THE USE OF 'LIVE' VERSUS PHOTOCOPIED SCRIPTS IN THE FIRST PHASE SAMPLE OF MARKING STANDARDISATION

Michelle Meadows

ABSTRACT

Prior to beginning marking, all examiners attend a Standardisation Meeting at which they are trained in the application of the mark scheme. Following the meeting, a First Phase Sample script review is conducted. This sample consists of ten scripts (five photocopied scripts and five 'live' scripts selected from the examiners' marking allocation). The scripts are marked by the examiner and forwarded to the Senior Examiner who then re-marks them. The Senior Examiner provides feedback to the examiner, including whether he/she is cleared to continue marking. This paper reports a study to investigate whether, in a number of GCE units, the type of script ('live' or photocopied) affected the size of marking discrepancies between the Senior Examiners and examiners and/or Senior Examiners' decisions to clear examiners to mark.

It was found that there was a greater discrepancy between examiners' and Senior Examiners' marking of photocopied than 'live' scripts. This was possibly because examiners' marking of photocopied scripts was compared with a pre-set mark allocation, while the Senior Examiners' marking of 'live' scripts was influenced by the marks and comments of the original examiner. Senior Examiners' decisions to clear examiners to mark were influenced by discrepancies in examiners' marking of both photocopied and 'live' scripts. Photocopied scripts provide the Senior Examiner with a valuable method of comparing examiners. It is recommended that the effect on standardisation of including 'live' scripts only in the First Phase Sample, as planned from January 2004, should be carefully monitored.

INTRODUCTION

To standardise marking a Standardisation Meeting is held, except in very small units/components in which all scripts are marked by the Principal Examiner. The main aim of the meeting is to ensure that the mark scheme is understood and applied consistently by all examiners, so that the marks awarded to a candidate are not in any way dependent upon which examiner is doing the marking. The meeting is normally held three to five days after the examination and is attended by the Principal Examiner, other Senior Examiners (if appointed) and all Assistant Examiners, including reserve examiners appointed to cover emergencies. The Chair of Examiners or the Chief Examiner may also attend.

Following the Standardisation Meeting a First Phase Sample script review is conducted. The purpose of this process is to achieve a common standard of marking between examiners. The First Sample consists of ten scripts (five photocopied scripts[1] and five 'live' scripts). Photocopied scripts are pre-selected by the Principal Examiner and Subject Officer and are normally

---

[1] There are exceptions to the practice of using photocopied scripts, for example in GCSE Design and Technology: Graphic Products, where the size of the samples and the colour used by candidates make photocopying work impractical. In these cases ten live scripts are used.

distributed at the Standardisation Meeting. Examiners pick the sample of 'live' scripts from their marking allocation. Both the photocopied and 'live' scripts are carefully selected, from a range of centres, to ensure maximum coverage of optional questions, as well as a reasonable cross-section of the mark range. The examiner should also include in, or add to, the sample any 'problem' scripts which raise difficulties in applying the mark scheme, with an accompanying letter indicating the nature of the problem.

The sample of scripts is fully marked by the examiner and forwarded to the Senior Examiner within two days of the Standardisation Meeting. The Senior Examiner then re-marks the scripts. The Procedure Guidance File for Pre-Standardisation and Standardisation Meetings states that the Senior Examiner responsible for reviewing the scripts should:

(i)   be satisfied that each examiner is adhering to all the instructions and interpreting the mark scheme correctly;

(ii)   point out to each examiner any errors being made and give advice on specific points of difficulty and misinterpretation;

(iii)   check for signs of inconsistency and point them out;

(iv)   check the accuracy of the addition of marks;

(v)   check the appropriateness of the comments made.

On completing the re-marking the Senior Examiner completes a First Phase Sample Form, a copy of which is returned to the examiner with the scripts. The Senior Examiner's response may take one of three forms:

(i)   the marking is satisfactory, and the examiner is cleared to mark. This response is also made by telephone;

(ii)   the marking must be reviewed in the light of the comments made, but no additional First Phase Sample is required;

(iii)   the marking must be reviewed in the light of the comments and amendments made. The problems are specified and an additional First Phase sample of ten scripts from the examiner's allocation must be submitted and approved before marking may proceed.

There are number of disadvantages of using photocopied scripts during standardisation:

(i)   the selection and preparation of these scripts takes time and can significantly lengthen pre-standardisation meetings;

(ii)   resources are spent on the duplicate marking of scripts by examiners rather than the marking of 'live' scripts;

(iii)   the cost of photocopying and delivering the scripts to the Standardisation Meeting is significant.

