

## Criteria, comparison and past experiences: how do teachers make judgements when marking coursework?

Victoria Crisp

**To cite this article:** Victoria Crisp (2013) Criteria, comparison and past experiences: how do teachers make judgements when marking coursework?, *Assessment in Education: Principles, Policy & Practice*, 20:1, 127-144, DOI: [10.1080/0969594X.2012.741059](https://doi.org/10.1080/0969594X.2012.741059)

**To link to this article:** <https://doi.org/10.1080/0969594X.2012.741059>



Published online: 12 Dec 2012.



Submit your article to this journal [↗](#)



Article views: 3509



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

## **Criteria, comparison and past experiences: how do teachers make judgements when marking coursework?**

Victoria Crisp\*

*Research Division, Cambridge Assessment, Cambridge, UK*

*(Received 4 November 2011; final version received 5 October 2012)*

The process by which an assessor evaluates a piece of student work against a set of marking criteria is somewhat hidden and potentially complex. This judgement process are under-researched, particularly in contexts where teachers (rather than trained examiners) conduct the assessment and in contexts involving extended pieces of work. This paper reports research which explored the judgement processes involved when teachers mark General Certificate of Secondary Education (GCSE) coursework. Thirteen teachers across three subjects were interviewed about aspects of their marking judgements. In addition, 378 teachers across a wider range of subjects completed an associated questionnaire. The data provide insights into the way that criteria are used, the role that comparison plays in the process and the importance of various professional experiences to making assessment judgements. Findings are likely to generalise to ‘controlled assessments’ which have replaced coursework in the GCSE.

**Keywords:** coursework; assessment by teachers; assessment judgements; marking; GCSE

### **Introduction**

In the assessment of General Certificate of Secondary Education<sup>1</sup> (GCSE) coursework and other similar assessments, the teacher becomes the examiner. The judgement processes by which they determine the marks to award student work are central to such assessments. Yet, the judgement processes involved, both in assessment by teachers and (to a lesser extent) in the marking of traditional exams by examiners, are under-researched relative to other aspects of assessment. A fuller understanding of the judgements involved in marking should provide a sound basis for decisions regarding assessment training and for evaluating the validity of assessment by teachers. This is particularly pertinent given the recent change from coursework to controlled assessments in GCSEs (see Qualifications and Curriculum Authority [QCA] 2007, for information on the nature of controlled assessments).

In the National Criteria Glossary of Terms, coursework was defined as, ‘all types of activity carried out by candidates during their course of study and assessed for examination purposes’ (Secondary Examinations Council [SEC] 1986, 1). Whilst this definition is in some ways still correct, over time the term has perhaps developed a narrower meaning. ‘Coursework’ has tended to be used to refer to

---

\*Email: [crisp.v@cambridgeassessment.org.uk](mailto:crisp.v@cambridgeassessment.org.uk)

work conducted as part of GCSEs and A levels,<sup>2</sup> specifically for the purpose of it being assessed as part of the qualification but carried out over a period of time. It is likely to consist of one or more projects or tasks specifically conducted for this purpose, rather than being work selected from that completed more naturally as part of the course. Additionally, certain types of assessed work that would meet the definition are not usually referred to as 'coursework'. For example, within some vocational courses or vocationally related qualifications for which candidates submit a body of work conducted in the classroom or on work placement, this work is more likely to be referred to as a 'portfolio'. Further, the 'controlled assessments' which have been introduced in GCSEs and are also in use in 14–19 Diplomas (Principal Learning components) could fall within the definition of coursework above, but it might be confusing to refer to them as such. The current research focused on internally assessed work within the GCSE and as a consequence, on project work conducted with the explicit intention that it will be submitted for assessment.

The involvement of students' own teachers in high-stakes assessment is somewhat controversial. In some countries, this is not accepted or used, whilst in others, assessment by teachers is generally trusted and constitutes a significant part of national assessments (e.g. Sweden). There are also examples where it has been introduced relatively recently as a new assessment type. For example, performance assessment was introduced in USA in the 1990s as part of educational reforms (see Baker and O'Neil 1994) and school-based coursework assessment has been introduced in Singapore in the last decade (see Chong 2009). In the UK, assessment by teachers has always been a feature of vocational assessments where it is generally accepted as the best way of assessing certain skills. It is in the context of general or 'academic' qualifications where its role is debated. That said, it has now been a part of the assessment of general qualifications for some time (e.g. Certificate of Secondary Education<sup>3</sup> [CSE], A level, GCSE).

From the time of their introduction, many GCSEs involved a coursework component. The aims of using internally assessed work in this high-stakes context were to enhance validity by gathering wider evidence than can be assessed with written examinations alone, and to encourage teaching and learning of valued skills that are more difficult to assess in written exams (SEC 1985). GCSE coursework was assessed by teachers using criteria and guidance stipulated by the relevant awarding body, internally moderated across classes within schools and then externally moderated by examiners trained by the awarding body. (This process remains the same for many, but not all, controlled assessments.)

In recent years, due to a number of perceived concerns about coursework (such as the burden on teachers and students and the authentication of student work), the examinations regulator at the time (the QCA) was asked to review the value of coursework (Department for Education and Skills 2005), including whether it was still providing useful learning experiences. The QCA review involved surveys of teachers, students and parents and consultation with appropriate assessment experts. Some concerns were found (e.g. risks of plagiarism, burden, too much structure or support from teachers which might be limiting the learning experience), with views sometimes varying between subjects. However, amongst teachers there was general agreement that coursework had a positive impact on teaching, learning and assessment with greater benefits than drawbacks (QCA 2005). Nonetheless, the review led to the replacement of coursework with controlled assessment. These assessments

often still involve teachers in the marking process and thus the findings of the current research are likely to be broadly applicable to this new assessment form.

