

Remarking of Examination Answer Scripts – Finding a Standard for Quality Assurance

By

Joe Cesare, Gauteng Department of Education, cesare@netactive.co.za

Prof. Coert Loock, University of Johannesburg, coertl@uj.ac.za

Prem Govender, Gauteng Department of Education, preg@gpg.gov.za

ABSTRACT

The value of remarking of examination answer scripts is often limited to knowing how many changed symbols and *ad hoc* identification of the most glaring errors made by the initial markers.

In any marking action performed by people there will always be the occasion where the marker has to make a judgment call on whether or not to award a mark for a specific answer. To be fair towards the candidates the markers often use a system of marking per question to minimize the effect of this subjectivity and attain better overall consistency. It is a generally accepted educational principle that there may be small variations in marking by different markers, even if they use the same memorandum. However, the question still remains of when is the difference acceptable and when not.

With this paper the authors examine the changes in marks during remarking of a large number of subjects in a large-scale examination such as the Senior Certificate Examination in Gauteng over a number of years. The analysis is done in terms of changes in two-percent intervals on raw marks to determine a standard against which individual subjects can be measured. Raw marks are specifically used as to eliminate the effects of standardization. The analysis is spread over a number of years to determine patterns and eliminate once-off individual problems.

The effect of deliberate interventions introduced by the province to improve the quality of marking is shown in terms of changes in remarking patterns. The remarking of the subject Physical Science of one cycle is analyzed in detail to show the nature and frequency of common marking errors made during initial marking, with reference to the relationship between the nature of the question (multiple choice, calculations, etc.) and the type of error.

INTRODUCTION

This particular project initially started off as a search for a quality assurance tool to use during the marking process of the Senior Certificate Examination. The initial question posed was “how do we improve on the quality of marking?” Because remarking shows the “mistakes” made during the initial marking, it seemed a logical starting point.

It is interesting to note that there is very little information on the result of a very widely used concept that is within the general public domain. It is not as if remark results are classified as restricted, but rather as something one does not really want to highlight as it is indicative that the marking process may in fact not be as accurate as quality assurance bodies would like the general public to perceive. It is a characteristic of high volume - high stakes examinations that the public, the media and the politicians show a great interest in the examination, not because of the intrinsic value, but because of the potential value to promote specific agendas.

Misinterpretation of the remarking process could very easily erode the confidence of the public in an examination system. It was also quite an experience to observe the reactions of examiners and markers when we started discussing our analysis of remarking with them.

THE CONCEPT OF REMARKING

The concept of allowing remarking an assessment task is widely applied, not only at the levels of academic institutions but also in industries where formal examinations are conducted. Where remarking is a formal process or part of a formal process, there is a policy or guidelines that governs the process. These policies usually include reference to “where the student is not satisfied with the result obtained” or words to that effect.

While the particulars of these policies are unique for every organisation or institution, the requirements to qualify can be grouped into main categories:

- There is some monetary fee attached. The fees are normally affordable and related to administrative costs and markers remuneration involved, but there are instances where there are marked differences in fee structure applied by different assessment bodies for identical examinations, e.g. for remarking Senior Certificate Gauteng Department of Education charges R70 per subject and the IEB charges R 400 per subject (GDE, 2006; IEB, 2006)
- In the majority of organisations, the fee is refundable if the remark results in a grade/symbol change (DOE, 2005; GDE, 2006). There are however institutions that do not make a refund, particularly web-based programmes (Purpletrain, 2006).
- The application should usually be accompanied by a motivation for the request (City University London, 2006) or the request be supported by a governing body (University of Papua New Guinea, 2006). More often than not there is a qualifying criterion attached, e.g. having a certain minimum score, or be within a certain range for a pass or distinction (Unisa, 2006), result in a qualitative change of the student’s academic status (University of Addis Ababa, 2006) or be applicable to failed candidates (Open University of Hong Kong, 2006).

