Student perceptions of standardised testing: Survey exploring the attitudes of university undergraduate students towards the traditional methods of student evaluation in formal education

Matthew Taylor

Supervised by: Adam Goody

May 2013

Acknowledgements

I would like to thank my research supervisor Mr. Adam Goody, for his guidance, encouragement and useful critiques of this research work. I would also like to thank Dr. Catherine Fritz, for her advice and assistance in developing this study. My gratitude is also extended to Dr. Steve Dempster for his help in designing the research questionnaire and collecting the data. Finally, I wish to thank my family and friends for their support and encouragement throughout my study.

Student perceptions of standardised testing: Survey exploring the attitudes of university undergraduate students towards the traditional methods of student evaluation in formal education

ABSTRACT

This study explored student perceptions of standardised testing in education. The educational literature surrounding student assessment indicates that current testing practices are severely flawed. This is particularly concerning due to the high-stakes that are attached to exam outcomes. The purpose of this study was to assess whether student perceptions of standardised testing are consistent with the concerns raised by previous research. An online questionnaire was used to assess the overall attitudes of University students towards testing and the issues surrounding the assessment system based on their experiences throughout education. There were several significant patterns in student perceptions including four main factors identified by a principle components analysis which are analysed in this paper. The findings cast major doubts over the validity of standardised testing as well as raising serious concerns about the influence that standardised testing has on students throughout their education. The concerns raised by students emphasise the need for change in the ways in which standardised testing is used in education.

KEY WORDS:	High Stakes Testing	Formative Assessment	Summative Assessment	Validity	Reliability

Introduction

Student Perceptions of Standardised Testing

Standardised testing (ST) has become an instrumental part of the English educational context for over the last three decades (Department for Education, 2011). During this time, the British Government has placed great importance on the need to assess pupils in order to provide nationally comparable information summarising students' achievement for students, teachers and parents. This information is also used by educational institutions and business organisations that may use this information to inform their selection of students for a course or job. In 2007, the General Teaching Council (GTC) stated that school-children in England were the most tested in the world, facing an average of 70 tests and exams before the age of 16 (Asthana, 2007; House of Commons (HC), 2008a). Up until 2009, all children in state schools were externally tested in English and Mathematics at the ages of 7, 11 and 14, with science also being tested at age 11 and 14. However, in late 2008 it was revealed that key stage 3 'SAT' exams were to be abolished giving greater precedence to teacher's judgement as a means of tracking pupil's progress. In addition to this most pupils are tested at the age of 16, undertaking GCSE qualifications in approximately eight to twelve different subjects. Those who go on to study in higher education continue to be assessed predominantly through standardised testing whilst working towards A-Levels or degree qualifications.

ST refers to a form of assessment which is conducted within formal and specified procedures designed to ensure comparability of results between different schools and between different test occasions (Task Group on Assessment and Testing (TGAT), 1987). This form of assessment allows the performance of a student to be evaluated in terms of the national population sample for their age range. In order to ensure standardisation, marking instructions are provided by means of a set of criteria against which performance can be measured. The criterion referencing system provides a score or grade that is awarded purely on the basis of the quality of the performance, irrespective of the performance of other pupils. This function of testing is no doubt, extremely valuable at certain times during a student's education. However, the growing dependence on ST in evaluating the performance of pupils, teachers and schools has led to concern over whether it is fit to effectively serve these multiple functions simultaneously (DfE, 2011).

In a previous research project (Taylor, 2012), I explored student perceptions of education by conducting in-depth interviews with two university students about their overall experience of education. One of the main issues explored in the interviews focused on student attitudes towards standardised testing which revealed some intriguing results. The interviewees raised several concerns about the current assessment system which are summarised below:

- ST provides a fair means of comparison between students for low level abilities such as memorisation. However, in terms of higher level skills, ST fails to address these more valuable competencies and thus the value of these comparisons is undermined.

- ST is predominantly a memory test and fails to distinguish between those who understand the material and those who rote learn the material.
- ST should assess understanding rather than memory and should take effort into account.
- Coursework and class-work provide a better reflection of students' understanding.
- There are factors other than ability and understanding that can have a significant influence on test performance such as anxiety and time constraints.
- ST encourages superficial learning rather than mastery and thus learning material is often not retained beyond the exam.
- Student assessment should reflect the demands of later life which formal exam conditions fail to do. Coursework provides a better reflection of these demands with less time constraints and resources available to aid with answering.

Assessment is central to the education process with important consequences for students, as well as for other educational stakeholders, being made on the basis of test performance. However, students' perceptions of testing in education are vastly under-researched (Struyvena, Dochya, and Janssensa, 2005). Therefore, students' perceptions of assessment in education was judged to be the most appropriate and important area of my previous research that required further exploration. I will now give an overview of how some of these findings link to the previous research into testing, along with some other issues raised in the educational literature. This will provide the context for the current study which builds upon these links and attempts to gauge the ways in which these issues are perceived by students who have been subject to the testing culture.

The value of testing when used effectively is heavily substantiated by a plethora of research (Spitzer, 1939; Epstein, Epstein & Brosvic, 2001; Morris & Fritz, 2000, 2002; Morris, Fritz, Jackson, Nichol & Roberts, 2005). Early research by Spitzer (1939) recognized that immediate recall, in the form of a test, was effective in aiding the retention of learning. In addition, research into spaced retrieval practice has exemplified how testing can be used to enhance learning (Landauer and Bjork, 1978; Morris et al., 2005). To maximise the learning benefits of testing, feedback must be provided immediately for pupils to fix erroneous ideas assimilating any corrections (Epstein, Epstein & Brosvic, 2001).

The current testing procedures in schools in England was largely shaped by The Task Group on Assessment and Testing (TGAT, 1987) who were asked to advise on the national assessment system within the new national curriculum which was introduced in 1988 (Green & Oates, 2007). The vision of TGAT was to utilize the educational value of testing predominantly through the use of formative assessment to enhance learning. It has long been recognised that assessment can support learning as well as measure it (Black & William, 2003). The distinction between different forms of testing has been researched for over forty years with particular attention given to the conflict between summative and formative types of assessment (Bloom, Hastings and Madaus, 1971).

Bloom et al. (1971) first noted a contrast between summative testing; administered at the end of teaching episodes for the purpose of grading students, and formative evaluation; which focuses on enhancing students' learning by providing feedback to facilitate future improvement. There is an abundance of credible evidence that formative assessment (or assessment for learning) is one of the most effective pedagogical approaches for improving learning, nurturing skills and cultivating self-regulated learners (Black and Wiliam, 1998; Clark, 2012, Lam, 2013). The initial roll-out of national assessment intended to capture the potential of assessment as a pedagogical tool as well as recording pupils' progress (HC, 2008b). However during the 1990's, an increasing emphasis was placed on parental choice of school as incentive for schools to improve. This, accompanied by the emphasis on test results as the main source for school and teacher accountability, led to a shift away from formative assessment towards mainly summative assessment (HC, 2008b). This shift was reinforced by the Every Child Matters Policy (DfES, 2003), which emphasised the need to ensure all pupils achieve the minimum required standards indicated through test results.

Validity and Reliability of Standardised Testing

In order to assess the value of ST we must first address the meaning of the concepts of validity and reliability in relation to educational assessment. Validity is an evaluative judgement of the degree to which empirical evidence supports the *adequacy* and *appropriateness* of the *inferences* and *actions* based on test scores (Messick, 1989). Testing on only a single occasion for a subject and via a single testing method (written exams) is unlikely to provide an accurate reflection of a student's overall knowledge and ability. Reliability refers to whether students of roughly the same ability will consistently score similarly on the tests used. Again, the way in which information is obtained by ST (on a single occasion via a single method), as well as the impact of formal testing on students (e.g. anxiety), increases the likelihood of other factors confounding the results. Thus differences in test scores may reflect factors such as the ability to cope with test anxiety and sleep deprivation rather than an actual difference in ability.

Roderick and Engel (2001) suggest that the pressures brought about by high-stakes testing inevitably produces failure for 'low attaining' students who feel that the gap they have to close is too great. One example of how this can happen is students adopting self worth protection strategies such as Laddishness. Covington (2000) describes how the emphasis on testing in education encourages students to define themselves in terms of their exam performance. Jackson (2003) highlights how failure to perform well can produce feelings of shame and anxiety. Thus, in order to maintain a positive self concept, pupils may adopt 'laddish' behaviours, such as messing around in class and withdrawing effort, as a defence mechanism in order to provide an excuse for failure. However, the negative impact of ST is not confined to low attaining students; high ability students can also be affected. Weiner and Carton (2012) suggest that perfectionism can exacerbate test anxiety and lead to performance avoidance which inhibits test performance.

