

Back to the future: A return to coursework explored

BETHAN MARSHALL

King's College London

ABSTRACT: This paper examines the King's Oxfordshire Assessment Project, which took place in the UK and explored the use of coursework as a means of terminal assessment. In particular, it considers the findings of the six English teachers who were involved. In a standards-based curriculum all six teachers supported 100% coursework. The paper looks at how, in the ritual of atomised standards-based assessment, a basis for holistic coursework can be maintained. It considers the importance of guild knowledge within the system as well as the need for structural support. It considers also some of the difficulties that can be found with this type of assessment.

KEYWORDS: Guild knowledge, 100% coursework, standards-based assessment, holistic assessment.

INTRODUCTION

The project I am about to describe started life in England in 2004 and ended three years later in 2007. Though completed nearly five years ago, the findings are still relevant in that it looked at the introduction of a summative 100% coursework examination of thirteen-year-olds as a possible alternative to the tests taken in England at that time for pupils aged fourteen. (These were the key stage 3 tests which assessed a pupil's progress over the first three years of secondary education. Key stages 1 and 2 took place at seven and eleven.)

As we shall see, the inclusion of some form of course-based assessment is always pertinent to English teachers. Yet, in England, the Secretary of State for Education, Michael Gove, announced on 10th September 2012, that he was reintroducing a type of exam for sixteen-year-olds that would eliminate the last form of coursework that existed and return us to terminal exams at the end of a two-year course.

The King's Oxfordshire Summative Assessment Project (KOSAP) was undertaken with Maths and English teachers because these were two of the three subjects assessed at this age, the other being science. The KS3 tests it was due to replace were part of the Labour government's desire to assess both English and Maths as part of the standards-based reform that was introduced in England in 1989 under a Conservative government. Standards-based reform has now become common currency in Western countries. The introduction of Programme for International Student Assessment (PISA) and Progress in International Reading Literacy Study (PIRLS) has made the desire to gauge where pupils are in standardised tests a key to many educational initiatives. Significantly they contain no coursework whatsoever.

In 1990, for example, the New Zealand Qualifications Authority (NZQA) introduced the National Qualification Framework (NQF) and later, in 1999 – the National Certificate in Educational Achievement. Achievement was assessed through a matrix of Unit Standards rather than through individual subject areas like English. In 2008

Australia introduced the National Assessment Programme – Literacy and Numeracy (NAPLAN). Pupils in years 3, 5, 7 and 9 had to take a national test in reading, writing and language conventions including spelling, grammar and punctuation as well as numeracy. Locke (2007) has viewed such changes as part of a general trend towards the managerialism of education as seen in the UK and Australia (Locke, 2007), a kind of post-Fordism where everything could be ticked and measured and where everything is accountable.

The New Zealand English teachers...like their counterparts in England and Australia, had been asked to implement a new curriculum document which was partitioned into strands and tied, outcome by outcome, into a “progression” of levels viewed by many educators as flawed. (p. xvii)

Certainly the move towards standards-based reforms altered the climate into which new teachers came in to schools, because English as a subject has tended to view progression, and the knowledge gained therein, as diverse and complex rather than linear and atomistic.

And it was this aspect of subject English, coupled with the idea that it was difficult to write something on demand, that had encouraged teachers in the past to favour 100% coursework. Quoting a paper by Mellon, written in 1975 on behalf of the National Council of Teachers of English, Freedman, for example, cites the claim that answering a written question in exam conditions is an artificial exercise:

We all know that it is difficult enough to devote half an hour’s worth of interest and sustained effort to writing externally imposed topics carrying the promise of teacher approbation and academic marks. But to do so as a flat favour to a stranger would seem to require more generosity and dutiful compliance than many young people can summon up. (Mellon, 1975, p. 34, cited by Freedman, 1991, non-paginated)

English teachers not only like coursework but have also found ways of encouraging it within state and national frameworks (Marshall, 2011). 100% coursework, for example, was introduced in England in 1964 for English in the national test for sixteen-year-olds but was abolished in 1992, the last entrants being 1994. Course-based assessment was introduced in Queensland for all subjects in 1971 and still continues in addition to the NAPLAN tests. Various trials have taken place in New Zealand, the United States and Canada of some kind of portfolio or course-based assessment. The problem with this kind of course-based assessment is that it does not provide clear standards-orientated criteria of the kind now looked for by Western governments, which like to itemise the skills acquired, often individually (see for example Marshall, 2002 and 2011; and Locke, 2007). Rather they take a more holistic sense of what a pupil can do.

