National Curriculum Tests and the teaching of Thinking Skills at Primary Schools – parallel or paradox?

Hanneke Jones, Newcastle University

The drive to raise standards in core curriculum subjects, and the teaching of thinking skills are both aspects of the UK government's education policy for England. This article is based on findings from a questionnaire-based research project which investigated the relationship between National Curriculum tests, which are an important element of the ‘Standards Agenda’ on the one hand, and the teaching of thinking skills in primary schools on the other. Analysis of the data indicated a negative impact of these tests on the teaching of thinking skills. In examining the results, the principles underlying both the National Curriculum tests and the teaching of thinking skills are discussed, and found to conflict considerably. It is argued that pedagogies aimed at the development of higher-order thinking skills cannot be expected to flourish in an education system which continues to be dominated by tests and league tables.

Key words
Teaching thinking skills; statutory tests; performativity

Introduction
Since the 1990s, National Curriculum tests have been an important element of education policy in England, and over the same period there has been considerable interest in the teaching of thinking skills. Interest in the relationship between National Curriculum tests and the teaching of thinking skills in Northumberland had resulted from data from a questionnaire survey carried out in 2004 to investigate practitioners’ use and perceptions of the teaching of thinking skills in this Local Authority (Jones, 2008). The data analysis had suggested, for example, that the teaching of thinking skills was felt by some to raise standards, but also that ‘performance pressures’ left little time for the teaching of thinking skills (Jones, 2008).

As a result of this interest, replication of the 2004 survey was carried out among Northumberland primary practitioners with year groups taking the National Curriculum tests. The objectives of this research were: to investigate the teaching of thinking skills among this specific group of practitioners, and to investigate their perceptions of the relationship between the teaching of thinking skills and the tests.
The data obtained in this survey would also serve to make comparisons possible between the practice and views of primary practitioners conducting the National Curriculum tests, and those of the wider group of practitioners who had responded to the 2004 survey. The findings of this research are analysed in the present article, and used as a starting point for a further exploration of the relationship between such initiative as the teaching of thinking skills, and the Standards Agenda in a wider sense.

To begin with, an overview will be given of both thinking skills education, and the UK government’s use of statutory tests in England. This will be followed by a summary of the data collection, analysis and findings. In order to explain the findings, the discussion will focus on the differences in epistemology and purpose which underlie the teaching of thinking skills on the one hand, and the National Curriculum tests on the other. In the conclusion a review of the tests and league tables is argued for, in line with the Cambridge Primary Review (Alexander, 2009), and the potential role of thinking skills approaches in alternative assessment methods is discussed.

The teaching of thinking skills
General aims, methods and impact
The development of pupils’ cognitive skills has been a focus of international research interest for decades, and in the UK a specific surge of interest in this area took place in the 1990s, leading to McGuinness’s report for the DfEE (McGuinness, 1999), and inclusion of thinking skills in the National Curriculum for England (DfES and QCA, 2000). In the latter, thinking skills are categorised as information-processing, reasoning, enquiry, creative thinking and evaluation skills, with the suggestion that the use of these thinking skills will enable pupils to ‘focus on “knowing how” as well as “knowing what” – to learn how to learn’ (DfES and QCA, 2000, p.22).

Moseley et al. (2005) give an overview of over forty frameworks and classificatory systems dealing with ‘the processes and products of thinking and learning’, many of which have been used as a basis for programmes or approaches to teach thinking. These programmes share the aim of the development of cognitive skills and the belief that this can be achieved through education. However, specific aims vary widely. For example, the aim may be subject-specific, as in the CASE programme for the teaching of thinking skills in Science (Adey et al., 1995); the aim may be the maximisation of pupils’ learning potential, as in the Instrumental Enrichment programme (Feuerstein et al., 1980); or the aim may be to prepare pupils for democratic citizenship, as in the Philosophy for Children programme (Lipman, 2003). As Moseley et al. point out (2005, p.26), thinking skills programmes and
approaches draw on a range of psychological, philosophical, social and pedagogical
sources and do not share any unifying or overarching theory.