It was recommended by TGAT (1987), that in order to enhance the educational validity of assessment and broaden the range of abilities that are measured, it should exploit a wide range of methods in the delivery, working and response. In order to minimise curriculum distortion, TGAT proposed that moderated teachers' ratings be used in conjunction with a variety of standardised assessment tasks including practical assessments, as well as written tasks. It was also recognised that the use of standard procedures must be limited in assessing pupils' work in order to avoid narrowing the range of attainments being assessed.

Impact on Students' Learning Approaches

Bloom, Engelhart, Furst, Hill and Krathwohl (1956) believed that education should focus on mastery and the promotion of higher order thinking, rather than simply transferring facts. Mastery learning focuses on the excellence of skill development, with students given the opportunity to gain proficiency in each unit of learning before proceeding to a more advanced learning task (Bloom, 1985). Higher order thinking skills refer to the skills at the top end of Blooms taxonomy (Bloom et al., 1956) that require metacognition including analysis, synthesis and evaluation. Bloom et al. (1956) suggested that students should be encouraged and given the opportunity to practise these high order skills, however, it was estimated that 80-90 % of the questions that pupils encounter require them to think only at the lowest possible levels. Low order thinking skills include basic knowledge and understanding that rely on basic skills such as rote memorisation. Anderson (1994) also found that the majority of exam questions are aimed at assessing knowledge at the lower end of Bloom's taxonomy.

Different pupils have different skills and abilities, and progress at different speeds. Einstein wrote "Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid" (as cited in Kelly, 2004). This metaphor captures the absurdity of assessing all students (who possess a range of different strengths and weaknesses) by a single inflexible test. Although some pupils may be less gifted than others in terms of the skills that are academically assessed, the increased pressure on students due to high-stakes testing may be detrimental even to those who do possess the potential to achieve academic success. Furthermore, the pressure on teachers to focus on exam preparation often means that they cater instruction to the learning needs of the average student in order to save time. This results in a failure to differentiate lessons to address the needs of students who need additional support or those who need to be challenged further.

Educational experts also emphasise the importance of developing self-regulated learners through education (Gardner, 1963; Zimmerman, 1990; Winne & Hadwin, 2008, Clark, 2012). These learners are guided by motivation to learn (Brophy, 1998), and possess the metacognitive skills to enable them to select effective learning strategies and provide self feedback to aid their individual improvement (Zimmermen, 1990). Winne and Hadwin (1998) outline the cycle that self-regulated learners undertake when learning. The cycle involves analyzing the

task at hand, setting themselves goals, devising a plan and enacting strategies to achieve their goals.

Self-regulated learning can be encouraged by involving students in the decision making processes regarding their own education. For example, involving students in shaping their individual learning goals and determining what indicates success for them individually, can lead to students taking more responsibility for their own learning (Zimmerman, 1990; Main, 2008). However, ST offers little flexibility in relation to students' involvement, and thus, fails to promote self-regulated learning (David and Neitzel, 2011).

The Impact on Teaching

Whitten (2011) analysed the issue of learning from and for tests stating that testing has become ubiquitous in formal education. This highlights the ever-present nature of testing and how ingrained it has become within young people's learning development. He stated that the tests used in education are rarely viewed as an opportunity for learning, suggesting that teachers are often unaware of the positive effects that testing can have on learning through spaced retrieval practice. The growing emphasis on testing and increased pressure on schools to improve their exam results may counterproductively encourage teachers to engage in a variety of classroom practices that are not in the best interests of their pupils (HC, 2008b). These practices are aimed directly at improving test results and include teachers narrowing the curriculum and 'teaching to the test'. These practices have been found to be widespread phenomena in schools (HC, 2008a). This narrow approach often promotes memorisation of facts over mastery and the development of higher order thinking skills (Taylor, 2012). The danger of this is that the current education system holds schools accountable for the wrong thing. ST focuses predominantly on low level cognitive skills such as rote learning (Bloom et al., 1956; Anderson, 1994). Despite higher order skills having greater value in the development of young people, teachers who prioritise these skills over rote memorisation may be scrutinised if they produce poorer test results compared to teachers who 'teach to the test'.

'Teaching to the test' refers to teachers focusing heavily on the content that is likely to be tested and using the format of the test as a basis for teaching. This includes narrowing the focus of lessons, coaching pupils on examination technique, question spotting, and repetitive practice testing. The sole focus of this approach is to maximise marks in the test; and since this approach is successful in doing so; increased test scores do not necessarily represent an improvement in real academic performance. There has been much speculation in the media that tests are getting easier due to the vast improvements in test scores over the past decade, particularly since the implementation of the A-star grade. However, a further report from the House of Commons (HC, 2008b) suggested that this improvement is not due to tests getting easier, rather that teaching and learning has become increasingly narrowly focused on achieving test results.

The aim of summative assessment is to provide a record of each student's overall achievement in a specific area of learning at a given time (GTCE, 2004). This

type of assessment when used for accountability purposes creates what is referred to as high-stakes testing. High-stakes testing carries important consequences for pupils, teachers and schools. Exam results are often used as the sole determining factor for awarding qualifications and selecting pupils for higher education institutes, as well as for evaluating school and teacher performance. However, research has highlighted the negative consequences of high-stakes testing including increased test anxiety (Benmansour, 1999; Pollard et al., 2000; Ashcraft & Ridley, 2005) and increased pressure from parents and teachers on students to do well (Davies & Brember, 1998).

Narrowing the Curriculum

Since the 1988 Education Reform Act (ERA) all schools in England and Wales have a legal duty to provide a balanced and broadly based curriculum which:

a) Promotes the spiritual, moral, cultural, mental and physical development of pupils.
b) Prepares children for the opportunities, responsibilities and experiences of adult life (ERA, 1988, Sec 1.2).

The high stakes attached to ST and consequent narrowing of the curriculum bring into question the extent to which these aims are successfully achieved. In addition, the Every Child Matters Policy (DfES, 2003) states that every child should have the support they need to: be healthy; stay safe; enjoy and achieve; make a positive contribution; and achieve economic well-being. However, education prioritises the 'achieving' outcomes over the other aims due to high-stakes testing; this can come at the expense of students' enjoyment as well as the development of a broader skill set (Harlen & Deakin Crick, 2002). High-stakes testing practices can also be detrimental to the attitudes and enthusiasm towards learning that is necessary in developing self-regulated learners who will engage in lifelong learning (Davis & Nietzel, 2011). This links to the research into developing self-regulated learners highlighting the overlap between many of the issues surrounding ST.

Gardner (1983) disputed the traditional idea that there is a singular form of intelligence; instead arguing that intelligence has many forms that are not easily measurable. Gardner's theory of multiple intelligences outlines eight types of intelligence that all humans possess to varying degrees. These included verbal/ linguistic, logical/ mathematical, bodily/ kinaesthetic (motor skills), visual/spatial (problem solving), Intrapersonal (self regulation) and interpersonal (social skills). He viewed the different intelligences as equally important; but acknowledged that in western societies verbal/linguistic and logical/mathematical are prized over the others, thus limiting holistic development. The emphasis on quantifying achievement means that skills that are easily assessed often take precedence in assessment due to other forms of learning being less amenable to measurement (Stenhouse, 1975). Gardner's work has, however, been heavily criticised for the subjective and arbitrary nature of his defined multiple intelligences (Sternberg, 1983), as well as the lack of empirical evidence to support the claims made (Visser, Ashton & Vernon, 2006).

Bracey (as cited in Strauss, 2011), stressed that there are many qualities that standardised tests cannot measure. These qualities include; creativity, organisation, critical thinking, resilience, persistence, curiosity, reliability, enthusiasm, empathy, self-awareness, self-discipline, leadership, courage, compassion, resourcefulness, honesty and integrity. Many of these virtues are of equal, if not greater value than academic capabilities, in developing well rounded students, who are prepared for the opportunities and responsibilities of adulthood. In 2012, the employment organisation The Confederation of British Industry called for students to be given a broader education. They suggested that schools have become 'exam factories' that fail to equip young people with skills needed for life and work (BBC News, 2012).