There is a growing body of literature on assessment judgements in the context of marking by examiners and higher education assessors (e.g. Crisp 2008, 2010; Cumming 1990; Greatorex and Suto 2006; Lumley 2002; Milanovic, Saville, and Shuhong 1996; Vaughan 1991). These have provided various insights into aspects of the processes involved, such as in relation to the marking process, the social context, and how features of work affect decisions. However, only one previous study has explored the judgement processes involved when teachers assess GCSE or A level coursework (Morgan 1995, 1996), and this was specific to mathematics and had a strong focus on the discourse of mathematics rather than on judgement processes.

Marking research has been able to identify various 'cues' or aspects of responses that appear to be central to marking decisions (e.g. Cumming 1990; Milanovic, Saville, and Shuhong 1996; Vaughan 1991), and that these usually relate closely to marking guidance. However, differences in the aspects considered important have sometimes been found. For example, Vaughan (1991) found differences between raters in what they focused on in the assessment of writing (e.g. organisation, content, grammar, handwriting). Also, Cresswell (1997) found that the criteria used in grading judgements were modified in response to the nature of the material under consideration. Evidence from assessment in higher education has found that assessors use similar features of student work to each other and have similar understandings of quality but may give different evaluations of the same piece of work (Grainger, Purnell, and Kipf 2008). Another area of interest has been the relationship between mark scheme criteria and intuitive impressions of student work and whether this is always straightforward (e.g. Lumley 2002).

Sadler (1989) proposed that there is a pool of potential criteria available for any assessment, composed of two subgroups: manifest criteria, which are consciously attended to; and latent criteria which are unconscious but may be triggered by something unexpected, enter consciousness and become part of the set of manifest criteria, at least temporarily. The manifest criteria, in particular, are likely to have commonalities with the written marking criteria but it may not be possible to adequately represent the full pool of manifest and latent criteria in written assessment criteria. Features such as handwriting could be latent criteria that are generally not important to assessment but may become manifest criteria if triggered (e.g. by observing work with untidy handwriting). Equally, latent criteria may represent qualities that it is appropriate to assess but that are less common or less clearly articulated in written criteria. The notion of manifest and latent criteria seems plausible given the evidence that different assessors sometimes focus on different criteria and that Cresswell (1997) found that the criteria used in grading judgements were modified in response to the nature of the material under consideration.

The role of comparison in assessment judgements is of interest. Sadler (1989) asserts that teachers often find it difficult to make isolated judgements of quality without referring to a range of students' work. Additionally, Laming (2004) argues that judgements are not independent but are always relative or comparative. Thus, the evaluation of a piece of student work might involve comparison to other pieces of work (actual or remembered), to the marking criteria (or an internalised understanding of the criteria) or to a notional model answer. Laming suggests that there is a tendency to use model answers as a basis for comparison and that these

may differ between individuals based on past experiences. In the context of coursework marking by teachers, there could be more likelihood of differences than in other assessment contexts, given that: teachers generally only have the work of their own students to provide a reference point for comparison; and teachers may have more varied professional experiences in relation to assessment than a group of examiners who have received centralised training. However, the availability of exemplar coursework projects and various exam-board training and guidance may support teachers in providing appropriate reference points for use when marking.

In the context of the continued debate over whether teachers can be trusted to assess students fairly, it is notable that the processes underpinning teachers' judgements during marking have been subject to relatively little investigation. Further research would help to develop a more comprehensive model of how teacher assessment could be optimised, which may potentially be relevant to a range of contexts. Robbins (1998, 28) argued that, 'the nature of judgements being made and the reasons for them need to be made explicit if the dependability of coursework and the credibility of examinations is to be maintained and improved'. The current research aimed to understand the judgements involved in coursework assessment in order to contribute to this aim. Drawing on the literature, the research aimed, in particular, to explore the role of comparison, of criteria and the experiences that inform assessment judgements. The continued development of a better understanding of marking processes could suggest ways to improve consistency in judgement-based assessments.

## **Method**

This research involved interviews with a small number of teachers to explore relevant themes around the assessment judgements in an unrestricted way, followed by a survey to explore whether the various aspects of the marking judgements identified are widespread.

### ***Interviews***

Thirteen teachers were interviewed about their coursework-marking practice. Each of these participants was teaching one of three specific GCSE courses: English/English Literature (three teachers), Information and Communication Technology (ICT) (five teachers) or Geography (five teachers). The teachers represented seven schools and ranged in their teaching experience from 2 to 22 years.

The teachers were interviewed about their coursework-marking practices either individually or sometimes as a group where several teachers taught at the same school. The interviews were semi-structured in nature and covered themes apparent in the extant literature on exam marking. The interviews included questions on the thought processes involved in marking, how written criteria are used, features of the work that teachers attend to, the various past experiences that enable teachers to make marking judgements and the role of comparison.

Whilst the interview responses will not be representative of all relevant teachers, the participants were carefully selected to represent three contrasting subjects and a number of different schools.

### ***Questionnaire survey***

A questionnaire was used to survey a larger number of teachers across a wider range of subjects in order to validate and expand on the interview findings. The respondents were 378 secondary school teachers involved in coursework marking. Most of these teachers were recruited by asking examiners running training sessions to distribute the questionnaire to course delegates. A smaller number of the respondents were recruited by 'snowballing' with a teacher known to the researcher asked to pass the questionnaire to colleagues.

The participants ranged from 0 to 40 years of experience in teaching (mean = 14.9 years, standard deviation = 10.9 years). The teachers were grouped into four bands by teaching experience (0–5 years, 6–10 years, 11–20 years, 21 years or more) to allow comparison of responses.

Thirteen main subject areas or subject combinations were represented by the teachers, ranging across the arts, sciences, humanities, technology, English, business and social sciences. There were nine subjects with at least 10 respondents and these groups were used to explore possible differences between subjects. Amongst those teachers who indicated their gender, there were more female respondents than male (216 and 133, respectively) – perhaps connected to a predominance of female Health and Social Care teachers, the most heavily sampled subject.

The majority of responding teachers taught GCSE and/or A level courses. Some respondents taught BTEC<sup>4</sup>, OCR Nationals<sup>5</sup> or Entry Level as well as GCSE or A level courses, and a small number of teachers taught Entry level courses only. Some of the analyses explored whether responses to the survey varied by the qualification taught but no differences of interest were found.

