstream science curriculum in most countries still gives so much priority to the abstract conceptual bases of the sciences that Lyons [22] in Australia, Lindahl [23] in Sweden and Osborne and Collins [24], in England, found that by the end of compulsory schooling students reported that they had experienced school science as a subject with abstract content that had little relevance to everyday life. Nevertheless, the pressure to establish that relevance is now part of the strategy in the more developed countries to counter unpopularity of science as a field of interest or for a future career. As science teachers gradually adapt their pedagogical practice and their assessment processes to comply with these socio-scientific intentions, they may unintentionally create barriers for those students who bring less social capital to their classrooms. The question for science assessment is the need to clarify its focus—core scientific concepts and understanding; situated scientific understanding; or application—and to identify social capital barriers that each aspect of science learning and assessment may create for some students.

3. Gender and Science Achievement

While TIMSS data present strong evidence of gender difference in science achievement at fourth grade and eighth grade, these differences vary according to country and cultural background. While little difference exists in Australia, England and the US between boys and girls at fourth grade, by eighth grade boys outperform girls in Australia and the US [21]. Many different interactions between grade, gender and test data occur across countries. TIMSS’ analyses are not able to interpret the reasons for these gender differences.

The interaction between science reform, assessment and gender, however, is well-exemplified in historical changes to science curriculum over the last three decades. School-level Physics is one

example. Physics for girls has been problematic for two reasons. The first reason is participation, far fewer girls usually undertake the subject. A pattern of girls' choosing a wider range of subjects in senior schooling, both in and out of science disciplines, has persisted for some time [25]. Research indicates this is due to a mix of reasons including perception of use of the area of study in a future career, general liking or disliking of the subject, and perceived likelihood of success.

A further reason is that assessment patterns in girls' achievements in Physics differ across contexts—in some contexts their performance is consistently higher; in others, their performance is consistently lower [25]. The case of the Victorian Certificate of Education (VCE) in Australia provides an example of how contextual change in curriculum emphasis and in assessment can affect gender achievement in Physics. Reform of the VCE in the late 1980s involved a major revision of the academic components of the final two years of schooling and assessments leading to certification for competitive selection into university and employment. Victoria's political climate, with a strong feminist and environmental presence, was also conducive to significant reform. The McClintock Collective (a network of gender-aware teachers), and other interest groups were well represented on the Science Committee responsible for developments to reform Physics curriculum content, pedagogical approaches and assessment procedures. Hildebrand [26] has discussed the strategy of the McClintock Collective as being drawn from post-structural feminism, informed by research studies that suggest girls prefer to learn concepts in their social context rather than abstractly, and to give structured and extended response questions. Girls were considered likely to benefit from variety in assessed tasks spread throughout the course of study, rather than a single terminal examination.

The Collective's voice was not the only one on the Science Committee pressing for a broader sense of the purpose of school Physics and, as a result, a number of new aims were adopted for the Physics curriculum, including:

1. becoming aware of Physics as a particular way of knowing about the world which interacts with the setting, both social and personal, within which it is pursued;
2. understanding some of the practical applications of Physics in present and past technologies, examining the usefulness of such technologies as well as problems associated with them;
3. developing capacity and confidence to communicate knowledge of Physics.

Lists of aims are common for science courses and it is equally common for a number of these aims to then be ignored in the assessment process. This is often because they require modes of assessment that differ from traditional forms. In the case of VCE Physics (and other sciences at this senior level of schooling) these aims were, however, used to introduce two new forms of assessment—Work Requirements and Common Assessment Tasks (CATS).

Work Requirements consisted of six to eight tasks undertaken throughout the last two years of Physics in high school. They were intended to make explicit to students what they are learning. In this sense they served both metacognitive and formative assessment roles. Guidelines for the tasks provided students with opportunities to express their learning in various ways including posters, case study reports, student-designed investigations, a file of changing ideas, and so on. Satisfactory completion of the tasks was judged by classroom teachers supported by peer moderation, also a powerful professional development experience for the teachers.

CATS were common assessment tasks across schools spread throughout the second year of Physics that provided the basis for a graded assessment of the comparative quality of each student's learning work and were used for certification and university selection. Both CAT 1, an extended practical investigation, and CAT 3, a research project of Physics in a social context, required an extended report graded by the classroom teacher with peer moderation. CAT 2, comprehension and application of Physics, and CAT 4, explanation and modeling in Physics, completed under "test conditions", were set and assessed externally. This way there was genuine endeavor to ensure that each of the intended aims of the course were addressed, and hence valued, in assessment.

As a result, participation in and completion of Physics by girls increased, with mean scores for girls becoming significantly greater than for boys, and girls matched or exceeded boys in "A level" passes. The content of Physics in the VCE curriculum was very similar to that in the earlier curriculum, but the new modes of assessment and their distribution throughout the course of study meant not only that assessments may suit girls but also that teachers had to adopt new approaches in their teaching and pedagogies for its implementation. Unfortunately, when the student workload of this new VCE curriculum was assessed after several years of operation, it was considered too heavy for teachers and students. CAT 3, the research project, with its opportunity for extended prose favoring girls, was dropped, once more tipping the assessment stakes back in favor of boys.