Although many of the issues discussed so far are well researched, very little of this research focuses on the perceptions of the students. Students are often overlooked when it comes to testing; despite them providing a valuable insight into the assessment system. The majority of the literature surrounding the perceptions of standardised testing focuses on the attitudes of teachers. Much of this research has raised concerns about the use of ST (Herman & Golan, 1991; Aydeniz, M., & Southerland, 2012). In a survey of teachers in the United Kingdom, only 7% of teachers agreed that end of key stage tests are educationally worthwhile (Neil, 2002). The Times Educational Supplement (TES) provides plethora of articles by teachers and educational researchers arguing for and against ST which can be used to gauge the attitudes of educational experts. However, there is no such opportunity for students' attitudes to be gauged, and thus, academic research has a responsibility to provide an avenue for students' insights which can be used by policy makers to enhance education.

The most comprehensive research into student perceptions of assessment was a meta-analysis by Struyvena et al. (2005) which focused on the views of students in higher education. Students were found to have negative attitudes towards traditional forms of assessment indicating a detrimental effect on the learning process (Sambell, McDowell & Brown, 1997). These perceptions were found to play a significant role in students' study behaviours and approaches to learning. The three main learning approaches identified were surface learning, deep learning and strategic learning. Surface approaches to learning involve little personal engagement and often entail routine memorisation with a limited understanding of learning material (Entwistle, McCune, & Walker, 2001). Conversely, deep learning approaches comprise of active engagement and the motivation to understand and master the learning material. Struyvena et al (2005) also identified an additional learning approach which has been produced by the shift towards ST. This approach was categorized as a strategic or achievement approach to learning involving well organised study methods to achieve the best possible grades whilst exerting the minimum amount of effort that is required to succeed. This often comes at the expense of mastering the learning material which is not a necessity for performing well in these exams.

The purpose of the current study was to investigate the extent to which the views found regarding ST were held amongst a larger student population. The attitudes being explored were grouped into three categories when to help address the main issues raised in both my earlier research as well as the issues highlighted

in the literature review. These categories were the foundation for each of the three main research questions for the study which are outlined below:

1. *Do students consider ST to be a valid and reliable method of assessing students' development?*

2. *What effect does ST have on the learning approaches adopted by students?*

3. *To what extent do students perceive ST as narrowing the curriculum and limit students' overall development?*

Method

Participants

Participants were obtained through a volunteer sample of undergraduate students who responded to an online survey circulated by email at Lancaster University and via facebook to students from a number of other universities in England. One hundred and sixteen students of varying ages responded, the majority of the participants were in the 18-22 year age range. Two respondents failed to disclose any of their demographic data as well as failing to answer some of the questions regarding ST. Consequently, these two participants who failed to provide responses for over a third of the questionnaire items were removed from the analysis. Thus, the analyses reported are based on 114 people: 85 women and 29 men.

Methods and Design

A survey was designed and administered online for ease of access, assured anonymity, and ease and accuracy of data collection. The survey was fairly quick to complete taking approximately 10 minutes. Participants were asked thirty six questions using Likert type responses relating to the three research questions (E.g. '*A single exam for each subject effectively assesses the knowledge I have gained throughout the course*'). The Likert type response options included; strongly agree, agree, neither agree nor disagree, disagree and strongly disagree. In addition to these items, respondents were also asked six questions requiring them to select one or more options from a list of four choices. These questions addressed the validity of testing; enquiring about which factors should be assessed in education and those that are reflected by ST. In terms of learning approaches, respondents were asked to choose which of four approaches they focused on the most, second most and third most when revising. Finally, the students were asked to provide some demographic information including their age, gender and the academic department in which they studied (see Appendix A for details of all of the items).

The order of the non-demographic survey items was randomised with approximately half of the Likert type items stated in favour of ST (with agreement representing positive attitudes towards standardised testing) and the other half stating a negative view regarding ST. Several of the items included were directly contrasting statements to help provide an indication of the reliability of the survey responses. All of these related pairs correlated highly (all correlations $r > 0.64$) suggesting that the respondents were considering their responses to each item rather than merely 'satisficing' (Krosnick, 1991); being less thoughtful in their comprehension and judgement of survey items. Satisficing can lead to potential bias, particularly if survey items are leading. As a result of controlling for both of these issues in this survey, the data collected should provide a valid representation of student attitudes.

Procedure

Prior to the main research, a pilot study was conducted using a small group of students from Lancaster University. The participants used for the pilot study consisted of three students from a psychology in education class and three students from different faculties at the University. This mixture of participants was used to provide information from students who had studied research methods and questionnaire design as well as from those who were representative of the average University student in regards to their knowledge of research methods. The selected participants were informed of the purpose of the study and asked to read through the initial survey questions, identifying any issues they had with any of the survey items. This helped to ensure the content validity of each survey item, and controlled for any unintentionally leading or ambiguous items. Participants were also asked to estimate how long the survey would take to complete and were encouraged to provide any additional feedback.

The responses proved to be extremely valuable in developing an unambiguous, balanced range of items for the finalised survey. It was also highlighted by the participants that the survey should take no longer than ten minutes in order to encourage people to complete the survey and maintain the interest and engagement of respondents when completing the survey. Consequently, a secondary phase of the pilot study was run in which two students were asked to complete the survey ensuring they considered their responses to each item. It was found that the survey took 9 and a half minutes on average which was just short enough to maintain attention based on the response from the first phase of the pilot study.

Respondents were informed of the general aim of the study before beginning the survey as it was anticipated that this would not affect their responses and would help in gaining their interest. Students were also assured that the data they provided would remain anonymous and that completion of the survey was completely voluntary. They were also informed that they could use the email address which was provided to ask any questions that they had regarding the study or findings. Prior to completing the survey a description of the term ‘standardised testing’ was provided to clarify which aspects of testing the term was being used to refer to. The survey was circulated to students from a number of Universities in England who were asked to answer a number of questions regarding their attitudes towards standardised testing. Participants were also asked to provide some demographic data with the majority of survey items focusing on one of the three research questions.

Having collected the data, a decision was taken to remove the impartial responses (‘neither agree nor disagree’) from the Likert scale data for the initial analysis of the questionnaire responses. This decision was made due to the lack of any additional information that these responses were providing. In the initial inspection of the data it was evident that the distribution of responses either tended to be fairly balanced between agreement and disagreement with few impartial responses, or weighted significantly towards either agreement or disagreement. None of the items had a majority of neutral responses, and so these responses were

removed in order to make the reporting and interpretation of these items easier. Thus all the reported proportions of each response are calculated from only the responses that were analysed. The impartial responses were, however, included in the factor analysis.

In addition, the responses to the survey item addressing learning approaches when revising were transformed into an overall score for each of the of the four approaches. For each respondent, the approach selected as the main focus of revision was given a score of 5, the second main focus scored 3, and the approach least focused on scored 0; with the remaining approach scoring 1. Combining the responses to these three survey items provided an overall score representing the extent to which each factor was prioritised by students during their revision. This measure was taken to allow for an easier comparison to be made taking into account all the responses for these items.

Results

The demographic data of respondents on the whole, illustrated that the sample was fairly representative in terms of University undergraduate students in the UK (Higher Education Statistics Agency, 2012). There was one main exception to this with the distribution of gender within the sample significantly skewed; almost three quarters of the sample were female (see table 1). The distribution of respondents by academic subject area was reasonably representative of subject intakes. However, there were a greater proportion of social science students in the sample than would have been expected based on department intake. Over half of the sample studied in the arts and social sciences department (see table 1). This may have been a consequence of the total number of females in the sample who are perhaps more likely to study in this department. Alternatively this may represent the greater variety and number of subjects available in the arts and social sciences category.

Table 1
Reported Gender and Department

	Revised Academic Department				
	Arts & Social Sciences	Business & Management	Health & Medicine	Maths & Science	Total
Male	11	7	0	11	29
Female	47	14	9	15	85
Total	58	21	9	26	114

The majority of respondents were in either their first, second or third year of University with over 95% of the sample aged between 18 and 22 years of age. Although certain demographic populations were severely under-represented in the sample, these respondents were still included in the analysis. The decision to include these respondents was made as it was not deemed to be detracting from the main purpose of the study, in exploring whether concerns about ST were widely held amongst students. Nevertheless, there were not enough representatives from all of the categories within each demographic to allow for exploration or comparisons to be made on this basis. Thus this data was only used to provide some background information for the sample. The reported use of ST at different levels of education yielded similar results across each of the educational stages. Approximately 90% of the sample reported having been assessed through ST during their time in primary education, college, and university, and only 3 of the 114 respondents reported not having been assessed by ST during their experience of secondary school.