Indeed, the use of five photocopied scripts in the First Phase Sample will be discontinued with effect from the January 2004 examination series. There are, however, advantages to their use:

(i)   Senior Examiners spend less time re-marking photocopied than live scripts;

(ii)   the marking of photocopied scripts can be compared across examiners and they provide a 'control'. They enable the Senior Examiner to discern consistent misinterpretations of the mark scheme. For example, all examiners may mark a photocopied script differently from the Senior Examiner. The Senior Examiner would then question his/her own interpretation of the mark scheme or the explanation of the scheme during the Standardisation Meeting, rather than the examiners' marking. Photocopied scripts are a particularly useful tool where several teams of examiners mark the scripts for a component.

2

Anecdotal evidence from Subject Officers suggests that Senior Examiners are more likely to agree with examiners' marking of 'live' than photocopied scripts. There are two likely explanations for this. First, examiners' marking of 'live' scripts is not compared to a pre-set mark allocation, as it is with photocopied scripts. There is less room for flexibility in Senior Examiners' appraisal of the examiners' marking of photocopied scripts. Second, it is likely that Senior Examiners' marking of 'live' scripts will be influenced by the marks and comments of the original examiner. Murphy (1979) compared the reliability of examiners' marking when scripts had the previous examiners' marks and comments on them or had had them removed. Removing the marks and comments made a considerable difference to the extent to which the marks agreed. The differences between the marks were approximately twice as great when the original marks were removed.

Subject Officers also suggested that Senior Examiners are aware that photocopied scripts are associated with greater mark discrepancies and so are less concerned by these discrepancies, making the photocopied scripts less useful than 'live' scripts.

This paper reports a study to investigate whether, in a number of GCE units, the type of script ('live' or photocopied) affected the size of marking discrepancies between the Senior Examiners and examiners, and/or the Senior Examiners' decisions to clear examiners to mark.

METHOD

Principal Subject Managers, in consultation with subject staff, were asked to suggest GCE subjects in which there appeared to be differences in the effectiveness of photocopied and 'live' scripts in standardisation. Their beliefs about the relative usefulness of 'live' and photocopied scripts in standardisation and ideas as to how the research should be taken forward were also collected.

Photocopied scripts were reported to cause standardisation problems in units in Religious Studies, Media Studies, English Literature and Geography. To assess fairly the relative effectiveness of the use of 'live' and photocopied scripts, units in which photocopied scripts had *not* been considered problematic were also included in the research. Subject Staff comments suggested that the use of photocopied scripts in standardisation was more problematic in 'Arts' than 'Science' specifications. It was also suggested that the extent of any problems might vary across offices. Photocopied scripts have been used in standardisation in the Manchester office for many years, but the practice is relatively new to staff in the Guildford office. Hence a range of subjects, administered across different offices, was studied (see Table 1 for the units included in the research). The number of Senior and Assistant Examiners marking scripts from each unit is also included in Table 1. Information from the First Phase Sample forms of these units for the June 2002 examination series was collated and analysed.

RESULTS AND DISCUSSION

First Phase Sample forms for some units were unavailable (ESC4, MED2, MED3 and MED4). A First Phase Sample had been considered unnecessary for RS05 because only two examiners marked scripts from this unit.

To facilitate comparison across units, the difference between the marks allocated to the scripts by the Senior Examiner and examiner was converted to a percentage of the maximum mark

possible which varied from 20 to 100 across the units (see Table 1). A discrepancy of one mark resulted in a five per cent difference in LTA1 (max. mark = 20) but only a one per cent difference in PHB5. Across-unit comparisons of mark discrepancies should therefore be treated with caution.

Table 1. Number of Examiners Marking Scripts and Maximum Mark by Administering Office, Subject Type, Subject and Unit

		Office*	Specification	Unit**	No. of Senior Examiners	No. of Examiners	Maximum Mark
Subject	'Science'	M	Biology A	BYA1	9	49	75
				BYA2	8	40	75
				BYA3	4	16	75
				BYA5	7	29	75
		G	Environmental Science	ESC1	2	7	70
				ESC2	2	5	70
				ESC3	1	7	70
				ESC4	-	-	-
		G	Maths A	MAME	3	15	60
				MAP1	5	24	60
				MAP2	2	8	60
				MAP3	3	14	60
		M	Physics B	PHB1	2	8	75
				PHB2	2	11	75
				PHB4	2	8	75
				PHB5	2	9	100
	'Arts'	M	History	HS03	9	62	40
				HS1B	3	5	50
				HS2N	3	8	50
				HS4E	2	6	50
		G	English Literature A	LA2W	6	24	20
				LTA1	14	78	20
				LTA3	8	64	40
				LTA4	9	40	40
		G	Media Studies	MED1	6	40	60
				MED2	-	-	-
				MED4	-	-	-
				MED6	-	-	-
		M	Religious Studies	RS01	1	6	80
				RS04	1	5	80
				RS05	-	-	-
				RS06	1	4	80