Concerns with the lack of girls' participation and performance in science have resulted in reforms in other countries. A different approach was taken by Thailand when it reformed its science curricula in the mid 1970s. Its solution to the failure of girls choosing Physics was simple—Physics, Chemistry and Biology were all made mandatory for the Science stream in the senior years and the non-Science stream was made comparatively less attractive. The result by the 1980s, at least in the large metropolitan region around Bangkok, was that Thailand became the first country to report equal gender participation *and* achievement in Chemistry and Physics [27]. About 10% to 14% of girls and boys in this region undertook Physics and achieved equally in these senior years. Although communication skills were not emphasized, classroom laboratory and inquiry-based learning were the foci of the instruction and assessment, with dexterity and report writing of experiments as components of the assessment along with theoretical tests, forms and modes that balanced the assessment preferences of girls and boys. The Thai solution of removing choice at these senior levels of schooling is clearly not "simple" for a variety of reasons in other countries. Nor do other countries have the supply of female Physics and Chemistry teachers that exist in Thailand.

These two case studies demonstrate that curriculum and assessment changes have a clear impact on scientific achievement of boys and girls. Again, the decisions to be made are the foci of the Science curriculum and the best forms of assessment to identify the knowledge and skills that students have achieved. Different forms of assessment not only suit different types of learners, as this section has argued on the basis of gendered approaches to assessment, but they also identify different forms of knowledge. A strong argument for future Science curricula and their assessment is that all students, both boys and girls, should develop the range of scientific knowledge identified through different assessment forms.

4. Language and Linguistics Demands of Science Education and Assessment

In this section we consider equity issues associated with linguistic demands of science education and assessment. Scientific curriculum reforms have focused not only on the goal of deeper conceptual scientific understanding and application in real contexts, but also on scientific communication as a critical component of science knowledge. In Queensland, Australia, every science curriculum has a communication or language component [28]. Reforms in scientific curriculum, a focus on extended responses, and the general demand of language in education and assessment, affect the way that students from diverse backgrounds will be able to demonstrate their knowledge and understanding. Assessment for learning research highlights the need for clarity of expectations in assessment to guide students' work. The literacy and linguistic demands of science education are complex and change across areas of scientific study [29]. Interestingly, an Australian research study has shown that when curriculum and assessment literacy and linguistic requirements are made explicit, science teachers provide better direct instruction in the language of the discipline than their peers in non-science subjects [29].

In the 1970s democratic movements in many more developed countries led to a much wider cross-section of a country's young persons staying at school for secondary education. It became clear that the language in which science is expressed and recorded in text books posed a problem for its learning for many of this wider student population. Gardner [30] in Australia and Johnstone and Cassels [31] in Scotland drew attention to the fact that this was not just a matter of the highly technical words of science, but much more about the many everyday words that, in say, Standard English (or any other language,) take on quite precise and different meanings in the discourse of science.

Gardner *et al.* [32] examined the special linguistic character of science with a study of junior secondary students' ability to comprehend the meaning of and to use the great variety of logical connective words (LCA) used in science texts and science assessment items to link a proposition with another idea. They found a significant relationship between students' LCA knowledge and socio-economic status. Words like *thus*, *respectively*, *in addition*, and *hence* are comprehended differently among students of different SES.

Sutton [33] in England and Munby [34] in Canada drew attention to the centrality of language in science classrooms, but it was Lemke [35,36] who drew the attention of linguists such as Halliday and Martin [37] to the more linguistic aspects of science that now are a very active area of research interest. Inevitably this recognition of the complexity of science language heightens concern that equity issues associated with language learning itself will also be significant in students' science learning and in their responses to changes that occur in science assessment.

An obvious consequence of these language studies was the debate about the validity of written tests and their wording to measure scientific understanding. Harlen [38] and Murphy [39] suggested that inadequate performance on such tests did not necessarily reflect lack of learning, since the purpose of the assessor's task may be read differently by students, and within different groups of students, again opening the possibility of equity issues. Harlow and Jones [40] in New Zealand followed up suggestions by Messick [41] and Fensham [42] that interviewing students may illuminate the processes that underlie item response and task performance. Harlow and Jones administered 24 science items from the TIMSS test to a population of Year eight students and scored these using the TIMSS rubrics.

A large number of this population, representative of gender and achievement levels, and including a small number of two ethnic minority groups—Maori and Asian—were then interviewed about the test questions, their written responses to them, and the strategies they used to answer the items.

The students' range of scores on the written test were similar to those of the NZ national sample for TIMSS, but when the interview results were compared with the written responses, the overall scores increased in three of the five science content areas. Among the sub-groups, NZ European boys and girls, and Maori and Asian girls showed significant increases, that is, many students were able to demonstrate more science knowledge through oral responses than on the test forms. Results for Maori boys showed little or no difference in the two assessment contexts, perhaps because English was the medium of the interview.