There was a general sense of agreement from students regarding the use of ST in providing data that can be used to compare students. Almost 90% of respondents agreed to some extent with the value of ST in providing this function. In addition, 85% of respondents agreed to some degree that the tests used for GCSE's and A-Level's are challenging yet achievable for the majority of students. However, the majority of students agreed that exams fail to effectively assess students with differing abilities. This suggests that although students view exams on the whole as

being challenging and achievable for most of the students who take them, testing all students with the same exam and via a single method may fail to provide an accurate representation of every pupil's abilities.

In terms of the perceived validity of testing, Students were asked which two academic factors are the most important outcomes of education, and which two are assessed the most by ST. Student attitudes towards what education should aim to assess conclusively identified understanding and effort as the most important outcomes. Only one respondent did not select understanding as one of the main two aims of education, and 75% of students selected effort as one of their responses (see Figure 1). In contrast to these findings, memory and exam technique were viewed as the two factors that were best reflected by exam results (see Figure 2). Memory was selected by 77% of respondents as one of the two factors best reflected by exam results, with the second most common response being exam technique (72%). Understanding was selected as one of the two responses by only 39% of students, and effort was selected by only 10% of respondents. A similar, but less extreme pattern was found in the responses to which single factor is most accurately represented by exam performance; memory (50%) and exam technique (35%) were selected by over 85% of students.

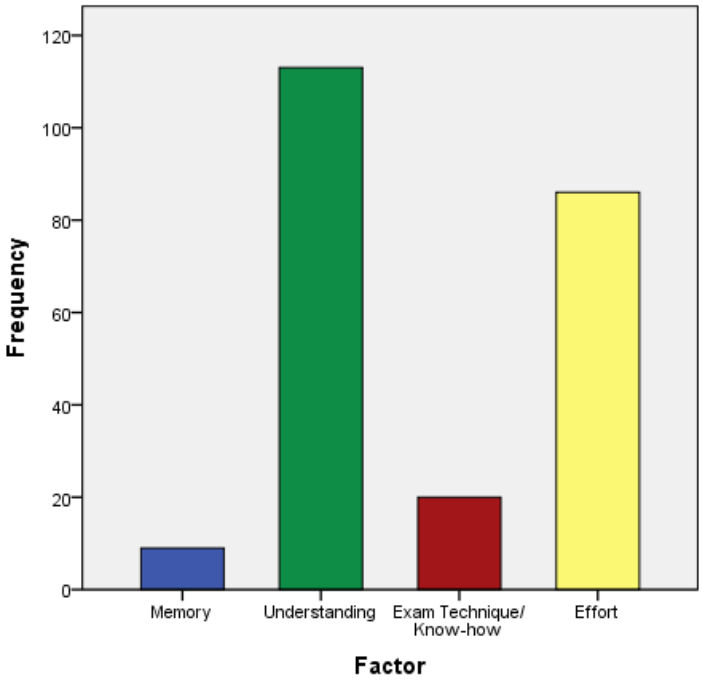

Figure 1: Bar chart showing student perceptions of what education should aim to assess?

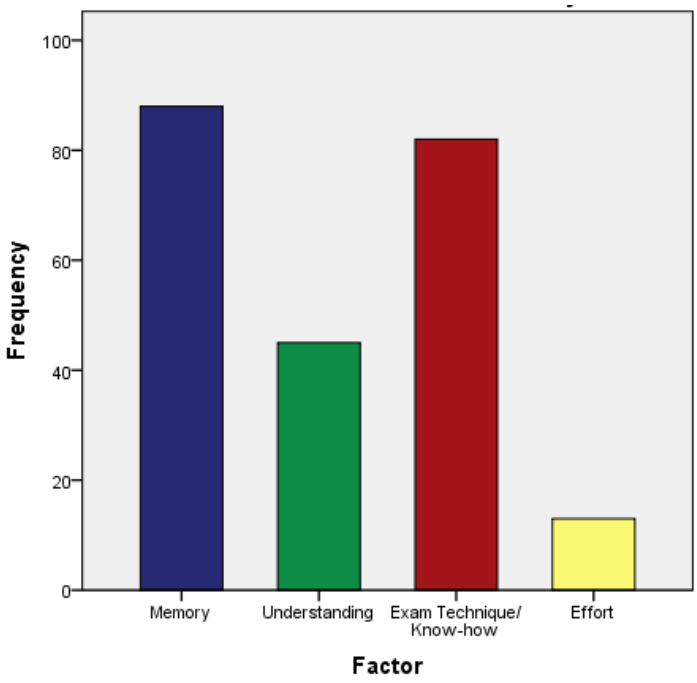

Figure 2: Bar chart showing student perceptions of what standardised testing assesses?

In regards to the use of a single test occasion for administering standardised exams, 85% of respondents agreed to some extent (with 35% selecting strongly agree) that a single exam is not enough to assess what is learnt in a year. This was also reflected in the item addressing this issue from the opposite direction with 80% of participants disagreeing (17% strongly disagreed) that a single exam for a subject effectively assesses the knowledge gained throughout a year. Furthermore, respondents also reported that their understanding was poorly reflected by exam performance, with coursework and classwork being viewed as more accurate measurements of understanding. One particularly interesting finding was that 51% of students moderately agreed and 7% strongly agreed that they could perform well in ST without having a deep understanding of the subject.

In terms of non-academic factors which can impact test performance, 33% of students strongly agreed and 48% moderately agreed that test anxiety has a detrimental impact on their test performance. Similar results were found in terms of the responses to items addressing similar issues; 89% of students reported both feeling anxious during exams and finding exam periods very stressful. In addition, 50% of students reported that they do not have enough time in exams to complete answers to the best of their ability, and 63% agreed to some extent (28% of these agreed strongly) that they had often been sleep deprived when taking standardised tests.

In relation to the effect of ST on student learning approaches, 91% of respondents said the main purpose of the examination process is to gain a good grade, whilst only 40% of students viewed exams as an opportunity to learn and improve. In response to the items addressing the conflict between mastery and

superficial learning, 93% (32% of whom selected strongly agree) reported focusing on what will be on an exam rather than gaining a deep understanding of a whole topic or enjoying learning. Moreover, 55% of respondents moderately disagreed, and 25% strongly disagreed, that they retain information for a long time after an exam.

In terms of the approaches undertaken in revising for exams, a score was calculated for each of the four learning approaches. This was done using a scoring system by which the approach selected as the highest priority when revising scored 5 points, the second most prioritised scored 3 points and the third most prioritised scored 1 point. This provided a total score for each approach which indicated the extent to which students focus on each approach when revising; these scores are displayed in Figure 3. The findings indicated that memory was prioritised the most, although there was evidence that students also focus on understanding information more than organising information or using past exam papers to prepare for tests (see Figure 3).

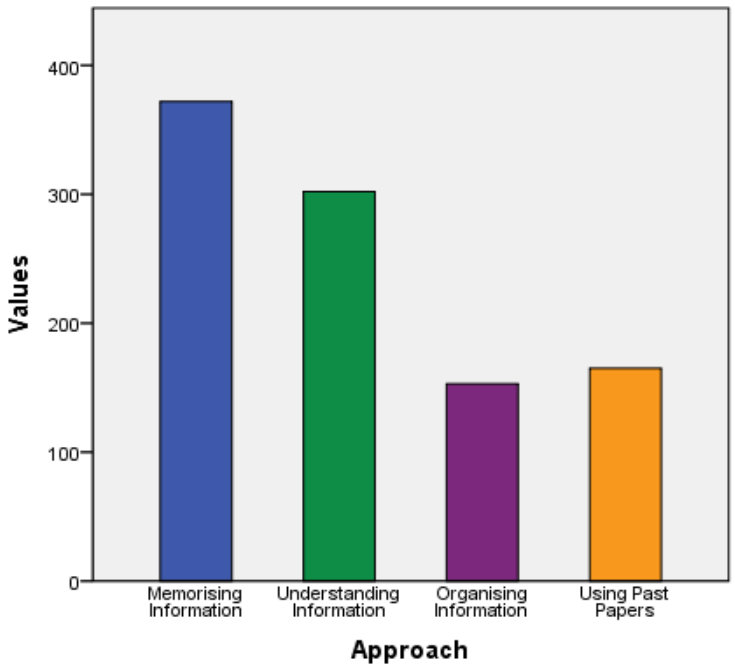

Figure 3: Bar chart showing the use of learning approaches in revision for ST

In terms of the extent to which ST narrows the curriculum and limits pupils' overall development, there was a stark contrast between the attitudes surrounding what the aims of education should be and what is assessed in reality. Around 95% of students agreed to some extent that education should aim to promote the spiritual, moral, cultural and physical development of students, and 99% agreed that education should help prepare young people for the opportunities, responsibilities and experiences of adult life (37% and 50% of respondents, respectively, strongly agreed with these curriculum aims). However, in terms of the realistic promotion of

these aims, 86% and 80% of students, respectively, disagreed to some extent that exams help to promote these broader aims of education.