The questionnaire was drafted based on themes in the existing literature and was informed by insights from the interviews. Most of the questions required teachers to respond using a Likert scale (e.g. never, occasionally, sometimes, usually, always), whilst some invited 'yes' or 'no' answers. These questions were supplemented by a number of free response questions. The themes covered included:

- strategies such as searching for evidence of criteria and making comparisons;
- the points at which teachers have an overall impression of the quality of a piece of work and what features influence this impression;
- whether teachers feel they have an internalised notion of the quality of work required for different grades or bands; and
- the importance of various past experiences to their marking practice.

Whilst the sample of survey respondents cannot be confirmed as fully representative of the population of appropriate teachers and there could be a bias towards teachers keen for professional development (and thus attending training), it does represent a range of school types, subject areas and (as the training courses targeted were in various locations across the UK) a range of geographical areas.

### **Results**

The results are presented under a number of themes. Within each section, complementary data from the interviews and survey are presented where both methods inform that theme. The interview data provide an overall representation of

the issues, and the survey data then provide detail on the extent to which processes/practice are similar or different across teachers and subjects.

### *Comparison*

When teachers were asked during interview whether marking involves an element of making comparisons between students' coursework pieces, there were clear differences between the three subjects. The English teachers felt that comparisons were an important part of the process. One English teacher commented that, as well as using the criteria, comparing a piece of coursework with another that is fairly similar helps to 'fine-tune' her judgement and to decide the exact mark. Most of the Geography teachers reported that they did not knowingly make comparisons between students when marking, that they treat each coursework piece as a separate piece of work and would only very occasionally think back to projects marked earlier. One Geography teacher reported that once she has marked a few coursework projects, she does begin to think about how new pieces compare and puts them into rank order as she marks. She expressed the need to get the rank order right as well as aligning to the mark scheme. The ICT teachers expressed mixed views on the role of comparison. Three ICT teachers felt there was an element of comparison, comparing the piece being marked to ones of similar quality that have already been marked, whilst the other two reported that they simply compared against the mark scheme.

In the survey, several questions asked teachers how frequently they make comparisons between pieces of work, see Table 1.

A third (35.3%) of teachers reported that they 'always' compare the work of different students within a class, and around 90% of teachers' responses suggested that this was at least 'sometimes' the case. The pattern of responses suggests that within-class comparisons of students are more important in some subjects than others, as shown in Figure 1. For example, comparisons were reportedly particularly frequent for Physical Education (PE) coursework, and rather less so for Science. This is perhaps explained by the nature of internally assessed work in these subjects, with PE involving a substantial performance element (and thus comparisons between the performances of individuals may inform judgements) whilst Science coursework is likely to involve research or practical work which may be easier to directly compare to marking criteria.

Table 1. Percentages of responding teachers selecting each response to survey questions about making comparisons.

When marking coursework, how often:	Never	Occasionally	Sometimes	Usually	Always
Do you make comparisons between pieces of work from different students within your class?	1.6	9.1	23.5	30.5	35.3
Do you compare the work of the student with your recollections of the work of previous students?	9.3	14.9	28.1	29.4	18.3
Do you compare the work of the student with actual examples of coursework from previous students?	17.5	17.5	27.1	21.5	16.4



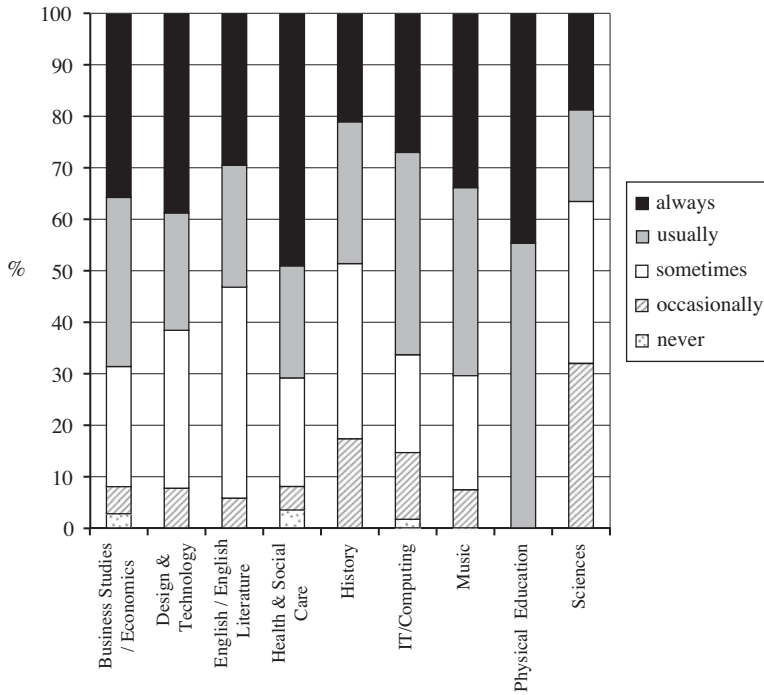


Figure 1. ‘When marking coursework, how often do you make comparisons between pieces of work from different students within your class?’ – Responses by subject.