The interviews indicated that there was very often something correct in incorrect written responses. For 14 of the test items students had more knowledge than they had written, although for seven items students who had "correct" written responses did not have an understanding of the concept being assessed. Of more interest are the ways the written test tasks were interpreted by students. For example, in a free response item that required a reason to be stated, many students did not give a reason in the written test, but half of them gave the reason in the interview. Again, a large number of students were found to have misinterpreted the task of one of the written items, confirming Harlen and Murphy's expectation. Similar misinterpretation of written items has been reported by other researchers [43–45]. A word, a phrase, or diagram, that is part of the question or a component in a set of multiple choice options can be the source of misinterpretation that means students' knowledge of the science may not be elicited.

Rather than heeding these warnings, written assessment in science remains the dominant form of international, national and local assessment of science learning and the amount of writing involved has increased substantially as context-based science becomes more usual (encouraged by the example of the OECD's PISA project and its use of more free response items requiring articulated prose). A feature of successive PISA Science studies has been the small gender difference in the students' science scores in many countries although the PISA Reading studies universally show that girls have much higher achievement than boys on their test items [20]. This finding has enabled critics of PISA, such as Sjøberg [46], to argue that gendered reading ability overshadows the PISA Science test's attempt to measure scientific knowledge. Although he misses the point that PISA Science is attempting to measure the students' ability to put their knowledge into practice in real world contexts, and not simply their recall of static science knowledge, he does highlight the fact that to measure this student ability requires linguistically more complex written items and responses.

Not surprisingly, the PISA reports consistently show that (a) students born outside the country of testing and with parents born outside; (b) students from lower socio-economic backgrounds; and (c) minority indigenous students (compounding cultural ways of knowing science, discussed in the next section), perform at lower levels than their culturally different and linguistically advantaged peers. National written tests for science confirm these findings. For the first of these equity shortfall groups the findings appear to be transitory, with second-generation immigrants being less affected. The other two equity group differences are more difficult to redress in terms of language literacy and its flow on effect in science learning.

For the case of inequity among Australia's Indigenous populations, McTaggart and Curro [47] noted that many of these students do not speak or hear Standard English outside the classroom. The standard language of the classroom is for them a second or third language, and in the science classroom they face even more subtle differences. Davidson [48] has found that such students may have the requisite knowledge and skills, but their lack of ability in the Standard English of assessment tasks is a barrier to their communication of their knowledge, including their understanding of multiple choice items.

Kaesehagen *et al.* [49] also found evidence for this literacy constraint, especially when the assessment task is phrased in two parts, common in mathematics and science tests, reflecting the research on LCA of Gardner *et al.* and the work of Harlow and Jones above. Thus, at the lower levels of schooling where assessment can be more formative in purpose, oral assessment of science knowledge may be more equitable than the current written modes. This in itself, however, may require teachers to develop new language skills. One solution would be to make greater use of multiple classroom-based evidence sources to demonstrate students' achievement. Digital technologies and e-portfolios may provide one way to record the science learning of all students more equitably.

5. Assessment of Science and Culture

The previous discussion focused on language in science from a linguistic perspective, and some interactions between language background and culture. Language and culture can have a much deeper interaction—the formation of science knowledge and experience, and demonstration of scientific knowledge in science assessments [50,51]. A core issue in standardized tests and international comparison tests is that, while they require test items to have cultural equivalence, this may not be establishable. Student achievement must be culturally-contextualized. Science is not “culture free” [50].

Much of the current research examining cultural difference compares achievements of groups such as Asian students, Asian-American students and other students, on the basis that the first two groups outperform other students on tests such as TIMSS and PISA or national or state-based science assessments. Simplistic interpretations of science achievement data that attribute differences to race or culture are flawed, however. Most differences, as noted, can be identified in terms of opportunity to learn, and the social capital of prior experience, the language of science and testing, and resources in the home. The higher performance of Asian students is often attributed to membership of a culture that values education, but may also be attributed to selectivity of schools and a highly competitive environment. By contrast, Asian countries such as China, Hong Kong, Japan and Singapore have been looking to the West to change pedagogy and assessment practices to focus more on problem-based learning and assessment and use of authentic and situated curriculum and assessment, and less on rote-based instruction. Comparison of science participation and achievement of Black and Hispanic students in the US with performance of Caucasian students are also a focus, as the last group significantly outperforms the first two. Again, differences are not due to racial background, but to the impact of social disadvantage, lack of resources in home and school, quality of teachers, and familiarity with standard language that many students from these cultural groups experience.

Validity of assessment for students from diverse cultural backgrounds encompasses a much deeper issue than disadvantage in socio-economic status. Different groups, especially indigenous groups, have

different ways of relating to natural phenomena and the environment, and science. Work in Australia with students at risk, especially Australian Indigenous students, has initiated place-based science education, working in the environment, with communities, and embedding Indigenous ways of knowing science [52]. Students' learning is assessed through performative assessments including digital photography, webpages and collaborative effort. Different ways of assessing have been designed to reflect different ways of knowing and learning and of interacting with others in the community.