Despite disparate aims and underlying theories, a number of similarities in the
methods used between different approaches are, however, apparent. Firstly, the
development of metacognition, or the process of ‘thinking about one’s thinking’ is
central to most thinking skills education (McGuinness, 1999). Although some would
argue (Smith, 1992) that thinking is inherent in all good education and should not be
taught explicitly, in most thinking skills pedagogies the raising of learners’ awareness
of their thinking becomes both an aim and a tool. Secondly, with the exception of
some early behaviourist instructional design programmes as described by Moseley et
al. (2005, p.44), thinking skills programmes tend to be based on either constructivist
theory, social-constructivist theory, or both. Essentially, in constructivist theory,
knowledge is regarded as constructed by the learner, whereas in social-constructivist
theory, knowledge is seen to be constructed by learners in a process of social
interaction. In many thinking skills programmes this theory base is apparent, as
learners are encouraged to search for and construct meaning together, and activities
tend to be open-ended, giving learners the opportunity to develop what Edwards
(2005) calls their intellectual autonomy. Thirdly, and following on from the social-
constructivist basis of many approaches to the teaching of thinking skills, thought is
regarded as closely related to language and interaction as Wells (1999) and
Alexander (2006) have also pointed out. As a result, pupil talk and dialogue tend to
be central to thinking skills education. Within this dialogue, however, the role of the
educator is seen to be as crucial as in any other teaching. As Alexander (2006)
points out, ‘learning to a specific cultural purpose requires intervention and support,
pupil engagement and teacher intervention’.

Complexities in the measurability of the effect of thinking skills programmes
Despite this, Higgins et al. (2005) found in their meta-analysis of the impact of
thinking skills programmes, that a substantial impact on both pupils’ cognitive
development and curricular outcomes can be identified. Adey et al have identified a
positive impact of the Cognitive Acceleration programmes CASE and CAME on
pupils’ achievements in standardised tests (Adey et al, 1995). Data from the
previously mentioned 2004 survey was shown by Moseley to indicate some potential
correlations between specific approaches the teaching of thinking skills, and above-
expected results from National Curriculum tests (Moseley, 2005). These findings,
however, were not replicated in the data from the present 2005 survey, as Moseley
and the present author have described (2008 a and b) This confirmed the notion that
outcomes of thinking skills education cannot easily be measured in standardised tests.

The teaching of thinking skills in Northumberland
The teaching of thinking skills has been an important element of professional development in the Northumberland Local Authority since the early 1990s, under the name ‘Thinking for Learning’, hereafter called T4L. In this approach, which is clearly based on social-constructivist theory, teachers present their pupils with a ‘cognitive conflict’ situation and encourage them to engage in group problem solving, enquiry and collaborative thinking. T4L comprises both the Philosophy for Children approach (Lipman, 2003), in which children’s questions are the starting point for a collaborative Socratic dialogue, and a range of ‘thinking strategies’ (Higgins, Baumfield, & Leat, 2001). Thinking strategies include such techniques as ‘the Odd-One-Out’, in which pupils are encouraged to reflect on conceptual differences and similarities; ‘Mysteries’, which encourage pupils to engage in problem solving and enquiry; ‘Fortune Lines’ and ‘Living Graphs’ which encourage pupils to interpret and present narrative data using data handling techniques; and ‘Mind and Concept Mapping’, in which pupils use and create visual representations of their thinking. Both the thinking strategies and Philosophy for Children can be applied to one or more curriculum content areas. It is the aim of T4L to develop pupils’ higher-order thinking skills and use of metacognition, and to provide opportunities for a deeper understanding of learning content.

As has been mentioned, use and perceptions of T4L in Northumberland among practitioners who had a level of training in this field have been investigated by the present author (2008). Respondents reported using T4L strategies, on average, thirty times in an average 12-week term. While pupil engagement and both social and cognitive development were acknowledged as major benefits of T4L, a number of challenges were also reported, mainly related to planning and time.

Statutory pupil attainment tests in England
Nature and intended purpose
Ball (2003) uses the term ‘performativity’ in his description of the way in which, internationally, education policy reform has become dominated by the aims of transferring private sector methods and values to public sector organisations. Strict monitoring of achievement through tests and the publication of results in ensuing league tables are clear elements of this policy.
In England, pupils and teachers have been subjected to an assessment regime of this kind since the early 1990s by successive governments and for different political reasons. The phrase ‘the Standards Agenda’ was coined by New Labour when it declared the raising of educational standards as its first priority after coming to power in 1997, whilst taking over many of the assessment policies from the previous government (James, 2000). For many years, pupils took National Curriculum (NC) tests, typically at age 7 (end of Key Stage 1, Year 2), 11 (end of Key Stage 2, Year 6), and 14 (end of Key Stage 3, Year 9), before taking their GCSE exams at age 16. These written tests were carried out by individual pupils in reading, writing and maths (Key Stage 1 only) and additionally science (all other Key Stages).

