

Assessing English: A trial collaborative standardised marking project

SIMON GIBBONS
King's College London

BETHAN MARSHALL
King's College London

ABSTRACT: Recent policy developments in England have, to some extent, relaxed the hold of external, high-stakes assessment on teachers of students in the early years of secondary education. In such a context, there is the opportunity for teachers to reassert the importance of teacher assessment as the most reliable means of judging a student's abilities. A recent project jointly undertaken by the National Association for the Teaching of English (NATE) and the Centre for Evaluation and Monitoring (CEM) was one attempt to trial a model for the collaborative standardised assessment of students' writing. This article puts this project in the context of previous assessment initiatives in English and suggests that, given recent policy developments, now may be precisely the time for the profession to seek to be proactive in setting the assessment agenda.

KEYWORDS: Assessment in English, history of assessment, collaborative marking, moderation of writing, teacher development, teacher professionalism.

“Nothing” Dewey once wrote, “has brought pedagogical theory into greater disrepute than the belief that it is identified with handing out to teachers recipes and models to be followed in teaching” (Dewey, 1938, p. 170). For many in the English teaching profession, the heavy reliance on terminal tests, with clear but restrictive criteria for their successful completion, has been the mere handing out of “recipes and models” for teachers and pupils alike. For that reason they constantly look toward some form of course examination that is not driven by testing and a limiting rubric. This article will look at an attempt to assess English through course work as opposed to a terminal exam, in a project carried out between the Centre for Evaluation and Monitoring (CEM) at Durham University and the National Association of the Teachers of English (NATE).

It should be remembered, however, that attempts to rethink the assessment for English are not new. There have been many notable examples of assessing English, particularly through coursework, which have later been abandoned. The Certificate of Studies (English) offered by New Zealand's Waikato University, was a qualification devised for senior secondary-school students that was entirely based on course work over a two-year period, but it was abandoned in 2004, largely because of government interference (Locke, 2007). Vermont had a system, too, of course-based tests instead of the usual examinations that were given for students in years 4 and 11. But this, too, was later dropped and the traditional multiple-choice tests were reinstated. The most notable and recent furore over testing in England came about when pencil and paper tests were introduced for fourteen-year-olds in 1991, and we will return to this

shortly, but at the same time these tests were being introduced, 100% course work at GCSE for sixteen-year-olds was also being abolished. These exams had existed in one form or another since 1964, so they had been around for just over twenty-five years. There were two types of exam that were assessed just through course work. The first was what was known as a mode 3 CSE and the second was the old O-level GCE.

The CSE was taken by the majority of pupils, around 80%, and was aimed at the secondary modern school, where the bulk of pupils went when they failed their 11+ exam. The GCE was taken by the so-called elite who, having passed their 11+, went on to grammar school. Mode 3 was introduced at the same time as the ordinary CSE, in the mid-Sixties. Although the age of leaving was still only fifteen, many pupils stayed on for an extra year until they were sixteen, and so could take a leaving exam. Essentially a school could organize their own syllabus for their pupils, provided it was approved by an examination board. Some exam boards, possibly inspired by CSE, began to look at GCE as well. GCEs were a more prestigious award, but the exam boards, building on their clientele, the teachers, began to look at ways of assessing pupils through coursework alone.

The Joint Matriculation Board (JMB) was one such board. In 1964 they started upon a trial of an alternative syllabus, and like the current project, did so along with the University of Durham. The trial ran for six years, at which point they opened it up to everyone already taking GCE with the JMB. In 1978 they began to run it for the whole country. The JMB's rationale was simple, as expressed in the first interim report:

The GCE O-level examination in English language is under bitter criticism as conducive to dull and cramped teaching and to crabbed rote learning and practice. The lively interest which should be aroused by learning to read and write English is killed, so it is asserted, by the need to prepare for writing stereotyped answers. (Hewitt & Gordon, 1965, p. 1)

The key to what the JMB was doing was to find a legitimate method of moderating coursework so that a legitimate and reliable grade could be given. Again, as they wrote in 1965,

But if the teacher of English is to be free to teach his pupils English as he thinks he should teach them without regard to traditional examination, how can the examining board, whose *testamur* at the end of the course is required, be assured that by the end of the course those pupils have benefitted from this untrammelled teaching and learning to an extent which merits an O-level pass in English. (Hewitt & Gordon, 1965, p. 1)

By 1978, the JMB extended coursework even further. Most schools had now become comprehensive and the school leaving-age had risen to sixteen, which meant that all pupils, both GCE and CSE, were in the same school and had to take some kind of exam at sixteen. Recognising this, the JMB produced a new exam called the 16+ for all pupils. With the national curriculum coming in in 1989 with a new exam – the GCSE, which was also aimed at all pupils – it was an easy transfer.