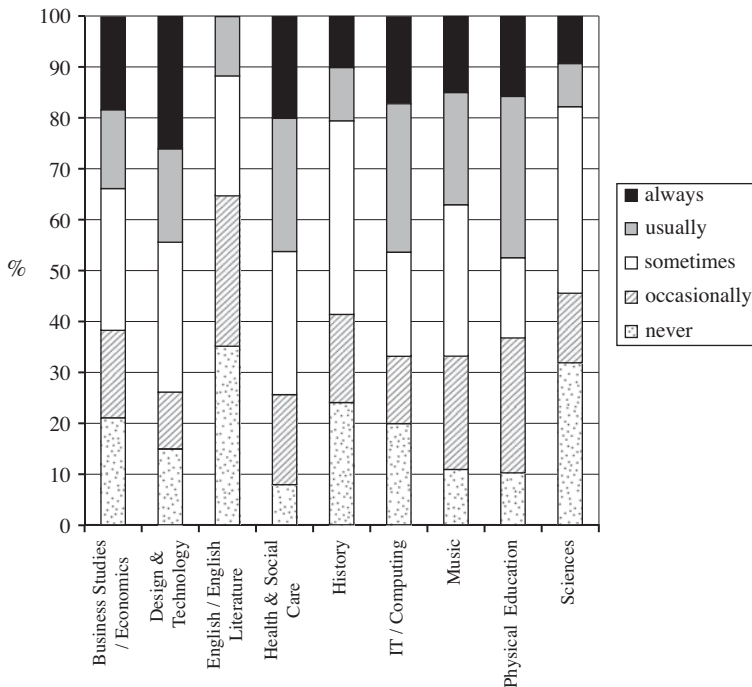


Figure 2. ‘When marking coursework, how often do you compare the work of the student with actual examples of coursework from previous students?’ – Responses by subject.



When asked about the frequency with which they make comparisons with recollections of previous students' work (see Table 1), responses indicated that this was fairly frequent but a little less important than comparing between current students.

The third question about comparisons asked teachers how often they compare work with actual examples of work from previous students. Responses were mixed and differences between subjects were found, as illustrated in Figure 2. For science and English teachers, comparisons with past work are reportedly less frequent than for other subjects. The responses relating to comparisons could suggest that for science, comparisons with current or previous work are less important than in other subjects. For English, comparisons between current students' work appear to be important but comparisons with the work of previous students seem to be less necessary, or just less practical given the extended textual nature of such work. Alternatively, less frequent comparisons with actual work for English could suggest stronger internalised notions of standards in this subject.

### ***Features affecting marks***

During interviews, the teachers were asked to describe their thought processes when marking or evaluating student coursework. Their comments often included reference to the specific features that would affect marks, as described in the relevant marking criteria. For example, English teachers commented on features such as good use of quotations and minimal use of narrative (in literature essays), and Geography teachers commented on project structure and aspects of data collection and representation. A few additional insights of interest arose. Depth of analysis and originality of points appeared to be important for English, especially for higher grades. These are represented in the marking criteria. Carrying out an independent project in Geography rather than a group project (which will be more teacher-led) was noted as something likely to gain students more marks for equivalent work because the student will have worked more independently. Another consideration, noted by a few participants, was a need for awareness of issues around plagiarism and being alert to possible indicators of copying.

### ***Active searching vs. observing and evaluating***

To explore the distinction between manifest and latent criteria (Sadler 1989), interviewees were asked whether marking involves actively searching for particular features, or reading through the work and evaluating the features observed, or a combination of both. Perhaps unsurprisingly, all teachers felt that a combination of the two strategies was used. One teacher described this in terms of knowing what it is that she wants to find in the work, and that when reading through features 'jump out' allowing her to 'mentally tick off' criteria as well as getting a general overall impression. Several others gave similar descriptions. Additional comments on this topic included that when a teacher is new to using a particular mark scheme, there is a tendency to be more focused on, and more led by, the mark scheme and that the combination of the two strategies resulted in a 'back and forth' process of going through the work to seek particular types of evidence not yet found.

Two questions in the survey were intended to explore the emphasis placed on pre-determined criteria or indicators relating to criteria (see Table 2). Perhaps unsurprisingly given the nature of most coursework mark schemes, many teachers

Table 2. Percentages of responding teachers selecting each response to survey questions about searching vs. observing.

When marking coursework, how often:	Never	Occasionally	Sometimes	Usually	Always
Are you actively searching for evidence of particular features/ qualities?	0	0.8	2.7	30.1	66.5
Are you looking through to see what is included and then evaluating this?	0.3	1.6	9.7	43.7	44.8

responded that they are ‘always’ actively searching for evidence of particular features/qualities when marking coursework. Teachers were also asked how often they look through the work and then evaluate what they find (see Table 2); this ‘finding then evaluating’ strategy was also reported to be frequent by many respondents.

### *Features that are not actively sought*

The interviewed teachers were asked whether there are sometimes features in coursework that they were not looking for but that come to affect their thinking or the marks awarded. Ten of the 13 teachers felt that such features could come to their attention. Several suggested that this sometimes related to content in student work that shows more initiative, for example, collecting extra primary data for Geography coursework without being prompted by the teacher. Such cases might result in extra marks being awarded where the marking criteria reward independent work. Two English teachers noted that features such as presentation, spelling and punctuation, essay length and ‘messy’ work are sometimes noticed. The latter was thought to give the impression that the student does not care about the work. Most teachers felt that although they might sometimes notice features in student work that they did not expect, these would not normally affect their process of going about marking, or the marks awarded (unless the marking criteria valued such features). One exception noted was that if a student has structured their work less conventionally, this could lead to a revised strategy for how to locate the relevant evidence.

In the survey teachers were asked: *Do any features/qualities other than those described in the mark scheme ever influence your evaluation of a piece of coursework?* Teacher responses suggested that this was not common but over 35% of teachers who responded to this question reported that this occurred at least sometimes. (The frequencies of each response were: never 36.7%, occasionally 24.6%, sometimes 23.6%, usually 11.8% and always 3.2%.) The reported occurrence of some use of other criteria is consistent with Sadler’s notion of ‘latent’ criteria and could potentially mean there is scope for features not represented in marking criteria to influence judgements occasionally. Considering the frequencies of response by subject (see Figure 3) suggested that features or qualities not shown in the mark scheme reportedly influence judgements less frequently in science and PE and more frequently in music. The former could be due to the less ambiguous nature of knowledge and understanding in science (at least at GCSE and A level), and perhaps due to the strong performance element in PE which means that marking criteria describe levels of performance skills in a broad, generic way. For music, it could be hypothesised that teachers feel that not all aspects of performance or

