

**Bulletin number: 12**

## **How do centres estimate pupils' attainment?**

Paper produced for  
Scottish Qualifications Authority  
by J Hart and C Howieson,  
Centre for Educational Sociology, University of Edinburgh

First edition: April 2005  
Publication code: AD2613

Published by the Scottish Qualifications Authority  
Hanover House, 24 Douglas Street, Glasgow G2 7NQ, and Ironmills Road, Dalkeith,  
Midlothian EH22 1LE

*The information in this publication may be reproduced in support of SQA qualifications. If it is reproduced, SQA should be clearly acknowledged as the source. If it is to be used for any other purpose, then written permission must be obtained from the Business Manager, Research and Information Services at SQA. It must not be reproduced for trade or commercial purposes.*

SQA is committed to the use of robust evidence in the development and evaluation of policy and its implementation, and carries out or commissions research across a range of topics to support this.

The publication of Research Bulletins allows us to disseminate the results of our research activity to practitioners, policy makers, parents, academics and anyone else who has an interest in the key role that qualifications play in economic growth and social inclusion in Scotland.

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
	Background and aims of the study	1
	Selection of subjects	1
	Selection of centres	1
	Analysis of estimates at national and centre level	2
	Issues covered in the research	3
	The interviews with staff	4
<b>2</b>	<b>Estimating — Policy and Practice</b>	<b>5</b>
	Policy on estimating	5
	The purpose of prelims	5
	The timing of prelims and estimating	6
	Prelim papers	7
	Prelim marking	8
	Arriving at estimates of performance	8
	Staff perceptions	9
	Subject-related factors	10
	Candidate numbers and class composition	10
	Staffing issues	11
	Submitting estimates	11
<b>3</b>	<b>SQA Support</b>	<b>12</b>
	What SQA does	12
	What SQA could do	13
<b>4</b>	<b>The Value of Experience as an SQA Appointee</b>	<b>14</b>
	Extent and use of appointee experience	14
	Perceived advantages of appointee experience	14
<b>5</b>	<b>Conclusions</b>	<b>16</b>
	Issues arising from the investigation	16
	Recommendations arising from the investigation	19
	<b>Appendix 1: The Interview Protocol</b>	<b>21</b>
	<b>Appendix 2: The Centres Visited</b>	<b>26</b>

# How do centres estimate pupil's attainment?

A paper produced for SQA by the Centre for Educational Sociology, Edinburgh University

## 1 Introduction

### Background and aims of the study

The estimates submitted by centres are essential information for SQA. They are used in determining grade boundaries and derived grades and to support appeals. Accurate estimates can also have some benefits for centres and even more benefits for candidates. It is important to SQA that estimates should be as accurate as possible, but the extent of the match between centres' estimates and candidate performance in the external assessment is known to vary between centres and between subjects. SQA therefore commissioned the Centre for Educational Sociology at the University of Edinburgh (CES) to explore how centres arrive at estimates and to attempt to identify approaches which lead to more accurate estimates with a view to improving SQA's processes and issuing guidance to centres.

The research design was determined in conjunction with SQA. This included the subjects to be covered in the interviews and the centres to be visited. The key elements of the investigation involved analysis of SQA data on estimates to identify subjects and centres for inclusion and a programme of interviews in the selected centres.

### Selection of subjects

In the first phase SQA provided data on estimates for 2002-03 and 2003-04 and this was analysed jointly by CES and SQA to agree which subjects should be examined. The intention was to include a range of subjects and types of assessment (practical, essay-based etc). The subjects were also selected with a view to ensuring that there were high enough numbers of presentations to give useful correlations between estimates and results. On the basis of these criteria the following subjects were selected:

- ◆ Biology
- ◆ History
- ◆ English
- ◆ Information Systems (IS)
- ◆ French
- ◆ Mathematics (Maths)

### Selection of centres

It was initially agreed that 10 centres should be included in the study but this was later extended to 11 to ensure a balance of subjects among the centres in the sample. The objective was to select centres that included departments with high, average and low levels of accuracy in estimating. The purpose of getting this spread was twofold: (i) to investigate whether the level of accuracy in estimating revealed in the national data could be explained by particular

approaches to making estimates in the centres involved in the study; and (ii) to assess the judgement of the interviewee about the department's success in estimating compared with the actual rating for the department in the national data.

Other considerations in selecting centres were to ensure a geographical spread across Scotland and coverage of both rural and urban catchments. It was also decided to include an independent school and a college.

A threshold of 20 presentations in the 2003 diet at both Intermediate 2 and Higher in each of the selected subjects was initially used in selecting centres. It was premised that centres with reasonably large numbers of presentations would be more able to give generalized responses than would be likely with smaller presentations. It was felt centres presenting small numbers would be more affected by issues related to specific candidates or groups of candidates, a view that appeared to be borne out in the investigation.

In the end, however, the threshold of 20 had to be lowered for some subjects to secure the geographical coverage: in some cases a presentation of 16 or more, usually in one subject at one level, was accepted. In two cases a much lower presentation in one subject at one level was accepted (details of this are given in Appendix 2).

## **Analysis of estimates at national and centre level**

CES undertook analysis of data for the 2003 diet to identify centres for inclusion on the basis of their accuracy in estimating in the selected subjects. The SQA data gave the numbers of grades and bands estimated by every centre in every subject and the numbers of grades and bands actually achieved by candidates in each centre. They also gave correlations between estimates and achievements. Considering this data, it was evident that, at national level, correlations between estimates and results for all subjects in the 2003 diet at Higher was 0.68 for grades and 0.69 for bands. French produced the highest correlations (0.80 and 0.83), whilst English produced the lowest correlations (0.59 and 0.61). A similar pattern emerged at Intermediate 2 where the correlation between estimates and results for all subjects was 0.58 for both grades and bands. French again had high correlations (0.71 and 0.70) although Biology had slightly higher; and English again had low correlations (0.43 and 0.45) although at this level the correlations for History were similar.