The Australian program reflects the work of Nelson-Barber, Solano-Flores, Estrin and Trumbull [53–55] who have explored Indigenous ways of knowing in science for American Indians and valid assessment for more than two decades. These researchers highlight that building on the prior knowledge of American Indian students is necessary not only to link their experiential understanding with the Western science curriculum but also to recognize and value the prior and cultural scientific knowledge that these students have. Despite their early work, the researchers note that more recent policies such as NCLB entrench cultural bias and marginalize students from different cultures. Culturally-based curricula and pedagogy require culturally-valid assessments, with more work needed on developing such instruments [54]. “Cultural validity” should be a core assessment concept to ensure equity for students in science that avoids an assimilationist perspective [54,55]. Aspects of standardized science assessment administrations that may be culturally-inappropriate include language, on-demand expectations, and instructions [55].

Nelson and Estrin [53] indicated that American Indian ways of knowing should be incorporated in science standards (curriculum expectations) and assessed in context, consistent with fundamental constructivist approaches to science education. The new Australian Curriculum has implemented a cross-curriculum priority of Aboriginal and Torres Strait Islander histories and cultures. As an example, a science curriculum component for Year seven students includes “investigating how Aboriginal and Torres Strait Islander knowledge is being used to inform scientific decisions, for example care of waterways” [56]. How this cross-curriculum priority could be incorporated in or could inform forms of assessment, whether the priority creates an assimilationist perspective of culture rather than recognizing cultural difference, and how the inclusion of such content will affect the science learning of Aboriginal and Torres Strait Islander students, are yet to be examined. Overall, empirical research has not compared the nature and depth of the scientific knowledge that students from indigenous backgrounds hold and can demonstrate in culturally-appropriate assessments *versus* standardized western culture tests. As Lynch noted, ““science for all” is not equal to “one size fits all”” ([57], p. 622). “One-size-fits-all” assessments are not equitable.

6. Students with Disability and Assessment

Equity, bias and assessment for students with a disability are relatively recent areas of assessment research as in the past many students with a disability, particularly severe physical or intellectual impairment, did not attend school or were educated in special institutions. Often the learning expectations for students were not high. The move to inclusive schooling, with as many students with a disability as possible learning in “mainstream” classrooms with peers without a disability, has seen a change in these expectations. A major purpose of NCLB, as noted, was for teachers to have high

expectations for all students. Accountability for learning outcomes for students with a disability has stimulated recent research in appropriate assessment modes and approaches to allow students with a disability to demonstrate their knowledge in areas such as science. Consider how traditional assessment modes in school could impact on the capacity of Stephen Hawking to demonstrate his scientific excellence.

Disability is defined very broadly in most nations, and by the OECD [58], to include physical, emotional and intellectual conditions. Such conditions vary in severity and impact. Disability includes dyslexia and autism spectrum disorder as well as visual impairment or hearing impairment. Most definitions include learning difficulties that affect the rate of progress of learning but are less easy to classify than other disabilities. Systems and teachers need to consider science education for these students, how assessment should occur, and most importantly how assessments can be framed to enable students to demonstrate their science knowledge in the most enabling way.

A common approach to assessment of students with a disability is to provide a range of “accommodations” to support the student with a disability to use the same assessment form or test used by other students.

Standard forms of accommodation include enlarged print or Braille versions for students with vision impairment, amanuenses for students unable to write, assistive technologies, and time allowances (more time, breaks). We noted at the beginning of this article that for the layperson, the concept of “fair” is all doing the same assessment. Clearly this view would not extend to an expectation that a student who is blind would read a written test paper. Immediately, some change to the assessment must occur.

Contentious issues in accommodations for students with a disability are the provision of additional time allowances and the reading of assessment information to students. The former is the question of how much time should be allowed, to notionally put the student with a disability on an equitable footing in assessment with students without a disability. The issue raised in the reading of assessment items is whether reading skill or comprehension of language is the knowledge being assessed, or whether the accommodation “fundamentally alter[s] the nature” of the test [59]. Given the previous discussion of the role that language and linguistics play in science curriculum and assessment, this is clearly an important issue for students with a disability. The concern in the provision of these types of assessment accommodations, often expressed, is that students with a disability may gain an unfair advantage over students without a disability [10]. The concern never appears to be expressed that students with a disability may already have cultural, experiential or linguistic disadvantage. Any equity concern that additional time may give an advantage to students with a disability is easily addressed. Research demonstrates that all students may gain from additional times or extra adjustments in assessment, not only students with a disability but also students from diverse language and cultural backgrounds [60,61]. Science assessments should focus on the scientific knowledge and related skills being assessed, not on speed of performance. As Sireci, Scarpati and Li [60] noted, if all students do better with more time, it is not the time accommodation that is unfair, but that the time conditions imposed on all students could be too stringent.

