Is the glass half full or half empty? What examiners really think of candidates' achievement.

Jackie Greatorex

University of Cambridge Local Examinations Syndicate

Disclaimer
The opinions expressed in this paper are those of the author and are not to be taken as the opinions of the University of Cambridge Local Examinations Syndicate or any of its subsidiaries.

Is the glass half full or half empty? What examiners really think of candidates' achievement.

Abstract

Grade descriptors are descriptions of the knowledge and skills likely to be found in performance at a particular grade in public examinations in the UK. They can be seen as an articulation of examiners' tacit (unarticulated) knowledge about what distinguishes one grade from another. This tacit knowledge is expressed in grade descriptors in an abbreviated form to make short snappy statements about each grade which will be useful as a quick reference in an Award Meeting and for candidates and users of the qualifications. They are also written in a positive manner. Yet examiners do not necessarily view achievement at each grade positively. Therefore grade descriptors are limited as an expression of examiners' tacit knowledge. This paper endeavours to use the discussions that took place in Kelly's Repertory Grid interviews with Accounting Examiners and a qualitative analysis to develop more valid descriptions of examiners' tacit knowledge which might give us further insights into examiners' tacit knowledge. This is an exploratory pilot study. Bearing this caveat in mind it was found that Accounting Examiners tend to see achievement in both positive and negative terms. This and other aspects of examiners' tacit knowledge are discussed.

Introduction

For A level examinations, decisions are made about where grade boundaries should be set at Award Meetings. The Code of Practice (QCA, 2000) requires that a variety of evidence, including the professional judgement of senior examiners and statistical evidence, is used to make these decisions. Senior examiners inspect scripts to maintain the standard expected at each grade from previous years when they decide where the boundaries for grades A, B and E should be placed. One of the types of evidence, listed in the Code of Practice, which can be used, is grade descriptors. These are descriptions of the qualities that examiners expect candidates to exhibit at judgementally awarded grades. They are used to remind the examiners of the qualities that are expected at each grade. Other grade boundaries are determined arithmetically. Although the Code of Practice applies only to UK examinations, similar procedures are used to grade A levels which are offered internationally.

Christie and Forrest (1981) argue that the model of subject competence used in the British examination system is *limen referencing*. In this system, cut scores (or grade boundaries) are not fixed in advance, rather they are arrived at based on the scrutiny of candidates' responses. Examiners read scripts on a range of marks until they reach a limen, or threshold, below or above which they are confident of the grade the candidate deserves. The range of marks where it is not clear which grade the performance deserves is called the zone of uncertainty. The grades do not have an explicit definition of the achievements that they imply, so grade descriptors are no more than a useful indicator. That is, the UK examination system is not a fully criterion referenced system.

Sir Keith Joseph's announcement in 1984 about the introduction of GCSEs (General Certificate of Secondary Education, examinations generally taken by young people who are 16 years old and over) included a reference to grade related criteria – the criteria that students must meet to be awarded a particular grade (Gipps, 1990). There were problems with the criteria that were developed, e.g. they were too complex, so they were dropped (Gipps, 1990). The difference between descriptors and criteria is that descriptors are *expectations* and criteria are *requirements*. The inclusion of grade descriptors in syllabuses became common after the introduction of new GCSE syllabuses in 1988. But it was not clear in some syllabuses where the grade descriptors came from, how they were developed and whether they were related

directly to examination performance. Anecdotal evidence suggests that grade descriptors were not generally developed using evidence about students' performance – the descriptors were apparently based upon committee decisions and experts' expectations of examination performance but in the absence of more information it is difficult to comment on the validity of the grade descriptors. In other contexts, however, we know more about how attempts have been made to describe candidates' performances. Massey (1982) used strict statistical criteria (referred to here as a mastery levels analysis) to identify questions which were mastered by candidates who had achieved one grade, but not by the candidates who had achieved the grade below. He used this to describe what candidates could achieve at judgementally awarded grades in O level Chemistry. In the SRAC grade descriptors project, examiners noted down what they believed distinguished performance at different grades. The notes were developed into grade descriptors and the descriptors were validated against scripts for a range of A level subjects (SRAC, 1987). Pollitt and Murray (1996) developed grade descriptors by using Thurstone Pairs for examiners to rank order the scripts. To elicit descriptions of the differences between higher and lower performance Kelly's Repertory Grid (KRG) was used. Pollitt and Murray (1996) argued that examiners distinguished between higher and lower performance by using their own personal criteria. In this paper this is called a discriminator model of performance. Grade descriptors are an articulation of senior examiners' personal constructs about what distinguishes one grade from another. But grade descriptors are just one indicator of what distinguishes one grade from another. Other sources of information are found in the syllabus, question papers and mark schemes.