The accuracy of estimating in departments was judged by working out the average correlations between grades estimated and grades achieved and bands estimated and bands achieved and comparing the correlations for the departments to be visited. A code was then noted beside the department giving a broad indication of its success at estimating relative to the national correlations for the subject. These were used as a guide to ensure that, as far as possible, the centres selected included departments with high, average and low correlations between the estimates submitted and the results.

The centres visited and the subjects covered at each were as follows:

<b>State secondary schools</b>	<b>Location</b>	<b>Subjects covered</b>
Centre 1	(south west)	French; History; Maths
Centre 2	(west)	Biology; French; IS
Centre 3	(north east)	History; IS; Maths
Centre 4	(west)	Biology; English; IS
Centre 5	(west )	English; French; History
Centre 6	(north east )	English; History; IS
Centre 7	(east )	Biology; Maths; IS
Centre 8	(east )	French; History; Maths
Centre 9	(central)	French; English; Maths

<b>Independent school</b>	<b>Location</b>	<b>Subjects covered</b>
Centre 10	(east)	Biology; English; History
<b>Further education college</b>	<b>Location</b>	<b>Subjects covered</b>
Centre 11	(west)	Biology; English; Maths

More information about these centres is given in Appendix 2 of this report.

Because it was decided to select schools with entries of 20 or more at two levels, the schools selected were mainly large schools. They also tended to have good pass rates across the curriculum, although this was not related to the accuracy of their estimates.

SQA contacted the selected centres to inform them of the research and invite their participation — all agreed to take part. CES then contacted centres to arrange the interview programme.

## **Issues covered in the research**

The second phase took place in parallel with the selection of centre. CES held discussions with relevant SQA managers and selected Customer Account Managers to refine the issues to be explored and to develop the topics which the interviews should cover — these were policy, practice and perceptions. The schedule was amended slightly after the first centre visit and this amended version is attached to the report as Appendix 1. As the interviews progressed less stress was placed on certain questions, which were not yielding much data and more emphasis was given to other areas of the schedule. The shift was away from general reflections on estimating towards more practical discussions about the mechanics of arriving at estimates.

The interview protocol can be thought of as being in five sections, as follows:

- ◆ Factual questions on the centre and the department.
- ◆ Questions about the estimating policy and practice in the centre.
- ◆ Questions about SQA materials and other support available to teachers/PTs.
- ◆ Questions about influences and pressures on teachers in the context of estimating.

- ◆ Invitations to reflect on the process.

This report concentrates on these headings.

## **The interviews with staff**

The interviews were carried out over a two month period, October and November 2004. Interviews were conducted with three principal teachers in each of the 10 schools and three section leaders in one college. The visits to the centres were arranged by the SQA Co-ordinator in each of them. The SQA Co-ordinators were not formally interviewed but they did raise issues in informal discussion with the researcher and these are reported where relevant. On reflection, it would have been useful to have included SQA co-ordinators in the interview programme since they have a key role with regards to many aspects of the estimating process.

The interviews were scheduled to last 50 minutes, which was the length of a period in most of the centres visited. In the event the interviews varied in length from 30 to 90 minutes, but most were completed within the scheduled time. The school departments varied in size, from 2 to 9. Most appeared to be stable in staffing terms, with a majority of experienced staff but this was not the case in at least three departments. We discuss the impact of staffing (including the effect of the McCrone agreement) later in the report.

In the main, the principal teachers and section leaders interviewed saw themselves as combining the roles of leaders, subject experts and managers, although the balance varied considerably from interviewee to interviewee.

All of the staff interviewed were very co-operative and open. Most indicated an interest in the investigation and an enthusiasm for it. There was general and unsolicited support for SQA commissioning research of this kind.

## **2 Estimating — Policy and Practice**

### **Policy on estimating**

Subject managers were asked whether the policy they followed in arriving at estimates was set by the education authority, the school or the department. This question was asked to test out the possibility that departments were being pressurised to estimate higher or lower than was justified by pupil performance for some reason associated with authority or school targets.

Almost without exception interviewees felt that the department set the policy on estimating and that the policy they set was based on specialist subject knowledge. By ‘policy on estimating’ they meant the nature, amount and origins of the evidence on which the estimates were based. As will be seen from the sections below, this varies in some significant ways, not only from subject to subject, but also from school to school.

In all the centres visited, it appeared that all departments were expected to conduct prelims at times set by senior managers, usually in consultation with subject managers. It could be said that in this way senior managers were setting policy related to estimating, but none of the interviewees saw this as an issue. Although there was evidence of a growing interest in departmental results among senior managers in schools and local authorities, and an increase in the use of targets in managing schools, there was no evidence at all of pressure on departments to estimate high or low. If these changes are having any effect relative to estimating, it is to move attention away from estimating and towards better ways of ensuring that more students are successful in the courses they undertake. In the departments visited, this might focus on the policy of admitting learners to courses at different levels or on ways of using assessment to sustain motivation among students and help them to concentrate their efforts where it would have the most effect on their examination performance. With pressure to work on these areas of departmental or section management, improving the accuracy of estimates as such did not emerge as a high priority target for subject managers in the centres visited.

### **The purpose of prelims**

Whilst the subject managers interviewed were aware in a general way of the importance of estimates, few gave any indication that they had a detailed understanding of how SQA uses estimates; a number of those interviewed made statements which suggested that they may have misapprehended some of the processes involved.

For many it appeared that providing estimates was simply an administrative chore and for their staff estimating did not have the same importance which SQA gives it in its literature or in its practices. On the other hand, there was no evidence that the requirement to run a prelim was an issue with the interviewees. Although the setting and marking of prelims was described by most interviewees as burdensome, all accepted it as an unavoidable burden and one with a sound pedagogical basis. Prelims were seen by most interviewees as a valuable opportunity for candidates to experience a replication of the final examination a few months in advance.

The prime focus of the prelims for interviewees, therefore, tended to be on giving candidates a reasonable view of what they were capable of at the time of the assessment and what they needed to do to maintain or improve that level of achievement through the remainder of the course. Since this is closely related to the issue of maintaining motivation in these students, this led to a wide range of approaches to the question of how difficult the prelim should be — an issue which is picked up later in the report.