A number of changes in the testing arrangements have been introduced in recent years: since 2005, teacher assessment, in part informed by the statutory tests, is now the main indicator of reportable pupil attainment in Key Stage 1. A consultation project is under way at the time of writing, to replace end of Key Stage tests for Key Stages 2 and 3 with ‘single level' twice-yearly National Curriculum tests (DfES, 2007), and in 2008 the DCSF announced the abolition of end of Key Stage 3 tests (DCSF, 2008). Despite these changes, however, the Department for Education and Skills (now the Department for Children, Schools and Families, DCSF) has stated firmly that ‘the framework of tests, targets and performance tables which have helped drive up standards so sharply over the past decade will not be compromised’ (DfES, 2007, p.2). Secretary of State Ed Balls (DCSF, 2008) reaffirmed the Government’s conviction that ‘the end of Key Stage 2 tests are the only objective measure of attainment in primary schools for parents, head teachers and the public’, and thus are here to stay for the foreseeable future. This intention was reiterated in the remit letter sent by the Secretary of State to Sir Jim Rose, in which the latter was appointed to carry out an ‘Independent Review’ of the primary curriculum in England (DCSF, 2008).

The combined test results for English, mathematics and science at the end of Key Stage 2 in England are published in performance tables by the DCSF. As parental choice in enrolment is influenced by a school’s position in the league tables, the test results have the potential to impact strongly on the financial status of schools and thus on the quality of education they can provide. Targets and test results are also frequently related to the performance management of individual teachers, and they are major factors informing the OFSTED inspection process of schools. As a result, wider pedagogical objectives at many schools and in many classrooms may thus well be subjugated to the narrow aim of raising National Curriculum test scores. Consequently, the tests can be said not only to assess achievement in the National
The actual effectiveness of the Standards Agenda, of which the tests are an integral part, has been much lauded by the government, but much disputed by others. In 1997 the Labour government set the target of 80% of pupils at the end of Key Stage 2 to reach at least NC level 4 for English (compared to 63% in 1997), and 75% in mathematics (compared to 62% in 1997) by 2002. These targets were reached in the 2007 tests, when 80% of pupils reached the required level in English and 77% in maths. However, this apparent success has been called into question by a number of studies. Tymms (2004) demonstrated that for both English and mathematics at the end of Key Stage 2, there had indeed been a rise in results between 1995 and 2000, but that actual attainment measured in the Key Stage 2 results had generally levelled out by 2000. This confirmed an earlier prediction by James (2000, p. 357). Since Tymms’ study, the levels reached have stayed relatively constant (Galton, 2007). In interim reports for the Primary Review, the rise in Key stage 2 test results between 1995 and 2000 has been called ‘seriously misleading’ by Tymms and Merrell (2007), and the actual reliability of the test results as a measure of pupil achievement has been called into question by Harlen (2007).

Alongside this disputed success of the tests in raising standards, serious negative implications of the tests and the Standards Agenda have been identified in a number of studies. Firstly, effects of stress on both pupils and teachers are described by James (2000, p. 347). Harlen and Deakin Crick (2003) identified a negative effect of testing on motivation for learning, and Hilton (2006) and Benady (2006) report pupil stress, peer intimidation, disaffection with the curriculum, and a ‘stifling’ of children’s creativity. Secondly, effects on the curriculum were found by a number of researchers: Boyle and Bragg (2006) identified a distinct narrowing of the curriculum in both Key Stages 1 and 2, which was reiterated by Tymms and Merrell (2007). Turner-Bisset (2007) has also described how performativity has seriously reduced creativity in the primary curriculum in England. Hilton (2006, p.39), in her study on the effects of the test on the teaching of English, identified not only a narrowing of the curriculum, but also a shift of concentration from higher-order reading comprehension skills to lower-order forms of literacy. Wiliam (2001) explains this point in arguing that improvements in test results can only be achieved at the expense of other elements of the curriculum, as teachers focus on those skills which are expected to come up in
the next set of tests. This has direct resonance with data from the present study, as the following analysis will make clear.

**Data**

*The 2005 survey and analysis*

Questionnaires (see Appendix 1) were sent out to all teachers with Year 2 and Year 6 classes within the Northumberland Local Authority in June 2005, just after the completion date of the National Curriculum tests. In this questionnaire, respondents were asked to indicate the typical frequency in which they used each of a number of given thinking skills strategies and the subjects to which these strategies were applied. They were also asked what they saw as the largest benefits and the largest challenges in using T4L with a year group taking the National Curriculum tests. Next, respondents were asked to indicate, on Likert-type scales, their views regarding the following issues: the importance of metacognition; how much thinking skills teaching their current class had received in previous years; the extent to which the use of thinking strategies contributed to increased National Curriculum test results; and to what extent the test results gave an adequate measure of thinking in English, maths, and science. Finally, further comments regarding the relationship between T4L and National Curriculum tests were invited.

Responses were received from 104 teachers in total, 67 of whom were teaching Year 2 classes, and 37 of whom were teaching Year 6 classes. This represented an overall return rate of 25%. Respondents worked at 38% of First Schools and 40% of Middle Schools. With such a return rate, the responses can, of course, not be said to be representative of all teachers in this Local Authority who were teaching Year 2 or Year 6 pupils. Nonetheless, it is felt that the views expressed warrant the exploration presented in this paper.