There are of course limitations with grade descriptors. There can be limitations of language and definition, especially when qualifying words are used. Appendix 1 is an example of grade descriptors from an Information Studies GCSE specification which was examined in summer 2000. In these grade descriptors, what is meant by the phrases like 'some degree of accuracy' , 'reasonable accuracy', 'considerable accuracy', 'considerable degree of accuracy' and 'complete accuracy'? These words might make grade descriptors 'vague'. Using qualifying words does not fit well with a discriminator model of performance. However it can also be argued that some people might use qualifying words to distinguish between different execution of the same skill at different levels and that the language in grade descriptors has meaning to a community of examiners and teachers within a subject area.

The aim of this paper is to explore the personal constructs (tacit or unarticulated knowledge) which Accounting Experts (AEs) have of judgementally awarded grades E, A and B. This knowledge is partly given in the syllabus, assessment criteria and in some cases, in published grade descriptors. However there might be other knowledge which is as yet unarticulated knowledge.

Studies to describe grade descriptors are discussed above. These are an attempt to elicit examiners' personal constructs (tacit knowledge) and to turn them into grade descriptors which are usually written in a positive light to convey what candidates know, understand and can do. The descriptors follow the optimistic principle that the glass is half full. But candidates' achievement has not always been perceived in this way. Selkirk (1988) explains that CSE and 'O' level grading systems were norm referenced. He adds that when educators were asked to develop grade related criteria the Examination Boards were keen to describe performance in more positive terms than previous common practice. He adds that:

Subject Working Party (SWP) members found it very difficult to draft statements in positive terms for those who had achieved grade C 'O' Level and 4 CSE because the nature of the evidence produced by candidates in 'O' level and CSE examinations was in response to questions which were designed to spread candidates over the mark range rather than to elicit from them what they knew, understood and could do (Selkirk, 1988, 24).

He goes on to explain that eventually grade descriptors rather than grade criteria were developed. These grade descriptors were also based upon the principle of indicating what candidates were expected to achieve rather than describing what candidates were expected not to know. Indeed marking in GCSE was about

rewarding a candidate for what they knew, understood and could do (what they had mastered) rather than taking off marks for mistakes or omissions (Selkirk, 1988). These principles now also apply to A levels and to public examinations generally in the UK. It may be the case however, that examiners' personal perceptions of achievement at different grades might not be as positive as suggested by the examination system e.g. in the content of grade descriptors. In an attempt to learn more about examiners' personal constructs and make their knowledge more transparent, this study focused on what examiners verbalised as the difference between achievement at adjacent grades during Kelly's Repertory Grid interviews. Initially these interviews were part of a study to develop grade descriptors.

The details of the study to develop grade descriptors have been reported elsewhere (Greatorex, 1999a). In summary, from the population of candidates of an A level Accounting examination who achieved a particular component grade and attempted a particular question, a random sample of 50 candidates was chosen. For example, from all candidates who achieved grade A on component 2 and who attempted question 1, a sample of 50 candidates was selected at random.

A mastery levels analysis as described by Massey (1982) was used to identify questions which discriminated between the achievement of candidates who achieved one component grade and those who achieved the adjacent grade. If a question discriminated between the achievement of two adjacent component grade groups and the higher grade group had averaged over a threshold, say, 50% of the marks, the higher grade group was considered to have mastered that question. Scripts from candidates who achieved the mean mark on this question for their grade group were used in the next stage of the process.

Greatorex (1999a) argued that the grade descriptors developed from statistical and qualitative evidence from scripts together with the AEs' tacit knowledge were valid as they were grounded in candidates' actual performance and examiners' tacit knowledge.
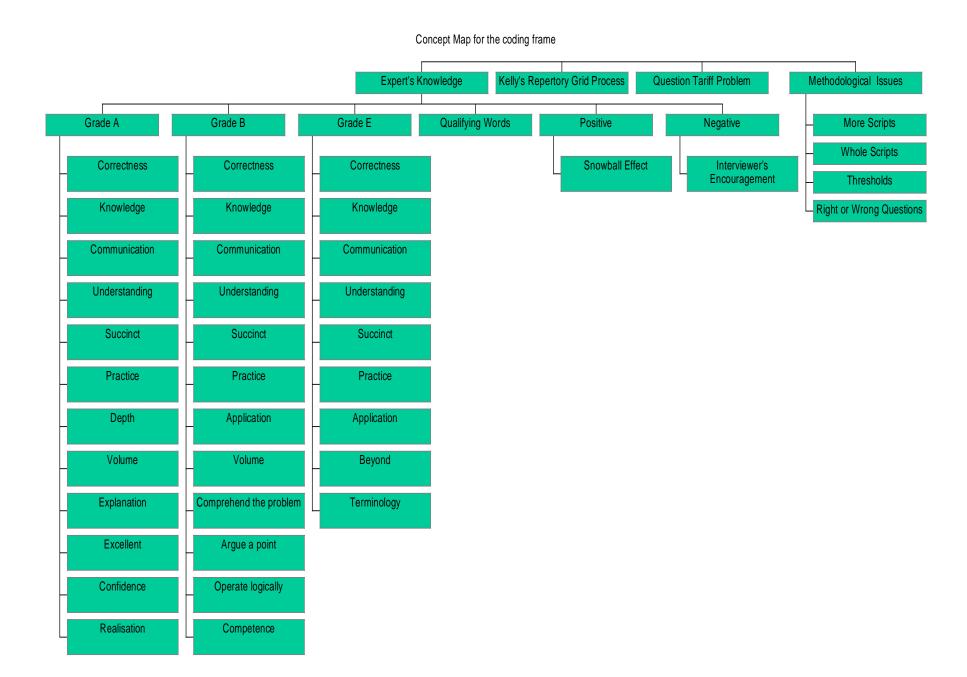