Despite finding that 93% of students acknowledged that non-academic skills are equally important to academic skills, a similarly high proportion of students thought that ST prioritised academic achievement over non-academic skills. It was also found that 88% of students disagreed to some extent that exam focused education provides sufficient opportunities for the development of these non-academic skills. Furthermore, 92% of respondents indicated that ST fails to consider skills that are less convenient to measure through written tests. Finally, 68% of students said that the amount of ST in education should be reduced, and an even more resounding 88% of respondents agreed that the focus on ST in education should be diluted to focus on the other skills and values.

In addition to these findings, an exploratory factor analysis was run in order to examine whether the variation in attitudes could be accounted for by particular collections of attitudes. Table 2 reports the results of the principal components analyses with varimax normalized rotation. In selecting the variables for the analyses, only items which used likert type responses were included due to the level of data being ordinal compared to nominal for the excluded items.

Table 2
Initial and Rotated Eigenvalues and Rotated Explained Variance

Component	Initial Eigenvalues	Rotated Loadings	
	Total	Total	% of Variance
1	9.081	5.088	14.133
2	2.815	4.266	11.850
3	2.173	3.997	11.102
4	2.044	2.762	7.672

The analysis (see Table 2) identified four components which had eigenvalues of greater than two which was deemed an appropriate cut off point. This decision was aided by observing the scree plot of factors, as well as considering the impracticality and potential overlaps in attempting to define the ten factors with eigenvalues greater than one. Once this decision had been made, varimax rotation was used to maximise the variance explained by the four factors. Table 2 reports the initial eigenvalues for each factor as well as the percentage of the variance explained by each factor in the rotated model, and the rotated eigenvalues. Table 3 presents these four factors, reporting the Cronbach's alpha, and percentage of the variance explained by each factor. The items loading heavily on each factor (loading of .5 or greater, or -.5 or lower) are also reported.

Table 3
Item Composition of Student Attitude Factors

Item	Loading
Factor 1: Valuable skills measured by ST (14.1%), *α*=.81	
I gain a deep understanding of the material that I revise when preparing for exams	.63
Exam-focused education provides sufficient opportunities for the development of non-academic skills	.62
I retain the information that I have learnt for an exam for a long time after I have completed the exam.	.53
Exams help to promote the spiritual, moral, cultural, mental and physical development of students	.58
I view exams as an opportunity to learn and improve	.58
The main purpose of the examination process is to gain a good grade	-.63
Focus on test results encourages superficial learning	-.62
I focus on learning what will be on an exam more than on gaining a deep understanding of the topic and enjoying learning	-.56
I can perform well on standardised exams without having a deep understanding of the subject	-.52
Factor 2: Validity of test occasions (11.9%), *α*=.83	
Coursework usually provides a more accurate reflection of students' understanding of a topic compared to testing	.76
Coursework and classwork allow me to demonstrate my knowledge more effectively than in exams	.75
A single exam for each subject is not enough to assess what I have learned in a year	.58
Coursework provides a more realistic reflection of the challenges in later life than standardised testing	.57
My understanding of a topic is more accurately represented through examination rather than coursework	-.73
A single exam for each subject effectively assesses the knowledge I have gained throughout the course	-.52
Exam performance is a true reflection of a student's understanding of a subject	-.50
Factor 3: Confounding variables (11.1%), *α*=.77	
Exam periods are a very stressful time for me	.71
I feel anxious during exams	.63
I feel rushed in exams due to the time limits	.61
Measuring knowledge should be the sole focus of assessment in education	.60
I have often been sleep deprived when taking standardised tests	.55
Exams fail to assess effectively the level of students with differing abilities	.55
I feel relaxed whilst sitting exams	-.61
I have enough time in exams to complete the answers to the best of my ability	-.51
Development of non-academic skills such as social and emotional skills is equally important to academic development	.61

Factor 4: Need for broader assessment (7.7%), *α*=.65	
Education should prepare young people for the opportunities, responsibilities and experiences of adult life	.59
Educational assessment fails to consider skills that are less convenient to measure through written tests, such as initiative and creativity	.54
Education should promote the spiritual, moral, cultural, mental and physical development of students	.54
Education should dilute its focus on testing to increase its focus on the development of other skills and values	.51

The four factor model accounted for 45% of the variance in student perceptions of standardised testing. The first component received high loadings on items indicating that ST encouraging mastery and deeper understanding of learning material, with substantial loadings for items suggesting education provides ample opportunity to develop non-academic skills in addition to academic competence. I have named this component; *Valuable skills measured by ST*. The second factor received very high loadings from items suggesting coursework provides a better reflection of understanding, and substantial loadings on items addressing the validity of testing on a single occasion. I have named this component *validity of test occasions*. The third factor is received high loadings from items regarding the effect of test anxiety, stress and sleep deprivation with substantial loadings from items indicating time constraints hinder test performance. This factor has been named *confounding variables.* The fourth factor received substantial loadings from items regarding the need for education to assess a broader range of skills. These items addressed the need to assess less academic skills that are less easily measured as well as the necessity to prepare students for adulthood. This final factor was named *need for broader assessment.*

Cronbach's alpha statistics were calculated for each of the four factors to assess their internal reliability. These statistics reported good reliability for the first two factors (*α>.8*) and was acceptable for the *confounding variables* factor (9 items, *α*=.77). The Cronbach's alpha for the fourth factor, *need for broader assessment,* indicated only marginally acceptable reliability (4 items, *α*=.65) which may be due to the range of different issues addressed by this factor with only four items loading heavily (>.5).In addition, the items loading heavily on each factor were used to calculate the mean scores of each factor. This was done by totalling the average likert values for each of the positive loadings and the reversed negative loadings for each factor. The totals were then divided by the number of loadings to create an average likert score between 1 and 5 (1= strongly agree, 5= strongly disagree). This provided an indication of the overall tendency for agreement or disagreement with the items associated with each factor. Mean scores closer to 1 indicated agreement with the positive loadings and disagreement with the negative loadings for each factor; whereas values closer to 5 indicated the opposite pattern. One Sample T-tests were then run for each of the four factors to determine whether the average scores were significantly different from the neutral value (3= neither agree nor disagree). The results of these t-tests are shown in table 4 reporting the t values, means, standard deviations and significance values for each factor.

Table 4
T-test statistics for the four factors explaining student attitudes towards ST

Factor	*t*	Mean	Standard Deviation	Significance Value
Skills measured by ST	10.09	3.58	.60	<.01
Validity of test occasions	-9.51	2.35	.72	<.01
Confounding variables	-10.67	2.4	.59	<.01
Need for Broader Assessment	-19.53	1.96	.56	<.01

Table 4 indicates that respondents tended to slightly disagree that ST measures valuable skills. The findings also suggest an overall agreement in the sample in relation to the invalidity of ST, as well as the negative impact of confounding variables on testing. Finally, the results show that students moderately agreed with the need for broader assessment. All of the factors were found to show highly significant differences between the mean scores and the neutral value (all significant at the $p<.01$ level).

Discussion

The findings of this study have raised some key issues in relation to the value and use of ST from the perspective of students who have experienced this kind of testing throughout their time in education. In regards to the methods used for ST, the use of a single test occasion was largely viewed by students viewed as invalid and unreliable. Exams were also not viewed as an accurate reflection of the demands likely to be faced in adulthood, with coursework being considered a more reliable approach to assessment. There was also evidence from the students' perceptions emphasising the detrimental effect of the confounding variables that can influence test performance. Generally, around 80% to 90% of students raised concerns about the influence of anxiety and stress, with a considerable number of students also subject to sleep deprivation when taking exams.