All responses were entered into a database. Questionnaire responses relating to the frequency of use of various T4L strategies were calculated over an average 12-week term. Responses related to benefits and challenges were categorised by a colleague and checked by the present author; inter-rater reliability was found to be around 90%. In order make comparisons between the responses to both the 2004 and the 2005 survey possible, identical categories were used to those of the 2004 survey, with one exception: an additional challenge category, related to the tests, was created. This was due to a large number of reported challenges which related specifically to the National Curriculum tests. Responses reporting the main benefits...
of teaching T4L in classes undertaking the tests were sorted into cognitive and metacognitive benefits; engagement benefits; curriculum benefits; social and communicative benefits; affective benefits; teacher benefits; and inclusion benefits. Responses reporting the main challenges were categorised into time-related challenges in the use of T4L in teaching; time-related challenges in the planning and preparing of resources for T4L; NC test related challenges; and other challenges.

There are some caveats, related to the validity of questionnaire-based research, which need to be considered. The views of respondents are necessarily expressed in a way which is limited, and there is a danger that these are misunderstood and misinterpreted by the researcher. There is also a danger that findings can be oversimplified. However, the questionnaire was seen as a useful device (Cohen et al., 2003) to gather the views of a large number of practitioners for this research project. As the responses seemed to indicate a large amount of shared understanding of terminology, and as a high rate of inter-rater reliability was established in the coding of responses, the risk of misinterpretation was felt to be limited.

**Data from the 2004 survey**

Some data from the 2005 survey just described were compared with the use and views of the much wider teacher respondent group which had been surveyed in 2004. These data have been described in their own right by the present author (2008). Comparisons were possible as the majority of respondents to the 2004 survey had also been working within the primary age range, and near-identical categories were used to sort the data from both surveys, as described in the previous section.

These comparisons, however, are complicated by three factors. Firstly, participants in the 2004 survey had been selected by the Head Teacher on the basis of having had some training in T4L. Although a high percentage of 2005 respondents also had a level of T4L training, it could be argued that the 2004 respondents group had a particularly high level of commitment to T4L, which could make a comparison of use and views between the two groups less valid. Secondly, the 2004 respondent group may have included a number of the Year 2 and Year 6 teachers who took part in the 2005 survey. As a result, the two groups of respondents cannot be said to be completely distinct, with the implication that any differences found might have been larger if it had been possible to separate the two respondent groups. A third caveat is that only 64% of respondents to the 2004 survey worked across the primary phase. As their use and views of T4L had, however, not been found to differ substantially from those of the 17% of Early Years practitioners, or the 19% of respondents who
worked with older pupils, this caveat is likely to be of least importance. Overall, it is felt, that while a level of caution needs to be maintained on the basis of these caveats, comparison of the two datasets can supplement our understanding of the effect of the statutory tests on T4L classroom practice.

Findings
Findings on both the reported use and views of T4L are reported here. In each section, the views of the respondents to the 2005 survey will be reported first, followed by a comparison, where relevant, to the views of respondents to the 2004 survey. Not all respondents claimed to use T4L, and benefits and challenges were only stated by users. Therefore, benefits and challenges will be given in percentages of T4L users in each survey, which can then be compared. These findings will be followed by a section on the further views of respondents to the 2005 survey regarding the relationship between T4L and the statutory tests.

Use of T4L
Of the 104 respondents, 94 reported using at least one of the T4L strategies in their teaching. The most frequently used strategies were the Odd-One-Out, and Mind and Concept Mapping (each on average 3.5 times per user per 12-week term), and, to a slightly lesser extent, Philosophy for Children. Other strategies were reported to be used less than twice per term on average.

The 94 users reported using at least some T4L strategies, on average, sixteen times per 12-week term (see Fig. 1). A comparison can be made here with the respondents to the 2004 survey, who, on average, reported using T4L thirty times per 12-week term. The much lower use amongst 2005 user respondents may signify an effect of the statutory tests on classroom practice, and will be considered further in the discussion.

Fig. 1: Average use of T4L strategies per 12-week term.

Reported benefits of T4L
All 94 users of T4L in the 2005 survey mentioned at least one major benefit of T4L. A total of 153 major benefits was mentioned; these were grouped into the benefits categories described previously. Cognitive and metacognitive benefits were reported
by 71% of users (67/94); engagement benefits were reported by 24% of users (22/94); curriculum benefits were reported by 23% of users (21/94); social and communicative benefits were reported by 22% of users (20/94); affective benefits were reported by 13% of users (12/94); teacher benefits were reported by 5% (5/94); and inclusion benefits were reported by 3% of users (3/94).