Equity for students with a disability is still a major issue for the international programs, TIMSS and PISA, that assess student science performance. These tests exclude students who are “intellectually or physically disabled” and cannot “perform in the ... testing situation” from inclusion

in the sampling ([62], p. 25). Only small, but varying, percentages of students are excluded from each country for these reasons [62], fewer than would be expected on the basis of students with a reported [reportable] disability in each country. While a shortened version of the test is available for students with a disability and more flexible test-taking conditions are allowed, few countries have opted to use this alternative [63].

No guidelines are provided with the PISA and TIMSS tests regarding accommodations for students with a disability, and the expectation is that countries will need to exclude students who would be very difficult or resource intensive to test [64]. The consequence for most countries is that unless improper sampling is occurring, students with a disability will be taking the standard test forms. In countries such as Australia, with strong policies of inclusion of all students in mainstream schooling, this may be a further explanation for the negative tail in the results that are solely identified as relating to disadvantage and ethnicity [65]. In recognition that these international science studies do not address the worldwide inclusive education agenda and are not fair for students with a disability, accommodations for future administrations are being investigated [66].

As NCLB provides that students with a disability may be assessed using alternative modes of assessment, alternative assessments and establishing their equivalence to standard test forms, have become a focus of considerable recent research in the US. These can include interviews, digital recordings, validated teacher observations or checklists, presented as evidence of student learning through portfolios [67].

Research-based evidence on appropriate assessment modes and conditions for students with a disability is scant [10,68], with little recent research looking at the interaction between different forms of disability, student knowledge, and achievement through assessment accommodations such as time allowances. One promising area of research emerging from the US is adjustments to multiple choice items to modify the cognitive demands of the item for students with intellectual disability. Contrary to stereotypes that may exist, students with intellectual disability can develop conceptual understanding. Their difficulty in learning and assessment may be limited memory, particularly working memory capacity [69–71]. The cognitive demands, or load, of a multiple choice item may mask whether a student with disability has scientific or mathematical understanding. Cognitive demands also reflect the language and linguistics demands of science test items previously discussed for students of different language backgrounds. Research by Elliott and colleagues [72,73] has demonstrated that items can have a reduced number of distractors and simplified language structures, while maintaining reliability and validity in assessing the intended conceptual learning focus.

7. Conclusions

In the heady euphoria of late 1960s Robert Hein, Director of the Elementary Science Studies project in the USA, claimed that science should be the easiest subject to teach in primary school because it only required observation and talking—powers that the great majority of young learners already could bring to school. This naturalist view of science may well still be a good way to start school science, but it does not get learners far into the abstract language of science that has been invented to systematize and explain these observations. Furthermore, visual observation is only one methodological tool that science uses in its ongoing investigation.

Although the abstractness of science content can mean that learning science is less linked to prior social capital than a number of other subjects, its language is to a greater or lesser extent a foreign language, and this makes it susceptible to variation in social capital and hence to a lack of equity across all students. Furthermore, how this language is expressed in the assessment of science learning—all too often written questions and written responses—can affect the equity of its measure of understanding science. The numerous examples of inequity that are discussed above concerning the style, format and modes of assessment in science call for much greater variety in the presentation of assessment tasks and in the manner of response by which students provide evidence of their learning.

Among the more developed countries there has been concern since about 2000 about negative attitudes towards science among students, and existing approaches to the assessment of its learning in school has given it the reputation of a “difficult” subject. This is counterproductive at just the time when a sound basis in science and the way its knowledge is increasingly interacting for good or ill with the lives of individuals, and societies and as international citizens. Newer science curricula encourage and expect science to be taught and learnt in this interactive sense. Their intentions for learning offer the opportunity for students to develop interest in, and to learn science from the situations in their own personal experience of these interactions. These opportunities open science to the widest populations of students regardless of their associated sources of educational advantage and disadvantage.

These curricula need to be matched by explicit delineation by curriculum authorities of the associated learning expectations, and of the assessment policies and practices that address the various sources of inequity to which we have drawn attention. Only then will we begin to move in science education from *Which Child Left Behind?* towards *No Child Left Behind*. As curriculum authorities take such official action, teachers will be encouraged and, indeed required, to use a variety of assessment tasks and modes for their classroom formative and summative assessments. These will optimize for all students both their ongoing engagement with and their learning of science as powerful and useful knowledge for living in the diverse and inter-connected biophysical world we all share.

This article has demonstrated that assessment in science is vulnerable to capturing social capital of students from diverse family backgrounds that relates to the gender of students, their cultural experiences, their language, and their capacity to engage with assessment demands. The discussion shows that the issues identified in this article have been raised over some past decades, usually without resolution. Trends to national and international standardized measures of science, and comparisons of performance on these, may be exacerbating rather than addressing these many issues of equity in science achievement.

Most comparative research examining differences in student achievement works from the gold standard of performance on assessments that reflect Western science curriculum and the language and experiences of students who speak the first language of their country and are from well-resourced homes. Implicitly or explicitly, sources of inequity in assessment result in difference being identified as a deficit from this norm, not in terms of scientific knowledge and understanding.