It may be that SQA managers have made it harder for departments to decide on the level and content of the prelim by changes to what constitutes appropriate evidence for appeals. The impression gained from the interviewees was that there is now a good deal of uncertainty in centres about what SQA will or will not accept as appropriate or adequate evidence. Indeed a number of interviewees noted there appeared to be variation in the view emerging from SQA depending on the subject in question. If there was a prevailing view among interviewees on the issue of estimates and appeals evidence it might be stated thus:

*Previously you could run a prelim, knowing that it represented a snapshot of attainment at a particular time or stage of the course. On the basis of that you could make an 'other things being equal' estimate that the candidate would probably achieve similarly in the exam and SQA would accept that evidence on that basis if you put in an appeal. Now the evidence has to be based on an exact replication of the final assessment and that can't be done through the prelim, so the prelim isn't really serving the same purpose and there is no obvious way to produce the kind of accurate estimate which SQA seems to want.*

Interviewees described a range of ways of dealing with this issue and these are described later in this chapter.

## **The timing of prelims and estimating**

The approach taken to the nature and timing of prelims across the centres visited was fairly uniform. The schools used what might be considered the traditional times of November/December and January/February for prelims, with new National Courses tending to be examined in January/February. Departments tried to mirror the final external assessment as far as they could at these times, given that candidates would not have completed the full course by either of these dates. In one of the school visited, one department managed to hold a prelim from which they derived estimates in March, by which time their candidates had completed the course.

In a number of schools, the earlier November/December prelim date was reserved for candidates in S4. This meant that S4 candidates for new national courses would be sitting prelims two months earlier than those in S5/S6 and this could create difficulties with regards candidates taking Intermediate 2 courses in S4. In at least one case this led to the same prelim being used with S4 candidates in November/December and with S5/6 candidates in January/February. The interviewee concerned acknowledged that this breached security. This issue did not affect most schools in this study.

Given the tendency to view prelims as contributing to learning and teaching (rather than being about the need to submit estimates), this meant that the traditional timing of prelims was not an issue to the interviewees. However, from the point of view of estimating, most interviewees did raise timing as an issue, because the prelim usually coincides with the

completion of the second of three course Units. Some departments felt they were able to take account of this; others that it simply ruled out the possibility of achieving accurate estimates. Since classes had not completed the course by the time of the prelim there were difficulties when it came to developing estimates which would support successful appeals. For many, perhaps most, of the interviewees, it was at this point that their own processes and those of SQA were most at odds.

Most of the departments visited ran a third Unit test of some kind after the prelims. These usually included questions which required the recall or application of knowledge and/or skills from other parts of the course. Only one department was able to carry this out within a timescale which allowed them to use the mark from this test in reaching estimates, but most put aside this evidence for use should there be appeals. While some departments were continuing to try to address the problem, most appeared to be resigned to the difficulty and were taking no particular action to address it.

Two interviewees talked about plans in the school to give more flexibility to departments on the timing of prelims — although the implication was that departments would still hold prelims. In one of these cases, the idea was to allow departments to set their own prelim dates; in the other it appeared that senior management were trying to resolve problems associated with ‘prelim leave’ by doing away with the prelim diet for Standard Grade and replacing it with a system of subject arrangements to be carried out within timetabled classes. The departments interviewed were concerned that they would not be able to find big enough blocks of time in the timetable to allow them to replicate the final examinations.

## **Prelim papers**

The source of the prelim papers varied. Two thirds of interviewees used commercial papers, although in almost every case these were altered in some way; the other interviewees made up their own papers by adapting past SQA papers or using a combination of source materials. In the sample, Mathematics PTs were most likely to make up their own papers; English and French PTs least likely. Given the number of centres visited, however, it is unlikely that any significant conclusion can be drawn from this.

Of the commercial papers, P&N was the most heavily used across subjects and was particularly favoured by the English and French PTs. Perfect Papers and Pegasus papers were also widely used and in Biology, DART papers were mentioned by two subject managers. Again the small numbers make it hard to draw any conclusions about this and the comments by interviewees, even in the same subject, varied on the extent to which they found the different publishers’ papers to be suitable or accurate or to need amendment. Virtually every interviewee who used a commercial paper amended the paper or the marking scheme which accompanied it, but there was no way of finding out whether the same changes were being made. More importantly, perhaps, were the comments on SQA’s view of commercial papers. In particular, some interviewees were convinced that SQA had a confirmed view such as, ‘We hear that only papers by such and such a publisher are acceptable for appeals.’

A small number of the interviewees had one paper which was used over again each year or two papers which they alternated. These papers were amended in the light of performance and feedback on appeals with a view to producing an ultra-reliable instrument of assessment. There was no evidence from the data that this was producing particularly accurate estimates.

## **Prelim marking**

Approaches to marking also varied — although less so. Consistency was the driver in almost all cases. Where the candidature was not too big, single teachers (the PT or an experienced teacher, or a member of department who was also an SQA marker) might mark all responses to a single paper. In bigger departments the practice tended to be that single teachers would mark all the answer to a particular section or question in the paper. Only in a small number of cases did teachers mark their own class sets. All departments used marking schemes of some kind.

For most of the subject managers, the aim of the assessment was to produce a realistic assessment. They hoped that SQA feedback on appeals would confirm that they were doing this or assist them to make improvements in this respect. However, views on the purpose of prelims could affect the approach taken to marking prelim scripts. For a minority of interviewees the purpose of the prelims was to give candidates a shock and prompt them to work that bit harder towards the final exams so their approach was ‘to mark hard’. Their perception that their marking standards were over-demanding tended to be confirmed by SQA feedback on their appeals, but this was not seen as a reason to change their practice. Other interviewees expressed the opposite view and said that their practice was to set and mark prelims which would encourage candidates and not put them off. To this end they encouraged a degree of leniency in prelim marking. Again SQA feedback on appeals had confirmed that the marking was too lenient, but again there was no intention to change the approach on this account.