- Remarking applies only to written final examinations and not the practical or course work (Addis Ababa University, 2006; GDE, 2006; Technikon North Gauteng, 2006).
- Remarking is often one part of an appeals process, such as found with GSCE and A-levels (Teachernet, 2006)
- There is a time limit attached to an application being made.
- A different marker, usually a more senior marking official, will do the remarking.
- The candidate usually receives the higher of the normal and remarking scores (DOE, 2005, IEB 2006). There are a few institutions where the candidate will receive the remark score as final, regardless of the previous score (Technikon North Gauteng, 2006, University of New England, 2006, Addis Ababa University, 2006).
- Remarking is often a component of a results enquiry process, where the enquiry can trigger re-checking of grades, re-moderation of course work or remarking of written papers. This is the process applied for GCSE and A-levels and seems to be a widely followed model (Teachernet, 2006). This process stands in contrast to the procedure followed in South Africa, where rechecking and remarking are two distinct processes, with a viewing process and an appeal process two consequential processes that may follow a remark (DOE, 2005)
- The majority of assessment bodies do make a distinction between re-checking or verification (essentially a clerical checking of addition and computation) and remarking by a different, usually more senior, marker.

THE RECHECKING PROCESS

The majority of assessment authorities allows for a re-checking process that is separated from the remarking process. While re-checking concentrates on clerical to ascertain if all work marked and marks added and computed correctly, the same actions are automatically performed during the remarking process. Very often a score will change after remarking, not because of a marking error but due to a clerical error that could have been picked up during a re-check. In the South African scenario the candidate has to choose either one of the two options. As the cost and effort required from the candidate is not significantly different ((R 12 for a recheck and R 70 for a remark (GDE, 2006)), many candidates opt for a remark. This tendency has major implications for the examining body in terms of time, logistics and infrastructure.

Our research shows that despite deliberate mechanisms to prevent transcription and arithmetical errors, they still occur and are responsible for a percentage of mark changes with remarking.

It seems as if there are two main “culprits” that manage to bypass to control measures such as checking adding and transfer of marks by another person than the marker. The first of these is mental fatigue. Senior certificate marking is always a high volume of work that must be completed in a short space of time under extreme pressure deadlines – normal “human error” will definitely start to exact its toll. In a Physical Science script the number of digits a marker has to read, mentally interpret correctly, check on correct placements in formulas, check on manipulation and calculation and then allocate the correct number of marks, add up 36 subsections

into nine question totals and transfer these correctly to a script cover and add the total, is a mind boggling number of mental computations. If the marker then has to transcribe 106 marks to candidate 1606060160 and 160 marks to the next candidate number 1606060166, the chance of making a mistake becomes definite.

The second culprit, not always recognised, is the role of language when using numbers. It is a natural tendency to think in one's mother tongue when doing simple mental arithmetic. In Afrikaans the number 69 is pronounced as nege(9)-en-ses(6)tig and in English as six(6)ty-nine(9). Writing 69 as 96 thus becomes a common error (then you can still add the problem of 6 looking like 0 when the hand gets tired!)

THE REMARKING PROCESS

One of the outstanding aspects of remarking is probably the fact that the examiner or a more senior marker is charged with the remarking of a script. This aspect could be considered to be a double-edged sword. On the one hand the remarker will be able to pick up on deviations/alternatives much easier and consequently evaluate the candidate's response more accurately – the candidate may thus gain or lose due to the marker's ability.

The influence of the markers' experience and ability varies from subject to subject – the effect is more pronounced in subjects where there has to be interpretation of the candidate's response, as found in essay type questions and questions where analysis and synthesis are required. Where the response is purely stating of facts or simple calculations the marker's ability plays a much smaller role.

The advantage seems to always be with the candidate. With few exceptions qualification authorities give the candidate the benefit of obtaining a higher mark with remarking and ignore the score if the remark results in a lower score.

When results are standardized the initial "incorrect" scores are used to determine adjustments. The remark results are subject to the same adjustment process as the original scores. While the numbers involved in remarking are such that it may be insignificant, it would be interesting to see if there would have been the same adjustments had the remarked scores been used.

In order to cancel out the effect of any standardization adjustments, it was decided that for the purpose of this research the candidates' raw marks would be used.