A comparison of the benefits reported by the 2004 and 2005 user groups (See Fig. 2) shows that relatively more cognitive and curriculum benefits were mentioned by 2005 users, whereas relatively more engagement, social, effective, inclusion and teacher benefits were mentioned by the 2004 users. This may indicate that teachers with pupils taking the National Curriculum tests tend to focus on a narrower, more curriculum-linked set of priorities than teachers working with pupils from a wider range of year groups. This seems to resonate with Wiliam’s point that teachers working with pupils who are taking the tests tend to focus mainly on what is relevant to test outcomes (Wiliam, 2001).

Figure 2: Benefits of T4L as mentioned by percentages of users.

**Challenges in teaching T4L**

Whereas at least one major benefit of teaching T4L in a class taking the statutory tests was reported by all 94 users, major challenges in using T4L with classes taking the NC tests were mentioned by only 69% of users (65/94). This is perhaps an indication that 31% of user respondents did not see the combination of teaching T4L and carrying out the statutory tests as problematic. A total of 108 challenges were reported and grouped into the challenges categories discussed previously. Time-related challenges in the use of T4L in teaching were mentioned by 50% of users (47/94); time-related challenges in the planning and preparing of resources were mentioned by 24% of users (23/94); specific NC-test related challenges were mentioned by 17% of users (16/94); challenges in teaching and facilitating T4L were mentioned by 10% of users (9/94); and other challenges were mentioned by 3% of users (3/94).

Figure 3 shows a comparison of the types of challenges mentioned by user respondents to the 2004 and 2005 surveys. This indicates that finding the time to plan and prepare T4L sessions was seen as a challenge by similar percentages of teachers in both groups. Finding the time to teach T4L, however, was mentioned by relatively more 2005 user respondents, alongside the test-related challenges.
mentioned by this group only. In contrast, intrinsic challenges in the teaching and planning of T4L itself were reported by relatively more 2004 user respondents. This may be interpreted as evidence that time pressures, already identified by many 2004 respondents, were even larger for teachers working with the test year groups. It could also indicate that teachers who, in the main, were not teaching year groups taking the test, could reflect more easily on the wider aspects of their teaching, such as their teaching and planning practice. This will be considered in more detail in the discussion.

Figure 3: challenges in teaching using T4L as mentioned by percentages of users.

**Further views**

In this section, which is based on the 2005 questionnaire only, views stated in response to Likert-style questions will be reported first, followed by further comments made on the links between using T4L and the National Curriculum tests. Where percentages are given, these are taken from the total respondent group (104).

**Responses given to Likert-style questions**

Respondents reported having a high commitment to metacognition: almost 90% of respondents (93/104) said that it was very or extremely important that pupils discussed their thought processes. With relation to the contribution of thinking skills teaching to an increase in National Curriculum tests results, however, responses were more muted: only 23% (24/104) felt that T4L contributed ‘very much’ or ‘extremely’ to a rise in test scores. 51% (53/104) of the respondents felt that this contribution was moderate or less, and 26% of respondents (27/104) answered this question with ‘don’t know’ or did not answer it.

The extent to which NC test results gave an adequate measure of thinking in English, maths, and science was considered to be moderate or less by 82% of respondents (85/104), whereas only 7% of respondents (7/104) felt that this was a good measure. 12% of respondents (12/104) answered this question with ‘don’t know’ or did not answer it.

**Further comments**

21 respondents added further comments regarding links between T4L and NC tests.

Many of these referred to the beneficial outcomes of using T4L in terms of test results mentioned in the above section, such as:
P4C has certainly helped to maintain reading skills, and
it helps children to prioritise their thinking.

In relation to this, two respondents warned that T4L should be taught in earlier year
groups leading up to the NC test year, for example:

I think that the biggest benefit is primarily when T4L is taught in all years in Key
Stage 2, rather than just in Year 6.

A number of respondents expressed their frustration at not being able to spend more
time on T4L:

I wish I had more time to use T4L – everything in Y6 is really urgent!

In contrast to that, some respondents reported the ability to resist pressures to spend
excessive amounts of time on test revision:

I don’t particularly focus on SATS but T4L makes for a better classroom climate
generally, and

Good teachers teach children not to the test: T4L in Y6 is a given.

Some respondents pointed out that the higher order thinking skills taught in
T4L are not assessed in NC tests, for example:

Children’s comprehension i.e. higher order thinking skills are much better –
unfortunately this is not tested in NC tests – the reading comprehension tests
require a much simpler level of answer at Y2: perhaps we are making the children
too clever!!