## **Arriving at estimates of performance**

Approaches to setting estimates were also varied. A number of possible components or stages in setting estimates emerged across the interviews. These were:

- ◆ carrying out a formal and standardised assessment of the candidate’s achievement at some fixed point in the year (normally about two thirds of the way through the course) based as closely as possible (given the timing) on the final external assessment for the course
- ◆ using the professional judgement of class teachers on the likely performance of individual candidates or groups of candidates in the final external assessment to evaluate the results of the formal assessment and set the pass mark and cutting scores
- ◆ carrying out a formal and standardised assessment of the candidate’s achievement in the final third of course, also based on relevant aspects the final external assessment for the course
- ◆ applying rules, or perceived rules, regarding what SQA will accept as appropriate evidence for appeals

Although it might be expected that all these components would be used in some combination to arrive at estimates of candidates’ performance, this was seldom the case in the departments covered by the research. For reasons of timing and consistency, most departments separate out the first component, taking steps to ensure validity and reliability, and simply using the results of the formal assessment to generate estimates. Sometimes this was done in an unmediated way, using the raw marks emerging from the prelim and simply applying ‘SQA cutting scores’; but in other departments what amounted to a pass marks procedure was carried out and cutting scores were being set based on judgements about the ease or difficulty

or fairness of the paper. In reaching this judgement use might be made of the professional judgements of teachers based on their expectations about particular candidates or groups of candidates. In one department this was formalised to the extent that there was a form of pre-prelim estimating.

In a small number of cases the approach of using raw marks derived from a single assessment was seen as having limitations and teachers were given some scope to use other information. Here teachers might be able to produce an alternative estimate based on a wider portfolio of evidence if a case could be made to the Principal Teacher/Section Leader. In this model, teachers/lecturers had a defined level of discretion regarding the basis for their estimates and this was moderated by the Principal Teacher/Section Leader. In one case, individuals appeared to be given full discretion in this respect in spite of the interviewee's view that this meant that most estimates which emerged were unrealistically optimistic.

In a small number of departments the sequence of teaching and testing was combined with the timing for the submission of estimates so that the evidence for the post-prelim, third Unit test could be taken into account, but most of the interviewees saw that as very difficult to administer and a number indicated that they doubted that it would not produce benefits which would justify the effort involved. In these departments the evidence from any post-prelim assessment was simply kept in reserve to be used in the case of appeals.

One interviewee took the radical step — based on close analysis of past performance — of consistently estimating higher than the prelim results on the basis that most of the students in that department would in fact produce evidence to match that estimate between the prelim and the end of the course. Other staff to whom this approach was outlined, rejected it firmly — some on principle and some on practical grounds.

However, the different approaches to setting estimates did not appear to make a difference to the accuracy of estimates. Analysis indicated that the various approaches were not consistently associated with high or low correlations between estimates and results in the departments/sections concerned.

## **Staff perceptions**

It was evident from the interviews that subject managers' views on what constitutes an appropriate level of accuracy in estimates varied. For example, one interviewee was working on the basis that to estimate to within a band was success and was able to give percentages of 'accurate to within a band either side', 'accurate to within two bands either side' and so on. On this basis the interviewee felt that the department was good at estimating — a view which was at odds with the rating from the 2003 results. Nevertheless, this might be seen as a reasonable judgement from the perspective of the interviewee or the department.

Staff's views of the purpose of estimates also had a bearing on the accuracy of estimates. As we noted earlier in this chapter, the subject managers interviewed concentrated on the pedagogical value of prelims with a focus on giving an assessment of the performance of individuals. This focus carried over to departments' review of the accuracy of their estimating in the light of actual performance by candidates. Here the key judgement often seemed to be more about the performance of individuals than about overall statistics. In most cases any review of success in estimating in the light of results appeared to be very informal, often at the level of individual teachers. Most interviewees thought their department was 'okay' or

'pretty good' at estimating but only three produced or referred to any statistical analysis. Appeals results were more likely to be the subject of analysis.

## **Subject-related factors**

The question of what makes it easy/difficult to produce estimates in each subject was explored to some extent with all interviewees, but on the basis of the analysis of the accuracy of estimates at national level it was given particular attention in the interviews with staff responsible for French and English. The analysis of SQA data had shown that across Scotland French had the highest level of agreement between estimates and results while English had the lowest. Most interviewees with whom this was raised (including those for French) were surprised by this, assuming that Maths would have the highest correlation. The possible explanations which were offered included: the language courses have been quite stable over time; they are centred on one set of skills and knowledge and reinforce a lot of the same learning over and over; most assessments in the subject are made on the same criteria; and the external assessments are very predictable. There was no opportunity to test these out to any extent within French or between French and other subjects. The observation is offered as a possible basis for further investigation.

As we have already indicated, marking schemes were used in prelim marking and these schemes vary according to the character of what is being assessed. At one extreme, some papers in some subjects require very detailed marking schemes while at the other the marking simply requires elaborations on the published grade descriptors. The impression which came through from the interviews was that it may be easier, paradoxically, to achieve consistency in the latter case, although this may also be related to the predictability of the assessment instruments.

## **Candidate numbers and class composition**

Most interviewees did not see the number of presentations as having a significant effect on the possibility of reaching accurate estimates. However two interviewees did point out that, statistically, where a department has large numbers of candidates the effect of random discrepancies (eg from special circumstances) will be less noticeable than in smaller presentations.

Two phenomena which related estimates to the composition of classes emerged from the interviews.

The first was the case where a department, for whatever reason has consistently large numbers of high achieving candidates. Here the number of potential fails will be smaller and the scope for variations within the A grade is greater, so that the pass/fail and grade correlations are likely to be higher.

The opposite case is one where a department runs a course for small numbers of candidates who enter for a wide variety of reasons and with very different levels of commitment to the course. The example given was of Intermediate courses for mixed groups of S5/6 students to the course: since the attitudes of these students may vary considerably over the year, it was stated, final results are more likely to be at odds with estimates, giving low and inconsistent correlations. The issues here relate to school or department policies on admissions to courses, which were not part of this investigation. These appear to be driven increasingly by a pressure

on centres to put all candidates into the appropriate courses from the outset. Interviewees reported on the concern of school and authority managers to eliminate failure and get good results and in a number of cases reported that this was limiting student choice and closing down the chance for young people to try things out and take limited risks.