Students generally felt that there were insufficient opportunities for the development of non-academic skills. This may be partly explained by the observation by students that education currently ignores skills that are less convenient to measure. The findings suggest that ST plays a significant role in the goal orientation and learning strategies that are adopted by students, with convincing evidence that ST encourages superficial learning at the expense of mastery. Furthermore, over 90% of students agreeing indicated that they focus on what is likely to be on an exam rather than mastering a subject or enjoying learning. It was also alarming that over 75% of students reported that they do not retain information learnt for a long time after the exam.

Over 90% of students supported the curriculum aims which promote the spiritual, moral, cultural and physical development of students, as well as preparing them for the opportunities, responsibilities and experiences of adult life. However, in terms of promoting this holistic development, students felt that exams were ineffective in doing so. The student attitudes provided conclusive evidence in favour of reforming ST. Almost 70% of students agreed that the amount of ST should be reduced, whilst approximately 90% of students supported the need to dilute the focus on testing in order to broaden young people's development.

The factor analysis revealed an intriguing pattern of student attitudes with four main underlying issues identified. These included the skills measured by educational assessment, the validity of ST, the impact of confounding variables and the need for broader assessment in education. It appears that the majority of students tended to slightly disagree that that ST measures valuable skills. There were also negative perceptions regarding the validity of ST with concerns being raised over the ways in which tests obtain samples of a student's knowledge and ability. Furthermore, the influence of confounding variables on test performance was acknowledged, highlighting a general consensus amongst students that issues such as stress, anxiety, time constraints and sleep deprivation can all influence performance on standardised exams. There was also a general agreement amongst students that there is a need for education to promote a wider range of skills and values, preparing students for adulthood. Students acknowledged that ST prioritises academic skills at the expense of these broader skills.

It is clear from previous research that testing is an essential component of education. The use of ST is extremely valuable in providing teachers, parents, and the general public with important information about the academic performance of students (House of Commons, 2008). Although the student perceptions of ST are fairly negative, the main issues that were highlighted related to how the current system misuses ST rather than an inherent negativity towards testing itself. Students highlighted that there are several confounding factors that can influence exam performance; however, it is reasonable to assume that factors such as anxiety and sleep deprivation would be less frequent if educational achievement was not solely dependent on test performance for so many high stakes purposes.

In order to effectively evaluate the use of ST one must consider the purpose of education considering that ST is the predominant form of assessing pupils' development. The process of education aims to develop skills, increase knowledge and understanding, and develop the competences of young people (TGAT, 1987). The five outcomes suggested by the Every Child Matters Policy (DfES, 2003); as well as the aims of education outlined by the national curriculum (ERA, 1988), emphasise the need for a broad spectrum of skills and values to be promoted by schools. These skills go beyond academic achievement to encouraging good physical and mental health, promoting moral and cultural development, and enjoyment of education. There also needs to be a shift towards a more equal balance between the development of academic skills and the cultivating of 'immesurables' (Toshalis & Nakkula, 2012). The ultimate goal of this broad approach is to produce well rounded young people who are prepared for the responsibilities and opportunities of adult life and are able to make positive contribution and achieve economic well-being.

Despite these goals being in place within official documentation, in reality, the shift towards ST has made addressing each of these skills and values extremely difficult. Generally speaking the aims that students identified as the most important reflect those that are documented in educational policy. However, the high stakes that are attached to testing undermine these aims by holding pupils, teachers and schools accountable predominantly on the basis of exam results. This emphasises the disparity between the planned curriculum and the curriculum as experienced by students which differs significantly (Struyven et al., 2005).

Students evidently feel that understanding and effort are more valuable academic skills than memory and exam technique and should be prioritised in appraising student achievement. However, in practice, the material that is assessed directs what is taught (Brady, 2012), and the format of exams become the basis for teaching (The National Centre for Fair and Open Testing, 2007). The high stakes attached to testing encourages adverse learning approaches amongst students which are encouraged by teachers' focus on exams. These approaches include rote memorisation and superficial learning strategies rather than mastery and thus students often fail to retain what they have learnt beyond an exam. This limits the value of testing for all students including those who perform well. The adoption of these approaches discourages the development of self-regulated learners who are capable of higher order thinking skills. This leaves many students unable to solve problems in real-world situations (HC, 2008a). The negative attitudes reported in this study support Struyven et al.'s (2005) suggestion that high achievement can often

mask students' dissatisfaction with the assessment system, as well as a lack of understanding or mastery of the test material.

There is no doubting the learning value that testing can have if used in the right way. Spitzer's work (1939) on immediate recall and the retention of learning provided an early indication of this with more recent research having developed our understanding of practice and recall (Epstein, Epstein & Brosvic, 2001; Morris et al., 2005). However, the majority of students reported that they fail to retain examined material in the long term and view testing as means of obtaining a good grade rather than an opportunity to learn. This indicates that testing may direct learning away from approaches that foster long term retention of learning material.

The main function of testing within the current education system is to obtain a sample of an individual's knowledge and abilities; thus enabling generalisations to be made about that individual's knowledge and ability as a whole. The findings suggest that the majority of students stress Whitten's (2011) suggestion that tests are rarely viewed as an opportunity for learning. This may be due to an overemphasis on the summative functions of ST. The findings of the current study indicate that students hold many of the same concerns that have been highlighted in the educational literature regarding the validity and reliability of the generalisations made from ST. The confirmation of these concerns brings into question the use of ST in providing school accountability.

The aim of ST for school accountability is to provide a fair, straight-forward measure of school performance, thus helping to maintain standards. However, policy makers tend to overestimate what tests can do (Koretz, 2008). Externally set, standardised exams can only cover either a small area of the curriculum in detail, or a broad area very lightly (HC, 2008b). Tests are not designed to reflect everything that students have learned; they only provide a sample of knowledge to help estimate a student's overall achievement. It is therefore argued that it is inappropriate to use these estimates by themselves in evaluating schools or students.

Policy makers, as well as society in general, appear to be so desperate for schools to be accountable for young peoples' learning, that too much emphasis is placed on test performance. This is due its capability to provide nationally comparable data of children and schools, which encourages teachers to 'teach to the test' and focus on superficial learning strategies rather than mastery and student motivation and enjoyment (which can enhance learning). This may produce better test results, but provides a biased sample of students' knowledge that does not reflect their overall understanding and mastery of a subject. Students acknowledge the importance of measuring higher level thinking over rote memorisation; however, high-stakes testing encourages students to focus on the latter in order to be successful in this kind of assessment.

In 2007, of the 4158 schools in England that had students taking GCSE exams, the correlation between the proportion of students in a cohort achieving a grade C in at least 5 subjects and the contextualised value added (CVA) measure was 0.27 (Wiliam, 2010). The CVA indicates the contribution of individual schools to their students' performance. This is calculated by comparing the performance of

students to their preceding performance, at a previous school for example. The low correlation suggests that the school attended contributes only around 7% of the variance in student outcomes indicating that the use of exam results as a means of school accountability is unfit for this function. Policy makers must find a balance between providing valid and reliable measures of achievement, ensuring accountability, facilitating learning, and achieving public understanding, confidence and trust (HC, 2008b).

In addition to the shortfall of ST in providing a valid measurement of student ability, it also takes attention away from the development of less academic skills and preparing students for adulthood. The five desired outcomes for young people (DfES) encompass a broad spectrum of qualities from being healthy and safe to achieving economic well-being. The findings from this study indicate that exams focus predominantly on measuring academic achievement; although even its use for that function has validity issues. This creates an imbalance between the development of academic skills and the development of non- academic qualities such as creativity or enthusiasm for learning. This is a result of the high stakes attached to testing which direct the attention of schools and teachers onto exams rather than broader student development.

School performance must be viewed in the wider context rather than merely focusing on achieving the highest possible test results. Test results are not a valid measure of school performance, because they only measure one narrow aspect of it. Assessment should encompass a broad range of skills both academic and non-academic to encourage students overall development. Therefore, the measures used in assessing the quality of schools must acknowledge the extent to which schools meet the broader aims of education. Schools should be evaluated in terms of how well they foster the development of well rounded students and prepare them for the opportunities and responsibilities of adulthood.

Despite the fairly conclusive evidence, that from the perspective of students there is a need for a change in the way students are assessed, proponents of ST may still argue that the current assessment methods are successful in raising the standard of pupil achievement. This stems from increases in test scores which have been observed following the introduction of high-stakes testing (GTCE, 2004). However, this can be largely attributed to teachers and students becoming more familiar with the test requirements and narrowing the curriculum, rather than the improvements reflecting any actual change in the quality of students' learning. Nevertheless, ST is to a certain extent necessary and serves important function in providing nationally comparable information as well as summarising students' achievement and progress for students themselves, teachers, and parents, as well as institutions that may use this information to inform their selection of students for a course or job.