Method

Kelly's Repertory Grid is used to make people's personal constructs (views or knowledge) more explicit. In this case, examiners' views of the distinguishing characteristics of performance at each component grade were identified. Two Senior Accounting Examiners (AEs) compared a triad of scripts. The triad consisted of two scripts from a judgementally awarded grade (say component grade A) and one script from the grade group below (say component grade B). The scripts were selected because they exemplified differences in performance on a given question, as identified in the mastery levels analysis. The scripts exemplified mean performance for the given item and component grade. The examiners described the similarities between the A grade scripts and how they were different from the B grade script with reference to the given question. This was repeated with one or two other triads of scripts for a particular question to ensure that as many as possible of the distinguishing characteristics of each grade were identified. The exercise was repeated for other questions where mastery was achieved and for each judgementally awarded grade. Two AEs might be considered to be a very small sample, but it should be remembered that these two examiners constituted 40% of the people who made decisions about the Accounting grade boundaries at the Award Meeting in the summer of 1998. Also this is an exploratory stage of research into examiners' tacit knowledge and in such situations case studies are often used. For example, Anderson (1990) reports that that he studied how one woman acquired the skill of using a machine in the initial stages of his work to develop a theory of expertise.

Greatorex (1999a) described how grade descriptors were developed from the Kelly's Repertory Grid interviews. For this paper a more detailed analysis of the interview data was undertaken to explore AEs' tacit knowledge of candidates' performance. The Kelly's Repertory Grid interviews were recorded and transcribed for component 2. It is beyond the scope of this paper to consider the results from more than one component.

The categories that were used to code the data arose from the data itself. Some of the AEs' discussion related to which questions were being used and how many scripts were available. These data were removed from the analysis. The rest of the data were coded as *KRG process*, *Experts' Knowledge*, *Methodological issues* and *Question Tariff problem*. The *KRG process* data were discussions where the AEs worked out what the candidates had done. The data in the *Experts' Knowledge* category were coded for each judgementally awarded grade (E, A and B), that is, when the AEs were talking about grade A, this part of the transcript was coded as grade A. The data within each grade were coded in relation to the characteristics the AEs mentioned, e.g. *application* and *understanding*. Some of these categories were linked i.e. the AEs saw these characteristics together in the candidates' performance. Examples of these links are given in the results and discussion section. The quotations in each category were summarised to provide a description of each grade. The *Methodological issues* data are points which were raised by the AEs about the research method. The *Question Tariff problem* data were discussions about right or wrong questions, further details are given later. Further categories are given in the concept map for the coding frame on the next page.

It could be argued that the codes for each grade did not only arise from the data but were also a product of the method that was used. For example, the grades were a product of the method and purposes of the original research, see Greatorex (1999a).

.

Concept Map for the coding frame

| Expert's Knowledge | Kelly's Repertory Grid Process | Question Tariff Problem | Methodological Issues |

| Grade A | Grade B | Grade E | Qualifying Words | Positive | Negative | More Scripts |

Snowball Effect — Positive

Interviewer's Encouragement — Negative

Whole Scripts

Thresholds

Right or Wrong Questions

Grade A
- Correctness
- Knowledge
- Communication
- Understanding
- Succinct
- Practice
- Depth
- Volume
- Explanation
- Excellent
- Confidence
- Realisation

Grade B
- Correctness
- Knowledge
- Communication
- Understanding
- Succinct
- Practice
- Application
- Volume
- Comprehend the problem
- Argue a point
- Operate logically
- Competence

Grade E
- Correctness
- Knowledge
- Communication
- Understanding
- Succinct
- Practice
- Application
- Beyond
- Terminology

Results and discussion

The results of the mastery levels analysis and the analysis of the KRG interviews to develop grade descriptors are given in Greatorex (1999a). The results that are given and discussed here are based on a qualitative analysis of the KRG interview data. The concept map for the qualitative data is given on the previous page.

Methodological issues

During the interviews the AEs raised some points about the method that was being used. One AE was concerned that the AEs were only looking at a small number of scripts for each grade:

AE1: I know we're looking at various bits, but we really ought to look at more scripts, I think, in a perfect world, so that we see a number of differences.