What the JMB came to realise, over a period of time, was that the best way of ensuring a reliable method of moderating was to have multiple readings of the text and some kind of trial marking. To begin with, the JMB and its successor, the

Northern Examination Association Board (NEAB), demanded that all teachers with examination classes assess, bi-annually, trial-marking material, consisting of folders of work of candidates from the previous year. As early as 1970, Rooke and Hewitt, working for the JMB, noted: "Experience has shown that it is essential for groups to meet together to discuss the results of trial marking and procedures for assessment." (Rooke & Hewitt, 1970, p. 14.) They went on to recommend that, on the extension of the scheme, "Provision must therefore be made for groups to meet for discussion" (Rooke & Hewitt, 1970, p. 14).

The folders always contained one or two candidate's work that were difficult to assess, for example a C/D borderline. All teachers marked these folders blind, then met with the rest of the department in order that a school grade could be decided. The individual scores and the school's agreed grades were sent back to the board. A standardisation meeting was then held by the board, in which the grades, agreed by a Review Panel, were given out. The Review Panel was made up of practising teachers, who, as with the original experiment, had been chosen for the accuracy of their assessments, through the trial marking.

The system by which the actual work of pupils was assessed was similarly rigorous. To ensure the reliability of these judgements, checks and double-checks were introduced. All candidates were marked both by their own teacher and another member of department. Where there was any disagreement, or when the candidate was on the borderline between two grades, their folder was submitted for scrutiny by the whole department.

The whole school entry was then moderated to ensure that the candidates' work was placed in the correct rank order, from grade A to U, before sending them to the exam board. Here, the work was moderated by a member of the Review Panel. All Review Panel members worked with partners. When one panel member moderated a school's entry, the other checked their judgement. The Review Panel members had the power to alter a school's grades, either up or down, if they felt that they had placed more than 50% of the candidates on the wrong grade. A "C" could become a "D" or a "B" or an "A". (The rank order of individual candidates could be changed only when the Review Panel members felt that an individual candidate had been wrongly graded by at least two complete grades.) The work of the vast majority of candidates was, therefore, read by at least five different English teachers before a final grade was awarded. One final check was built into the system. A sample of the cohort was sent to an Inter School Assessor. This teacher marked the entry blind and then sent their grades to the Review Panel. Again, if there was a serious discrepancy between the Assessors' grades and the school's, the panel members would moderate the school's entry.

The exam boards returned all course work to the schools, after they had been externally moderated, with comments on any adjustments that had been made as well as on the quality of the work. In this way, the whole process of exam board's decisions and moderations was entirely made by ordinary teachers whose own pupils were being examined. Moreover, a national network began to develop where the teachers were firmly in charge, but learning constantly from the dialogue that was created by the process.

But all this was abandoned when, in 1991, the then Prime Minister, John Major, announced that 100% coursework was to be no more. At more or less the same time, the pencil and paper tests for fourteen-year-olds (the KS3 SATs) came in. When the national curriculum for England and Wales was announced in 1989, the three assessments, which were to be made of pupils at ages seven, eleven and fourteen, were designed to be tasks taken over a period of time. The KS3 English tasks, designed by the Consortium of Assessment and Testing in Schools (CATS), were to be taken by pupils over about three lessons. In other words they were to be as like normal lessons as possible. There was even time for drafting and redoing a piece of work.

While the trials for these were going quite well, the government, instead of starting the KS tasks with secondary teachers, who were familiar with the testing process, determined that the first official national curriculum tasks would be for KS1 – for seven-year-olds. And these were a disaster. Along with John Major's tirade against 100% coursework at sixteen, then, the KS tasks were abandoned and tests were brought in. A new contract was given to the Northern Examinations and Assessment Board (NEAB) but this only lasted a year and, in 1992, the contract was given to the University of Cambridge Local Examinations Syndicate (UCLES, now Cambridge Assessment).

Now pupils' comprehension was to be tested by multiple-choice questions; an anthology of literature and, for the first time, a Shakespeare play were also added to the items to be assessed. English teachers had had enough. In a protest begun by the London Association of the Teachers of English (LATE), and later adopted by NATE and finally, in spring the following year by the unions, teachers boycotted the SATs. What is interesting to note in hindsight is the type of response to the tests, and the teaching required for them, when teachers eventually saw the SATs exams their pupils would have had to take. "A more didactic approach/More time spent on 'bitesize' (superficial) responses to literature (implying a 'right' answer)/Less time to develop individual responses" (Cooper & Davies, 1993, p. 566). All these indicated a very different approach to the way in which teachers felt they organised their classrooms at the time.