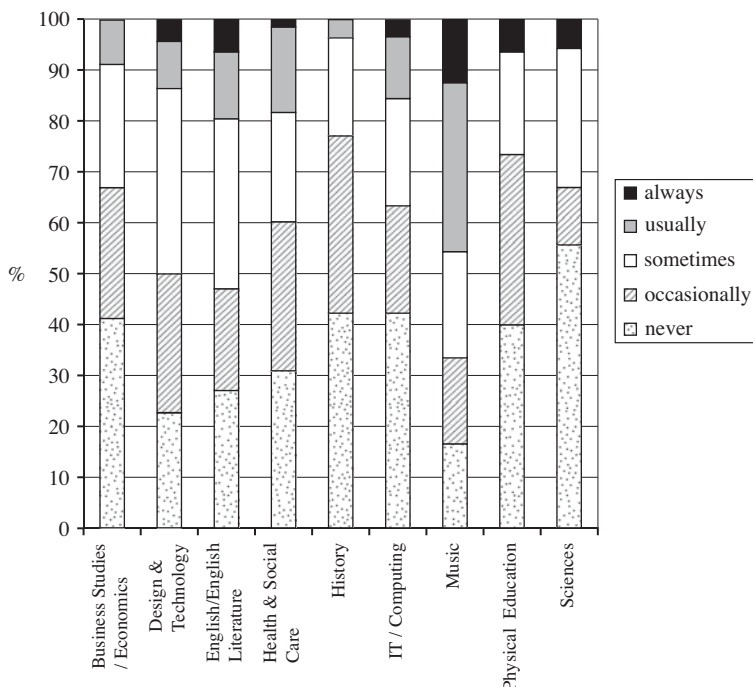


Figure 3. 'Do any features/qualities other than those described in the mark scheme ever influence your evaluation of a piece of coursework? If so, please give details.' – Responses by subject.

composition can be fully represented by marking criteria (although the criteria do tend to deal with a range of aspects such as accuracy, fluency, expression and style).

An open response space was also provided and teachers were asked to give any examples of features or qualities, other than those in the mark scheme, that influence their evaluations. A total of 166 examples were given by 137 teachers. The most frequent responses (occurring at least 10 times) related to clear presentation of student work, the quality of writing (or writing style) and originality. These were apparent for teachers from a variety of subject specialisms. Other features mentioned by at least six teachers were extra research or study, the level of independence shown by the student vs. the help required, and logical/coherent work.

The surveyed teachers were asked the extent to which they agreed or disagreed with the statement: 'When I read a piece of coursework, features or qualities that I wasn't really looking for sometimes become obvious and then I look for evidence of these as I continue reading'. Over 50% of teachers agreed with the statement. (The percentages of each response were: strongly agree 4.2%, agree 51.8%, neither agree nor disagree 34.4%, disagree 8.5%, and strongly disagree 1.1%.) These responses generally support Sadler's view on the use of criteria from an available pool. However, taken in combination with the responses to the previous questions on this theme, the use of latent criteria may be relatively infrequent. This could indicate that the precision of the mark schemes controls the criteria used to a fair extent.

### ***Judging overall quality***

When asked about their thought processes during marking, several of the interviewed teachers noted that they tend to have a rough idea of the quality they are expecting in a particular student's coursework based on having observed the student doing the work and having seen one or more earlier drafts. Whilst the initial expectation of the quality of work seems to provide a starting point for evaluation, there was a feeling that this did not constrain their overall judgement of the final work produced. Several teachers noted how they are trying to form an overall impression or 'feel' for a piece of work when evaluating it. This was described by one teacher as trying to, 'hold the whole answer in my head drawing on every aspect of it'. Some teachers felt that they could gain a first impression of quality from an initial scan or even from reading the first page and that this gives them a rough idea of the level of the work which they then adjust up or down as they read more of the piece.

Judging the work against their internalised understanding of the marking criteria and direct use of the mark scheme were mentioned by some interviewees in relation to judging overall quality. Participants seemed to vary with regard to whether they marked against individual criteria or sections thus leading to an overall mark, or whether they determine an overall level first and only then check this against the written criteria.

The survey asked teachers: *Do you think that you hold mental representations of what coursework on different grades/bands/levels is like?* All survey respondents answered this question, with 67.2% responding 'yes', 3.2% responding 'no' and 29.6% answering 'partly'. Further exploration of the responses showed a progressive pattern that those with greater teaching experience were more likely to report that they have internalised notions of work on different grades, bands or levels (see Figure 4). This could indicate real differences in how accurate or well-formed their notions of standards are, or could reflect their confidence in their understandings of standards.

Survey respondents were asked whether they have an approximate idea of the quality of work before marking from seeing earlier drafts: Before you start the final marking, do you have a rough grade/band/level in mind that you are expecting from seeing earlier drafts of the students' coursework? Over 50% of teachers responded

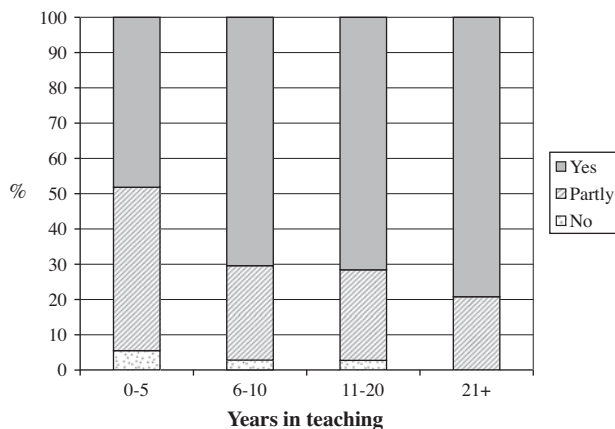


Figure 4. 'Do you think that you hold mental representations of what coursework on different grades/bands/levels is like?' – Responses by years of experience.

that this was 'usually' the case. (Percentages of responding teachers giving each response were: never 4.5%, occasionally 9.8%, sometimes 23.5%, usually 51.6% and always 10.4%.)