About 6 scripts had been used for each grade and question to provide some variety. But it was sometimes difficult to identify scripts with the right grade and score on a question and there was a limit on how much time could be devoted to the study. However the scripts that had been used could be seen as a case study which represented prototypical (or average) performance for that grade and question. It should be possible to generalise from these cases as they were taken as a representation of the average from the statistical results of the mastery level analysis. Also in the KRG procedure AEs described the characteristics evident in scripts until there were repetitions in the qualities that they noticed in the performances. It was at this point that the assumption was made that most of the characteristics that distinguished the achievement of two different grades for these AEs had been identified. It might be that this process did not lead to an exhaustive list of distinguishing qualities.

The same AE also wanted to look at whole scripts rather than individual questions:

AE1: I also felt that eventually we should look at the script as a whole to try and see that it fits in.

It is understandable that this AE wanted to consider whole scripts as this is what happens in Awarding Meetings. Generally grade descriptors are used to facilitate making holistic judgements. But the adapted method of KRG used in this paper involves considering only a part of a student's performance and the whole of a candidate's performance is likely to be greater than the sum of the parts. However, Pollitt and Murray (1996) argued that examiners looked for specific features in performance to differentiate between performance at different levels or grades. The mixture of the mastery levels analysis to identify which questions the AEs should look at to find the discriminating features and Kelly's Repertory Grid to contrast higher and lower performance was designed to follow this discriminator model and should have been a more efficient way of finding discriminating features than using whole scripts. After all, for candidates on adjacent grades there is likely to be a good deal in common when their whole scripts are considered.

Sometimes, although the questions had met the statistical criteria for mastery, it was difficult for the AEs to talk about similarities and differences between performance. For example:

AE2: I suppose a lack of knowledge, there's a lack of knowledge. It's difficult to say with a one mark question.
INTERVIEWER1: Yes, really.
AE2: It's right or wrong and they didn't have to show their workings. Some of them just wrote down 5 or 6 and get their money, so to speak.
INTERVIEWER1: Yes.

At least the AEs could see which items distinguished between grades in terms of the likelihood of whether a given question was right or wrong.

There were other difficulties. To understand the following example it is necessary to know that grade O was the fail grade below the pass grade E.

AE1: It's difficult because the way that I am doing this with E and O, I am going to end up with E looking like A in that sense, you know, good knowledge and everything. That's not true.
AE2: That's right. The way that we are at the moment, we're not comparing bad and very bad, we're comparing as if E was good, and that first guy we looked at has got good marks.
AE1: All the Es we have looked at have got good marks on that question because we have worked on this 75% mastery, haven't we?

It was explained that we had experimented with a 75% threshold and other lower values in the statistical criteria (see Greatorex, 1999a, for more details). It was also explained that candidates who achieved different grades might master different types of questions. As the AEs sometimes described the context within which the candidate was performing, if the same words were used to describe A and E performance there should still be a difference between the description of E and higher grades because the context would be different. Despite this, when comparing grades B and C, the same AE was concerned that B would look like A.

AE1: As I say, the danger here is that if you use C then it makes B look like A.

He was later convinced that they had described A and B differently, for example:

INTERVIEWER1: I mean in your A grade you were saying things like showing 'enthusiasm' and a 'motivation to give a full answer'.
AE1: Yes. These are not. They are showing competence but no flair.

Despite these difficulties, towards the end of the interviews the AEs thought that the mastery levels analysis had worked:

AE1: You have actually found us some really nice discriminating questions.

This enabled them to see the qualitative differences between the performance of candidates gaining adjacent grades:

AE1: In both the ones that you have given us, there are clearly differences between Bs and Cs.

This is a particularly encouraging comment as some examiners say that they have a clear concept of what constitutes A and E grade performance but that B performance is harder to identify and is just not quite as good as A grade performance.

The Kelly's Repertory Grid Process

The AEs were asked to compare, say, two A grade scripts with a B grade script and to describe the similarities between the two A grade scripts and how they differed from the B grade script. To do this the AEs studied the scripts to work out what each candidate had done, then they made inferences based on the evidence in front of them and on their personal knowledge. After the AEs had considered the scripts they

identified what, say, A grade candidates had mastered which B candidates had not. They repeated the process at the B/C and E/O boundaries:

AE1: They [B] can understand a formula and apply it. Whereas C can't.

AE2: You've got comprehension at the E grade which perhaps they don't have at the O grade.

These are examples of where the Kelly's Repertory Grid procedure was working well.

Experts' Tacit Knowledge - Describing each grade

Tables 1 to 3 contain examples of what the AEs said about the scripts. These quotes are categorised and the heading for each category is given in the first column of the tables. The final column contains a summary of the characteristics which were considered to distinguish achievement at adjacent grades. In most cases these summaries are based on many more quotes than the examples that are presented. The grade descriptors which were developed from Kelly's Repertory Grid interviews and the mastery levels analysis are available in Greatorex (1999b). They were validated in the summer 1999 Award Meeting. It should be stressed that the descriptions of each judgementally awarded grade in tables 1 to 3 have NOT been used to award marks or grades to students' work.