At the same time, teachers were boycotting the SATs, they were also protesting about the changes made to GCSE. Mike Lloyd, a teacher from Birmingham, was at the time campaigning to keep 100% coursework. Advocating for the what was known as the Save English Coursework Campaign, Lloyd petitioned 4000 schools (almost all the secondary schools in the country), both independent and maintained, and received an 85 per cent return. Of this 85 per cent, 95 per cent wanted no more than 20 per cent timed testing (Lloyd, 1994), a precise reversal of John Major's dictat. Lloyd's campaign, however, was to prove totally unsuccessful.

Sadly, also, the SATs boycott lasted only two years and in 1995 the tests came in. The next fourteen years, before they were finally withdrawn in October 2008, were dogged by reports of how much they narrowed the curriculum, including one by David Bell, then chief inspector for Ofsted.

Teachers...continue to take the [KS3 tests] seriously and to prepare students as fully as possible.

However, this is part of the problem. Many teachers spend too much time preparing for the tests. Consequently, the curriculum narrows significantly in year 9 [fourteen year olds]. Few departments continue to allocate time to promote pupils' wider reading and the range of work covered is much more limited than in the previous two years.

An additional problem is that too much time in year 9 is based around a narrow view of skills needed by the pupils for the tests. In many schools too much time is devoted to test revision, with not enough regard given to how pupils' skills could be developed in more meaningful ways. Too little emphasis is placed on developing pupils' ability to work independently and to think creatively. Instead, work tends to consist of a great deal of teacher instructions and completion of practice papers. As one inspector commented, this can lead to 'dependency on teachers [that] sometimes prevents higher achievement' (Bell, 2005, cited in Mansell, 2007, p. 59-60).

Teachers felt constrained by the tests. It was for this reason that groups of teachers continued to look for ways in which pupils could be reliably assessed using coursework. One such project was the King's Oxfordshire Summative Assessment Project (KOSAP), which worked with English and Maths teachers from three Oxfordshire schools. The project was originally funded by the DfES but later by Nuffield. Here project teachers got together portfolios of a pupil's work, gathered over a year. The procedure they followed in order to arrive at a final grade was not unlike the NEAB. Portfolios of work were marked and levelled by the individual teacher who taught them. A sample was then blind-marked by rest of the department and given a level. These portfolios were sent to the other two participating schools, who again blind-marked them and gave them a level as well. Altogether, there were nine portfolios to assess, each school assessing three from their own school and six from the two others. There was broad agreement as to the levels given to the portfolios.

Another such project has been the one on which the rest of this article will be focused – between NATE and the CEM at the University of Durham. Here, however, the focus has been on marking individual essays, followed by multiple-marking. In this respect, it is more closely allied with the work James Britton did with LATE in the 1950's and 60s. Britton was also concerned with the old O-level exams, and had been since the early Fifties; LATE had indeed introduced an alternative O-level in the early 1950s (Gibbons, 2009).

In fact, Britton had conducted a number of trials during that decade on teachers' marking. The main dilemma for him was that teachers tended to give different marks for the same essay. He first conducted research in 1950, three years after LATE was formed, and reported it in the *Report on the Meaning and Marking of Imaginative Composition* (LATE, 1950), which in some ways demonstrates that assessing English was already a concern for some. What is interesting about this research is that Britton asks the teachers marking the work what criteria they might use when assessing. The teachers suggested criteria that looked not unlike the current Assessing Pupils' Progress categories. (These involve a number of writing and reading assessment foci and will be discussed in more detail later). For example, one suggested they should consider,

Quality of imagination shown in detail (number, variety, value of idea)
Structure of a sentence

Precision in language
Total effect.
(LATE, 1950, p. 1)

And another,

Imaginative conception (what the writer has made for himself from the material given him).
Literary technique. Extent to which his mastery of vocabulary, sentence structure, etc. enables him to express his imaginative conception in words.
Practical equipment – spelling, punctuation, handwriting.
(LATE, 1950, p. 1).

In fact Britton chose not to use these and the group arrived at two more: “a) pictorial quality and b) creativeness” (LATE, 1950, p. 2). Unfortunately, however, he found that “In spite therefore of our most careful preparation (months of discussion) we clearly did not agree on the qualities required of good imaginative composition” (LATE, 1950, p. 3). This then became had asked children to write for an hour. Now they asked them to write a hundred word piece and changed the criteria again. This time they asked for:

- 1) General impression (By your own personal method; by impression rather than by analysis in search of particular characteristics).
- 2) To what extent can the reader experience what is presented (i.e. see, feel, hear etc.)
- 3) Originality of ideas. To what extent is the writer’s view of the subject distinctive (i.e. as compared with the ideas of the group as a whole.)
- 4) Feeling for words. To what degree does the writer use words a) strikingly AND b) effectively? (LATE, 1950, p. 3)

Again though, teachers disagreed. What is interesting also is that Britton analysed all the results quantitatively putting them through an elaborate factor analysis. This, along with the APP style assessment, makes James Britton’s research similar to the work carried out by the CEM and NATE, for, as we shall see, the CEM too wanted the English essays divided into categories and carried out quantitative research on the results.