This may be of interest to SQA in so far as SQA's policy statements on estimates may influence admissions policies and the climate in which they are made. .

## **Staffing issues**

As we noted in the introduction, most of the school departments involved in the research appeared to be stable in staffing terms, with a majority of experienced staff, but at least three appeared to be in a state of some chaos due to the difficulties of replacing staff on sick leave or maternity leave or both. One interviewee judged that both the estimating and the results for that department in the previous year had been poor due to that problem. The inclusion of colleagues in the department with additional responsibilities (eg management or pastoral responsibilities) was the subject of a question, because it had been anticipated that this might be a factor in maintaining an understanding of standards across the department. It was not seen as an issue. The impression given was that in most cases these colleagues were experienced teachers who understood the demands of the subject and contributed fully to the work of the departments they were in.

The re-structuring which has been introduced following the McCrone report was having a significant effect in three of the schools. Two of the interviewees were in new posts of PT Curriculum and Guidance and had only been months in post and in one case, a new McCrone PT, who came from a different subject to that which was the subject of the interview, asked to sit in on the interview to gain subject information. It was clear even from this small sample, that the McCrone arrangements will create issues for SQA when it comes to the issues surrounding estimating. The effect of McCrone will be varied across the country, but will mean that traditional subject Principal Teachers will be replaced. SQA will need to consider how their replacements can best be supported in this aspect of their work.

## **Submitting estimates**

Most schools set a date on which estimates had to be submitted to the schools' SQA Co-ordinator and around half of those interviewed were aware of this date, but not of the date by which estimates had to be submitted to SQA.

The interviewees who were fully aware of SQA requirements tended to be in schools which submitted estimates as late as possible. Of these, only three said that they would occasionally use the time to adjust estimates, normally where a candidate had performed significantly better in class tests taken after the prelim than in the prelim itself. This appeared usually to be associated with special circumstances. Of the departments which submitted early, a few were clear that estimates could be altered, but only one claimed to have used this facility — also in the case of significant improvement associated with special circumstances at the time of the prelim.

### 3 SQA Support

By and large the view of SQA support in relation to estimating was positive, but there appeared to be considerable confusion about the nature and provenance of various publications, about what SQA did/didn't publish and about what the status of different publications might be — eg how confidential they were. Some discussions would end up with statements like, 'or maybe these were the ones published by HSDU' or '...but I think this is confidential, so we don't use it.'

#### What SQA does

All interviewees were aware of the Course Arrangements documents and the Grade Descriptors for their subjects and it appeared that in most cases individual teachers had access to their own copies of these in departmental handbooks, or other similar documents. Usage varied from 'all the time' to 'as necessary'. A number of more experienced interviewees felt that the arrangements and grade descriptors were part of their thinking by now and therefore used them less frequently than at first.

Views on the usefulness of the grade descriptors varied by subject. It appeared that where the characteristics covered were very generic — eg for essay answers — they were found most useful and might even be used with candidates. Some interviewees felt that they were lacking in detail, but this was not a general complaint. In two interviews, however, this call for more precision appeared really to be a call for a change in the standards set for formal accuracy in the subject.

The general trend of the comments was in support of more concrete exemplification. Marked examples, published marking instructions and the professional development events were all rated highly by those who were aware of them or had used them.

A number of references were made to materials available on SQA's website ([www.sqa.org.uk](http://www.sqa.org.uk)). Only one interviewee indicated that there were problems of access to the Internet within the centre and one suggested that the SQA website was hard to navigate, but since interviewees were not specifically asked about access or the website it is hard to judge what the value of these responses is. Overall it appeared that the interviewees were evenly divided between those who used SQA's website regularly, those who were familiar with the website but seldom used it, and those who did not use the website. Again, however, the small number of interviewees, together with the fact that five out of the thirty were computing specialists makes it difficult to draw specific conclusions from this other than that SQA should be aware that a significant number of subject managers do not appear to see the website as an obvious reference source. This often emerged when interviewees were asked about their awareness of specific items of SQA support and answered negatively, adding 'Oh, that will be on the SQA website.' It also manifested itself in responses which indicated that SQA information came to them through the SQA Co-ordinator, who would get it off the website.

Principal Assessor (PA) reports and subject update letters were well known by all interviewees and the majority seemed to circulate these and discuss them in departmental meetings. In the case of the PA report, the topic of discussion, according to a number of interviewees, was 'Does this apply to us?' and that might be linked to an informal review of

passes and appeals. A number of interviewees were critical of aspects of the PA report, however, indicating that it was repetitive and insufficiently detailed to help them.

As might be expected very few of the interviewees had contact with the PA for their subject, while about half had had contact with the Qualification Manager. In the main these interviewees knew the Qualifications Manager by name and where this was the case spoke very positively of the assistance they gave.

## **What SQA could do**

Most interviewees felt that SQA was already doing a lot and did not have an immediate or very convinced answer to the question, 'What more could SQA do?' Two suggestions were made by a number of interviewees, but in each case they were qualified by statements to the effect that the interviewee was not convinced that what was being proposed was actually realistic.

One was the proposal that SQA should publish a national prelim. This was offered by eight interviewees although most offered it with accompanying remarks such as, 'but that would probably be too expensive.'

The other was the request from a third of interviewees — for more or fuller feedback on appeals evidence. Could this be more detailed? Could it highlight what was in line with SQA requirements as well as what was not? Could feedback be given in successful cases as well as unsuccessful ones. These comments were in line with the general trend to look for more concrete forms of support. Again this was offered with the rider that it was probably not possible for SQA to do this.