The number of comments that were made in each category has not been given in the tables below as it is the meaning of the comments that is important rather than the number.

Table 1 Distinguishing characteristics of E grade performance

Characteristic	*Example(s) of quotes*	*In their performance an E grade candidate is likely to:-*
Communication Skills and Aesthetics	There is a sort of neatness.	exhibit presentation skills such as neatness, legible handwriting, precise wording and clarity; present their answer in a way which is understandable for the reader;
Correctness	It's almost in the correct answering of the question. They've focussed down a bit more on the problem.	use the correct formula and / or the correct process to answer the question; give the correct answer; focus more on the problem; select the correct information from the question;
Terminology	I feel a better familiarity with accounting terminology, that they use the correct terminology… He knows the right words, the right jargon.	use and illustrate familiarity with the correct terminology;
Application	He's got the right formula, he's got the right application of it.	apply formulae;
Knowledge	It's knowledge………	illustrate their accounting knowledge;
Understanding	They understand it, the others don't. You've got comprehension at the E grade.	show understanding and comprehension of the level accounting domain;
Succinct	Both [Es] have been more economical in terms of getting to the answer.	utilise a correct and economical route to reach an answer;
Beyond the requirements of the question	Yes, one of them has, sort of presented the formula, actually both of them have presented the formula, which we didn't ask for.	substantiate their answer by going beyond the requirements of the question;
Practice	Rehearsed.	show evidence of rehearsal and practice.

Some of the categories were related to one another, for example, *correctness* was related to *succinct*:

AE2: They've adopted the correct approach, or the fastest route to get the answer.

There were also a links between:-
- *correctness* and *terminology;*
- *correctness* and *application;*
- *correctness* and *knowledge;*
- *application* and *knowledge.*

The order of the codes in the table is determined by whether the codes were linked to one another. So for example, *correctness* linked to *terminology*, *application* and *knowledge.* These codes follow one another in the table. Other than these links there was no significance to the ordering of the codes in the tables.

Table 2 Distinguishing characteristics of A grade performance

Characteristic	Example(s) of quotes	In their performance an A grade candidate is likely to:-
Correctness	Correct technique and given the correct answer. Accurate calculation.	use the correct technique to gain the correct answer; calculate answers accurately;
Knowledge	There is a level of knowledge which is reflected in the answer.	exhibit a level of knowledge;
Depth	Greater depth.	illustrate greater depth;
Explanation	The Bs tend to just merely say this is not so, rather than explain why like the As.	explain by giving a reason why; give a full explanation;
Volume	Again the As have fully answered the question.……More points, more ideas. Yes. It's not quite total quantity but there is a, there is more from the As generally.	fully answer the question e.g. by presenting more points and ideas;
Communication Skills and Aesthetics	It is clearer, precision, more precise.	communicate clearly; provide precise answers;
Excellent	Excellent.	achieve excellence;
Succinct	Very succinct, really this guy in particular.	provide a clear, compact and succinct answer; use a simple, succinct and direct route;
Confidence	There is almost a confidence thing with A.	show confidence;
Practice	Practice is an aspect here, it's less practice that they [B] are showing.	show evidence of practice;
Understanding	A understands all aspects of it [the question].	understand all aspects of question;
Realisation	They have realised when they have answered the problem fully.	realise when they have answered the problem fully.

There was an interesting paradox within the characteristics that were noted in the A grade performances. The A grade candidates were showing *depth*, fullness or a sense of *volume* to their answers but they were also described as *succinct* or using an economical route to answer a question. It could be that the candidates answer the question thoroughly but that they were able to do this in a *succinct* manner. There were also some links between the different categories. There was a link between *correctness* and *knowledge:*

AE1: There is a level of knowledge which is reflected in the answer. One is right and one is wrong and the difference is either 2 or 0. It is that sort of right/wrong question.

There were also other links between:-
- *depth* and *knowledge;*
- *practice* and *knowledge;*
- *explanation* and *volum*e;
- s*uccinct* and *explanation.*

Table 3 Distinguishing characteristics of B grade performance

Characteristic	Example(s) of quotes	In their performance a B grade candidate is likely to:-
Practice	So well practised.	appear practised;
Correctness	The problem is that B is imperfect but C on the one I am looking at is very imperfect. B has some of the ideas but he is carrying them out imperfectly.	give some ideas but carry them out imperfectly; make some calculation errors; offer an imperfect answer;
Operate logically	B is able to do some element of novelty, to add, to take cognisance of change.	show cognisance of change; make some progress in a novel situation; work/think through/around a novel problem logically;
Comprehending the problem	He sees the problem.	'see' a novel situation/problem; illustrate that they have fully read the question;
Knowledge	Yes, a knowledge difference between the B and C.	show a level of subject knowledge;
Application	It's application of the knowledge.	apply knowledge; apply a standard/routine formula;
Understanding	I think that they understand what they are talking about. The content is much better grasped here.	illustrate that they understand syllabus content (e.g. a routine formula);
Volume	Answering fully in a routine situation, that's what would define B.	provide a full answer in a routine situation; finish a question; answer the question i.e. conform to the terms of the question;
Communication Skills and Aesthetics	Yes, they are economical in their application of space.	use the space provided in an appropriate manner; approach the question in a methodical and systematic fashion e.g. numbering the points in their answer 1 to 4 when they have been asked for 4 points;
Argue a point	Able to reason and to put together an argument.	be able to reason and to put together an argument;
Succinct	Succinct.	be succinct;
Competence	They are showing competence.	show competence.