By 1964, when yet again Britton looked at English teachers’ assessment, however, he had abandoned such segmentation of pupils’ work although he still used quantitative analysis. He carried out work in conjunction with an exam board where he asked teachers (unlike the exam boards, who required detailed analytic marking) to rapid-impression-mark pupils exam essays. He found that,

The system of multiple marking employed in this experiment, used to mark essay scripts written in a public examination of a GCE Board, gave a greater reliability and validity than the system of marking of that Board, rigorous though it was. (Britton, 1964, p. 27)

He was still interested, though, in the comments teachers made on the marking they did. He asked them all to write “brief notes on the criteria upon which they had based their assessments” (Britton, 1964, p. 19). These he classified into the following categories: “a) involvement, b) organization c) mechanical accuracy” (Britton, 1964,

p. 23). The section on mechanical accuracy was given little overall weight by the markers within the general impression of the piece, but it was mentioned. Here again, therefore, his approach was not dissimilar to the NATE/CEM project to which we shall now turn our attention.

THE NATE/CEM ASSESSING WRITING PILOT PROJECT

The project run jointly by the National Association for the Teaching of English (NATE) and the Centre for Evaluation and Monitoring (CEM) was initially launched at a LATE Saturday Conference in December of 2008 at the British Film Institute, following an approach from Professor Peter Tymms to the Chair of NATE's secondary committee, Simon Gibbons. CEM is part of Durham University and is particularly known in England for the assessment and monitoring systems it provides for schools, although it is also involved in educational research and continuing professional development for teachers.

Although coming from different starting points, both parties had concerns about the method of assessment being recommended for English teachers within the classroom. The current recommended strategy – Assessing Pupils' Progress (APP) – involves a number of writing and reading assessment focuses. Teachers are encouraged to assess student reading and writing against these focuses, using the descriptions to set individual targets. This is proposed as a form of assessment for learning, though many would see it as often more summative in its use than formative. Underlying APP is an idea about what constitutes good writing, and this certainly has grown from National Strategy ideas that are, at heart, influenced by a genre approach to writing, and supported by a focus on the use of a variety of grammatical features in writing as a means of judging quality. In the wake of the abolition of SATs, we saw a danger that APP would simply become accepted practice, even though its status is non-statutory in terms of policy. Alternatively, it might be that we were just waiting for new look SATs to arrive in some guise. We felt in this context it would be worth trying to encourage English teachers to volunteer to trial an alternative model of assessment rather than simply wait for policy-makers to impose one.

In our initial discussions we established some key principles for the project. Agreeing that teachers, as the experts, ought to have control over the assessment, it was decided that teachers would set and mark work for their individual classes. As a means of looking at reliability of marking and standardising judgments, assessed pieces would be distributed across colleagues involved and second-marked. Through this process the CEM would be able to look at the extent to which there were shared understandings of assessment criteria, and to what extent judgments might be said to be reliable. Critically, the teachers involved would have a say in putting together the criteria that would be used to judge quality in writing. We felt it crucial that we didn't start from APP statements or National Curriculum level descriptors; we wanted the teachers to have a discussion about what ought to be assessed in writing and how this might be done. We wanted those involved to have some sense of involvement in the production of, and thereby a sense of ownership of, the assessment model being trialled. We genuinely wanted this to be a "bottom-up" model of development.

At the launch of the project at the LATE conference, and subsequently through e-mailing LATE members, teachers were invited to participate in an assessment

initiative which would involve key stage 3 classes (students aged between 12 and 14). Given that the intention was to start on a small scale, we aimed to involve about twenty-five colleagues. The project was intended to be a starting point, trialling a process to see the value in it, rather than trying to launch a fully-fledged system. Therefore, it was felt that the best way forward was to involve interested colleagues in the teaching, setting and assessing of an individual piece of work for a class. Clearly, assessing an individual piece of student work is not the most effective way to assess that student's attainment or ability, but it would be a manageable way to trial the material, the process of multiple-marking, and to make judgments about taking the work forward.