Key: * M = Manchester, G = Guildford; ** data for shaded units were unavailable

Statistics presented in Table 2 show the average mark difference by unit and subject. The mark allocated by the examiner was subtracted from that allocated by the Senior Examiner. Hence, a negative difference in marks indicated that the examiner chose a more generous mark than the Senior Examiner. A positive difference indicated that the examiner chose a less generous mark than the Senior Examiner. The type of script being marked (photocopied or 'live') did not seem to be associated with more or less generous marking on the part of the examiner compared to the Senior Examiner.

For further analyses the direction of the difference in marks was ignored to avoid differences across scripts cancelling each other out, for example an examiner might be five marks more lenient than the Senior Examiner on one script, but five marks more severe on another script. This examiner would score a mark difference of ten, not zero. Statistics presented in Table 3 show the *absolute* average difference by unit and subject.

Across most of the units studied there was a greater difference between the marks allocated by the Senior Examiner and examiner when photocopied, rather than 'live', scripts were being considered (see Table 3). Paired t-tests demonstrated that the difference was statistically significant in Biology, Maths, History, Media Studies and Religious Studies, and in both the Art and Science subjects taken overall. The photocopied scripts were also associated with a greater mark discrepancy in Environmental Science and Physics, but not to a statistically significant extent.

The data collected have a hierarchical, nested structure. Scripts were grouped by examiner; examiners were grouped by Senior Examiner; Senior Examiners by unit. This structure is the same for each unit. To ignore the hierarchical nature of the data risks overlooking the importance of group effects, and may also render invalid the statistical techniques used to study the effects of 'live' versus photocopied scripts. A three level model was constructed; level one consisting of the ten scripts included in the first phase sample, level two being the examiners and level three being the Senior Examiners. The absolute difference in the marks allocated by the examiner and the Senior Examiner was modelled, dependent on whether the script was photocopied or 'live' (see Table 4). The difference was significantly larger when photocopied, rather than 'live' scripts were considered. The effect of 'live' versus photocopied scripts on the absolute mark difference varied significantly across Senior Examiners, examiners and the scripts (see Appendix 1). This may reflect differences in the way in which sample scripts are chosen. Some examiners may emphasise 'problem' scripts in their sample, others may focus on ensuring maximum coverage of optional questions and so on. Indeed the method of choosing scripts may be wholly appropriate to the unit they are marking.

Table 2. Mean and Standard Deviation of Mark Difference by Unit and Subject

'Science' Specifications	Unit	'Live' scripts		Photocopied scripts	
		Mean	S.D.	Mean	S.D.
Biology A	BYA1	-.33	6.75	2.08	7.52
	BYA2	.13	8.08	-.34	8.57
	BYA3	-1.83	13.47	3.67	12.26
	BYA5	1.48	8.59	7.63	11.51
	Overall	**.01**	**8.51**	**2.78**	**9.77**
Environmental Science	ESC1	5.95	6.22	1.71	6.42
	ESC2	-4.29	7.35	-2.14	12.26
	ESC3	1.43	10.07	7.14	13.70
	Overall	**1.35**	**8.76**	**3.13**	**11.48**
Maths A	MAME	-2.00	4.33	.56	8.35
	MAP1	.69	6.59	2.85	4.46
	MAP2	-1.46	12.80	10.63	7.92
	MAP3	1.41	7.69	-3.69	7.85
	Overall	**-.11**	**7.40**	**1.80**	**7.92**
Physics B	PHB1	.00	4.16	-2.00	4.02
	PHB2	-3.39	15.87	-.61	10.21
	PHB4	-4.50	8.52	-2.83	7.35
	PHB5	2.78	5.17	-1.67	9.29
	Overall	**-1.34**	**10.22**	**-1.66**	**8.20**
'Science' Overall		**-.12**	**8.52**	**1.94**	**9.32**
'Arts' Specifications	Unit	'Live' scripts		Photocopied scripts	
		Mean	S.D.	Mean	S.D.
History	HS03	2.29	25.67	-1.17	27.04
	HS1B	.40	6.69	-7.00	18.94
	HS2N	4.50	10.46	-2.25	19.43
	HS4E	-16.33	19.49	-2.67	23.45
	Overall	**.99**	**23.73**	**-1.69**	**25.46**
English Literature A	LA2W	6.88	30.89	-19.32	37.87
	LTA1	12.71	47.06	.67	13.47
	LTA3	4.15	16.55	.94	11.95
	LTA4	7.69	25.09	8.99	16.92
	Overall	**8.44**	**34.07**	**-.03**	**20.21**
Media Studies	**MED1**	**-2.35**	**27.93**	**-5.48**	**36.03**
Religious Studies	RS01	-1.67	7.53	-9.38	19.44
	RS04	-.63	6.17	5.00	28.84
	RS06	-22.50	23.14	-15.94	35.60
	Overall	**-7.32**	**15.92**	**-6.33**	**26.84**
'Arts' Overall		**4.61**	**30.75**	**-1.44**	**24.34**