Although we will discuss it in more detail later on in this article, it is worth introducing the concept of “guild knowledge” at this point (Sadler, 1989), for it might be argued that this is what teachers were using when assessing pupils work. Royce Sadler, who introduced the term, was describing the kind of processes that teachers go through when attempting to assess a piece of work, where the criteria for assessment appear unattributable or vague but are nevertheless successful. He argues that teachers make “qualitative judgements” (1989, p. 127) based on the summation of a piece of work’s merits rather than on individual, pre-determined aspects of it. In other words,

they use “guild knowledge”. We will return to this when considering how the English teachers in KOSAP assessed.

KING’S OXFORDSHIRE SUMMATIVE ASSESSMENT PROJECT

As has been previously stated, standardised tests were introduced in England in 1992 and, although they were boycotted for two years, they became part of the climate and culture of standardised school assessment in England (Marshall, 2008). At the time the project was started, pupils still took a test at Key Stage 3 for fourteen-year-olds. These were finally abolished in October 2008. This project, as has been said, was intended as an alternative to the tests and, indeed, in its first year was financed by the DfEE, though in the final two years it was funded through the Nuffield Foundation.

It was also carried out before Assessing Pupil Progress, a system first introduced as a pilot in 2006, and later as a way to assess pupils at KS3. This provided a matrix of pupil level with assessment foci, whereby you receive a level for each focus. Thus it breaks down the levels into discreet parts. In writing, there are eight foci and in reading there are seven. Although with each focus “the judgement is” intended to be “made in a holistic way” (Qualifications and Curriculum Development Authority [QCDA], 2009, p. 5) the overall effect of the grid is atomistic. Because it was introduced in 2006, the teachers comment on it as we shall see.

KOSAP took place in three schools in Oxfordshire, all noted for their work in formative assessment. The project, however, was on summative assessment. Two articles have already been published that look at the validity and reliability of the way in which the teachers assessed the coursework. One looked at summative assessment in general, the other at whether or not course-based summative assessment was reliable and, more particularly, whether or not it might enhance classroom learning (Black, Harrison, Hodgen, Marshall & Serret, 2007; Black, Harrison, Hodgen, Marshall & Serret, 2011). This article looks more at the teachers’ views of what coursework might offer the subject of English. Writing about summative coursework assessment then may seem somewhat quaint now in England but, as we shall see, it tells us much about how English teachers view 100% coursework. Four out of the six English teachers who took part in the project had never experienced a course-based exam, having taken and taught the standardised key stage tests at 7, 11 and 14. Yet the fact that our project was offering some alternative to standardised, timed tests was viewed sympathetically by them.

All of the schools operated on a system of portfolio assessment in English, and there was still some course-based assessment at GCSE, the exam for sixteen year olds. 40% of the English GCSE was conducted through coursework-based assessment. As we shall see, this was to prove significant. Yet it should also be remembered that this project included Maths teachers. Although this article will concentrate entirely on English teachers, the questions that were asked of the teachers had to be applicable to Maths teachers, too, and in many respects Maths teachers made strange bedfellows with their English colleagues. The Maths teachers, as Mellon again suggested, were more sceptical and regularly gave their pupils summative tests.

Answering multiple-choice questions without a reward in mathematics or science lesson may be one thing. Giving of the self what one must give to produce an

effective prose discourse, especially if it is required solely for the purposes of measurement and evaluation, is quite another. (Mellon, 1975, p. 34, cited in Freedman, 1991 non-paginated)

Their starting points then were very different and the project had to accommodate a Maths teacher's sensibility as well as an English teacher's one.¹

The research took place in year 8 (twelve and thirteen-year-olds) as opposed to year 9, because that was the year in which they traditionally took the KS3 tests. Three interviews were carried out with the participating teachers. In one of the schools, one teacher came late to the process and so was interviewed only twice. In all other schools, both teachers were part of the project throughout the three and a half years. The transcripts of interviews and of discussions were analysed, using a coding scheme partially derived from theory and partially grounded in the data (Glaser & Strauss, 1967). Reliability in the application of agreed codes was crosschecked between pairs of team members. The names of the teachers and schools have been anonymised.

Holistic versus atomistic

What is significant about all of the English teachers is that they preferred holistic rather than atomistic assessment of pupils; this might be because of the way in which they viewed English. This, despite the fact that when the teachers began on the KOSAP project they were still doing the high-stakes KS3 tests or SATS exam, as they were called. This had an intricate mark scheme that could be described as atomistic. Linn (2000) has commented:

As someone who has spent his entire career doing research, writing and thinking about educational testing and assessment issues, I would like to conclude by summarising a compelling case showing the major uses of tests for student and school accountability during the last 50 years have improved education and student learning in dynamic ways. Unfortunately, that is not my conclusion. Instead, I am led to conclude that in most cases the instruments and technology have not been up to the demands that have been placed on them by high-stakes accountability. Assessment systems that are useful monitors lose much of their dependability and credibility for that purpose when high stakes are attached to them. The unintended negative effects of the high-stakes accountability uses often outweigh the intended positive effects. (p. 14)

Linn's scepticism over high-stakes tests was certainly found amongst English in KOSAP and goes some way to describing their views on the validity of the assessment at fourteen. Daniel, for example, said that, "SATs are completely unreliable and random. We've had very little faith in the consistency in marking SATs. Our SATs results this year...were just ludicrous" (DG3). Others contrasted this with the type of assessment they would prefer: "I mean I would be completely in support of having a kind of coursework approach to KS3" (EB1). This is because, "It's very artificial in an exam to ask the students to respond creatively to a very, very dry stimulus" (EB1).