In the event there were around thirty teachers who expressed an interest in being involved in the project, and about twenty of these came to the initial meeting, which was also attended by members of NATE's secondary committee, which included academics from King's College London and a Local Authority English advisor. Through discussion, the general principle of the work was established, and some possible writing tasks and a draft mark scheme were considered. The discussion of the task to be set and the way in which it should be marked sparked interesting debate. What was clear was that those present were relishing the opportunity to argue over basic principles of teaching, learning and assessment, rather than how to implement best an imposed method or strategy. That might have been the inevitable consequence of a room full of volunteers giving up an evening, but it was precisely this spirit of professional autonomy that we hoped the project would engender.

The meeting achieved the aim of agreeing on a written task and mark scheme. The written task would be loosely based on The National Gallery "Take One Picture" initiative. Here, a single picture from the Gallery's collection is selected, which can be approached in a number of ways by the teacher, leading to a number of possible written outcomes (autobiographical writing, descriptive writing, empathetic writing, and so on.¹ Such an approach was taken because it would provide some sense of a common context, but individual teachers involved would be free to tailor the material and final writing task to their particular classes. The mark scheme for the task went through a number of versions. Initially a single scale with descriptors of features of writing was proposed, but it was felt that such broad judgments might obscure particular strengths and weaknesses in an individual's work, thus not allowing for comments on the writing to have sufficient specificity. After a second draft proposing four criteria for assessment was discussed, a final version indicating three areas for assessment was agreed upon, as shown in Table 1:

Mark	General Writing quality/impact on reader	Writer's choices	Cohesion and coherence
1	Low	Use of simple words to convey some sense of meaning	Cohesion and coherence hampered by little or no attention to punctuation or paragraphing
2	Some sensitivity to the needs of the reader.	Appropriate vocabulary used to convey meaning.	General cohesion/ coherence achieved through accurate use of full stops, capital letters

¹ Further details can be seen at the National Gallery's dedicated webpages at <http://www.takeonepicture.org/>.

		An emerging sense of form, purpose and audience shown through writer's choices	and question marks to demarcate sentences Sentences developed in a logical sequence.
3	Writing is lively, thoughtful and engages the reader, though perhaps not consistently across the piece.	Some adventurous vocabulary choices, and choice of words for specific effects, influenced by form, purpose and audience. Some variation in sentence structure for effect.	Spelling and use of full stops, capital letters and question marks generally accurate. Paragraphs beginning to be used to contribute to overall shape/cohesion.
4	Writing consistently engages and sustains the interest of the reader.	Writing style and vocabulary choices are adapted appropriately to suit form, purpose and audience Where appropriate to the style of writing, a range of sentence structures are used	Accurate use of full stops, capital letters and question marks to demarcate sentences. Evidence that punctuation is used to clarify meaning Overall coherence enhanced by use of paragraphs and linking devices
5	Writing is confident and assured, showing conscious awareness of the needs of the reader.	Ideas and descriptions are very well developed. Stylistic choices are clearly made for particular effects, in relation to audience, form and purpose.	Punctuation and paragraphing consistently accurate, and used consciously to overall clarity and cohesion of the piece Writing has overall shape and cohesion and coherence.

Table 1. Final agreed mark scheme for the NATE/CEM project²

In agreeing on these three areas for assessment, we wanted to foreground what we called “General writing quality/Impact on the reader” to encourage a judgment on the writing that did not prioritise areas of technical accuracy, and to suggest that in a piece of writing, the overall effect may be more than the sum of its parts. In then making a judgment on “Writer’s Choices” and “Cohesion and Coherence”, we hoped to draw attention to linguistic and grammatical choices and rate the writing in these areas.

Once agreed, the pilot project began. Teachers involved were given guidance about how to approach the picture task and suggestions as to a possible two-lesson sequence leading to a variety of potential written tasks. Classes taking part completed a written piece, either by hand or word-processed, and the class teacher assessed each piece, giving a mark out of five for each of the three criteria, and filling in comments on a standard form. At this point, each piece of work was scanned (if necessary) and sent electronically to the CEM in Durham, who then redistributed blank, anonymised versions of each piece of work to the others taking part. The pieces generated by the original classes involved were broken up when sent to the second marker, so that a

² Tables and figures from the NATE/CEM project are reprinted with the permission of CEM. The full report of the NATE/CEM Collaborative Standardised Assessment Project will be available from <http://www.cemcentre.org/>

teacher involved might receive half a dozen scripts from five different colleagues. Each participant thus second-marked around thirty of the scripts. Again, in second-marking, each piece was given a mark out of five for each of the criteria, comments added to a marksheet and all information returned to the CEM. Within this process, two pieces of work were selected at random which were to be assessed by all involved in the project as “control pieces”. The CEM then analysed the results, looking at areas such as the relative harshness or leniency of the different markers involved, and the relationship between an individual teacher’s marking of their own student’s work compared to her marking of other class’s writing.

