

Developing aspect-specific assessment criteria for examination answers and coursework essays in psychology

James Elander, Department of Psychology,
London Guildhall University.

Abstract

This article describes the development of assessment criteria for specific aspects of examination answers and coursework essays in psychology. The criteria specified the standards expected for seven aspects of students' work: addressing the question, covering the area, understanding the material, evaluating the material, developing arguments, structure and organisation, and clarity in presentation. The criteria were designed to promote student learning, increase the reliability of marking, facilitate research on marking, and promote reflective practice among markers. Student feedback indicated that, despite being published in course materials, the impact of the criteria on students was much less than it could be, but the criteria supported other initiatives to promote student learning, including more structured feedback to students on coursework essays, and generic skills teaching on essay writing. The criteria were used in research to analyse the judgements made by markers of examination answers. That research revealed substantial individual differences among markers and more systematic differences between first and second markers. Some staff had reservations about specifying the assessment criteria in such detail, but the criteria provided a stimulus for staff reflection on the process of marking and agreeing marks. The development of the criteria opened up a number of avenues for further work on student learning and assessment and for further research on the psychology of marking.

Introduction: What are assessment criteria for?

Assessment criteria usually take the form of brief descriptions of the type of work that is expected at each grade band, as in the example given in Table 1. They have two main purposes. One is to allow students to understand how their work will be assessed

and how it could be improved. The other is to improve the quality of marking, which means increasing the reliability and validity of marks awarded as a measure of the standard of students' work.

Assessment criteria serve important pedagogic aims. Partington (1994) suggested that the criteria for assessment

**Table 1. An example of assessment criteria for examination answers
(adapted from Miller *et al.*, 1998).**

Grade	Criteria
A	Excellent critical and conceptual analysis; comprehensive survey of relevant issues; well argued; well presented; relevant reading effectively incorporated.
B	Good critical and conceptual analysis; good survey of relevant issues; satisfactory presentation; relevant reading effectively incorporated.
C	Rather more descriptive than critical and conceptual; analysis lacks clarity in parts; evidence of relevant reading but not always effectively used.
D	Perfunctory; largely descriptive; disorganised and lacking in detail.
E	Perfunctory; almost entirely descriptive; narrow in conception; poorly organised.
F	No evidence of understanding; little evidence of a serious attempt.

should be specified well in advance as part of the published syllabus. Having the criteria attached to the assignment topic in advance helped students to explain their grades and prepare for further assignments (Miller *et al.*, 1998). 'Wherever possible, students have a right to know how their essays will be marked and, generally speaking, students have a right to know afterwards the basis on which their marks were awarded' (Miller *et al.*, 1998, p.113).

A more specific pedagogic reason for publishing assessment criteria is to promote 'deep' approaches to learning. 'Deep learning' or 'deep processing' involve attempts to understand the material rather than memorise and reproduce it (Marton & Saljo, 1976). Longhurst and Norton (1997) described five qualities of coursework essays considered by psychology tutors to characterise deep learning. These were: addressing the question throughout the essay, clearly organised essays with structure appropriate to the question, quality and relevance of argument, depth of understanding in relation to underlying psychological issues, and evaluation of theoretical concepts and research evidence. Those qualities usually feature prominently in assessment criteria, and publishing the criteria is

an opportunity to remind students about the importance of adopting a deep approach to learning.

Students may not always assimilate the published criteria at face value, however. Research by Lin Norton and colleagues showed that psychology students had their own views about what tutors look for in coursework essays, believing that markers were impressed by strategies that did not involve deep learning and were not part of the published assessment criteria (Norton *et al.*, 1996a; 1996b; 1999). Even where students accept the importance of the criteria, they may still have difficulty translating the published criteria's rather abstract description of desirable qualities into concrete ways they can improve their own work. Longhurst and Norton's (1997) research showed that students may appreciate what the criteria are, but still misunderstand how tutors will apply them when marking their essays. One of the methods suggested by Norton (1990) to help clear up misconceptions among students about what is important in essay writing was for tutors to state their criteria more explicitly, and to make efforts to explain to students what is meant by the qualities set out in the criteria. Another was for tutors to use a format for

written feedback to students that is focused on the criteria.

Turning to the second purpose of assessment criteria, the reliability of university marking in psychology has been the subject of a good deal of research, much of which was reviewed by Newstead (1996). The results have been fairly mixed, and some of the findings indicate considerable scope for improvement. Laming (1990) examined the marks awarded by pairs of markers to answers in a university examination over two years. The correlations between the two marks ranged from .47 to .72 for one year and from .13 to .37 for the second.

Newstead and Dennis (1994) examined the marks awarded by 14 external examiners and 17 internal markers to six answers to a single examination question ('Is there a language module in the mind?'). The coefficients of agreement between markers were .46 for the external examiners and .58 for the internal markers. In Dracup's (1997) analysis of marking over the range of units in a psychology programme, the correlations between marks awarded by first and second markers ranged from .47 to .93 for compulsory units, and from -.28 to .94 for optional units with smaller numbers of students. Caryl (1999) examined the reliability of marking in second year psychology examinations over five years. Reliability overall ranged from .75 to .87, but there were considerable variations in reliability between years and between areas of psychology.