Additionally, although this was not a direct answer to this question, many interviewees who were also appointees emphasised the value of discussion of the subject and the standards. This also came up in some of the responses on the professional development workshops, for example: 'They highlight common problems and bring teachers back to discussing the concepts behind the subject. This kind of continuing dialogue with teachers is very important. It feeds back into teaching.' Although SQA cannot take on responsibility for this kind of development, it may wish to bear this in mind when considering the format of events involving subject teachers and might also want to draw this finding to the attention of the Scottish Executive Education Department, Learning and Teaching Scotland or whichever other agencies can take steps to meet this need.

## **4 The Value of Experience as an SQA Appointee**

### **Extent and use of appointee experience**

Most of the departments/sections interviewed had some staff with experience of marking or examining Standard Grades, SCE Highers or new National Courses at one or more levels. Most of the experience was with Standard Grade, but a number of the interviewees were markers, setters or examiners for new National Courses and a small number had carried these duties through from SCE Highers.

The extent to which the subject managers interviewed knew whether or not staff were appointees varied considerably. A small number of interviewees clearly saw working for SQA as a matter for the individual concerned and of no interest to them in their role as subject manager.

By and large, the presence of appointees in the departments was seen as desirable and useful, although only a few interviewees appeared actively to encourage their staff to become markers. One interviewee appeared to see it as a mark of status, the kind of experience which validates subject leadership. On the other hand another interviewee argued that SQA should be breaking down such perceptions and taking steps to spread the advantage among younger staff.

### **Perceived advantages of appointee experience**

Two interviewees (in different subjects) made the point strongly that they thought marking was literally and metaphorically unrewarding but other interviewees who were involved in marking, or other appointee work, or had staff who were appointees, made no reference to payment as an issue and generally saw involvement as a useful process. Most comments were to the effect that participation in marking gave an insight into ‘what is acceptable’ or ‘what SQA wants’, and there were several comments along the lines of, ‘It clarifies the grey areas’.

This view of marking as giving insight into what answers are acceptable seemed to be imply that this was ‘hidden’ knowledge — either because SQA is looking for something special that can’t be accessed any other way, or because marking teams keep changing the requirements in subtle ways. However, even where the knowledge gained from marking was seen as important, it was not always fed back into the department and where it was, it was usually done by informal processes. Only a small number of interviewees said that they gave appointees a role in the department which would ensure that their experience was made use of. This usually meant putting that colleague in charge either of the setting or the marking of the relevant paper in which they had appointee experience. Sometimes this could not be done because the colleague who was an appointee held another role in the school, such as Depute Head, and could not be asked to take on that level of additional responsibility in the department.

Interviewees felt that no matter how open SQA was about its processes, it would always be advantageous to be an appointee. This was for two reasons: first because the sheer volume of material which would have to be published to cover all that was gained by being an appointee would be unmanageable from a teacher’s point of view (and probably from SQA’s); and

secondly — and possibly more importantly — because a lot of the advantage in being an appointee came from meeting and talking with colleagues from other centres. From this it appeared that some of the perceived advantage in being an appointee may not relate directly to understanding the standards being set by the marking process, but may relate more to learning and teaching issues. In other word, the appointee system is making a considerable but patchy contribution to the quality of learning and teaching around the country.

## **5 Conclusions**

The subjects that were investigated provided a fair cross-section of the curriculum, covering a reasonable range of syllabus patterns and forms of assessment. The centres selected offered a range of accuracy in estimating and also covered a variety of locations and catchment areas.

### **Issues arising from the investigation**

#### **Policies on estimates**

The question of who sets policy about the development of estimates, turned out not to be an issue for interviewees and there was no evidence of any pressures on staff which would lead them to do other than produce estimates which truly reflect candidate performance as accurately as possible. The departmental approach to the role of the prelim in informing and motivating students, however, did have an effect on the kind of estimates submitted.

#### **The purpose of estimates**

In the main the subject managers interviewed did not appear to have a clear understanding of how SQA uses estimates or to see it as something which should be of particular concern to them in carrying out their responsibilities. Their concerns were focused on how to improve their candidates' success rates. Most interviewees said that prelim marks were used in the centre for reporting on students' progress but few talked in terms of using the estimates for any other formal purpose. Although candidates were informed in a general way of the estimating process, estimating as such did not appear to be an important feature of teacher-student discourse.

#### **Approaches to estimates**

One of the main aims of the study was to examine whether high/low correlations from the national data would match with particular approaches to making estimates revealed in the interviews. Although approaches to estimating varied across the centres and across subjects, no simple matches emerged between the approaches to estimating as described by the interviewees and the ratings derived from the national data. It is also true to say that the range of factors identified as affecting the estimating processes and practices adopted by the different departments, and the extent to which these reflect matters such as school or departmental ethos, makes it unlikely that a single approach could be identified which would guarantee an improvement in accuracy.

#### **Sector differences**

The interviews at the independent school were useful but none of the issues raised there appeared to be peculiar to the independent sector. The visit to the college, on the other hand, threw up some of the specific approaches associated with the college sector, in particular issues associated with the nature of the college year, which tends to start later than the school year and problems associated with night school classes. It would probably be fair to say, however, that while these might merit further investigation by SQA to ensure that its procedures are taking full account of the college sector, there did not appear to be any issues of principle in the college visited which were not also found in the schools.

## **Reviewing the accuracy of estimates**

There are issues about how the accuracy of estimating is to be judged and more work should be done on this if departments are to be encouraged and supported to review their performance in estimating. Subject managers' judgements on the level of accuracy of estimating by their departments/sections appear to be often at odds with SQA's analysis but their reasons for this make sense from a teaching perspective.

For the subject managers the key judgement is often more about the performance of individuals than about overall or statistical accuracy, but even their views on how to judge how accurate a set of estimates is varied. This made it difficult, if not impossible, to draw any consistent conclusions about the relationship between the interviewees' perceptions of success in estimating and the correlations prepared for the investigation. SQA needs to consider how to address this gap between its awarding body perspective and the teaching perspective on estimating as part of its strategy on communication about standards.