There were some links between different categories. There were links between the categories *operate logically* and *comprehending the problem*:

AE1: I think that this is a very good definition of B, wouldn't you? Able to react to some extent and see a novel situation and to work through logically.

There were also links between:-
- *volume* and *communication skills and aesthetics;*
- *knowledge* and *application;*

- *application* and *understanding*.

There were some aspects of the E grade performance in Table 1 which made E grade performance sound as good as A grade performance. For example, in Table 1 for grade E 'use the correct formula and / or correct process to answer the question' and from Table 2 for grade A 'use the correct technique to gain the correct answer'. Occasionally the problem is reduced by the inclusion of a description of the context of task that the candidates were required to undertake. These descriptions of context make statements about the relative demand of the task, for example, 'basic formula'.

Qualifying Words

It was evident that the AEs used qualifying words to describe the differences between one grade and another. For example:

AE2: Yes, that's what I am saying. The content is <u>much better</u> grasped here.

AE2: <u>Superior</u> application of knowledge.

AE1: E has <u>greater</u> knowledge. (author's emphasis)

This does not fit easily with the discriminator model of performance which suggests that to distinguish between levels of performance examiners look for distinguishing characteristics. It could be that the groups of characteristics which make up performance at a particular grade have some unique features and some features which apply to a greater or lesser extent. In the latter case qualifying words would apply. The characteristic of *correctness* consistently discriminated between adjacent grades. This suggests that AEs see achievement in Accounting, in part, as a function of *correctness*. It might be a 'catch all' but unhelpful concept as it does not explain what qualities distinguish between performance at each grade. This concept of *correctness* as a discriminator might have dominated Accounting more than would have been evident in another subject as in Accounting it is possible to be 100% correct and usually only in one way. In contrast in more subjective and qualitative subjects like English Literature and Arts subjects gaining 100% is not impossible but is less likely and can be demonstrated in a number of different ways. The other characteristics which consistently differentiated between grades were *communication skills and aesthetics*, *knowledge*, *understanding* and *succinct*. See the concept map earlier which illustrates which categories were associated with each grade. The characteristics in each of these categories were different for each grade. There were further categories which were distinct for each grade e.g. *terminology* for grade E and *depth* for grade A. The latter categories fit well with the discriminator model of performance. If the categories such as *terminology* and *comprehending the problem*, which are distinctive to a grade, are scrutinised more closely it becomes evident that these categories must apply to other grades, that is, it is not only grade E candidates who are familiar with *terminology*. AEs' personal constructs and the discriminating features of candidates' performance have two structures:-
- There are categories which are 'progressive' i.e. they can be used to distinguish between one or more grades e.g. *knowledge* and *understanding*. Here relative judgements or descriptions of the context e.g. the demand of the task might be used to distinguish between levels of performance.
- There are categories which are 'emergent' i.e. they emerge at a particular grade and are likely to be evident for all higher grades e.g. being familiar with *terminology* emerges at grade E but is also true for higher grades. Here AEs look for a characteristic which distinguishes a higher grade from an adjacent grade. It might be that at higher grades, either the characteristic discriminates significantly or perhaps it becomes implicit and examiners do not pay attention to it when they are trying to discriminate between performances. This is a suggestion which has been made by Pollitt and Murray (1996) and fits with the discriminator model of performance.

Is the glass half full or half empty?

A short while into the interview the two AEs made comments that suggested that rather than seeing achievement in a positive light they tended to see it in a negative light.

AE1: Here we're trying to define levels of error, sort of thing, and it's an imperfect product … the technical errors that you can make with a few numbers are considerable.
AE2: Between E and O you are looking at very bad compared to really terribly bad.

This links to the catch all notion of *correctness* that the AEs used to describe differences for each grade. Here the AEs are using a deficit model of achievement - rather than seeing the glass as half full they have seen it as half empty. Assessors having a deficit model of achievement or learning has been evident in other areas of education, for example healthcare courses in higher education (Greatorex, 1998). But does this negative view of achievement permeate through all their comments in the interviews? There were instances where the AEs tended to describe the errors of the lower grade performance (for example, O in the case of E/O) rather than what the higher (E) grade had mastered. This is evidenced by the interviewer encouraging the AEs to describe mastery rather than error:

INTERVIEWER1: Those things that you've been saying about having the wrong idea and circuitous route and things like that, can we turn them up the other way to say what's similar about the emergent construct in, you know, looking at those characteristics, only in a positive way, and if we try and talk about what they are in the E grade.