This discussion has demonstrated that difference may reflect bias in the assessment tasks, and in the curriculum’s construction. Attempts to remove or control for the cultural context of a student in assessment do not result in a culture-free zone [15,74].

A different perspective is to take each group as representing its own gold standard. The task for research in science assessment is to identify the nature of the science that the students know, its depth,

its richness, and to identify the construct-irrelevant variance due to the nature of an assessment task. Much more empirical research is needed into how students know and understand science, and the nature of this science, viewing language, gender, cultural experience constructively. As Luykx and colleagues noted, such research requires fine-grained qualitative analysis, necessarily with small groups of students, along with collaboration across disciplinary boundaries of “science educators, assessment specialists, linguists, anthropologists, discourse analysts, statisticians, and ... others’ ([50], p. 920).

Conflict of Interest

The authors declare no conflict of interest.

References

1. No Child Left Behind Act of 2001. Available online: <http://www2.ed.gov/policy/elsec/eag/esea02/index.html> (accessed on 1 August 2013).
2. Atkin, J.M.; Black, P.; Coffey, J. *Classroom Assessment and the National Science Education Standards*; National Academy Press: Washington, DC, USA, 2001.
3. Campbell, D.T. Assessing the Impact of Planned Social Change. In *Social Research and Public Policies: The Dartmouth/OECD Conference*; Lyons, G., Ed.; Dartmouth College, The Public Affairs Center: Hanover, NH, USA, 1975; Chapter 1, pp. 3–45.
4. Darling-Hammond, L. Evaluating “No Child Left Behind”. *The Nation*, 21 May 2007. Available online: <http://www.thenation.com/doc/20070521/darling-hammond> (accessed on 3 March 2013).
5. ARG (Assessment Reform Group). *Assessment for Learning: 10 Principles*; ARG: London, UK, 2002.
6. Black, P.; Wiliam, D. Assessment and classroom learning. *Assess. Educ.: Princ. Policy & Pract.* 1998, 5, 7–74.
7. Harlen, W.; Deakin Crick, R. Testing and motivation for learning. *Assess. Educ.: Policy Pract.* 2003, 10, 169–207.
8. Shepard, L.A. Reconsidering large-scale assessment to heighten its relevance to learning. In *Everyday Assessment in the Science Classroom*; Atkin, J.M., Coffey, J.E., Eds.; NSTA Press: Arlington, VA, USA, 2003.
9. Black, P.; Harrison, C. *Science inside the Black Box: Assessment for Learning in the Science Classroom*; GL Assessment: London, UK, 2004.
10. Cumming, J.J. *Valuing Students with Impairments: International Comparisons of Practice in Educational Accountability*; Springer: Dordrecht, The Netherlands, 2012.
11. Moss, P.; Pullin, D.; Gee, J.; Haertel, E. The idea of testing: Psychometric and sociocultural perspectives. *Meas.: Interdiscip. Res. Perspect.* 2005, 3, 63–83.
12. Broadfoot, P. Preface. In *Educational Assessment in the 21st Century: Connecting Theory and Practice*; Wyatt-Smith, C., Cumming, J., Eds.; Springer: Dordrecht, The Netherlands, 2009; pp. v–xi.
13. Lave, J.; Wenger, E. *Situated Learning: Legitimate Peripheral Participation*; Cambridge University Press: Cambridge, UK, 1991.
14. Saxe, G.B. The mathematics of child street vendors. *Child Dev.* 1988, 59, 1415–1425.