## **Subject differences**

The issue of subjects for which it is easier/harder to produce accurate estimates might be worth further investigation. The factors which emerged from the study were: the stability of the subject arrangements; a concentration on similar skills over time; a close link between Unit and external assessment; and the predictability of the external assessment. In so far as the factors concerned are extrinsic to the subject and can be affected by SQA (eg stability), they might be added to the considerations which affect decisions to review subject arrangements. However, if they are closely linked to the nature of the subject (eg the nature of the skills development) they would simply have to be taken account of in any review of how SQA should deal with different levels of accuracy in different subjects.

## **The timing of estimates**

The timing of estimate submission was regarded by most interviewees as problematic. Most commented on the difficulties of estimating performance before candidates had completed the course. In virtually all cases this meant setting prelims when the candidates had completed only two of the three Units of the course. However, for most interviewees, this issue was tied to the nature of the evidence which SQA would accept for appeals rather than to the timing of prelims. The impression given was that if the submission date for estimates was later and SQA was more flexible regarding the evidence that was accepted in appeals, the estimates would be more likely to be accurate. In most of the departments in the investigation, prelims would not be moved, but estimates would be based on the prelims plus some form of formally-generated supplementary evidence. The data available neither supported nor countered this view.

## **Role and conduct of prelims**

Prelims were universally seen as a good thing, but were more strongly linked with issues relating to learning and teaching than with estimating. Interviewees tended either to have a strong view about what made a good prelim or to be troubled by this issue. The link between a good prelim paper from the department's point of view and one which would be well received by those dealing with appeals at SQA was a matter raised by most interviewees — often with puzzlement. There was a feeling in some cases that SQA was not consistent year-on-year in the judgements it made in this respect, sometimes in relation to the same paper. What were understood to be SQA's views of certain commercially-produced papers was also

a persistent theme in the interviews. This was expressed in some cases as a view that it was well known that SQA favoured or did not favour certain commercial papers, although it did not prove possible to trace the origins of these views.

### **SQA's support on estimating**

The support given by SQA was generally welcomed and seen as good. However, there was some confusion about the nature and availability of some of the documents produced by SQA. The professional development workshops run by SQA were highly rated by those who had participated in an event and by those who knew colleagues who had participated. Where development managers were referred to as contacts, they were either positively or highly spoken of. No practicable suggestions for improvements which SQA could make in the area of support emerged.

### **Staffing and appointees**

Staffing, which it was anticipated might be a significant issue, turned out not to be. In particular, the interviewees offered no support for the idea that either inexperienced teachers or teachers with responsibilities outwith the department had any significant effect on the ability of the department to arrive at accurate estimates. The effects of the McCrone restructuring in schools with the changing role of the Principal Teacher, however, does raise issues about the potential loss of expertise in relation to estimating.

Whilst the presence of appointees in a department was mainly seen as valuable, less formal use was made of appointees than was anticipated and some interviewees saw the question of whether or not colleagues were appointees as largely a private matter — rather than a matter of professional and departmental interest. It might be useful for SQA to make clear its views on the deployment of appointees by local authorities and centres in ways which draw on and utilise the experience they have gained.

## Recommendations arising from the investigation

These findings suggest that SQA might consider the following in its work in supporting the estimating process in centres:

### Dialogue with teaching staff

- ◆ consider how to align SQA's awarding body perspective more clearly with the teaching perspective on estimating as part of its strategy on communication about standards
- ◆ consider publishing examples of approaches taken by centres for discussion or emulation (*These might range from a simple composite model of prelim setting, marking and pass-marking based on SQA's own practices to the approach of the interviewee who had taken the step of consistently estimating higher than the prelim results. It should be noted, however that this study does not give any evidence that this would directly affect the accuracy of estimating.*)
- ◆ seek new ways to increase dialogue with teaching staff on the issues related to estimating — review the use of chat groups or similar web-based communication as one channel (*eg consider whether more needs to be done to explain to teaching staff and subject managers the use to which estimates are put and the importance of estimates in the awarding process*)

### Enhancing centre expertise

- ◆ take any available steps to make it attractive to become an appointee and to make the cadre of appointees as inclusive as possible
- ◆ publicise its views on the deployment of appointees by local authorities and centres in ways which draw on and utilise the experience they have gained
- ◆ investigate the likely effects of McCrone on the management of assessment-related processes in schools with a view to establishing whether there is a need to give tailored support to the new managers in this area of their work
- ◆ investigate whether expertise in assessment to national standards and the setting of estimates could and should be included or strengthened in pre-service and in-service teacher training and professional development programmes for teaching staff and subject managers

### SQA support and support materials

- ◆ continue to review PA reports and associated documents in the light of regular feedback sought from teaching staff and subject managers on their utility
- ◆ concentrate on the use of concrete examples in documentation and on practical training materials designed to address the immediate needs and concerns of teaching staff and subject managers
- ◆ continue to run professional development events on all aspects of standards and how to use them — these should focus on practical issues and processes
- ◆ work with education authorities where possible, to maintain or extend the range of events on standards, marking, estimating, etc for teachers and subject managers

- ◆ consider whether further written guidance is needed on the technicalities of setting and assessing examination papers

### **SQA processes**

- ◆ avoid asking for estimates at earlier dates — or publicise the mechanism which allows centres to revise and re-submit estimates
- ◆ ensure that there is a corporate view on the use of this mechanism and that a consistent message is given about it from all units in SQA
- ◆ review what should constitute appropriate evidence for appeals and how consistently this is set across and within subjects.
- ◆ publicise SQA's position on commercial prelim papers addressing any misapprehensions revealed by this investigation

# Appendix 1: The Interview Protocol

**Date of Visit**

**Research Project for SQA:  
How Do Centres Estimate Pupil's Attainment?**

**School:**

**Subject:**

**Subj Head:**

**Presentations:**     @ Int 2  
                              @ Higher

**No of staff:**

**Appointees:**

**Policy:**

**Practice:**

*Other comment:*

**Sources:** What sources of evidence are used?

past papers  
commercially produced papers  
internally generated papers  
NAB materials  
Evidence from units

What is the difference between unit and course assessment in your subject?

Is that an issue?

How are IAs validated?

How are teachers' judgements reached/ moderated?