THE RESULTS

In the event, as is perhaps inevitable with a project like this that relies on colleagues’ good will in sacrificing their own time and energies, the final number taking part in the pilot was smaller than originally intended. The timing of the project in the summer term of 2009 meant that other priorities (examinations, marking of coursework and the like) prevented some of the original volunteers completing the teaching and marking of the sample piece. It was with regret that many participants expressed their decision to withdraw, and all expressed a desire to be kept informed of findings, and to be offered the opportunity to be involved in future projects. In some ways, given the pressures on teachers’ time and the sense over the last twenty years that the profession has had increasingly less say on its own destiny, the fact that enough colleagues remained on board to make such a non-funded, totally voluntary viable is no small achievement.

A total of 128 pieces of student work were submitted, and nine teachers were involved in the marking process (though not all markers had completed the work with classes of their own). The results offered some interesting points for discussion. The graph provided by the CEM on markers’ relative leniency is printed as Figure 1:

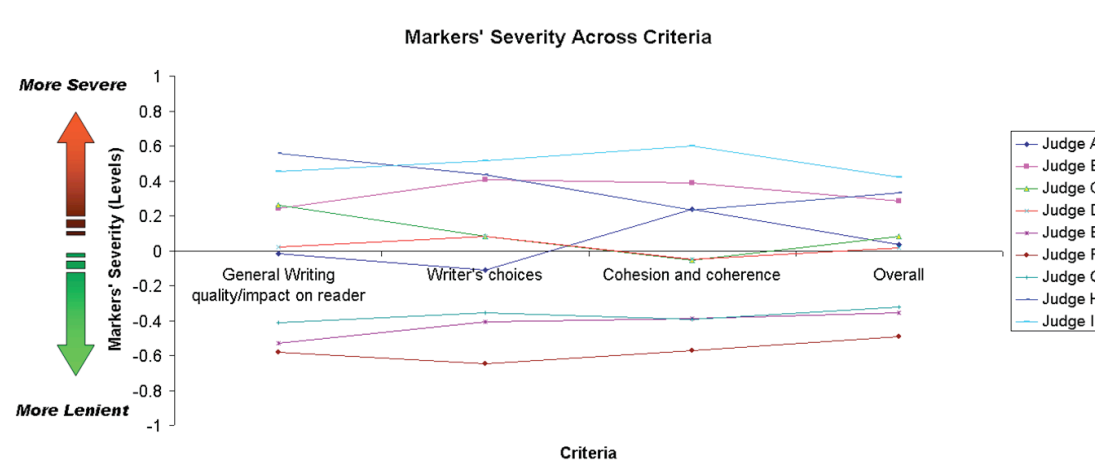


Figure 1. Severity of markers involved in the assessment project, provided by the Centre for Evaluation and Monitoring

On the graph, each line indicates the marking of an individual involved in the project across the criteria used in the mark scheme. Those markers whose lines appear above

the 0 axis are those that are relatively severe in their judgements in relation to the group as a whole.

The spread of marks revealed by the graph suggests that between the harshest and most lenient markers there is a difference of between one mark and a mark and a half. On a scale of only five this might appear to be relatively large, perhaps suggesting less common understanding about the quality of writing than we might assume a group of English teachers might share, though there are clearly questions to be asked about such a conclusion and there are ways the data might be interrogated that didn't form part of this initial pilot exercise. The relative flatness of the lines does suggest that marks awarded across the three criteria by each teacher were consistent across the pieces of work, and this could perhaps be taken as some evidence that the mark scheme worked in providing an effective relationship between the three identified aspects of student writing.

A comparison of the marking of teachers of their own students, as opposes to their marking of anonymous scripts (see Figure 2) did not reveal any conclusive trend, with some markers more lenient on their own students' work, whilst others were more harsh.