15. Stobart, G. *Testing Times: The Uses and Abuses of Assessment*; Routledge: Abingdon, UK, 2008.
16. Hill, S.; Fensham, P.; Howden, R. *PhD Education in Australia: The Making of Professional Scientists*; Report No.7, Science and Industry Forum; Australian Academy of Science: Canberra, Australia, 1974.
17. Fensham, P.J. Social content in chemistry courses. *Chem. Br.* **1976**, *12*, 148–150.
18. Solomon, J.; Aikenhead, G. *STS Education: International Perspectives on Reform*; Teachers College Press: New York, NY, USA, 1994.
19. Bennett, J.; Hogarth, S.; Lubbens, F. *A Systematic Review of the Effect of Context-Based and Science-Technology-Society (STS) Approaches in the Teaching of Science*; Department of Educational Studies, University of York: York, UK, 2005.
20. OECD (Organisation for Economic Co-operation and Development). *PISA 2006: Science Competencies for Tomorrow's World*; OECD: Paris, France, 2007; Volume 1 (Analysis).
21. Martin, M.; Mullis, I.; Foy, P.; Stanco, G. *TIMSS 2011 International Results in Science*; TIMSS & PIRLS International Study Centre: Boston, MA, USA, 2012.
22. Lyons, T. Different countries, same science classes: Students' experiences of school science in their own words. *Int. J. Sci. Educ.* **2006**, *28*, 591–614.
23. Lindahl, B. *Pupils Responses to School Science and Technology? A Longitudinal Study of Pathways to Upper Secondary School*; Göteborg Studies in Educational Sciences, 196 Acta Universitatis Gothoburgensis: Gothenburg, Sweden, 2003.
24. Osborne, J.; Collins, S. Pupils' and parents' views of the school science curriculum. *Sch. Sci. Rev.* **2000**, *82*, 23–31.
25. Brotman, J.S.; Moore, F.M. Girls and science: A review of four themes in the science education literature. *J. Res. Sci. Teach.* **2008**, *45*, 971–1002.
26. Hildebrand, G. Redefining achievement. In *Equity in the Classroom*; Murphy, P.F., Gipps, C.V., Eds.; Falmer Press: London, UK, 1996; Chapter 12, pp. 149–172.
27. Klainin, S.; Fensham, P.J.; West, L.H.T. Successful achievements by girls in physics learning. *Int. J. Sci. Educ.* **1989**, *11*, 101–112.
28. Doig, S.M.; Wyatt-Smith, C.M.; Cumming, J.J.; Ryan, J. The evolution of language education within official accounts of Queensland curriculum. *Qld. J. Educ. Res.* **1998**, *14*, 4–44.
29. *Literacy-Curriculum Connections: Implications for Theory and Practice*; Cumming, J.J., Wyatt-Smith, C.E., Eds.; ACER: Melbourne, Australia, 2001.
30. Gardner, P.L. *Words in Science*; Australian Science Education Project: Toorak, Victoria, Australia, 1972.
31. Johnstone, A.H.; Cassells, J.R.T. *Words that Matter in Science*; Royal Society of Chemistry: London, UK, 1985.
32. Gardner, P.L.; Schafe, L.; Thein, U.M.; Watterson, R. Logical connectives in science: Some preliminary findings. *Res. Sci. Educ.* **1977**, *17*, 97–108.
33. Sutton, C.R. Language and communication in science lessons. In *The Art of the Science Teacher*; Sutton, C.R., Haysom, J.R., Eds.; McGraw Hill: London, UK, 1974; pp. 41–53.
34. Munby, A.H. Some implications of language in science education. *Sci. Educ.* **1976**, *61*, 115–124.
35. Lemke, J. *Classroom Communication of Science*; Educational Resources Information Center: Washington, DC, USA, 1983.

36. Lemke, J. *Talking Science: Language, learning and values*; Ablex: Norwood, NJ, USA, 1990.
37. *Writing Science: Literacy and Discursive Power*; Halliday, M.A.K., Martin, J.R., Eds.; Falmer Press: London, UK, 1993.
38. Harlen, W. *The Teaching of Science in Primary Schools*, 2nd ed.; David Fulton Publishers: London, UK, 1996.
39. Murphy, P. Sources of inequity: Understanding students' responses to assessment. *Assess. Educ.* **1995**, *2*, 213–232.
40. Harlow, A.; Jones, A. Why students answer TIMSS test items the way they do. *Res. Sci. Educ.* **2004**, *34*, 221–238.
41. Messick, S. Meaning and value in test validation: The science and ethics of assessment. *Educ. Res.* **1989**, *18*, 5–11.
42. Fensham, P.J. Student response to the TIMSS test. *Res. Sci. Educ.* **1998**, *28*, 481–489.
43. Shapiro, B. *What Children Bring to Light: A Constructivist Perspective on Children's Learning in Science*; Teachers College Press: New York, NY, USA, 1994.
44. Lokan, J.; Adams, R.; Doig, B. Broadening assessment, improving fairness? Some examples from school science. *Assess. Educ.* **1999**, *6*, 83–99.
45. Wang, J. A content examination of TIMSS results. *Phi Delta Kappan* **1998**, *80*, 36–38.
46. Sjøberg, S. (2012) PISA: Politics, fundamental problems and intriguing results. *La revue. Res. Educ.* **2012**, *14*, 1–21.
47. McTaggart, R.; Curro, G. *Book Language as a Foreign Language: ESL Strategies for Indigenous Learners*; Queensland College of Teachers: Toowong, QLD, Australia, 2009.
48. Davidson, T. *The Language of Maths: Literacy and Numeracy in the Middle Years*; Far North Cluster of Indigenous Schools; DEST: Canberra, ACT, Australia, 2005.
49. Kaesehagen, C.; Klenowski, V.; Funnell, R.; Tobias, S. Where did I lose you? Accessing the literacy demands of assessment. *Prim. Middle Years Educ.* **2012**, *10*, 3–11.
50. Luykx, A.; Lee, O.; Mahotiere, M.; Lester, B.; Hart, J.; Deaktor, R. Cultural and home language influences on children's responses to science assessments. *Teach. Coll. Rec.* **2007**, *109*, 897–926.
51. Turkan, S.; Liu, O. Differential performance by English language learners on an inquiry-based science assessment. *Int. J. Sci. Educ.* **2012**, *34*, 2343–2369.
52. Wilson, K.; Stemp, K. Science education in a "classroom without wall": Connecting young people via place. *Teach. Sci.* **2010**, *56*, 6–10.
53. Nelson-Barber, S.; Estrin, E. Bringing Native American perspectives to the teaching of mathematics and science. *Theory Pract.* **1995**, *34*, 174–185.
54. Nelson-Barber, S.; Trumbull, E. Making assessment practices valid for Indigenous American students. *J. Am. Indian Educ.* **2007**, *46*, 132–147.
55. Solano-Flores, G.; Nelson-Barber, S. On the cultural validity of science assessments. *J. Res. Sci. Teach.* **2001**, *38*, 553–573.
56. Australian Curriculum, Assessment and Reporting Authority (ACARA). Science Foundation to Year 10 (v 5.0) Year 7 Science as Human Endeavour/Use and Influence of Science. 2010. Available online: <http://www.australiancurriculum.edu.au/Science/Curriculum/F-10> (accessed on 7 December 2010).