In subjects such as Physical Science and Mathematics, there are often more than one method that can be applied to get the correct answer, particularly in questions that require analysis and synthesis. Examiners usually try to cover most of the alternatives in the memorandum. Other alternatives are readily recognised and credited by markers. Because mark allocation is usually not an all or nothing situation, it does become a problem when an alternative is partially correct – two different markers may give different marks and both be able to justify why they awarded the particular mark. The same concept can be applied in most subjects.

Marking done by educators is not always the objective exercise that educators promote it to be. This is particularly true in high profile - high stakes examinations

across the world. The very fact that the candidates performance are often used to measure the success of the educator of the institution, gives the marker a vested interest (even if he is not marking his own candidates) and they will tend to err to the advantage of the candidate – this is particularly true in remarking borderline cases. Markers will tend to search for possible additional marks and not necessarily mark with the same strictness as during the initial marking.

It is an unfortunate fact that exam results are used to measure – what you get from your measurement depends on your intended outcome. If the marks are very high, it is often said that the examiners set the paper too easy, or the markers were too lenient, or the standard of examinations is too low (and the opposite is equally true). If a remarking action results in general increases or decreases, it can only add fuel to the fire, as was seen in the UK when there was an overhaul of the grading system (Demopoulos, 2005, Tomlinson, 2002). It is for this reason that in an ideal world there shouldn't be significant changes with remarking – which relates directly to the topic finding a standard for quality assurance.

THE VIEWING AND APPEAL PROCESS IN RELATION TO REMARKING

It is interesting to note that there are two schools of thought on allowing a candidate access to information (in this case the actual marked script). There are instances that allow access to original scripts under supervision on request (UP, 2006; DEFS, 2006; DOE, 2005) or allow photocopies to be accessed (DEFS, 2006), while in most this is not the case.

While access to the script would probably have a significant effect on the reduction of the number of remark requests, as shown by the high percentage of scripts that have no change in marks after remarking, one of the preventative issues in allowing open access to candidates would be the logistics involved in maintaining the integrity and security of the scripts in large scale examinations.

WHY USE REMARKING FOR QUALITY ASSURANCE?

There are basically three reasons why candidates apply for remarking:

- § Candidates on the verge of a higher symbol hope for a mark or two to get into the next range.
- § Candidates require higher symbols and/or scores to qualify for admission to tertiary institutions or specific faculties and for bursaries and financial assistance.
- § Candidates, their parents and educators do not trust the quality and consistency of the marking or the abilities of the markers and hence apply for a remark.

The whole concept of remarking is based on providing an opportunity to a candidate to have a script remarked and rechecked – unfortunately it has lately turned into a second chance for candidates that need higher symbols and higher scores for bursaries, etc.

In any marking action performed by people there will always be the occasion where the marker has to make a judgment call on whether to award or not to award a mark for a specific answer. The effect of this subjectivity is largely cancelled by using a system of marking per question so that no individual is solely responsible for the final

result. This gives better overall consistency and is considered to fair towards the candidates.

One of the aims and objectives of the Marking is to ensure that our clients perceive our marking process as accurate, fair and consistent. With the high profile of the Senior Certificate Examination, such perceptions have a spill over to education in general. It is a generally accepted educational principle that there may be small variations in marking by different markers, even if they use the same memorandum. Our aim is to **reduce the margin of error to acceptable levels**. The remarking process can be seen as a test for our quality assurance mechanisms used in marking.

In the past our analysis of Remark/Recheck results only indicated the number of remark applications that result in a symbol change and as this could be indicative of either very small changes or of major changes in the raw mark. In terms of our processes, practices and operations this did not add any value, as we could not determine where we are going wrong. This necessitated that our approach to remark analysis had to change. The statistics provided by the remark markers were also not always that accurate and could not be used to address shortcomings in the marking process.

ANALYSIS OF THE GAUTENG SENIOR CERTIFICATE REMARKING RESULTS

The challenge posed to the Marking Unit was to assure the quality of marking. Quality is an aspect that is difficult to measure in education, as we do not have a rigid and objective baseline from which to evaluate. As the baseline should not be influenced by activities outside the marking process, our departure point was to analyse remark results in terms of raw marks obtained, as to eliminate the influence of standardization on the final symbol awarded. The very nature of the marking task, coupled to the approximately 1.2 million scripts marked by 6,500 people, made using observation against checklists (as it is used during monitoring) an exercise that eases the conscience rather than add value. It is just too easy to 'hide' bad practice from a monitor that is not necessarily a subject expert.