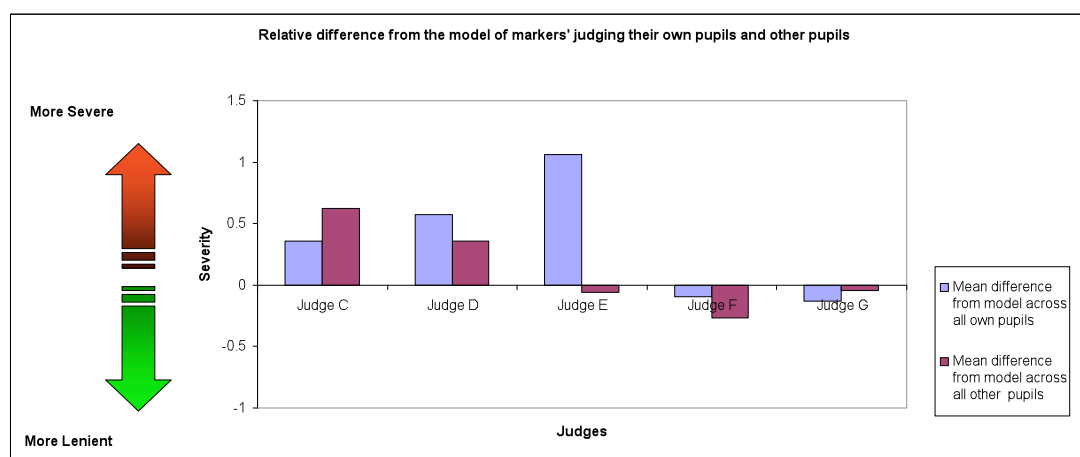


Figure 2. Relative difference of markers' assessing their own and other students, provided by the Centre for Evaluation and Monitoring

The group of teachers involved, including those who had eventually been unable to take part, met again after the project when the outcomes were shared and discussed. Again, the meeting witnessed lively conversation over the nature of the assessment, how judgments had been made and what possible ways forward there might be.

CONCLUSIONS

The success or otherwise of the project is of course open to interpretation. Clearly the project was small-scale, and as such the results can mean only so much in relation to what they might suggest about an effective way forward in proposing a system whereby students' work in English might be moderated and standardised. The apparent gap in some markers' relative severity might well be interesting to analyse in terms of the relative experience of the markers in question, or to see whether

particular types of writing were assessed particularly harshly by individuals or groups of teachers. An indication of possible developmental work might come from such an analysis.

In terms of a model of assessment, the CEM call the approach used in this pilot project Collaborative Standardised Assessment (CSA). Through such a process, it is suggested that alongside measures of an individual marker's relative harshness or leniency, it is also possible to provide statistical adjustment of student grades which allows for that relative leniency and information on the relative difficulty of a given task. There does seem to be genuine potential here for future work; the web based nature of the system allows for the possibility of teachers from across the country involved in cross-marking and moderation. On a large scale, this could give strong evidence of the reliability of teacher assessment, lending weight to its status in the context of an assessment system still in many instances dominated by the constraints of external testing or assessment systems.

There have long been calls for a system of properly moderated teacher assessment to be developed, and perhaps pursuing a web-based system might make this an affordable reality. Though this might not be in the same tradition as some other models of teacher or coursework assessment, it does offer practical advantages in terms of cost. In many ways, in fact, the project could be said to have more similarities with the work carried out by James Britton and his group of English teachers in the 1950s and 1960s, than with models of coursework assessment used in CSE and GCSE examinations in the 1970s and 1980s. One clear strength of this is the genuinely "bottom up" nature of the work, though there are certainly questions in such an approach about the nature of the task set, the method of assessment, and the ways in which the results might be best interpreted.

There are arguments to be had about whether one would want to use standardized individual tasks if such a project like this were extended, or whether it would be possible to use such a process to assess something like a portfolio of a student's work produced as part of their day-to-day English work. The latter would no doubt be less straightforward in terms of standardisation, but would no doubt be more satisfying to many English teachers and offer a more rounded view of the achievement of a given student. A suggestion from the CEM is that if the same students were involved in such a CSA process periodically, then development in writing over time might be tracked. What does seem critical to the approach taken in this project is the emphasis – made clear by the CEM – on the process being led by the profession. There are undoubtedly questions to be asked about whether the way the three assessment criteria were used in this project was actually the most effective way to judge writing, but the potential of such a project is that there is scope for that debate to be had, as it begins from the teachers deciding what to judge, not external agencies.

It does seem that there is a particularly important reason to be trying to extend projects like this in England in the current context. Recent policy moves, beginning with the 2007 version of the National Curriculum and the removal of Key Stage 3 testing, followed by the withdrawing of funding for the National Strategy programmes in 2011, have been part of an apparent shift to return more power to teachers and schools to have influence over curriculum, pedagogy and assessment. There are undoubtedly economic reasons for this shift in policy; the United Kingdom government's debt problems mean that there needs to be substantial savings in public

expenditure, and devolving power and removing costly, state-funded, centrally driven programmes are potential contributions to this cost-cutting exercise.