¹ It should also be noted that on the research team, I was the only one with an arts background and perspective. The four other people had a maths or science background, so I was the only person present who had examined a subject using 100% coursework.

In this respect there are two things wrong with the SATs. The marking is “unreliable and random” and the exam itself is “artificial”. This is contrasted with the desire for students to “respond creatively” and creativity is seen as important. Katrina comments, “It’s all about creating, releasing creativity of the moment” (KSE3). Although they use the term holistic rarely, the desire for valid assessment has a holistic feel in that it is very often contrasted against a “tick box” or atomistic means of assessing a pupil. Katrina’s comments are helpful in seeing how she developed this point. In her first interview she said that she,

was almost falling into a very checklisty trap erm, The only thing I could say is the assessed criteria and then you jump through these hoops and I knew it was a flawed thing. But I couldn’t articulate why. And now I’ve got a language to talk about that and it kind of holds a new set of discourse about assessment that I didn’t have before and I found that really, really valuable in kind of understanding you know some of the chaos. (KSE1).

And in her last interview she said:

I knew I hated the KS3 tests and found the preparation of students for these tests as one of the most unrewarding aspects of my job. However, through the project, I now have a much better understanding of why the tests are so problematic. (KSE3)

She asserted that the “Project has removed anxiety about delineating success only in terms of a neat, prescriptive check list” (KSE3). She knew she disliked “checklists” and the tests but now felt better able to “articulate” an alternative.

For this teacher, then, teaching English and being good at the subject were, therefore, not something that can be measured through an atomistic tick-list. Katrina, for examples, said, “I don’t mind marking. It’s when you’re marking in a very narrow way, where you’re not allowed to make assumptions that deadens” (KSE1). In other words, the whole is seen as greater than the sum of its parts. The atomistic does not capture the holistic. What is interesting is that she believed that *not* to make “assumptions” about a piece a child has written “deadens” the whole process. To mark, you read into what is written; you assume certain patterns. *Not* to do this narrows and eventually “deadens” the marking.

This can make marking “Tricky” according to Natasha. Many schools in England used the now-ceased National Literacy Strategy (NLS) in writing schemes of work and assessment. The NLS again tended to focus on the technical and grammatical aspects of English (Marshall, 2002). Focusing purely on the technical, however, can be problematic:

There is a difference between what you can actually...you could go through and underline all the connectives and say “yes” they are using those well and they are solid or they mark out an argument and then insight again. It’s not as tangible as that and that is what makes marking the hard part. You can read something and demonstrate insight...[you] go through and find where the writer has created a sense of atmosphere. It’s tricky. It’s not like you go through and find all the verbs. So there is a difference. (NC1)

Even those qualities which she thinks are a bit more “solid” have a nebulous quality. For instance, she commented that a good argument is “mark[ed] out” but in another essay the argument may just emerge and with sufficient “insight”; this might be a better way of approaching it. Or atmosphere. Clearly this can’t simply be judged by counting the verbs. She implied that some other criteria will have to be used.

Natasha felt, therefore, that there is a difference between the success criterion, for example, for using “connectives”, which is technical, and one for “insight”. One has “tangible” qualities, while the other is open to interpretation. It has a subjective quality that is hard to pin down and yet is there. “Insight” was a word she used in class.

We always look at the mark scheme and then again that doesn’t really give much of an idea of what the word actually means....I’d like to think I try, I do try and model it in a way. You know we will have a discussion...and I will say, “Stop, you have just shown me that.” It’s more of a kind of them doing it and then me saying, “Well, you may not realise it that that was what you were doing.” (NC1)

It may be that “insight” is associated more in her mind with what it means to be good at English. Other teachers spoke of “flair” and “confidence”. “Insight”, for Natasha, was something which can be seen when talking, writing or even when commenting on something a child has read. Whether or not a child has shown insight, however, was open to debate and this made marking “tricky” a decision but one which she made nevertheless.