Table 3. Mean and Standard Deviation of Absolute Mark Difference by Unit and Subject

'Science' Specifications	Unit	'Live' scripts		Photocopied scripts		Mean absolute mark difference	N
		Mean	S.D.	Mean	S.D.		
Biology A	BYA1	7.56	4.81	9.14	5.21	-1.58*	48
	BYA2	8.10	5.85	10.46	5.54	-2.36*	39
	BYA3	11.83	9.31	17.17	8.69	-5.34*	16
	BYA5	9.29	5.13	14.43	6.94	-5.14**	28
	Overall	**8.61**	**5.98**	**11.64**	**6.76**	**-3.03***	**131**
Environmental Science	ESC1	11.14	6.65	8.57	3.50	2.57	5
	ESC2	8.59	4.52	11.43	8.16	-2.84	4
	ESC3	7.96	7.18	13.27	6.89	-5.13*	7
	Overall	**9.11**	**6.21**	**11.34**	**6.31**	**-2.23**	**16**
Maths A	MAME	5.56	4.11	12.11	4.29	-6.55***	15
	MAP1	7.78	6.15	7.15	3.88	0.63	24
	MAP2	8.96	9.76	13.54	8.52	-4.58	8
	MAP3	8.08	6.00	10.77	8.68	-2.69	13
	Overall	**7.44**	**6.22**	**10.03**	**6.36**	**-2.59****	**60**
Physics B	PHB1	7.78	3.09	7.33	2.76	0.45	6
	PHB2	12.85	10.49	12.48	6.15	0.37	11
	PHB4	8.83	5.83	11.17	4.15	-2.34	8
	PHB5	7.89	2.62	11.22	5.89	-3.33	9
	Overall	9.70	6.98	10.93	5.28	-1.23	34
'Science' Overall		**8.51**	**6.21**	**11.12**	**6.44**	**-2.61***	**241**
'Arts' Specifications	Unit	'Live' scripts		Photocopied scripts		Mean absolute mark difference	N
		Mean	S.D.	Mean	S.D.		
History	HS03	23.95	18.14	42.54	26.19	-18.59***	62
	HS1B	9.00	5.03	40.00	14.70	-31.00**	4
	HS2N	12.00	5.66	26.75	13.94	-14.75*	8
	HS4E	23.67	13.35	27.33	5.47	-3.66	6
	Overall	**21.99**	**17.06**	**39.69**	**24.33**	**-17.70***	**80**
English Literature A	LA2W	30.68	19.96	46.59	38.77	-15.91*	22
	LTA1	20.93	45.87	14.50	16.90	6.43	70
	LTA3	13.86	11.32	14.79	11.84	-0.93	35
	LTA4	23.40	18.96	24.86	18.41	-1.46	36
	Overall	**21.27**	**32.82**	**21.18**	**23.19**	**0.09**	**163**
Media Studies	**MED1**	**28.87**	**19.53**	**49.73**	**23.57**	**-20.86***	**37**
Religious Studies	RS01	9.58	3.51	25.21	11.71	-15.63**	6
	RS04	18.13	8.07	33.75	7.43	-15.62	4
	RS06	30.00	18.31	46.56	12.84	-16.56*	4
	Overall	**17.86**	**13.20**	**33.75**	**13.70**	**-15.89***	**14**
'Arts' Overall		**22.26**	**27.12**	**30.41**	**25.54**	**-8.15***	**294**

Key: * p<.05, ** p<.01, *** p<.001, one-tailed

Table 4. Parameter Estimates for the Multilevel Model, Modelling the Absolute Difference in Marks Allocated, Dependent on Whether the Script was Photocopied or 'Live'

Fixed Effects	β	SE	p	Odds
Constant	3.194	0.286	<.001	24.386
Type of script	1.114	0.132	<.001	3.047
Random Effects	Var. Comp.	SE	p	
Level-three (Senior Examiners)	8.087	1.195	<.001	
Level-two (Examiners)	2.179	0.314	<.001	
Deviance	35357.720			

The feedback given to examiners is determined by the extent of the discrepancy between the marks allocated by the Senior Examiner and examiner. See Table 5 for a breakdown of the feedback given to examiners in this sample.