The externally-prescribed national assessment tests cannot assess pupils' achievement of all of the set attainment targets. It is only through teachers' appraisal, over an extended period of pupils' progress, that comprehensive evidence can be created (TGAT, 2008b). Yet currently, teachers judgements do not contribute towards any formal assessment of their students'. Teachers must work within the structures and limitations set by schools and educational policies. This limits the

extent to which they can use assessment effectively as a learning tool despite them being the only people whose actions directly affect students (Harlen & Deakin Crick, 2003).

The concerns raised by this study emphasise the necessity for the high stakes attached to ST to be reviewed and reformed if we wish to provide students' with the best chance of succeeding in education. The findings from this study suggest that most students are inhibited by ST some degree, whether it be because of test anxiety, or the shift towards superficial learning over mastery. A more valid approach is to use a wider range of assessment tasks involving a variety of contexts, and a range of response formats and styles. This provides more opportunities for pupils to demonstrate their ability, even if they are disadvantaged by any one particular assessment method. This approach gives them the best chance of producing their best work (Linn, 1992; Gipps, 1994). Furthermore, Gipps (1994) suggests that assessment should strive for conditions that are not threatening in order to reduce stress and provide conditions that help students perform to their full potential.

The national assessment system must therefore minimise the amount of information to be collected while maximising confidence in its interpretation (TGAT, 1987). The key to achieving this is placing more trust in the judgements of teachers. A reform could assess a broader range of skills, over a longer period of time (through teacher observations), providing more opportunities to demonstrate ability. Thus, the assessment of student development would be much less stressful as well as being far more valid and reliable.

Limitations

The purpose of this research was to document the perceptions of students regarding the use of ST in education. The sample was generally reasonable for this purpose: it was drawn from a population of young students living in Britain. However, the composition of the sample was predominantly female with three times as many females as males. The volunteer nature of the sample resulted in disproportionate representation of the gender groups, as well as some minor bias in some of the other demographic distributions such as ethnicity. Therefore the findings may be skewed towards white, female perceptions of testing. The purpose of this study was to evaluate the extent to which the attitudes towards ST identified in earlier research (Taylor, 2012) were held widely by students. Although there were some disparities within the sample, it was still fairly representative and sufficient in fulfilling the aim of the study.

The other main issue of the sample was the fact that all of the respondents were university students. This demographic are likely to have succeeded throughout their education by performing well in exams, considering they have progressed on to study at degree level. They are therefore more likely to hold positive views towards ST than the average person of their age group. The responses may also be subject to the social desirability effect with these students more likely to wish to maintain the status quo (Krosnick, 1991). It would also be reasonable to assume that successful students may not wish to undermine the educational tools which have facilitated their

success. Criticism of ST from these students could bring into question whether their own achievements are due to the flaws of testing rather than genuine excellence. However, this population of interest was selected intentionally as it was deemed that by selecting this target population, it would make the evidence even more compelling if the concerns surrounding ST were confirmed by the demographic that had the least motivation to criticise it.

Another limitation of the study was the limited number of issues that could be covered by the online survey. Given that the current research was unfunded and thus could not provide incentive for completion of the survey, the survey had to be short enough to hold the attention of un-incentivised respondents. Krosnick (1991) described how students can become fatigued, disinterested and distracted fairly quickly when responding to online surveys. Therefore, it was impossible to address all of the issues that have been raised surrounding ST in educational literature and ensure the quality of the data. The most significant issues have, however, still been discussed within this paper. One way in which the survey could have gained a more comprehensive insight, whilst still being manageable for students, would have been to narrow the focus of the research to one specific issue regarding ST. However, the lack of previous research into student perceptions of testing made it difficult to choose a single issue to focus on. It was necessary for this study to take a broad approach in gauging student perceptions in order to provide information that can inform future research, and facilitate more comprehensive exploration of the student perceptions towards specific issues in this area.

Future Research

The current study has provided a foundation from which further investigation into the use of standardised testing can build on and use to direct future exploration. Future research should aim to develop our understanding of testing from the perspective of students' who provide the most valuable insight into this area. The alternatives to ST should also be focused on to explore the most effective ways in which the assessment system can be improved. This study has also provided a tool which can be used in future research. The four factors identified by the factor analysis may be a particularly useful basis for directing further exploration. These factors can be used to explore the attitudes of students from different countries and education systems, as well as high and low performing students. This will enable researchers to identify whether the same factors drive the perceptions and attitudes of students from these different groups.

Conclusion

The current research aimed to assess the perceptions of students in relation to the validity of ST; the influence of ST on students' learning; and the extent to which ST aids the overall development of students. Testing undoubtedly has the potential to enhance learning if used in the right way. Nevertheless, we need to temper our expectations of the functions for which testing can be used effectively. Students generally acknowledged the value of testing as a tool for providing

comparable data; however, there are limits to the meaning that can be derived from standardised test scores. Therefore, the way in which testing is used needs to be changed if it is to be utilised as a pedagogical tool. Education must acknowledge what ST can and cannot do, and use it only for the functions that it can realistically perform.

The findings show compelling evidence that the negative impacts of high-stakes testing are widely acknowledged by students. The need for reform is evident given that student attitudes confirm the concerns that have been raised in my previous research, as well as concerns raised by educational professionals. These concerns were raised and warned against by TGAT (1987) in their initial advice to the government; although the advice regarding such issues has since been ignored during the shift towards high-stakes testing. The main functions of testing include measuring pupil attainment, teacher and school accountability and monitoring national standards. However, the evidence clearly shows that a single set of tests cannot validly achieve all of these purposes simultaneously. National testing system should be reformed to remove perceptions that it is imperative to pursue test results at all costs. This will remove some of the pressure that contribute to many of the validity issues surrounding testing, as well as providing more opportunities to implement a broader curriculum.

References

Anderson, L. (1994). Research on Teaching and Teacher Education. In Anderson, L. & Sosniak, L. A., (eds.), *Bloom's taxonomy: a forty-year retrospective*. Chicago IL: University of Chicago Press.

Ashcraft, M. H., & Ridley, K. S. (2005). Math anxiety and its cognitive consequences: A tutorial review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 315-327). New York: Psychology Press.

Assessment Reform Group (ARG). (2002). *Testing, Motivation and Learning.* Cambridge: University of Cambridge Faculty of Education.

Aydeniz, M., & Southerland, S. A (2012). A national survey of middle and high school science teachers' responses to standardized testing: Is science being devalued in schools? *Journal of Science Teacher Education, 23*(3), 233-257.

BBC News. (2012). *CBI Complains of 'Exam Factory' Schools.* Retrieved from, http://www.bbc.co.uk/news/education-20355664.

Belden Russonello & Stewart Research and Communications. (2000). *Making the Grade: Teachers' Attitudes toward Academic Standards and State Testing*. Washington, DC: Education Week. Retrieved from http://www.edcounts.org/archive/sreports/qc01/pdfs/qcresearch.pdf.

Benmansour, N. (1999). Motivational orientations, self-efficacy, anxiety and strategy use in learning high school mathematics in Morocco. *Mediterranean Journal of Educational Studies, 4*, 1-15.

Black, P., & Wiliam, D. (2003). In praise of educational research: formative assessment. *British Educational Research Journal 29*(5), 623–637.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5* (1), 7–74.

Bloom, B. S. (ed). (1985). *Developing Talent in Young People*. New York: Ballentine Books.

Bloom, B., Engelhart, M., Furst, E., Hill, W. & Krathwohl, D. (eds.). (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I - Cognitive Domain.* New York: McKay.

Bloom, B. S., Hastings, J. T. & Madaus, G. F. (Eds). (1971). *Handbook on the Formative and Summative Evaluation of Student Learning.* New York: McGraw-Hill.

Brady, M. (2012, March 22). What standardized tests should assess. *Washington Post*. Retrieved from www.washingtonpost.com/blogs/answer-sheet/post/what-standardized-tests-should-assess/2012/03/11/gIQAJzDfSS_blog.html.

Brophy, J., (1998), *Motivating Students to Learn*, Boston MASS, Mc Graw Hill.

Clark, I. (2012). Formative Assessment: Assessment Is for Self-regulated Learning. *Educational Psychology Review, 24*(2), 205-249.