Two open-response questions were included in the survey to gather insights into how soon teachers form an impression of the overall quality of a coursework project and what informs this impression. These issues are central to how judgements are made at an overall level. They also relate to the thoroughness of the consideration of student work. Firstly, teachers were asked: *After beginning the final marking, how soon do you form an impression of the overall quality of the coursework pieces? (e.g. after reading the first page, about half way through reading, only after reading the whole project).* Of the teachers who answered the survey, 367 gave a response to this question. In line with the previous question, a few teachers reported that they had an impression of overall quality before they began the final marking because they have seen earlier drafts and have observed the student working. Some teachers felt they had an overall impression after reading a small proportion of the work. For example, 25 teachers responded that the first page gave them this impression and 20 reported that they gain this impression quite quickly or early on. Many teachers reported that they have gained an overall impression of quality after reading a certain proportion of the work. For example, 101 teachers reported having an impression of overall quality once they have read half of a project. A large number of teachers (135) claimed not to have an overall impression until they have read the whole project. Some teachers noted in their comments that although they may form an impression on overall quality before reaching the end of the work, they always consider the full piece and often adjust this early evaluation as they see more evidence. For example, one teacher stated: 'Inevitably [I have an impression of overall quality] from the beginning but I try not to let this impression prejudice my judgement of the whole piece and all its components'.

The second open-response question on this theme asked: *What are you looking at initially to form this impression?* Over three quarters of the teachers (315) responded including one or more points. Many points related to features of work that would be expected to affect a teacher's overall impression of quality, such as the match of the work to the marking criteria, and the content (perhaps indicated initially by headings in a piece of written work). Also mentioned by a number of teachers were depth, accuracy or detail, the structure of work, whether the work answers the question or title set and the quality of analysis, evaluation, argument or interpretation. A few features that reportedly affected initial impressions might arguably be less anticipated by marking criteria (at least in most subjects). Some teachers noted, for example, that the clarity of the student's expression affected their impression of the work, and others that presentation was an influence.

### ***Importance of experience and training***

The interviewed teachers were asked about the impact of professional experiences and training on their marking practice. Working with colleagues, drawing on their experiences and discussing coursework examples were thought to be important, particularly when a teacher was new to teaching or to marking. Internal moderation experiences were also considered helpful. Externally run courses on coursework were regarded as useful and at one school, it was considered essential that the teacher leading the coursework marking attended this training and fed back to

others. One teacher felt their experience as an exam marker informed their marking of coursework, and another teacher made an interesting point that exam board training had suggested the need to mark according to the centrally set standards (i.e. to apply someone else's standards rather than necessarily his own) in order to achieve consistency of marking. Feedback comments from external moderation were noted as useful by some teachers, both in terms of how to mark work and in terms of ideas to suggest to their students in future.

When asked about the effect of experiences of seeing various examples of student work over time, this was usually considered useful in terms of providing a reference for current marking. Some teachers mentioned keeping old coursework examples for occasional physical reference. Experience with lots of coursework pieces ranging in the ability shown was considered important.

Survey respondents were asked about the importance of a number of experiences to their marking practice: *How important do you feel are each of the following to your understanding of the standards required for coursework? Please rank in order of importance (1 being the most important and 7 being the least important)*. The frequencies with which each type of experience was assigned each rank are shown in Table 3. If a teacher gave any tied ranks rather than ranking from 1 to 7 as requested, these responses were removed from the analysis.

Teacher responses were quite varied but some overall patterns in the relative perceived importance of different experiences can be observed. Exam board guidance and exam board courses were most frequently ranked as the first and second most important influences on coursework marking. School-based experiences such as working with other teachers, experiences of student work, and internal moderation were frequently ranked in the second or third positions of importance. 'Other courses' received low rankings, perhaps because of the lack of specificity of this item and because only a limited range of other courses may be relevant to coursework marking in a particular subject. Involvement in examining was often reported as of low importance as an influence, but this may be partly due to a fairly limited proportion of teachers being involved in examining.

Table 3. Teachers' rankings of the importance of different experiences to their understanding of standards.

Experience	Percentages of teachers assigning each ranking							Omit (or excluded)
	1 Most important	2	3	4	5	6	7 Least important	
A Working with other teachers	13.8	11.1	17.2	14.0	15.1	6.9	0.5	21.4
B Experiences of students' work	10.6	12.4	16.1	17.7	13.8	7.4	1.1	20.9
C Internal moderation meetings	5.6	12.7	15.3	16.9	16.1	8.5	2.1	22.8
D INSET courses run by the exam board	24.3	15.1	9.5	11.9	13.5	4.2	0.5	20.9
E Other courses	0.5	0.3	0.5	1.6	3.4	20.6	47.1	25.9
F Exam board guidance materials (e.g. exemplars)	19.6	21.4	14.8	10.8	8.7	2.9	0.8	20.9
G Involvement in examining	5.6	6.3	5.8	6.1	6.3	24.1	19.6	26.2



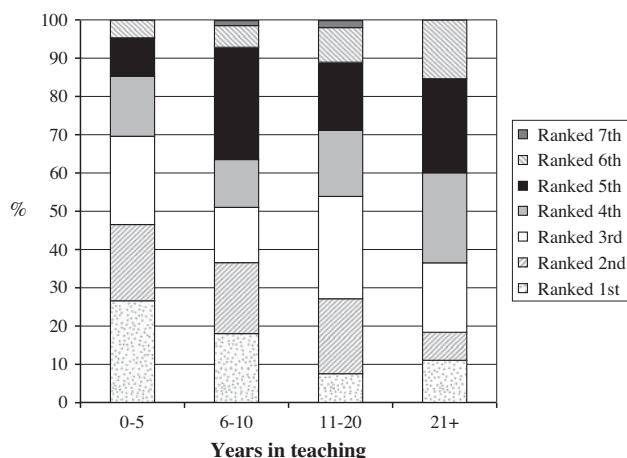


Figure 5. Teacher rankings of the importance of Experience A (working with other teachers) to their understanding of standards – Responses by years of experience.

For Experience A ‘working with other teachers’, there were differences in response depending on the number of years of teaching experience. As shown in Figure 5, working with other teachers appears to be more important for new teachers and becomes less important with additional years in teaching. This finding is consistent with theories of communities of practice.