How are less experienced/less secure teachers supported in reaching estimates?

Are there any conflicts of role (eg the need to encourage students v realism?)

Are candidates told estimates?      Yes / No

Are teachers subject to pressures?

***Other comment:***

**SQA support:**

*What support from SQA are you aware of in making estimates?*

*Aware of Use?*

*Course Arrangements documents*

*Grade Descriptions at A & C*

*Course Assessment Specifications*

*assessment instruments*

*marking instructions*

*exemplar candidate evidence*

*appeals exemplification materials*

*PA Reports*

*subject update letters*

*PA (visits)*

*SQA Professional development events*

*LA Prof development events / wkshops*

*Inhouse Prof development events*

*QM*

*Other*

*Utility?*

*Grade Descriptors?*

*What else could SQA do to assist/support centres with accuracy of estimating?*

***Other comment:***

**Internal processes:**

*Estimates used for internal processes?*

*When do you make estimates?*

*Do you submit them right away?*

*Do you review them before submitting them?*

*Are you aware that you can alter the estimates you have submitted if you submit them early?*

**Timing:**

*How do SQA's requirements affect your subject policies/procedures?*

*How do institutional / authority requirements affect your subject policies procedures?*

**Making use of estimates and results:**

*Do you (department/faculty/school/the authority) use the estimates as part of any review process?*

*Do you investigate the reasons for any discrepancies in a formal way?*

*Do the results influence your submission of Appeals?*

*Do the results and appeal results influence your approach to estimating?*

**Values & attitudes:** *How do you view the validity and reliability of estimates?*

*(a) in general in relation to appeals*

*(b) in relation to your subject?*

*How would you rate the department at estimating?*

AH

H

Int2

Int1

S Grade

*How would you rate the school at estimating?*

*Are there any factors which make it difficult/impossible to achieve accurate estimates?*

*What does success in estimating say about a teacher/department?*

***Other comments:***

## Appendix 2: The Centres Visited

The figures in bold indicate the number of presentations in 2003.

The letters in italics indicates the correlations between estimates and results.

		English	Maths	Biology	History	French	Info Systems
<b>Centre 1</b>	<i>Int 2</i>		<b>17</b> <i>High</i>		<b>45</b> <i>High</i>	<b>49</b> <i>Ave</i>	
<i>South west</i>	<i>Higher</i>		<b>54</b> <i>High</i>		<b>40</b> <i>High</i>	<b>16</b> <i>Ave</i>	
<b>Centre 2</b>	<i>Int 2</i>			<b>15</b> <i>Ave</i>		<b>202</b> <i>High</i>	<b>134</b> <i>Ave</i>
<i>East</i>	<i>Higher</i>			<b>115</b> <i>High</i>		<b>52</b> <i>V high</i>	<b>71</b> <i>Ave</i>
<b>Centre 3</b>	<i>Int 2</i>		<b>40</b> <i>Ave</i>		<b>9</b> <i>Low</i>		<b>79</b> <i>Ave</i>
<i>North east</i>	<i>Higher</i>		<b>62</b> <i>Ave</i>		<b>55</b> <i>Ave</i>		<b>58</b> <i>Low</i>
<b>Centre 4</b>	<i>Int 2</i>	<b>46</b> <i>Low</i>		<b>25</b> <i>High</i>			<b>23</b> <i>Low</i>
<i>West</i>	<i>Higher</i>	<b>120</b> <i>Low</i>		<b>44</b> <i>High</i>			<b>35</b> <i>Low</i>
<b>Centre 5</b>	<i>Int 2</i>	<b>83</b> <i>Low</i>			<b>16</b> <i>Low</i>	<b>172</b> <i>High</i>	
<i>West</i>	<i>Higher</i>	<b>173</b> <i>Ave</i>			<b>42</b> <i>Low</i>	<b>43</b> <i>High</i>	
<b>Centre 6</b>	<i>Int 2</i>	<b>62</b> <i>Low</i>			<b>25</b> <i>Low</i>		<b>11</b> <i>High</i>
<i>North east</i>	<i>Higher</i>	<b>105</b> <i>Ave</i>			<b>53</b> <i>Low</i>		<b>19</b> <i>High</i>
<b>Centre 7</b>	<i>Int 2</i>		<b>46</b> <i>Ave</i>	<b>17</b> <i>Ave</i>			<b>27</b> <i>High</i>
<i>East</i>	<i>Higher</i>		<b>63</b> <i>Ave</i>	<b>43</b> <i>Ave</i>			<b>19</b> <i>Low</i>
<b>Centre 8</b>	<i>Int 2</i>		<b>50</b> <i>Ave</i>	<b>44</b> <i>High</i>		<b>18</b> <i>High</i>	
<i>East</i>	<i>Higher</i>		<b>138</b> <i>High</i>	<b>110</b> <i>High</i>		<b>37</b> <i>V high</i>	
<b>Centre 9</b>	<i>Int 2</i>	<b>54</b> <i>Ave</i>	<b>40</b> <i>High</i>			<b>67</b> <i>V high</i>	
<i>Central</i>	<i>Higher</i>	<b>112</b> <i>Low</i>	<b>64</b> <i>Ave</i>			<b>39</b> <i>V high</i>	
<b>Centre 10</b>	<i>Int 2</i>	<b>121</b> <i>Low</i>		<b>92</b> <i>High</i>	<b>81</b> <i>Ave</i>		
<i>East</i>	<i>Higher</i>	<b>94</b> <i>Ave</i>		<b>38</b> <i>Ave</i>	<b>46</b> <i>High</i>		
<b>Centre 11</b>	<i>Int 2</i>	<b>53</b> <i>Low</i>	<b>45</b> <i>Low</i>	<b>8</b> <i>High</i>			
<i>West</i>	<i>Higher</i>	<b>138</b> <i>Low</i>	<b>42</b> <i>Low</i>	<b>18</b> <i>Low</i>			
<b>TOTALS</b>		<b>6</b>	<b>6</b>	<b>6</b>	<b>5</b>	<b>5</b>	<b>5</b>