The fact that students' degree classes are based on a number of assessments means that the reliability of degree awards is much higher than for individual units of assessment. Newstead and Dennis (1994) argued that the measurement error they estimated would lead to misclassification only for students who were very close to degree class borderlines, and when marks were averaged across all the units in Dracup's (1997) study, the correlation between first and second marks was .93. This may be

reassuring from the point of view of awarding degrees that reflect students' average level of performance, but we should still be concerned about the reliability of marking for each unit or assignment.

There is not much evidence from psychology teaching about whether using assessment criteria helps to improve the reliability of marking, but studies of assessment in other subjects suggest that it probably does. For example, Miller *et al.* (1998) describe research showing that in English composition, agreement between markers was greatest when, prior to marking, lecturers discussed the criteria that were likely to influence their judgements.

The reliability of marking in schools examinations appears to have improved considerably over time, possibly because of the introduction of measures to ensure comparability of standards, including detailed marking schemes. Hartog and Rhodes (1935) described the results that were obtained when a selection of School Certificate (the precursor of GCEs and GCSEs) scripts in history were remarked between 12 and 19 months after the first marking. Reliability coefficients were not given, but Newstead and Dennis (1994) concluded that reliability was almost certainly very low; despite the fact that only three categories were used (fail, pass and merit), nearly half of the scripts were assigned to a different category on being remarked.

In later studies, GCE scripts in a range of subjects were remarked, and the reliability coefficients were impressively high (ranging from .73 for English to 1.0 for Mathematics (Murphy, 1978; 1982). This was after detailed marking schemes had been introduced, along with other measures, and Newstead and Dennis (1994) interpreted the improvement in reliability as evidence that those measures had been successful. Psychology was not included in the GCE studies, but higher reliability was associated

with subjects examined by highly structured questions and lower reliability with essay-type questions: 'In fact, the difference in the style of the examining techniques tends to outweigh between-subject differences' (Murphy, 1982, p.62).

There is much less empirical evidence about the validity of marking than about reliability, because there is almost never a gold standard or external criterion against which marks can be compared. Concerns about the validity of marking in psychology are raised, however, by evidence of differences in degree classifications between institutions or between different years (e.g. Myron-Wilson & Smith, 1998; Smith, 1990), by evidence of gender biases (e.g. Bradley, 1984; Newstead & Dennis, 1990), and by evidence that marks may be affected by personal knowledge of the student (e.g. Dennis *et al.*, 1993).

In most cases the best approximation to a gold standard for marking is the assessment criteria themselves, which specify the qualities for which marks should be awarded. Recommendations about the way assessment criteria should be developed often emphasise that they should represent a summary of what differentiates work of different quality: 'By comparing outstanding and very poor quality examples, assessors can zero in on what key features make them different. Then tentative criteria can be refined and confirmed by applying those criteria to other samples representing high and low performance' (Quellmalz, 1991, p.330).

Assessment criteria, therefore, have a key role to play in both learning and assessment. This article describes the development and evaluation of a set of assessment criteria for coursework essays and examination answers in psychology that were designed to support student learning and achievement, improve the reliability and validity of marking, facilitate research on marking, and encourage markers to reflect on their performance.

The issues

There were several related issues that informed the development of assessment criteria in the psychology department at London Guildhall University. The first was the need to review learning outcomes for the course units of the undergraduate programme and ensure that the assessment criteria were consistent with those. Benchmarking statements were expected to set out subject skills and generic skills that students should demonstrate, with explicit progression in those skills from one level of the programme to another. We therefore wished to develop assessment criteria that incorporated the skills we believed were developed and assessed in the undergraduate programme.

The second issue was the need to support student progression and achievement. The university's very active role in widening participation has meant that an increasing proportion of students begin their degree courses with little confidence in their academic ability or study skills (Hall *et al.*, 2001). The development of the assessment criteria supported several initiatives designed to promote student learning. One of those was to provide more effective feedback to students about their coursework. We wished to formulate assessment criteria in such a way that feedback could be related to specific aspects of the criteria. Another initiative was to provide workshops and tutorials at Level 1 that focused on key generic skills, including essay writing and exam preparation. To be effective, these needed to relate as closely as possible to the ways in which students were assessed in the subject specific parts of the psychology programme. The publication of detailed assessment criteria would, we hoped, provide a focus for those workshops and tutorials, and increase the likelihood that discussions about skills were linked in students' minds with the type of work they would need to produce in psychology.