Katrina used the word also. Again it had qualities which cannot be attributed “mechanically”. This again gives a sense of why all the English teachers were much happier with holistic rather than atomistic assessment, as it was more congruent with their sense of what it means to know in English. “Insight”, for Katrina, can mean, “a kind of sharpness and precision of analysis, of language”. She commented:

Insight is also important. They can’t become of this sort of mechanical analysis. It’s got to be married up to some sort of crisp understanding...and the most sophisticated readers will understand what kind of message or a meaning behind, some sort of sense of authorial tension. (KSE3)

“In writing,” she added, “there will be a kind of adventurousness to it. Often imaginative writers subvert conventions or subvert questions. They’ll be technically accurate though not necessarily superb” (KSE3).

Guild knowledge

All these phrases such as creativity, imaginative, adventurousness, insight, flair in some ways imply an unstated knowledge of what being good at the subject means. It means that the teachers have a “guild knowledge” of English. Sadler (1989) came up with the term “guild knowledge” in trying how to determine how teachers assessed. This is particularly relevant to arts-based assessment such as English. Sadler felt that all written descriptions of criteria could be seen as woolly (Sadler, 2009) and that teachers do not use them in the strictest sense, either formatively or summatively.

The difficulties are perhaps best seen by using the analogy of a kaleidoscope. One tiny shift and the whole pattern changes. So it is with a piece of writing. Each word on the page is like a chip of glass capable, when placed along side others, of creating a myriad of effects. No starting point is the same. Alter one element and a whole new set of issues arise, many of which cannot be seen, or necessarily anticipated, until the new pattern emerges. For Sadler, the way in which teachers cope with the multiplicity of variables is by making what he calls “qualitative judgements” about pupils’ work. But he admits, “How to draw the concept of excellence out of the heads of teachers, give it some external formulation, and make it available to the learner, is a non trivial problem” (Sadler, 1989, p. 127).

This is the problem that Wyatt-Smith and Bridges encountered when they encouraged teachers to use exemplars of pupil work with their students:

I think to a certain extent that we’ve empowered students in the learning process because there’s not secret teacher’s business anymore in terms of what the expectations are, that students are becoming very au fait with the criterion and being able to apply them in their own work. (Wyatt-Smith & Bridges, 2008, p. 61)

And it is what appears to have happened amongst the KOSAP teachers when using peer assessment with their pupils – assessing each other’s work. Teachers were able to share with their pupils some sense of “guild knowledge” in the process of writing the assignments and they did it predominantly through peer assessment. Karen, for example, says that she used “peer assessment” all the time, as did Katrina

I also do quite a lot of peer and self-assessment....What’s been really interesting is watching how the processes between peer and self-assessment has actually will find their sensibilities about what a particular skill actually constitutes. So, the beginning we were very mechanical and quite tick boxy about, you know, use of variety sentences and sort of count up....And now they, they, they’ve kind of internalised it (KSE3).

So although she started with an approach that could be called “quite tick boxy”, she ended up with a class who “internalised” the process. She went on to explain this further:

I guess what, I would hope what they’re burning up is good knowledge and I guess in the case of some students that has happened. It has also legitimised their own sense of what quality is because I’ve sometimes said, you know, it might not fit in a box, why is this good or why is that bad, and I think (*Pause*) to me it works both ways round when you get assessed and then that piece of work gets assessed with strengths and weaknesses. In terms of the whole process it assesses really their ability to understand what quality is (KSE3).

The class gained “good knowledge” of “quality” that cannot be expressed in a tick box. They moved from counting the variety of sentences to recognising that “quality” is something more and this is a good thing. Although, immediately afterwards Katrina added, “their understanding of the particular concept like writing, sentence structure and analysis of language” (KSE3) she did so within the overall context of quality work, which is a very vague, non-specific term. Even “analysis of language” or sentence structure become less definable because they are predicated on “quality” and the “concept” of writing. Writing has become more abstract, more of a “concept” that

can be seen in many ways. In this sense she required that her class make artistic judgements. The students' engagement in discussion about quality writing was beneficial in that rather than arriving at clearly specified atomistic criteria they were possibly progressively formulating and negotiating criteria for responding to each other's work. And this is again consistent with the nature of English as a disciplinary field.

It is very like Eisner's view on art, which in a way is the antithesis of the tick-box approach, or what Dewey disparagingly calls "ledger entries" (2005, p. 44). For Eisner art was about, "Judgement in the absence of rules. Indeed, if there were rules for making such choices, judgement would not be necessary" (Eisner, 2002, p. 77). He goes on to write, "Work in the arts, unlike many other rule-governed forms of performance, always leaves the door open to choice, and choice in this domain depends upon a sense of rightness" (2002 p. 77). His notion of judgement and, too, the sense of rightness depend upon an appreciation of the aesthetic and of artistry. While this can appear somewhat elitist – there are those in the know who have artistic judgement and those who do not – it is possible that a "sense of rightness" may be more democratic. The aesthetic may be something which is negotiated; there may be more than one interpretation of the artistic. In this sense, then, the pupils in Katrina's class developed a "guild knowledge" of artistry and the aesthetic.