However, there may well be other, more principled motives behind the recent moves. There is a substantial amount of research evidence that points to the problems of top-down reform, seeing a limit to its ultimate effectiveness and pointing to the damaging effect it can have on teachers, in terms of their sense of professionalism, control of their destiny and motivation. High-stakes testing of students in England has certainly been one area where such top-down reform has been pervasive in the past two decades. It has been suggested that the effect of such continuous top-down reform has been, “to erode teachers’ autonomy and challenge their individual and collective professional and personal identities” (Day & Smethem, 2009, p. 142).

Certainly, the tendency of such reform has been to ignore teachers’ own “sense of passion and purpose” (Goodson, 2001, p. 49). There is substantial evidence, coming, for example, from reports such as those from the Institute for Public Policy Research (Brooks & Tough, 2006) and the House of Commons Children Schools and Families Committee (2008), of the damaging effects of high-stakes testing on the nature of curriculum and pedagogy, but it would also be sensible to say that there have been similarly damaging effects on teachers themselves and on their sense of themselves as professionals.

What seems to be the current context of a devolution of power back to teachers ought to be welcomed cautiously, however. Freedom after twenty years may not be a simple gift to use. It has been suggested (Day & Smethem, 2009) that both experienced and new teachers may face difficulties. Teachers who have worked through a long period of reform may be too disillusioned to embrace new found autonomy after having this eroded for many years. Teachers who have known nothing but centrally driven reform may have developed technical competency but not the ability to evolve their own effective practice.

This means a “reprofessionalisation” may well be called for, and part of the way this can happen is with English teachers working together in networks and groups, to address fundamental questions about the way their subject works. In the event, the most striking aspect of the project may well have been the willingness of those involved to embrace the work involved, knowing that they had a control over the process. The two meetings and the conversations they facilitated perhaps demonstrated the passion and commitment that exists in the English teaching community, and the readiness to regain autonomy over their professional practices. Indeed, the importance of the group engaging in debate was something we would argue must be preserved within any moderation system, even if the opportunities offered by technology are exploited. That the conversations in this project took place between experienced and new teachers, alongside university academics and assessment experts, suggests that, given the freedom, professional networks can be developed, within which teachers may bring their own passions and purposes to the process of proposing and enacting reform, rather than responding to it.

The NATE/CEM project was a first step in addressing the area of assessment of writing, encouraging teachers to develop their own professional understanding and skill through shared classroom work and subsequent discussion and evaluation. There may well be messages, even from such a small-scale project, about developing

expertise in the practice of teacher assessment. It is to be hoped that this project will be extended, but it's importance may go beyond the particular context of assessment, towards encouraging teachers to work together and be proactive in researching their own subject in order to improve the student's experience of English.

REFERENCES

- Britton, J. (1964). *The multiple marking of compositions* London. England: Her Majesty's Stationery Office.
- Brooks, R., & Tough, S. (2006). *Assessment and testing: Making space for teaching and learning*. London, England: Institute for Public Policy Research.
- Cooper, P., & Davies, C. (1993). The impact of national curriculum assessment arrangements on English teachers' thinking and classroom practice in secondary English schools. *Teaching and Teacher Education*, 9(5/6), 559-570.
- Day, C., & Smethem, L. (2009). The effects of reform: Have teachers really lost their sense of professionalism? *The Journal of Educational Change*, 10, 141-147.
- Dewey, J. (1938). *Experience in education*. New York: Simon and Schuster.
- Gibbons, S. (2009) Back to the future? A case study in changing curriculum and assessment: The Story of the London Association for the Teaching of English's alternative "O" Level English language paper. *English in Education*, 43(1), 19-31.
- Goodson, I. (2001). Social histories of educational change. *The Journal of Educational Change*, 1, 46-63.
- Hewitt, E. A., & Gordon, D. I. (1965). *English language: An experiment in school assessing (first interim report)*. Manchester, England: Joint Matriculation Board.
- House of Commons Children Schools and Families Committee (2008). *Testing and assessment*. London, England: The Stationery Office Limited.
- L.A.T.E. (1950) *Report on the Meaning and Marking of Written Composition*. Available in the L.A.T.E. Archive held at the London Institute of Education.
- Lloyd, M. (1994). *Save English coursework: Coursework in GCSE English*. Unpublished pamphlet widely circulated through NATE to schools.
- Locke, T. (2007). *Resisting qualifications reforms in New Zealand: The English Study Design as constructive dissent*. Rotterdam/Taipei: Sense Publishers.
- Mansell, W. (2007). *Education by numbers: The tyranny of testing*. London, England: Politico Press.
- Rooke, H., & Hewitt, E. (1970) *An experimental scheme of school assessment in Ordinary Level English language: Third report*. Manchester, England: Joint Matriculation Board.

Manuscript received: June 25, 2010

Revision received: October 5, 2010

Accepted: November 24, 2010