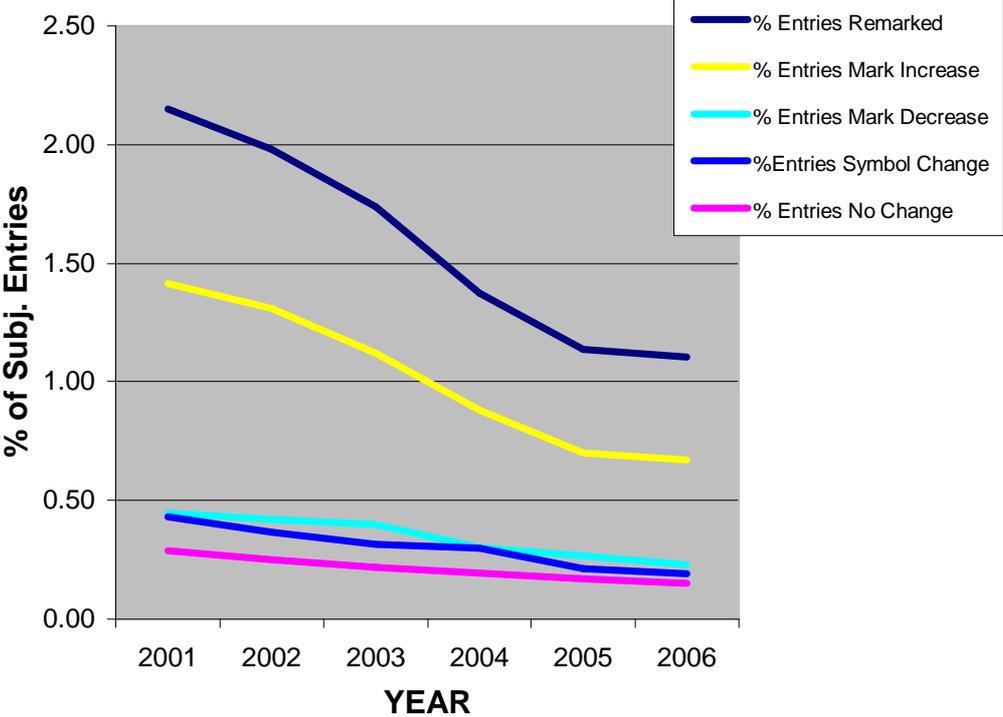
Using remarking as a vehicle to drive a quality assurance process seemed to be an option. We were thrilled when our initiative started showing positive results.