Finally one respondent expressed reservations towards T4L:

This is something the school is moving towards, but I need to be convinced of
increased (or at least not decreased) test scores to use the activities more often
with year 5 and 6.

Summary
In summary, the vast majority of respondents used T4L at least intermittently in their
classrooms, and saw the development of metacognition, an element which is
essential to the teaching of thinking skills but which is not assessed in National
Curriculum tests, as highly important. A similarly vast majority stated that the tests
did not adequately assess their pupils’ thinking in the core subjects. This, alongside
the many benefits of T4L which were indicated, shows that thinking skills teaching
was generally seen to be important for a range of pupil outcomes which are not
measured in curriculum tests.
Around a quarter of respondents identified a positive impact of T4L on test results, or mentioned curriculum achievement as a major benefit. This view was explicitly expressed by a number of respondents when asked for further views.

Although around 30% of users did not report any major challenges, 50% of users mentioned time pressures as a major obstacle, a response which was also explicitly mentioned by many respondents when asked for further views.

Discussion
In order to put the above findings into context, we will now focus on the differences in epistemology and purpose which underlie the teaching of thinking skills on the one hand, and the National Curriculum tests on the other. Having explored this, we will then discuss the mutual impact which T4L and the tests were seen to have on each other.

Differences
Apart from the obvious differences in purpose and procedure between thinking skills education and statutory pupil attainment tests, there are a number of tensions between the underlying premises on which each is based. These are representative of some of the dichotomies currently found within primary education in England, and which are in many ways similar to those pointed out by Edwards (2005). They also bear some relation to the contradictions identified by Alexander in the Primary Strategy and in other primary initiatives in England (Alexander, 2004; Alexander and Flutter, 2009, Alexander, 2009).

The first tension between the teaching of thinking skills and the National Curriculum tests is in the notions of learning and knowledge. The theoretical basis of the National Curriculum tests may be characterised as largely positivist, in that learning and understanding are seen as the transmission and (ideally) replication of a linear set of finalised, external facts and skills which can be measured in pencil-and-paper tests. In contrast to this view, in most forms of thinking skills education, knowledge and meaning are regarded as collaboratively constructed by pupils through a process of enquiry, interaction and discussion. Understanding can here be seen as ‘a means of the continuous transformation of both the self and the social environment’ (Wells, 1999, p.85). According to Wells and Claxton (2002), this is in line with ‘the best of current theory’. One respondent illustrated the role of discussion in both T4L and the tests with the following comment:
T4L is all about discussion, but in the tests the child is not allowed to discuss anything with us.

Secondly and related to the epistemological points above, individual subject contents in the National Curriculum tests are treated as compartmentalised items of knowledge. In T4L, on the other hand, a holistic view of knowledge is assumed and pupils are encouraged to build on prior knowledge, to make integral connections between subjects, and to engage and in ‘deep learning’ (James, 2000, p.357). James describes this kind of learning as ‘knowledge of concepts, principles and processes that can be applied in creative ways in novel contexts’. One respondent illustrates this by pointing out:

Children often find the format of the tests very limiting – not much scope for thinking…

A third dichotomy is related to beliefs concerning the optimisation of achievement, motivation and agency. Extrinsic motivation can be said to be an integral part of the way in which the Standards Agenda is implemented. The setting of targets by external agents, the publication of results in league tables, and the implementation of reward systems on the basis of pupil performance are all elements of this. In contrast, thinking skills education promotes self-regulation (Moseley et al., 2005) and is, in many approaches, based on the kind of pupil engagement which has been described by Dewey (1966), and self-direction. It is interesting to note in this context that intrinsic motivation, as advocated by Kohn (1999), was reported by 20% of questionnaire respondents to be a major benefit of Thinking for Learning. It is also noteworthy that many teachers in this Local Authority chose to undertake professional development and action research in T4L, at a time when they were also undergoing large amounts of compulsory training as part of the Standards Agenda.

**Mutual impact**

Having explored some of the underlying differences between thinking skills education and statutory attainment tests, we can now discuss further relationships at a pragmatic level between the two. In the classroom practice of all respondents the National Curriculum tests were a reality, and in most classrooms T4L was also reported to feature at least periodically. Data from the research seem to describe an interesting relationship. On the one hand, 24% of respondents mentioned that T4L improved test scores, and no negative impact from T4L onto the tests was reported. On the other hand 17% of users felt that the tests caused difficulties for the teaching of thinking skills, and no positive impact from the tests onto the teaching of thinking
skills was reported. In respondents’ eyes, it thus appears that a positive impact is made by T4L onto the tests results, whilst a negative impact is made by the tests onto the teaching of thinking, as illustrated in figure 4.