57. Lynch, S. “Science for all” is not equal to “one size fits all”: Linguistic and cultural diversity and science education reform. *J. Res. Sci. Teach.* **2001**, *38*, 622–627.
58. OECD (Organisation for Economic Co-Operation and Development). *Students with Disabilities, Learning Difficulties and Disadvantages: Policies, Statistics and Indicators*; OECD: Paris, France, 2007.
59. Cumming, J.J. “Falchenberg v NY State Department of Education, 2008, p.183” Case Discussed in Valuing Students with Impairment: International Comparisons of Practice in Educational Accountability; Springer: Dordrecht, The Netherlands, 2012; p.29.
60. Sireci, S.G.; Scarpati, S.E.; Li, S. Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Rev. Educ. Res.* **2005**, *75*, 457–490.
61. Zuriff, G.E. Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Appl. Meas. Educ.* **2000**, *13*, 99–107.
62. OECD (Organisation for Economic Co-Operation and Development). *PISA 2009 Results: What Students Know and Can Do—Student Performance in Reading, Mathematics and Science*; OECD: Paris, France, 2010; Volume 1.
63. Brzyska, B. Requirements/Accommodations for Special Educational Needs within Large International Surveys and Studies. In Proceedings of the 12th Annual Conference of the Association for Educational Assessment—Europe, Belfast, Northern Ireland, 10–12 November 2011.
64. Martin, M.O.; Mullis, I.V.S.; Foy, P.; Olson, J.F.; Erberber, E.; Preuschoff, C.; Galia, J. *TIMSS 2007 International Science Report: Findings from IEA’s Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*; TIMSS & PIRLS International Study Center, Boston College: Boston, MA, USA, 2008.
65. Zanderigo, T.; Dowd, E.; Turner, S. *Delivering School Transparency in Australia: National Reporting through My School, Strong Performers and Successful Reformers in Education*; OECD Publishing: Paris, France, 2012.
66. Ben-Simon, A. Improving Access to PISA for Students with Disabilities and Other Special Education Needs. Meeting of the PISA Strategic Development Group: 2011. Available online: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/PISA/GB/SDG\(2011\)3&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/PISA/GB/SDG(2011)3&docLanguage=En) (accessed on 11 January 2013).
67. Kleinert, H.; Browder, D.; Towles-Reeves, E. Models of cognition for students with significant cognitive disabilities: Implications for assessment. *Rev. Educ. Res.* **2009**, *79*, 301–326.
68. Mandinach, E.; Bridgeman, B.; Cahalan-Laitusis, C.; Trapani, C. *The Impact of Extended Time on SAT Test Performance*; College Board: New York, NY, USA, 2005.
69. Baroody, A.J.; Snyder, P.M. A cognitive analysis of basic arithmetic abilities of TMR children. *Educ. Train. Ment. Retard.* **1983**, *18*, 253–259.
70. Cumming, J.J.; Elkins, J. Lack of automaticity in the basic addition facts as a characteristic of arithmetic learning problems and instructional needs. *Math. Cognit.* **1999**, *5*, 149–180.
71. Finnane, M. The Role of Fluency in Mathematical Development: Factors Associated with Early Learning Difficulties in Mathematics. Ph.D. Thesis, University of Queensland, Brisbane, QLD, Australia, 2007.

72. Elliott, S.; Kettler, R.; Beddow, P.; Kurz, A.; Compton, E.; McGrath, D.; Bruen, C.; Hinton, K.; Palmer, P.; Rodriguez, M.; *et al.* Effects of using modified items to test students with persistent academic difficulties. *Except. Child.* **2010**, *76*, 475–495.
73. Elliott, S.; Kettler, R.; Roach, A. Alternate assessments of modified achievement standards: More accessible and less difficult tests to advance assessment practices. *J. Disabil. Policy Stud.* **2008**, *19*, 140–152.
74. Cumming, J.J. After DIF: What Culture Remains? In Proceedings of the 26th Annual Conference of International Association for Educational Assessment, Jerusalem, Israel, 14–18 May 2000.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).