Other patterns in the responses to this question suggested that working with other teachers is more important in subjects where marking might arguably be more subjective, such as music and English. Experience of students’ work appeared to be most important in music and least so in science. This may relate to the arguably more ephemeral nature of music, meaning that marking criteria need more support from examples of work in order to be fully understood.

## Discussion

The interviews and survey responses provide various insights into aspects of the judgement process involved when teachers mark GCSE coursework. Whilst coursework has now been replaced with controlled assessments, which usually involve more limited but broadly similar types of tasks, most of the current findings are likely to be relevant to teacher marking in this new context.

One of the themes of interest in this research was how marking criteria are used and strategies around identifying features of student work which indicate its quality. The current research suggests, unsurprisingly, that actively looking for evidence of particular criteria is frequent in coursework marking. This aligns with Sadler’s (1989) notion of manifest criteria and Vaughan’s (1991) findings regarding marking strategies focused on particular criteria. Teachers’ responses also indicated that a strategy of viewing the content of student work and then evaluating that content in relation to the relevant criteria was also frequent. These two strategies (seeking evidence of particular criteria, and looking at work and evaluating content) appeared to be part of the marking practice for most teachers, although some felt that one or the other was dominant.

There was some evidence that features that teachers were not actively seeking might come to their attention during assessment. Most of the given examples related



to potentially credit-worthy features such as showing initiative, originality, extra study and independent working. Teachers tended to feel that when unexpected features of work were noted, these did not change their marking process as such but might lead them to look out for this feature as they continued. However, such features reportedly only affected marks if this was legitimated by the mark scheme. This evidence gives some support for the presence of latent criteria that come into use as manifest criteria if triggered, as proposed by Sadler (1989). However, this does not seem to be a major factor.

A second key theme of interest in this research was the role of comparison. The current evidence supports the assumption that comparisons between students' work play a role in the assessment process for some teachers, though this may be more prominent in some subjects than others. Comparison was reportedly sometimes used to fine-tune the mark awarded and to check that the pieces are in the appropriate rank order. Comparison was found to be more important in English of the subjects represented in the interview data, and to be particularly frequent for PE of the subjects represented in the survey. The greater occurrence of comparison in English and PE may be due to the qualitative nature of the types of features valued in the marking criteria. For PE, the finding is consistent with Hay and Macdonald's (2008) research. As discussed earlier, Laming (2004) argues that all judgements are comparative. This would imply that when comparisons between pieces of work are not being used, there must be comparison against some other reference point. Laming's (2004) ideas link to Rosch's (1978) notion of prototypes and categorisation processes. This is thought to involve comparison between the representation of the current material and prototypical representations of responses that are typical of students achieving a particular level (which will have been developed through the experience of many examples of student work). Baird (2000) suggests that such categorisation processes may be at work during grading decisions in relation to examinations. Somewhat similar comparative or matching processes could underpin possible comparisons between the representation of the current student work and representations of the mark scheme. During interviews, some teachers described marking as occurring in light of experience of exemplar work used at training meetings and in light of feedback from the external moderator on previous years' student work. This provides some support for comparison of current coursework pieces to a prototype (based on previously seen examples) as a part of marking.

Linked to the role of comparison in judgements is the notion of how an overall impression of quality is gained. The current research suggests that many, but not all, teachers hold internalised notions of the standards expected for particular grades. These are likely to be helpful to their judgements, provide a reference point for comparison and assist with consistency across teachers if their notions of standards are similar. Mental representations of grades/levels were more commonly reported by teachers with greater experience.

The experiences that inform coursework marking were a further theme of interest in this research. The interview data suggest that school-based experiences with colleagues (such as involvement in internal moderation, working with colleagues), experiences of many examples of student work and exam board courses were important influences on their marking practice, helping in terms of providing a reference point for the standards to be applied. Guidance and training from the exam board, perhaps unsurprisingly, were considered more important than other factors by survey respondents. Working with other teachers was more important to

those with less experience, in line with theories of communities of practice. Communities of practice (e.g. Wenger 1998) are groups with a shared knowledge, repertoire of terms and tacit conventions. This theory could apply to how a new teacher joins the teaching community, and specifically the subject community in the school in which they teach and that they develop their understanding and skills in relation to relevant processes through ‘legitimate peripheral participation’ (Wenger 1998). Interactions with other teachers and with senior examiners at exam board training events may make them part of subject-related teaching communities of practice (concerned with a range of practices as well as assessment) and part of an extended community of practice of assessors conducting marking using their understandings of examiners’ views, as discussed at training and articulated through relevant documentation.

## Conclusion

Insights into the use of criteria, the role of comparison and the kinds of experiences that support marking have been gained through this research and may have implications for practice. Ensuring that teachers attend to appropriate features of student work and are not influenced by inappropriate ones relates to the validity of the inferences that can be made from coursework scores. We need to know that the appropriate qualities in the work have been valued during marking judgements in order to know that the scores reflect the knowledge, understanding and skills that were intended to be the focus of the assessment. In terms of identifying relevant features in student work (and relating these to written criteria), strategies sometimes varied but teachers appeared to attend in detail to appropriate features. Achieving a good level of marking consistency between teachers is highly desirable and relates to the notion of reliability. However, it could be detrimental to try to force all teachers to go about the complex process of marking extended pieces of work such as coursework using a particular strategy, unless there was strong evidence to suggest a problem with a particular strategy or practice. Further research could usefully explore whether certain practices (e.g. forming an overall impression of student work after only the first page, the extent to which comparisons are used) have consequences for marking accuracy and consistency.

Given that materials and training supplied by the awarding organisation are reported as important in informing coursework practices, there is a strong argument that the quality and availability of such provision need to be ensured. Recent increases in online provision of qualification-related training may assist in this regard. Further to this, whilst most qualification-specific training courses for teachers are due to cease in 2013 to minimise risks of inappropriate information being given (see Ofqual 2012), face-to-face training relating to controlled assessment will still be permitted where this is essential to preparing teachers to set, deliver and mark controlled assessments. This acknowledges the greater role of the teacher in such assessments and the importance of providing appropriate support. Also, given that comparisons and internalised notions of expected standards appear to play a role in judgements, it is probably important that teachers develop sufficient experience with a range of pieces of student work to provide the necessary point of reference. This means that the provision of marked examples by the awarding organisation is likely to be important along with school-based experiences such as involvement in internal standardisation. The latter would suggest that involving all members of a department

in internal moderation (rather than this process being undertaken by one senior department member) may have advantages.