Coolican, H. (2009). Introduction *to Research Methods and Statistics in Psychology* (5th edition). London: Hodder & Stoughton.

Covington, M. (2000). Goal theory, Motivation and School Achievement: An Integrative Review. *Annual Review of Psychology, 51*, 171-200.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*, 438–481.

Darling-Hammond, L., & McCloskey, L. (2011). *Assessment for Learning Around the World: What Would It Mean to Be 'Internationally Competitive?* Retrieved from http://standardizedtests.procon.org/sourcefiles/assessment-for-learning-around-the-world-what-would-it-mean-to-be-internationally-competitive.pdf.

Davis, D. S., & Neitzel, C. (2011). A self-regulated learning perspective on middle grades classroom assessment. *Journal of Educational Research, 104*(3), 202-215.

Davies, J., & Brember, I. (1998). National curriculum testing and self-esteem in year 2 the first five years: a cross-sectional study. *Educational Psychology, 18*, 365–375.

Department for Education. (2011). *The Framework for the National Curriculum: A Report by the Expert Panel for the National Curriculum review*. London: Department for Education.

Department for Education and Skills (DfES). (2003). *Every child matters*: *Presented to Parliament by the Chief Secretary to the Treasury by Command of Her Majesty, September 2003.* London: Her Majesty's Stationery Office.

Department for Education and Skills (DfES). (2004). *The Children Act: report 2003.* London: Department for Education and Skills.
Education Reform Act 1988, source:
http://www.legislation.gov.uk/ukpga/1988/40/section/1/enacted.

Entwistle, N., McCune, V., & Walker, P. (2001). Conceptions, styles, and approaches within higher education: analytical abstractions and everyday experience. In Sternberg and Zhang (Eds.), *Perspectives on cognitive, learning and thinking styles* (pp. 103-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Epstein, M.L., Epstein, B.B., & Brosvic, G.M. (2001). Immediate feedback during academic testing. *Psychological Reports, 88*, 889-894.

Gardner, J. W. (1963). Self-Renewal. New York: Harper & Row.

Gardner, H. (1983). *Multiple intelligences: the theory in practice*. New York: Basic Books.

General Teaching Council for England (GTCE). (2004). *Perspectives on pupil assessment: a paper presented to the GTC conference: New relationships: teaching, learning and accountability*. Retrieved from http://dera.ioe.ac.uk/14022/1/1104_Perspectives_on_Pupil_Assessment._New_Relationships__Teaching%2C_Learning_and_Accountability.pdf.
Gipps, C. (1994) Developments in Educational Assessment: what makes a good test? *Assessment in Education, 1*(3), 283-291.

Harlen, W., & Deakin Crick, R. (2002). A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning. In Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Harlen, W., & Deakin Crick, R.(2003). Testing and Motivation for Learning. *Assessment in Education, 10*(2), 169-207.

Herman, J.L., & Golan, S. (1991). Effects of standardized testing on teachers and learning- Another look*. CSE Technical Report #334*. Los Angeles: Centre for the Study of Evaluation.

Higher Education Statistics Agency. (2012). *Statistics - Students and qualifiers at UK HE institutions*. Retrieved from http://www.hesa.ac.uk/content/view/1897/239/.

House of Commons: Children, Schools and Families Committee (HC). (2008a). *Testing and Assessment: Third Report of Session 2007–08: Volume 1*. London: Stationery Office.

House of Commons: Children, Schools and Families Committee (HC). (2008b). *Testing and Assessment: Third Report of Session 2007–08: Volume 2.* London: Stationery Office.

Jackson, C. (2003). Motives for 'Laddishness' at School: Fear of failure and fear of the 'feminine'. *British Educational Research Journal, 29*(4), 583-598.

Jeffrey, B. (2002). Performativity and primary teacher relations. *Journal of Education Policy, 17*(5), 531-546.

Kelly, M. (2004). *The Rhythm of Life: Living Every Day with Passion and Purpose*. London: Beacon Publishing.

Koretz, D. (2008). *Measuring Up: What Educational Testing Really Tells Us.* Cambridge, MA: Harvard University Press.

Krathwohl, D. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory into Practice, 41*(4), 212-218.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236.

Lam, R. (2013). Formative Use of Summative Tests: Using Test Preparation to Promote Performance and Self-Regulation. *The Asia-Pacific Education Researcher, 22*(1), 69-78.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.

Linn, M. C. (1992). Gender differences in educational achievement. In J. Pfleiderer (Ed.), *Sex Equity in Educational Opportunity, Achievement and Testing* (pp. 11-50). Princeton, NJ: Educational Testing Service.

Main, K. L. (2008). Teaching beyond the Test: The Possibility of Success. *The English Journal, 97*(95), 46-51.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*, 3rd ed., (pp. 13-103). New York: Macmillan.

Morris, P. E., & Fritz, C. O. (2000). The name game: using retrieval practice to improve the learning of names. *Journal of Experimental Psychology: Applied, 6*, 124-129.

Morris, P. E., & Fritz, C. O. (2002). The improved name game: better use of expanding retrieval practice. *Memory, 10*, 259-266.

Morris, P. E., Fritz, C. O., Jackson, L., Nichol, E., & Roberts, E. (2005). Strategies for learning proper names: expanding retrieval practice, meaning and imagery. Applied Cognitive Psychology, 19(6), 779-798.

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist, 22*, 155–175.

Neil, S. (2002). *National Curriculum Tests: A survey analysed for the National Union of Teachers*. Warwick: University of Warwick Institute of Education/NUT.

Pollard, A., Triggs, P., Broadfoot, P., McNess, E., & Osborn, M. (2000). *What Pupils Say: changing policy and practice in primary education.* London: Continuum.

Reay, D. & Wiliam, D. (1999). 'I'll be a nothing': structure, agency and the construction of identity through assessment. *British Educational Research Journal, 25*, 343–354.

Roderick, M., & Engel, M. (2001). The grasshopper and the ant: motivational responses of low achieving pupils to high stakes testing. *Educational Evaluation and Policy Analysis, 23*, 197-228.

Sambell, K., McDowell, L., & Brown, S. (1997). 'But is it fair?': an exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation, 23* (4), 349-371.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641-657.

Stenhouse, L. (1975). *An introduction to curriculum research and development*. London: Heinemann.

Sternberg, R.J. (1983). How much Gall is too much gall? Review of Frames of Mind: The theory of multiple intelligences. *Contemporary Education Review 2*(3), 215–224.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment & Evaluation in Higher Education, 30*(4), 325-341.

Strauss, V. (2011, April 15). The Myths of Standardized Testing. *Washington Post*. Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/post/the-myths-of-standardized-testing/2011/04/14/AFNxTggD_blog.html.
Task Group on Assessment and Testing (TGAT). (1988). National Curriculum: A Report. London: DES.

Taylor, M. (2012). *Student Perceptions of Education.* Unpublished Manuscript, Lancaster University.
The National Centre for Fair and Open Testing. (2007). How *Standardized Testing Damages Education*. Retrieved from www.fairtest.org/facts/howharm.htm

Toshalis, E. and Nakkula, M. J. (2012). The integration of motivation, engagement, and voice in student-centered learning. *Students at the Center: Teaching and Learning in the Era of the Common Core project*. Retrieved from http://www.studentsatthecenter.org/topics/motivation-engagement-and-student-voice.

Visser, B.A., Ashton, M.C., & Vernon, P.A. (2006). g and the measurement of Multiple Intelligences: A response to Gardner. *Intelligence 34*(5), 507–510.
Webb, P. K. (1980). Teaching Methods: Learning Applications. *Theory into Practice*, *19*(2), 93-97.

Whitten, W. B., II. (2011). Learning From and For Tests. In A. S. Benjamin (Ed.), *Successful Remembering and Successful Forgetting* (pp. 217-134). London: Taylor & Francis.

Wiliam, D. (2010). Standardized testing and school accountability*. Educational Psychologist, 45*(2), 107-122.
Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In D.J. Hacker,

J. Dunlosky & A.C. Graesser (eds), *Metacognition in educational theory and practice* (pp. 277-304). Mahwah, NJ: Lawrence Erlbaum.

Winne, P.H. & Hadwin, A.F. (2008). The Weave of Motivation and Self-Regulated Learning. In D.H. Schunk, & B.J. Zimmerman (eds.), *Motivation and Self-Regulated Learning: Theory, Research, and Application* (pp. 297-314). New York: Routledge.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*, 3-17.