Table 1: Trends in Remarking

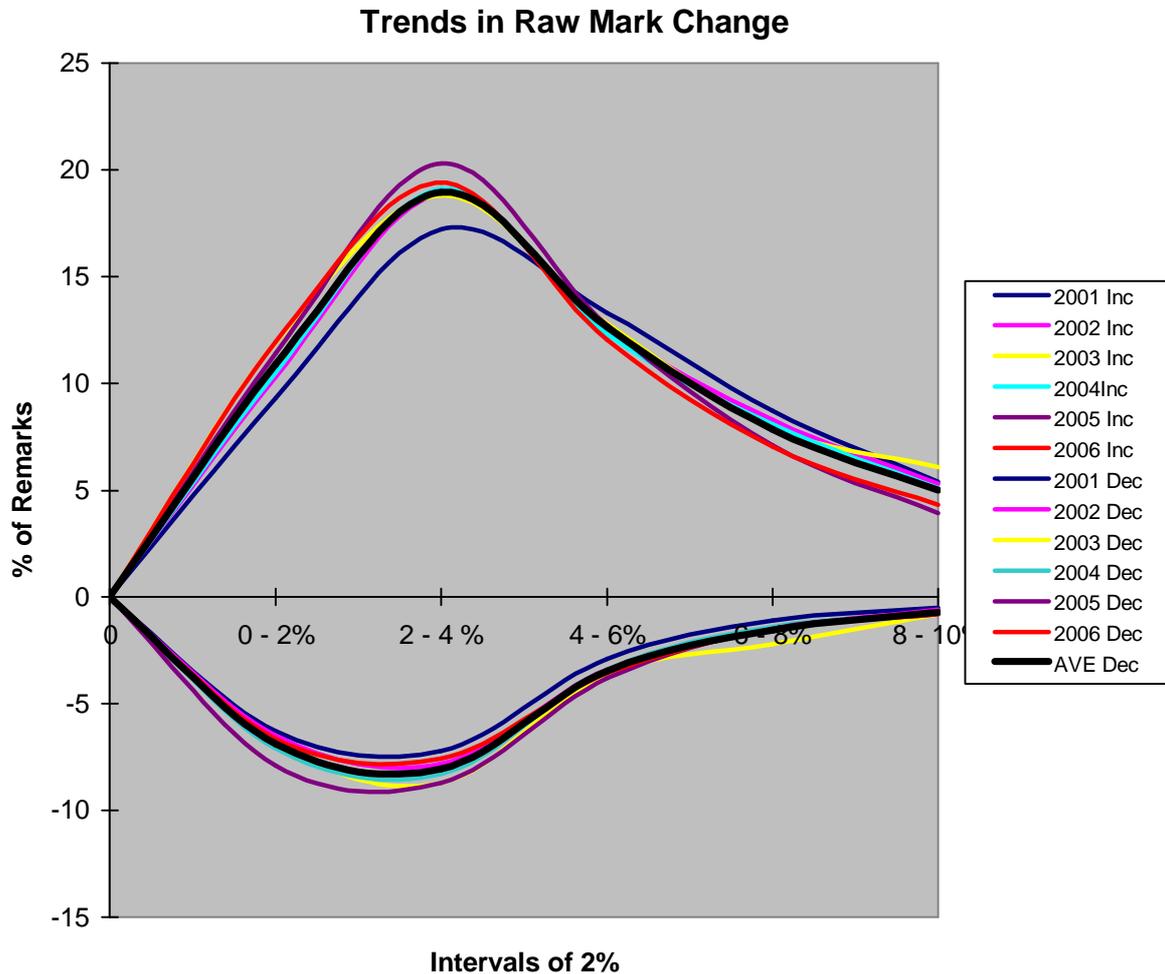
Year	Subj. Entries	Remark Applications	% Entries Remarked	%Entries No Change	%Entries Increase	%Entries Decrease	%Entries Symbol Change
2001	619488	13319	2.15	0.29	1.42	0.45	0.43
2002	621122	12301	1.98	0.25	1.31	0.42	0.37
2003	637904	11082	1.74	0.22	1.12	0.40	0.31
2004	664455	9136	1.37	0.19	0.88	0.30	0.30
2005	685047	8004	1.17	0.17	0.70	0.27	0.21
2006	697317	7694	1.10	0.15	0.67	0.23	0.19

For the investigation all subjects, levels and grades were combined. As entries increased significantly over the five years, we used percentage of entries as a criterion. The question is now: At what value would the percentage be acceptable and can we say that the quality of marking was at an acceptable level? That magical figure will indicate that marking was up to standard – we just have to determine the figure and agree on it. One of the aspects that this research shows is to perhaps extend it further to include a much bigger database, such as at a national level.

Graph 1: Trends in Remarking



In order to see where remarking had a significant effect, the remark analysis was extended to measure changes with remarking in 2% intervals, based on the raw marks. As can be seen in the graph, the tendency is towards a normal distribution around the 2 – 4 % interval. This brings us closer to an ideal situation that could become our standard.



Changing trends in remarking does not happen automatically. There has to be deliberate intervention actions and strategies. A very valuable contribution to the overall success of the marking process was making the detailed remark results available to Internal Moderators and Chief Markers. Not only could they determine how they actually fared, but could see in which intervals the problems were. During remarking sessions Chief Markers are now noting areas with major errors, not only those made by markers, but also sections of the work or memorandum that lends itself to inconsistent marking. Markers can now also be alerted to bad practices during the training session and put corrective measures in place to counter these. At the memo discussion the examiners can now not only engage in discussions about agreeing on the answers to accept, but also contribute towards meaningful to the marking process. Because we included analysis per subject paper and per subject, an added bonus from the year 2004 onward is that we are now able to pinpoint discrepancies to a specific paper.