The current research provides insights into the judgement processes underpinning marking by teachers, an area of assessment theory that is relatively under-researched. Whilst the data were collected in the context of GCSE coursework, many of the findings are likely to be broadly applicable to controlled assessments and to other contexts where teachers assess student work for high-stakes, summative purposes.

## Notes

1. GCSEs are qualifications, available in various subjects, taken by most students at age 16 years in England, Wales and Northern Ireland.
2. A levels (General Certificate of Education, Advanced level) are qualifications taken by many students at age 18 years in England, Wales and Northern Ireland.
3. The CSE was a predecessor to the GCSE. It was available from the 1960s to the 1980s.
4. BTECs are vocational qualifications offered by the awarding organisation Edexcel. BTECs are named after the organisation which originally offered these qualifications, the Business and Technology Education Council, a predecessor of Edexcel.
5. OCR Nationals are vocational qualifications offered by the awarding organisation Oxford, Cambridge and RSA Examinations.

## Notes on contributor

Victoria Crisp is a senior research officer at Cambridge Assessment and areas of research and publication include: issues in question difficulty and examination validity; investigating the effects of answer spaces on student responses; profiling the use and purpose of annotations in examination marking; and exploring the judgement processes involved in marking.

## References

- Baird, J. 2000. Are examination standards all in the head? Experiments with examiners' judgements of standards in A level examinations. *Research in Education* 64: 91–100.
- Baker, E.L., and H.F. O'Neil. 1994. Performance assessment and equity: A view from the USA. *Assessment in Education: Principles, Policy & Practice* 1, no. 1: 11–26.
- Chong, K.K. 2009. Whither school-based coursework assessment in Singapore? Paper presented at the International Association of Educational Assessment annual conference, September 13–18, in Brisbane, Australia.
- Cresswell, M.J. 1997. Examining judgements: Theory and practice of awarding public examination grades. PhD diss., Institute of Education, University of London.
- Crisp, V. 2008. Exploring the nature of examiner thinking during the process of examination marking. *Cambridge Journal of Education* 38, no. 2: 247–64.
- Crisp, V. 2010. Towards a model of the judgement processes involved in examination marking. *Oxford Review of Education* 36, no. 1: 1–21.
- Cumming, A. 1990. Expertise in evaluating second language compositions. *Language Testing* 7, no. 1: 31–51.
- Department for Education and Skills (DfES). 2005. *14–19 Education and skills white paper*. London: Department for Education and Skills. <https://www.education.gov.uk/publications/eOrderingDownload/CM%206476.pdf>.
- Grainger, P., K. Purnell, and R. Kipf. 2008. Judging quality through substantive conversations between markers. *Assessment and Evaluation in Higher Education* 33, no. 2: 133–42.
- Greatorex, J., and W.M.I. Suto. 2006. An empirical exploration of human judgement in the marking of school examinations. Paper presented at the International Association for Educational Assessment conference, May 21–26, in Singapore.

- Hay, P.J., and D. Macdonald. 2008. (Mis)appropriations of criteria and standards-referenced assessment in a performance-based subject. *Assessment in Education: Principles, Policy & Practice* 15, no. 2: 153–68.
- Laming, D. 2004. *Human judgment: The eye of the beholder*. London: Thompson Learning.
- Lumley, T. 2002. Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing* 19, no. 3: 246–76.
- Milanovic, M., N. Saville, and S. Shuhong. 1996. A study of the decision-making behaviour of composition markers. In *Performance testing, cognition and assessment*, ed. M. Milanovic and N. Saville, 92–114. Cambridge: Cambridge University Press.
- Morgan, C. 1995. An analysis of the discourse of written reports of investigative work in GCSE mathematics. PhD diss., Institute of Education, University of London.
- Morgan, C. 1996. The teacher as examiner: The case of mathematics coursework. *Assessment in Education: Principles, Policy & Practice* 3, no. 3: 353–75.
- Ofqual. 2012. *Exam board seminars: Final report*. Coventry: Office of the Qualifications and Examinations Regulator. <http://www.ofqual.gov.uk/files/2012-04-26-exam-board-seminars-final-report.pdf>.
- Qualifications and Curriculum Authority (QCA). 2005. *A review of GCE and GCSE coursework arrangements*. London: Qualifications and Curriculum Authority.
- Qualifications and Curriculum Authority (QCA). 2007. *Controlled assessments*. London: Qualifications and Curriculum Authority. [http://www.ofqual.gov.uk/files/qca-07-3208\\_controlledassessmentsfinal3.pdf](http://www.ofqual.gov.uk/files/qca-07-3208_controlledassessmentsfinal3.pdf).
- Robbins, J.H. 1998. *Improving the dependability of examination coursework and assessment: A discussion paper*. London: Qualifications and Curriculum Authority.
- Rosch, E. 1978. Principles of categorization. In *Cognition and categorization*, ed. E. Rosch and B.B. Lloyd, 27–48. Hillsdale, NJ: Erlbaum.
- Sadler, D.R. 1989. Formative assessment and the design of instructional systems. *Instructional Science* 18, no. 2: 119–44.
- Secondary Examinations Council (SEC). 1985. *Working paper 2: Coursework assessment in GCSE*. London: Secondary Examinations Council.
- SEC. 1986. *Working paper 3: Policy and practice in school-based assessment*. London: Secondary Examinations Council.
- Vaughan, C. 1991. Holistic assessment: What goes on in the rater's mind? In *Assessing second language writing in academic contexts*, ed. L.H. Lyons, 111–25. Norwood, NJ: Ablex.
- Wenger, E. 1998. *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.