AE2: Technical error.
INTERVIEWER1: I keep saying 'Can we turn it up the other way' or whatever, because I'm aware that we're trying to describe …
AE1: More positive.

There was other evidence that the AEs conceptualised achievement negatively:

AE1: The problem is that B is imperfect but C on the one I am looking at is very imperfect.
AE2: Yes that's right.
INTERVIEWER2: But when you say it is imperfect what is B getting that …
AE2: B has managed to do some of the second part.
AE1: B has some of the ideas but he is carrying them out imperfectly. But as I say we are looking at the difference between B and C, C is a long way short, C is not really, at least my C…

AE1: They are certainly imperfect, the Bs.
AE2: Well, they haven't got the full eight marks for a start.

But there were also more positive views of achievement, for example, for grade B:

AE1: There's nought wrong with that, it's bang right.
AE2: Excellent.

AE1: Correct.
AE2: Correct technique and given correct answer.

AE1: And given in a compact and clear fashion. He got to the heart of the matter. If you give them a question they answer it.

For grade E:

AE2: This one's got the full marks with the right answers.
AE1: The better ones have gone the right way.

It has already been mentioned that the AEs seemed to associate some characteristics with one another. These links have been listed above. All the comments about characteristics which were linked were positive. It appears that some positive characteristics had a *snowball effect*. That is, when the AEs saw a positive characteristic in an answer they sometimes noticed another characteristic. For example, when looking at A grade scripts the following comment was made:

AE2: Greater depth, knowledge as well.

In conclusion, at times examiners will see achievement in negative terms - the glass is half empty. This happened despite the researchers' attempts to encourage the AEs to describe candidates' achievement in positive terms and the established practice of describing achievement in positive terms for examination purposes in the UK. But there are also times when they describe achievement, even lower achievement like grade E, in positive terms – the glass is half full. However it does seem to be that the examiners see achievement as a continuum of *correctness* i.e. a continuum of correct to incorrect answers.

AE2: Very hard to see a difference between the As and the Bs except that B got it wrong.

Correctness is given by the similarity between the mark scheme and the candidate's answer. This is reflected in part of the views of one AE:

AE1 It's [answering examination questions] a bit like a crossword puzzle, if you will, you have to crack the code in the examiner's mind or the setter's mind, and once you have done that it's in.

It is this notion of correctness which has been mentioned for each grade.

A case of the question tariff problem?

A final point is that there are some questions which the AEs described as right or wrong situations i.e. candidates would get the question either right or wrong. For example:

AE2: Well, we are swapping the Bs [exchanging B scripts] but if they got 2 marks then they will be right and that's about it.

Another example was:

AE1: There is a level of knowledge which is reflected in the answer. One is right and one is wrong and the difference is either 2 or 0. It is that sort of right/wrong question.

The questions which have been described in this way tend to be questions where there were very few marks available. This issue of questions where there was either a right or wrong answer might warrant further investigation for this syllabus. It could be that some of the Accounting questions suffer from 'the question

tariff problem' where, say, for a two mark question, most people gain 0 or 2 marks rather than the majority of people gaining 1 mark. For further details see Bramley (2000).

Conclusion

A logical final stage of the analysis in this paper which has not been undertaken due to limited resources is asking the AEs to validate the descriptions of each grade. In future studies it is hoped that this can be undertaken by using a vignette approach. The descriptions of each grade would be presented to the AEs and they would be asked to comment on the results of the research and to suggest amendments. This would be repeated to ensure that the amendments were satisfactory to the AEs.

The grade descriptors which were developed using a mastery levels analysis and KRG were grounded in candidates' achievement and AEs' tacit knowledge (Greatorex, 1999a & b). This meant that the grade descriptors were perhaps more valid than other grade descriptors which were not based upon evidence. The results of this exploration describe the examiners' personal constructs (tacit knowledge) of the prototypical features of each judgementally awarded grade. Despite the established principle that performance should be described in a positive manner the AEs tend to describe achievement in both positive and negative ways. This means that grade descriptors are not necessarily a valid representation of AEs' tacit knowledge and / or candidates' performance as grade descriptors tend to describe achievement in positive ways but candidates' performance does not only have positive qualities and AEs' tacit knowledge suggests that there are positive and negative aspects to candidates' performance.

There is also a tension between the system of describing candidates' achievement in positive ways and the AEs' tacit knowledge which is both positive and negative. This too affects the validity of grade descriptors. It might be inevitable that examiners will view achievement in both positive and negative ways as there are obviously both positive and negative aspects to achievement. AEs also have a concept which links the correctness of the response to the grade that candidates are awarded. This concept of correctness might lead to the positive and negative descriptions of performance. This phenomenon might be specific to Accounting and / or subjects which depend heavily upon numeracy.