Figure 4: perceived mutual impact of tests/test results and T4L

The positive impact of T4L on test results, claimed by some respondents, was not supported by the statistical analysis mentioned before (Moseley and Jones, 2008 a and b). However, the perceived negative impact of the statutory tests onto the teaching of thinking skills as an aspect of a ‘broad and rich curriculum’ (DfES, 2006) warrants a further exploration. Very few educationalists would argue with the importance of pupils reaching their full potential in English, mathematics and science. However, the question is how the term ‘full potential’ can be defined, how this full potential can be reached, and how we can monitor such achievements. Using tests and targets as a main way to drive up standards (DfES, 2007, p.2) in a narrow range of skills can hardly be seen as excellent pedagogical practice, if it is our aim to support pupils to reach their full potential as well-rounded people. As Wiliam (2001, p2) points out:

When test scores at Key Stage 2 improve, we cannot conclude that education in Key Stage 2 has improved. We cannot even conclude that performance in English, mathematics and science has improved. All we can conclude is that the narrow range of skills tested in the Key Stage 2 tests has improved.

Most commentators would agree that improvements in the quality of education originate in the reflective use and development of pedagogic methods based on current theory. Some of these methods, such as Assessment for Learning (Black and Wiliam, 2001), Dialogic Teaching (Alexander, 2006), Creative Teaching (Craft, 2002), as well as methods to teach thinking skills are, in England, promoted by the Department for Children, Schools and Families (DfES, 2007, pp 14-18).

However, the integration and development of such methods requires a high level of teacher commitment, professional development, and a large amount of classroom time. As the majority of the respondents in this survey indicated, this cannot be realised while it is seen as a main priority by policymakers to raise achievement in the narrow range of subjects and skills assessed in the National Curriculum tests. As one respondent put it:
The LEA puts a lot of pressure on us to improve results year on year - there is little
time for anything but revision.

The impact of the tests on many respondents’ classroom practice was
illustrated not only by the large numbers of teachers who explicitly mentioned such
problems as ‘a lack of time for teaching and planning T4L’, but also, arguably by the
average number of T4L sessions per term which these Y2 and Y6 practitioners
carried out. Despite the many benefits of T4L which were recognised by this 2005
respondent group, average use was seen to be approximately half of that claimed by
2004 respondents. Integral challenges in teaching and planning T4L were also
reported by a smaller percentage of users in the 2005 survey than in the 2004
survey. An explanation for this reduction might be that pressures on teachers working
with year groups taking the tests are such that they can afford less time to be
reflective on issues which are not directly related to the tests. If we subscribe to the
notion that reflective teaching is essential for good quality education (Pollard, 2002),
an impact of the tests which allows teachers fewer opportunities for reflection on
pedagogy would, in itself, be detrimental.

The Department for Children, Families and Schools states in its introduction
to the Primary National Strategy: ‘High standards and a broad and rich curriculum go
hand in hand’ (DfES, 2006). This axiom can, as Alexander points out in his critique of
the Primary Strategy, only be realised if pupils, teachers and schools are freed from
‘the continuing pressure of testing, targets and performance tables’ (Alexander, 2004,
p.15). In England, summative assessment outcomes, whether they are arrived at
from end of Key Stage tests or, as currently piloted, from single-level tests, continue
to be tied to the reductionist and competitive system of league tables. While this
continues, quality of education will continue to be overshadowed by the needs of
schools to appear in as high a place as possible in the league tables. As a result, the
drive to achieve this will continue to drive the curriculum.

Conclusion
Although evidence from countries such as Finland (Curtis, 2004) shows that frequent
testing is not necessary for the achievement of very high educational standards,
there is a role for the evaluative purpose of summative assessment which Newton
has described (Newton, 2003). The ‘reform of the testing regime’ which Alexander
has eloquently argued for (Alexander, 2009. p. 23) is highly necessary. Such a
review could examine moderated teacher assessment of collaborative problem-
solving tasks, which has been suggested by Wiliam (2001). An assessment system
based on such tasks could feasibly provide effective evaluative school data and,
along with formative assessment of individual pupils, help rather than hinder the delivery of a broad and balanced curriculum.

The potential use of T4L-style thinking skills strategies for the purpose of such assessment methods could be investigated, as their potential for teachers' understanding of pupils' thinking is well documented (Baumfield, 2004). However, the use of such strategies to determine pupil achievement in terms of narrow National Curriculum level descriptors may well be problematic, as the higher-order thinking responses which are stimulated by these strategies are rarely quantifiable in the linear, pre-determined progression of learning which the National Curriculum represents (Jones and Steele, 2002). On the other hand, thinking skills' development of metacognition could make a contribution to the quality of pupil self-assessment, which Black and Wiliam (2001) have identified as an essential component of formative assessment, and the use of this in summative assessment could be explored.