It was found to be a useful tool for target setting to motivate markers. Subjects that set themselves formal targets after receiving the first analysis have made huge improvements. It was also used to recognize good performance and practices with subjects that did well.

ANALYSIS OF REMARKING RESULTS OF PHYSICAL SCIENCE

The analysis was done for the 2004 (889 candidates) and 2005 (798 candidates) Physical Science Higher Grade answer scripts of the Senior Certificate Examination. The average difference of the original and remark scores for the individual questions were determined. All scripts where there was a difference of two or more marks in any one question, or a difference of four or more on the total, were looked at individually to determine the nature of origin of the difference. The variety of different scenarios was just too diverse to use in pattern determination – mainly due to the unique nature of partially correct individual answers.

There were however some aspects that could be identified as ‘culprits’.

- Errors during carry-through of marks of incorrect/partially correct answers. This ‘error’ seems to be equally advantageous and disadvantageous to the candidates.
- Remark markers do not see a “fresh” script and there is a tendency to repeat the error of the initial marker. A significant number of cases were seen where the remark marker was “influenced” by the original marker to allow partially correct alternatives not provided for in the memo. This tendency was also observed in work that was moderated. The question may well be asked: Is the remark marker ‘moderating’ the original marking, or looking for places to add or deduct marks, or marking the script afresh, strictly according to the memorandum? The nature of the marker as educator will probably steer the marker towards advantaging the candidate.
- A complex marking memorandum makes marking difficult – too many alternative methods, for which different partial marks are given, is in an attempt to have a comprehensive memo and this may well confuse the marker – or is it perhaps that the subject knowledge of markers is so poor that we need to give step by step instructions for everything?
- Surprisingly there were errors in marking Multiple Choice Questions – these were however related to candidates not following instructions, e.g. two crosses and one not cancelled clearly, cancelling a question and changing the numbers of subsequent questions on the answer sheet, etc.
- A major factor, usually picked up on during rechecking, is that candidates answer questions in bits and pieces. While it is generally picked up during marking, carry-through of answers and also contradictions and double dipping often get missed during initial marking. This aspect seems to contribute significantly to changes in marks.
- Deteriorating handwriting and letters and digits that have to be interpreted rather than read is a major culprit and is equally applicable to markers and candidates.

Questions, where the average difference between original score and remarked score was in the order of ≥ 0.1 , were studied in detail. It is interesting to note that these relate to specific sections.

Table 2: Average difference between original and remarked scores in sections.

Physical Science HG Paper 1 - Physics			
Section	2004	2005	2006
1. MCQ's	0.02	0.01	0.03
2. Vectors	0.16	0.13	0.02
3. Forces	0.28	0.21	0.06
4. Motion	0.07	- 0.01	0.13
5. Laws of Motion	0.07	- 0.14	0.14
6. Energy	0.09	0.09	-0.13
7. Momentum	0.07	0.12	0.03
8. Electrostatics	0.01	0.24	0.02
9. Current and Resistance	0.03	0.03	0.07
Total	0.81	0.68	0.38

Sections 2, 3, 5: These sections had the common scenario that there are various alternatives that can be used, requiring a very complex memorandum, containing carry-through of incorrect/partially correct answers. It generally required higher order skills from candidate.

Section 7: Trust a candidate to really mess up something when a concept has to be expressed in words or symbols. Language skills of both candidate and marker are tested and both groups get it wrong.

Section 8: Once again candidates cannot express concepts in words. Carry through into more than one alternative method caused problems. An interesting note on the memo stated: "weight accepted this year, but not in future" – examiners themselves are not consistent in requirements and a marker may remember a comment such as this two years from now and mark it wrong even if it is acceptable in the particular memorandum of a different section.

Table 3: Average difference between original and remarked scores in sections.