It is a view echoed in Natasha's responses also. Although in her first interview she still had some affinity for a checklist of criteria – "I kind of see it in terms of boxes they are ticking and they are ticking the high boxes because they have got those words like insight and layers of meaning" (NC1) – by her second interview she remarked that those criteria needed to be more openly applied.

We are putting more of an emphasis on these sort of independent learning...being more creative, not necessarily giving them the success criteria – even the "must, could, should"...I think...that's also influenced our, our GCSE thinking as well because giving them a lot of scope...Creating the scope so you know, having open-ended success criteria, getting them to design the task success criteria themselves...I think that when we first started using them [must, could, should] it was like "brilliant, this is great" because it gets them to make independent choices themselves about what they are going to do. You know they want to do the "should" and the "coulds"...but at the same time it's still quite fixed. So the next step is to think about how can we open up those criteria. (NC2)

And this made her doubt the efficacy of the tick box what she called a tick box, the idea that she had certain learning outcomes in her lesson or specific criteria that she had covered. "There is a danger in that you approach it like that. 'Oh, they've done that, they put a paragraph in... they've done something interesting with verbs', so you start thinking in more of a tick-box way" (NC2). Like Katrina, she developed a sense of what "guild knowledge" might be and had communicated this to her pupils, albeit in an amorphous kind of way. What it means to have insight, for example, can no longer be ticked off. How the pupils use verbs has to be dependent on some form of what Eisner calls "judgement". The criteria have to be "opened up" so that there is the element of "choice" (2002, p. 77).

In this way, the formative process informs that summative product, particularly through peer assessment. Natasha, like Katrina, had established the same kind of

critical readings of assignments through pupils marking each other's work, "They are marking each others [drafts]" (NC2). So pupils had begun to acquire a sense of what the "construct" of a particular level or grade looks like through the act of reading each other's work in a critical capacity, and, in so doing, began to extend their aesthetic understanding. As Liz put it: "Ongoing assessment means that everything is valued. I think it would mean a richer curriculum because I think preparing for exams is reductive" (EB2).

Structural support

If ongoing assessment is to be seen as credible, however, it does need some form of moderation beyond an individual teacher's judgement and this means structural support of some kind. As Angela put it, "You've got to trust colleagues' judgement and you....But you've also by the same token to validate it, have some form of moderating it" (AS3). Moderation between teachers and schools was a very important part of the JMB/AQA 100% coursework, as it was, for example in Victoria until the early nineties. Both were stopped by right-wing administrations although, as has already been said, a kind of moderation in English coursework was maintained in England until 2010.

Natasha also felt that the decision-making had to be collective:

I think it's, it's made me consider more than I consider before the real importance of having a shared vision and shared practice amongst a group of people and when you are assessing, so that you have confidence in those assessments and so that those assessments are valid. So...not just school teaching one book and then giving back to one student, but doing whole moderation. (NC3)

Interestingly, and possibly significantly, teachers differentiated between what they called standardisation and moderation though both were important. Liz explained the distinction in the following way:

Moderation does not mean taking grades off and taking teacher's comments off work. To me moderation is you look at how a piece of work has been marked. And you look at the mark and the piece and comments and that's what you are moderating. Standardising is when you meet your, you're ensuring you are clear about what level four means. Moderating is to see whether, yourself and your colleagues can apply those levels consistently. (EB3)

Angela argued that, as teachers, you needed to do standardise as well, not the least because the process had the potential to provide examples of what you might do:

I think you've got to have materials and standardisation materials, they've got to have examples of tasks, so if you looked at the model....So not dictating, "this is the task, take the task, you should do this". Here are ideas for the sort of things you might want to include. (A3)

In so doing they echo both Klenowski & Wyatt-Smith and Klenowski & Adie who were looking in particular at the Queensland system of course-based assessment, but were also drawing more generalised conclusions. Klenowski & Wyatt-Smith believe that:

Moderation too is intrinsic to efforts by the profession to realise judgements that are defensible, dependable and open to scrutiny. Moderation can no longer be considered an optional extra and requires system-level support, especially if, as intended, the standards are linked to system-wide efforts to improve student learning. (Klenowski & Wyatt-Smith 2008, p. 1)

While Klenowski & Adie found that,

The initial stage of the research reported in this paper suggests that the practice at the local level of social moderation has the potential to fulfil an important role as a process for aiding teachers in ascribing value to student work through the use of standards that help them understand curriculum year level requirements and student achievement within year levels and in doing so attend to system level accountability. (Klenowski & Adie 2009, p. 2)

The Queensland authorities, until recently, felt that in order for a course-based system to work they should have the following procedures in place:

- syllabuses that clearly describe content and achievement standards
- contextualised exemplar assessment instruments
- samples of student work annotated to explain how they represent different standards
- consensus through teacher discussions on the quality of the assessment instruments and the standards of student work
- professional development of teachers
- an organisational infrastructure encompassing the QSA and schools to ensure the above takes place. (Queensland Studies Authority, 2009, p. 3)

Again, however, they have now introduced the standards-based NAPLAN alongside the coursework.