Describing achievement only in positive terms might be an attempt to help lower achieving young people feel that they have achieved something. However it might be that if descriptions of lower levels of achievement referred to deficits in achievement as well as the positive aspects of achievement this would indicate to students how they needed to improve to achieve higher grades. Of course this might not be desirable as it would be changing the function of grade descriptors from one of a number of indicators used in an Awarding Meeting to a diagnostic tool.

It was also found that there are both progressive and emergent characteristics in candidates' performance at different grades and in AEs' personal constructs (tacit knowledge) of performance. Further research might include whether some Accounting questions suffer from the question tariff problem and whether there are similarities between the tacit knowledge of examiners from different subjects.

References

Anderson, J. R. (1990) *Cognitive Psychology and its Implications,* (3rd ed.), W. H. Freeman and Company: United States.

Bramley, T. (2000) *The question tariff problem*, a paper presented at the British Educational Research Association Conference (BERA), Cardiff University, 7-9 September.

Christie, T. and Forrest, G. M. (1981) *Defining Examination Standards*, Schools Council Research Studies: London.

Gipps, C. (1990) *Assessment A Teachers' Guide to the Issues,* Hodder & Stoughton: London.

Greatorex, J. (1998) *Educational levels in a higher education and healthcare context,* unpublished PhD thesis, University of Derby.

Greatorex, J. (1999a) *Making the grade - a discriminator model of performance?* unpublished MEd dissertation, University of Bristol.

Greatorex, J. (1999b) *Making the grade – a novel method for developing grade descriptors*, a paper presented at the British Educational Research Association Conference (BERA), University of Sussex, 2-5 September.

Massey, A. J. (1982) Assessing 16 + Chemistry. The exposure - mastery gap, *Education in Chemistry,* September, 143 - 145.

Pollitt, A. and Murray, N. L. (1996) What raters really pay attention to, in M. Milanovic and N. Saville (Eds) *Studies in Language Testing 3, Performance Testing, Cognition and Assessment*: selected papers from the 15th Language Testing Research Colloquium, Cambridge University Press: Cambridge and Arnhem.

QCA (2000) *GCSE and GCE A/AS code of practice,* QCA: London.

Selkirk, K. (1988) *Assessment at 16*, Routledge: London.

SRAC. (1987). *Interboard Grade Descriptors Project.* The 1987 Studies. Procedure Document, Stranding Research Advisory Committee for the GCE Boards.

Appendix 1

An example of GCSE grade descriptions

OCR Information Studies GCSE 3428 examined in summer 2000

Grade G

Candidates collect and process information using given equipment with some degree of accuracy. They present information in different ways using comprehensible language and usual conventions. They understand the basic premise of storage, responsibility, accuracy and clarity. They can handle information in more than one form.

Grade F

Candidates collect and process information using appropriate equipment with some degree of accuracy. They present information in different ways using comprehensible and largely accurate language, and usual conventions. They understand the basic premise of storage, responsibility, accuracy and clarity. They can handle information in textual, graphic and numerical form.

Grade E

Candidates collect and process information using appropriate equipment with reasonable accuracy. They present information in different ways using comprehensible and accurate language, and usual conventions. They understand the basic premise of storage, responsibility, accuracy and clarity. They can handle information in a variety of forms.

Grade D

Candidates collect and process information from located sources, using appropriate equipment with reasonable accuracy and appropriate to the purpose. They present information in different ways using appropriate language and applying taught styles and conventions. They understand and can apply methods of storage, responsibility, accuracy and clarity. They can handle information in a variety of forms.

Grade C

Candidates select appropriate sources, information required and method of display for a given need with considerable accuracy. They present information in a variety of ways using appropriate language and applying taught styles and conventions. They understand and can apply methods of storage, responsibility, accuracy and clarity. They can handle information in a variety of forms. They can exercise a degree of initiative.

Grade B

Candidates are able to select sources of information, methods of display, information necessary for a defined purpose and process information with a considerable degree of accuracy. They can present information for a variety of purposes using appropriate language and applying taught styles and conventions. They understand and can apply methods of storage, a range of areas of responsibility, accuracy and clarity. They can handle information in a variety of forms. They can exercise a degree of initiative.

Grade A

Candidates are able to select sources of information, methods of display, information necessary for a range of purposes and process information with complete accuracy. They can plan the presentation of information for a variety of purposes and produce documents suitable for a specific audience using appropriate language and applying taught styles and conventions. They understand and can apply methods of storage, a range of legislation, regulations, guidelines and responsibility in detail with accuracy and clarity. They can handle information in a variety of forms. They can exercise a considerable degree of initiative.

Grade A*

Candidates are able to select sources of information, methods of display, information necessary for a range of purposes and process information with complete accuracy. They can plan the presentation of information for a variety of purposes and produce documents suitable for selected audiences using appropriate language, applying taught styles and conventions in a range of situations. They understand and can apply methods of storage, a range of legislation, regulations, guidelines and responsibility in detail, with accuracy and clarity. They can handle information effectively in a variety of forms. They can exercise a high degree of initiative and work largely without instruction.