Physical Science HG Pr. 2 - Chemistry			
Section	2004	2005	2006
1. MCQ's	0.01	0.02	0.03
2. Intermolecular Forces and Gasses	- 0.30	0.24	0.15
3. Inorganic Chemistry	0.00	0.06	-0.01
4. Reaction Rate	0.00	0.04	0.09
5. Chemical Equilibrium	0.39	0.04	-0.01
6. Acid-Base Reactions	0.04	- 0.13	0.01
7. Electrochemistry	- 0.06	- 0.02	0.09
8. Organic Chemistry	- 0.03	- 0.02	-0.14
Total	0.00	0.24	0.27

Section 2: Language ability of both markers and candidates are questionable.

Candidates use 'correct' and 'incorrect' words in an specific context and markers do not always pick up on finer points – haste in marking may play a role where markers are only reading the words and not the context or way it is used. Formulae are often manipulated in an unusual way and then seem incorrect at first glance.

Section 3: Writing balanced equations and getting it partially correct confuses markers, despite instructions – mark allocation varies from year to year and question to question and markers start ‘mixing instructions’

Section 6: More than one method can be used – candidates ‘combine’ methods, doing both correct and incorrect work in their total answer. This gives a problem if the candidate doesn’t get it correct and has to be awarded partial marks.

CONCLUSION

All indications are positive that remarking can be used as a tool to quality-assure marking processes. Not only does it take away the subjectivity that goes with monitoring and observation, but is in fact one of the few quantitative measuring methods in education that are not subject to the ability of the specific group of candidates.

Further research into the statistical patterns in remarking, with a broad database, will lead to accurate criteria with which to evaluate the marking process. Research using the existing national papers as base will enhance the position on a national level. The move in RSA towards national question papers will make it possible to then establish a national standard against which marking can be evaluated. As remarking is a process where marking errors can be corrected, candidates should not subject it to an attempt to “gain” from the system. We believe it should be affordable, but candidates should feel the full effect of the challenge and accept the outcome, even if it results in a lower score. This should help to so that only real cases of a difference of opinion on quality of initial marking are evaluated against the norm and further be linked to qualifying criteria.

LIST OF REFERENCES

- Addis Ababa University**, 2006. Registration Information.
www.aau.edu.et/Applyonline/Reginfo/Academicslist.php
- City University London**, 2006. University Information and Procedures
www.city.ac.uk/edc/programmes/apractice/proregulations
- DEFS**, 2006. Department of Education and Skills. Access to Scripts.
www.dfes.gov.uk/qualifications/
- Demopoulos, K**, 2005. Rise in number of exam papers regarded. The Guardian.
<http://education.guardian.co.uk>
- DOE**, 2005. Regulations for the conduct, administration and management of assessment for the Senior Certificate. Government Gazette (484) No. 28156. 21 October 2005. Pretoria. Government printers.
- GDE**, 2006. Circular 1 of 2006. Re-checking, Re-marking and the Viewing of the Answer Scripts of the Subjects Written in the Oct/Nov 2005 Senior Certificate Examinations.
www.education.gpg.gov.za/Legislation/circulars/circulars.htm
- IEB**, 2006. General Notes: Senior Certificate Examination.
www.ieb.co.za/Formal/FETC_remarks_2005.htm
- Open University of Hong Kong**, 2006. Course Results Query and Appeal.
www.ouhk.edu.hk/
- Purpletrain**, 2006. PurpleTrain.com. Examination Queries.
www.purpletrain.com/studserv/faq_exam.asp?ccode=0136
- Teachernet**, 2006. Appeals procedures for GCSE and A level examinations
www.teachernet.gov.uk/educationoverview/uksystem/examinationsandqualifications/appeals/
- Technikon North Gauteng**, 2006. Remarking of scripts.
www.tng.ac.za/operations/studadexam.html
- Tomlinson, M**, 2002. Report on the Outcomes of Review of A-level Grading.
www.dfes.gov.uk/
- University of Papua New Guinea**, 2006. Rules and Procedures for All Undergraduate Programs. www.upng.ac.pg/uni_rules.htm)
- University of Warwick**, 2006. Frequently Asked Questions.
www2.warwick.ac.uk/fac/soc/celte/welt/faqs/
- UP**, 2006. University of Pretoria. Access to Marked Examination Papers & Remarking of Examinations. www.ee.up.ac.za/undergrad/