It could be pointed out that a number of respondents commented on the fact that pupils who participate in thinking skills lessons tend to become more able to accept ambiguity and relish complexity. If policy makers were to credit the general public with this ability, the value of education would no longer have to be reduced to one-dimensional figures in league tables. Only then, schools and teachers would have a real opportunity to focus on helping pupils to develop as thinking citizens.

Acknowledgements:
I would like to express my gratitude to the late Rosie Turner-Bisset for her encouragement. Thanks also go to David Moseley, Francis R. Jones, David Leat and Robert Peers. The project reported on in this paper was made possible by a grant from the Newcastle University Faculty of Arts and Humanities research fund. Finally, thanks go to Northumberland Local Authority and the survey respondents. Authorisation was given for Northumberland to be named in this paper.

References


Moseley, D. and Jones (2008, a) Does Thinking for Learning live up to its name? Teaching Thinking and Creativity. Vol 8: 4

Moseley, D. and Jones (2008, b) Further exploration of professional development through Thinking for Learning in Northumberland schools Teaching Thinking and Creativity. Vol 8: 4


Appendix 1

Questionnaire used in the 2005 survey

(Thinking for Learning in Northumberland and trends in SATs results)

Dear colleague,

We are undertaking a county-wide survey amongst Year 2 and Year 6 teachers to see if there are any relationships between Thinking for Learning approaches and trends in SATs performance.

In this questionnaire, Thinking for Learning (T4L) is defined as

- Thinking Skills strategies infused into the curriculum, and
- Philosophy for Children

We are very aware of the time pressures on you, but we would be extremely grateful if you could take some time to complete this questionnaire.

The responses of those teachers who use T4L rarely or not at all, will be as valuable as the responses from those who do use T4L in their classrooms.

School____________________________________
Please complete or tick where appropriate

1. Please tick the year group you work with:
   Year 2       Year 6
   O                  O

2. Please tick the subjects you teach with Y2 or Y6 pupils:
   English    Maths  Science
   O                  O                    O

3. Have you received any training in Thinking for Learning?
   O None
   O Teaching Thinking Certificate
   O Philosophy for Children Level 1
   O Action Research please specify ______________
   O Philosophy for Children L 2 or 3
   O Other please specify ______________

4. Please indicate how important it is in your view that your pupils think about their thinking, i.e. discuss their thought processes.
   extremely    very much    moderately    not very much    not at all
   O                  O                  O                  O                  O

5. To the best of your knowledge, how much has T4L been used to teach the pupils currently in your SATs class during previous years?
   extremely    very much    moderately    not very much    not at all    don't know
   O                  O                  O                  O                  O                  O

The remainder of the questionnaire is about your use of T4L when teaching Y2 or Y6 pupils. If you feel that any of the questions are not relevant to your classroom practice, please write ‘n.a.’ next to the question.

6. Please indicate on the following page how often you use the following strategies/approaches in your classroom (please use bottom spaces to indicate any others which are important to you)
<table>
<thead>
<tr>
<th></th>
<th>less than once a term</th>
<th>roughly once a term</th>
<th>roughly once every half term</th>
<th>roughly once every few weeks</th>
<th>once a week or more</th>
<th>every day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mysteries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Odd-One-Out</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P4C/ Community of Enquiry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mind or concept mapping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fortune lines/living graphs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>diamond ranking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>card classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maps from memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thinking hats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>others – please specify</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. Please indicate in which subjects and how often you use teaching thinking (use a ‘T’) or philosophy (use a ‘P’):

<table>
<thead>
<tr>
<th></th>
<th>once a week or more</th>
<th>roughly once every few weeks</th>
<th>roughly once every half term</th>
<th>roughly once a term</th>
<th>less than once a term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maths</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. Please indicate how much, in your view, thinking strategies contribute to increased SATs scores.

<table>
<thead>
<tr>
<th></th>
<th>extremely</th>
<th>very much</th>
<th>moderately</th>
<th>not very much</th>
<th>not at all</th>
<th>don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
9. In your view, to what extent are SATs scores an adequate measure of the development of pupils’ thinking, in Science, Maths and English?

<table>
<thead>
<tr>
<th>extremely</th>
<th>very much</th>
<th>moderately</th>
<th>not very much</th>
<th>not at all</th>
<th>don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

10. In your view, what are the biggest benefits in using Thinking for Learning with a SATs class?

________________________________________________________________________

________________________________________________________________________

11. In your view what are the biggest challenges in using T4L with a SATs class?

________________________________________________________________________

________________________________________________________________________

12. Is there anything else you would like to mention regarding the link between Thinking for Learning and SATS, based on your classroom experience?

________________________________________________________________________

________________________________________________________________________