The KOSAP system had all of the Queensland measures. It held two moderation meetings – one half way through the project and one at the end. The first was seen, in a way, as a trial run for the second. In the first meeting the class teacher marked the portfolios and then they were sent to the other two schools for moderation. In the second meeting the work had been marked and levelled by the individual teacher. 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. In both meetings after discussion, they agreed on a grade for each of the students. Although there was some debate about a candidate in both sessions, the disagreements were overcome when it was decided that one could only mark what was in front of them and not take into consideration, for example, how many supply teachers the candidate might have had (Black et al., 2007, 2011).

Problems with course-based assessment

The English teachers on the course did find certain problems with what they were asked to assess. The portfolio they decided on included three reading, three writing and three speaking and listening assignments but they overlapped, so that there was an

assignment where reading and writing were assessed together and one where reading and speaking and listening were jointly marked. As has been said, this was moderated across the schools. The first difficulty was found in assessing reading through writing.

On the face of it, assessing reading through writing should not have been problematic. Reading, writing and talk should all be integrated in English. A piece of work should be assessed for what it is, rather than trying to segment it into different categories like reading and writing. But for Karen it was “Just impossible to do”, with her adding: “I don’t see how you can meaningfully assess both at the same time in terms of a written piece of work” (K3).

The problems for assessing reading through writing highlighted, for two of the teachers from the same school, the way in which their assessment of reading had become problematic. Daniel commented:

It really made us think about how we assess reading across KS3 and how we rely on essays. The kind of lit crit, analytical outcome, which it’s very hard to mark for reading....And it’s really difficult to separate out the writing from the....in the end we were just giving a kind of vague impression mark, which was overly influenced by their ability to write....we were assessing writing much more than we were assessing reading,,,,the writing was taking over. (DG2)

While Liz observed, “A huge part of our teaching ends up being about how to teach writing essays, not teaching a sophisticated reading response.” She added:

I think the effect will be that teachers will be able to comment much more widely on students working in English. So I feel like the focus in the past has largely been on writing and I think that the result of this project would be they would comment on their reading skills and on their speaking and listening. I think it will mean a much, you know, a much richer report. (EB2)

The conflation of reading and writing had for both these individuals become an issue. Yet the problem is that you cannot assess reading on its own. It is assessed through either writing or speaking and listening. In fact, part of the problem with both the testing system in England – GCSEs and the KS3 tests – was felt to be that there was too much assessment of reading albeit in written form. The KS3 tests, in particular, had two reading-type assignments – a comprehension activity and Shakespeare – and only one for writing – the short and long writing tasks. It was also decided in 2002 that the Shakespeare paper, which had been assessed for reading and writing, would only be assessed for reading.

In some ways this was the conclusion that Daniel came to, that is, that pieces of work should not be dually tested, that they should either be tested for reading or writing, even though the test of reading was done in written form. In his final interview, he commented:

I don’t think it’s difficult to do a piece of writing based on a piece of reading but I think the outcome makes it very difficult to assess two things simultaneously because they so kind of blur into each other. But it’s...very easy for teachers to end up assessing writing (PAUSE) and call it a reading assessment and it isn’t really. (DG3)

Interestingly, he has come to the conclusion that it is perfectly possible to write about what you have read; just that you should not assess for both in the same piece of work. Marking for writing as well as reading can confuse the teacher: the components “blur into each other”. He does not reduce the components still further, however. In fact, later on in the interview he remarked, “I think two pieces of reading, two pieces of writing assessment is enough. Whereas I think we were looking for three and I think that’s probably too many” (DG3). Far from wanting very distinct items to assess reading, it can all be done in “two pieces”. This may be a vestige of the system he currently had to operate in. Even though he wished the assessment to be holistic, he still compartmentalised reading and writing.