Table 5. Frequency of Types of Feedback Given to Examiners in the First Phase Sample

Feedback	Percentage of Examiners[2]
Examiner cleared to mark	43.2 (n=219)
Examiner to review marking in light of comments	40.6 (n=206)
Additional First Phase Sample required	16.2 (n=82)

It had been suggested that Senior Examiners are aware of the greater mark discrepancy associated with photocopied scripts and tend to take less account of these discrepancies when considering the feedback they give to examiners. To test this hypothesis a direct logistic regression analyses was performed with the feedback given to examiners as the dependent variable and mark discrepancy on photocopied and on 'live' scripts as the predictors.

A test of the full model with both predictors against a constant-only model was statistically significant, χ^2 (4, N=507) =218.88, p<.001, indicating that the predictors, as a set, reliably distinguished between the levels of feedback given to examiners. The variance in feedback accounted for is moderate, however, with McFadden's rho = .211. Both mark discrepancies on photocopied (Wald=31.99, p<.001) and 'live' (Wald=34.55, p<.001) scripts independently predicted examiner feedback, although the Wald statistic for the 'live' scripts is somewhat larger than that for the photocopied scripts. This suggests that Senior Examiners take mark discrepancies from *both* the 'live' and photocopied scripts into account when deciding the type of feedback given to the examiners, but a slight emphasis is placed on examiners' accuracy in marking 'live' scripts.

---

[2] Feedback was missing from the First Phase Sample forms of 34 of the examiners.

CONCLUSION

There was a greater discrepancy between the examiners' and Senior Examiners' marking of photocopied than 'live' scripts. This was possibly because examiners' marking of photocopied scripts was compared with a pre-set mark allocation, while the Senior Examiners' marking of 'live' scripts was influenced by the marks and comments of the original examiner. The Senior Examiner may have taken advantage of the extra information available when trying to judge the 'best' mark for the candidate (Massey and Foulkes, 1994). There was no evidence, however, to suggest that Senior Examiners were unconcerned by discrepancies in examiners' marking of photocopied scripts, making them less useful than 'live' scripts. Discrepancies in the marking of both 'live' and photocopied scripts predicted the type of feedback given to the examiner.

Despite the drawbacks associated with the inclusion of photocopied scripts in the First Phase Sample, it provides the Senior Examiner with a valuable method of comparing examiners. Senior Examiners may find the marking standardisation process undermined by the inclusion of 'live' scripts only in the First Phase Sample as planned from January 2004. It is recommended that the effect on standardisation of including 'live' scripts only in the First Phase Sample, as planned from January 2004, should be carefully monitored.

REFERENCES

Massey, A. and Foulkes, J. (1994) Audit of the 1993 KS3 science national test pilot and the concept of quasi-reconciliation. *Evaluation and Research in Education*, v8, p119-132.

Murphy, R.J.L. (1979) Removing the marks from examination scripts before re-marking them: Does it make any difference? *The British Journal of Educational Psychology*, v49, p73-78.

APPENDIX 1

OUTCOME OF MULTILEVEL MODELLING

Levels: 1. Script; 2. Examiner; 3. Senior Examiner.

$$\text{absdiff}_{ijk} \sim N(XB, \ \Omega)$$

$$\text{absdiff}_{ijk} = \beta_{0ijk}\text{cons} + 1.114(0.132)\text{photo}_{ijk}$$

$$\beta_{0ijk} = 3.194(0.286) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \ \Omega_v) \ : \ \Omega_v = \begin{bmatrix} 8.087(1.195) \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \ \Omega_u) \ : \ \Omega_u = \begin{bmatrix} 2.179(0.314) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \ \Omega_e) \ : \ \Omega_e = \begin{bmatrix} 24.875(0.490) \end{bmatrix}$$

$$-2*loglikelihood(IGLS \ Deviance) = 35357.720(5740 \ of \ 5740 \ cases \ in \ use)$$

The use of 'live' versus photocopied scripts...   Michelle Meadows