A second issue is poetry. None of the teachers talked about assessing the writing of poetry. It was not mentioned by any of the teachers in any of their interviews. However, students did write poetry. In Bishop Thomas School, for example, they wrote ballads. Yet, the idea of having poetry as an assessable part of the portfolios seems not to have occurred. It is possible that teachers believed it too hard to assess. If this is so, it demonstrates one of the difficulties of portfolio assessment and that is a kind of conservatism. You get pupils to do what you know you can assess and in this way it resembles a problem with exams. You test what you can in the time. While the range of work that it is possible to complete in a portfolio is far greater, and it does not stop you writing poetry, for instance, it may prevent you from counting it in the final grading. This in many ways echoes the Ofsted report on poetry teaching which noted that teachers did not count it in formal assessments (Ofsted, 2007).

The teachers *did* want reading to be assessed, however, but through speaking and listening. This posed a third problem and prompted an interesting debate for the two teachers in Bishop Thomas School, Liz and Daniel. Waverly School had set up a system of peer-assessed group work and the other two adopted it. Liz prepared a grid based on Natasha’s resources (from Waverly School), starting with simple questions and working towards the more difficult, “and each group had to sort of, well group analysis of each group member” (EB2). For Natasha the activity was very successful, but for Liz there were problems, namely how you got round a group of thirty pupils. However, she thought that the groups which she did hear showed good discernment:

They really went for it in their groups. I was really, really pleased with their response. I heard one group at the end really tussling over one student had shown a particular skill or not, which I thought was really positive. Obviously the difficulty then is when assessing them and say I’ve only assessed six students in about half a lesson....I don’t think it’s possible for the teacher. (EB2)

It also caused problems for Katrina, who while determining to keep the speaking and listening tasks, felt she needed to “get them better set up” (KSE3). The aim was to have small groups working together on a whole text that they had studied. What she should have done, she felt, was to keep an ongoing running record on how students performed as groups, and have this completed by the students themselves. This was how she believed Natasha operated.

But I didn’t use that as part of my assessment. I think I just disconnected the two things. But looking at the whole process from the assessment of the task to the final performance and evaluation – that would have been more helpful to me than thinking

at which point shall I assess this. Because then I ended up with a kind of unsatisfactory mismatch of marks. (KSE3)

Tellingly, both Liz and Katrina maintained that once they had sorted out some of the problems in the way in which they set the task up they would have no difficulty in assessing reading through this type of speaking and listening activity. Again, it is fairly reliant on the pupils themselves having some understanding of how they are assessing themselves. In fact, according to Liz, “Natasha says that it is and that they actually do it instead of [written] course work at GSCE” (EB2).

This was not universally the case, however. Speaking and listening activities caused more difficulties than anything else. Part of the problem was that although talk in the classroom was vital in stimulating ideas, it was difficult to capture and harder still to moderate. To begin with, it was ephemeral unless taped. Two activities were taped as part of the moderating process – one was a group activity on poetry, the other was a courtroom drama.

The poetry activity was hard to level because there was insufficient dialogue from each of the pupils and little opportunity for cross discussion. They were given a ballad and had to put it some kind of order. Pupils tended to focus on the text and just shuffle the bits of paper around so that, while it might have been a good classroom activity, there was insufficient talk to allow for assessing speaking and listening. Part of the difficulty with the courtroom drama was that, almost inevitably, not everybody had an extensive speaking part. Someone may have been good at speaking and listening but spoke only a little and was therefore difficult to assess.

A fourth problem was with the level descriptors. It was felt that the level descriptors were not specific enough to enable assessment at KS3. The GCSE criteria were much more explicit. In some respects, this gets to the heart of the problem with the introduction of coursework. While teachers wanted the assessment of pupils to be more holistic, they were still in many respects tied to the systems that they already had, even while often rejecting them.

For Katrina, the problem was not that the level descriptors were vague and the GCSE precise; it was rather that the level descriptors were “incoherent.” They did not “represent something that to an English teacher looks like a continuum” (KSE3).

You get these weird anomalies saying suddenly in level 7, you get this reference to handwriting. And you think like “Oh great, I’ve sort of taken that for granted and now and now I am worried about handwriting.” And so you think it’s the discontinuity which bothers me more than the lack of precision, because a lack of absolute precision is kind of what you’d expect from level criteria representing English (KSE3).

She concluded, “Sometimes it surprises me English teachers are so in favour of this amount of precision” (KSE3). Her solution, if it was one, was to look to the way GCSE coursework in England has, in the past, solved the problem: “If we look at the main criteria for GCSE it’s more detailed,” but she added, “There’s a real balance between, you know, representing the kind of continual English skills, which is necessarily a bit imprecise, I think, versus giving people real stability and security” (KSE3). Here she defines the problem between the holistic and the atomistic in terms

of level descriptors. She allows for the fact that the criteria for English are necessarily “imprecise”. In so doing she agrees with Sadler’s conclusion that all written descriptions of criteria can be seen as vague (Sadler, 2009). Yet she also says that teachers want “stability and security”, which suggests that they have become used to a certain “amount of precision” – the very tick-box mentality they are so keen to reject.

CONCLUSION

Re-introducing a 100% coursework exam to a generation of English teachers who had experienced nothing except standards-based testing was difficult. They might want to reject the strict atomistic criteria of standards-based assessment. Katrina, for example, thought that the Assessing Pupil Profile, that was due to replace the KS3 tests, was “horrendous” and should be “abolished” (KSE3). All might be committed to holistic assessment but there were vestiges of the standards-based curriculum lurking not very far beneath the surface: the split between reading and writing amongst two of the KOSAP teachers, for instance, and the yearning for an assessment that had a certain “amount of precision”. Even those who had assessed using 100% coursework did so twenty years before. Yet amongst the KOSAP teachers, there was a willingness to take on the demands that 100% coursework offered.

In particular, for the KOSAP teachers, it was a way in which “guild knowledge” could be internalised by their pupils. It made the peer assessment and the drafting process that arose from it a part of the everyday business of English teaching and improved pupil learning. Most importantly, it gave a holistic edge to the subject of English that was not found in the tick-box approach of KS3 assessment.

Of course there were problems with course-based assessment, particularly with speaking and listening, but the moderation process that teachers were involved in was also significant and, to cite Klenowski and Wyatt-Smith, “Moderation can no longer be considered an optional extra and requires system-level support, especially if, as intended, the standards are linked to system-wide efforts to improve student learning” (2008, p. 1).

In a world where, in England at least, we are going to have grammar tests for eleven-year-olds and terminal summative assessment for sixteen-year-olds, the days of course-based assessment seem long gone. This project *did* herald the end of the key stage 3 tests, though it is very unclear whether or not KOSAP itself helped in this process (Marshall, 2008). Certainly, the project has had no impact on the current Coalition government who just want exams. Yet if our research has shown anything, it highlights the fact that knowledge in English is amorphous or even vague but our ability to assess in this manner, that in some ways incorporates that vagueness, is not. These teachers, at least, were very reluctant to have a system of assessing pupils that ticked all the boxes but was ultimately reductive. Standards-based assessment may fulfil the recent PIRLS and PISA requirements, but this project appeared to assert that to truly assess and improve in English, something different was required. So the battle continues.

REFERENCES

- Black, P., Harrison C., Hodgen, J., Marshall, B., & Serret, N. (2007). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2), 215-232.
- Black, P., Harrison C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4), 451-469.
- Dewey, J. (2005). *Art as experience*. New York, NY: Perigree.
- Eisner, E. (2002). *The arts and the creation of mind*. New Haven, CT: Yale University Press.
- Freedman, S. W. (1991) Evaluating writing linking large scale testing and classroom assessment. *Centre for the study of writing, University of California Berkeley Occasional Paper 27*. Washington, DC: Office of Educational Research and Improvement.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.
- Klenowski, V., & Adie, L. (2009). Moderation as judgement practice: Reconciling system level accountability and local level practice. *Curriculum Perspectives*, 29(1), 10-28.
- Klenowski, V., & Wyatt-Smith, C. (2008, December). Standards- driven reform Years 1–10: Moderation an optional extra? Paper presented at the *Australian Association for Research in Education Conference*.
- Locke, T. (2007). *Resisting qualifications reforms in New Zealand: The English Study Design as constructive dissent*. Rotterdam, the Netherlands: Sense Publishers
- Marshall, B. (2002). Literacy by numbers. *Critical Quarterly*, 44(2), 103-110
- Marshall, B. (2008). The end of testing and future possibilities: An examination of the demise of Sats in England and possible alternative assessments. *English Teaching: Practice and Critique*, 7(3), 65- 73.
- Marshall, B. (2011). *Testing English: Summative and formative assessment in English*. London, UK: Continuum.
- Ofsted. (2007). *Poetry in schools: A survey of practice 2006/7*. London, England: HMSO.
- Qualifications and Curriculum Development Authority. (2009). *Assessing Pupils' Progress: Assessment at the heart of learning*. London, England: HMSO.
- Queensland Studies Authority (2009). *Student assessment regimes: Getting the balance right for Australia*. Retrieved January 7, 2011 from <http://www.qsa.qld.edu.au/assessment/586.html>
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144.
- Sadler, R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 45-63). Amsterdam, The Netherlands: Springer.
- Wyatt-Smith, C. M., & Bridges, S. (2008). *Meeting in the middle: Assessment, pedagogy, learning and students at educational disadvantage*. Evaluation for the Literacy and Numeracy in the Middle Years of Schooling Initiative –

Strand A. Retrieved August 28, 2012 from
<http://education.qld.gov.au/literacy/docs/deewr-myp-final-report.pdf>

Manuscript received: July 15, 2012
Revision received: October 3, 2012
Accepted: January 11, 2013