The third issue concerned the administration of marking and the use that was made of double marking. Previously the department had applied double blind marking of all work that contributed to students' degree classifications. Marking consumed a great deal of staff time, and consideration was being given to a system where only a sample of the work submitted for each assignment would be double marked, or where second markers would moderate rather than mark blind. A system of sampling or moderating depends on quite strong assumptions about the reliability of marking, however. If the sampling or moderation indicated serious concerns about the reliability of marks, all of the work for that unit would need to be double marked and the gains associated with the new system would be forfeited. We therefore wished to have available very detailed assessment criteria to support the reliability of marking: 'The absence of previously agreed criteria can actually nullify the hoped-for effects of double marking...Time-consuming double-marking ceases to be necessary if there are published mark schemes moderated by the external examiner' (Partington, 1994, pp.59–60)

The fourth issue was the development of research. We wanted the assessment criteria to provide a tool for investigating markers' judgements in research that treated marking as a psychological phenomenon in its own right. The findings might also help to improve the reliability and validity of marking if, for example, they identified biases that could be addressed.

The fifth issue was the desire to encourage reflection and debate among staff about marking and how it is undertaken. By developing criteria with detailed descriptions of specific aspects of essays and examination answers, we hoped to encourage markers to award marks that reflect a balance of relevant qualities and avoid judgements of the kind that have been used to illustrate the sometimes absurdly imprecise assessment of examination scripts:

'What a pedestrian attempt! Forty-five, I suppose. But at least he doesn't drop in a lot of leftist slogans. Okay, fifty two.'

'Quite well argued. Sixty-two, maybe. But he refuses to get to grips with the real issues. Fifty-seven.' (Husbands, 1976, in Laming, 1990, p.241).

To summarise the issues that informed the development of the assessment criteria, we wished the criteria to:

- a. Improve the quality of marking by increasing the correspondence between marks awarded and the specific learning outcomes for each unit of assessment.
- b. Provide a framework for more structured feedback to students about their work, and provide explicit links between generic skills teaching and subject specific assessments.
- c. Increase the reliability of marking in order to support the use of moderation or sampling of marks rather than double blind marking.
- d. Provide a tool for research on the psychology of marking.
- e. Stimulate staff reflection on marking.

The development process

Assessment criteria generally provide anchor points along a grading scale in the form of descriptions of the quality of work expected at each grade band, like the example given in Table 1. Those criteria are broadly similar to the ones used by many university psychology departments. They provide rather global descriptions that combine a number of aspects of students' work, including the breadth and depth of material in the essay, and the quality of argument, critical analysis and presentation. Following discussions at Boards of Studies and Standards Boards, the department decided to develop criteria for each of several specific aspects, or dimensions, of students' work.

The first step was to identify the aspects that would be specified. We aimed for a manageable number of aspects covering

Table 2. An attempted synthesis of marking schemes, adapted from the *Guidelines for External Examiners on Undergraduate Psychology Degrees, Appendix C* (The British Psychological Society, 1989).

Grade	Criteria/typical qualities
First class	<ul style="list-style-type: none"> ● Comprehensive and accurate coverage ● Critical evaluation ● Clarity of argument and expression ● Integration of a range of materials ● Depth of insight into theoretical issues ● Originality of exposition or treatment
Upper second class	<ul style="list-style-type: none"> ● Generally accurate and well-informed ● Reasonably comprehensive coverage ● Well organised and structured ● Addresses the question, some evidence of general reading ● Evaluation of material, good understanding of the material ● Clearly presented.
Lower second class	<ul style="list-style-type: none"> ● Generally accurate, though with some omissions and errors ● An adequate answer to the question, largely based on lecture material and required reading ● A good answer to a related question, but not the one set ● Clear presentation, no real development of arguments
Third class	<ul style="list-style-type: none"> ● Does not answer question directly ● Misses key points of information ● Contains important inaccuracies ● Sparse coverage of material, possibly in note form ● Assertions not supported by evidence
Pass	<ul style="list-style-type: none"> ● Very little appropriate or accurate material ● Cursory coverage of the basic material with numerous errors omissions or irrelevancies ● Loose structure ● Poor or non-existent development of arguments
Borderline (compensatable) fail	<ul style="list-style-type: none"> ● Some appropriate material, but poor coverage ● Evidence that the student has been to one or two lectures or done a bare minimum of reading ● Disorganised or sketchy essays ● Inappropriate material, lack of argument
Non-compensatable fail	<ul style="list-style-type: none"> ● Misunderstanding of basic material ● Complete failure to answer the question set or anything similar to it ● Totally inadequate information ● Incoherent presentation

nearly all of the important ways in which exam answers could vary in quality. Candidate aspects were available from published descriptions of good practice and educational research. The British Psychological Society's *Guidelines for External Examiners*, for example, contains an 'attempted synthesis of marking schemes' (British Psychological Society, 1989). The synthesis 'attempts to capture the main features of performance at different degree levels' and 'give an indication of typical performance at each class...' (p.24). The criteria or typical qualities given for each degree class are reproduced in Table 2. They include coverage of the area, critical evaluation, organisation and structure, addressing the question, development of argument, understanding of material, and clarity of presentation, as well as several other qualities and faults, although not every aspect or quality is represented for each degree class. As the document notes, 'assessment of degree classes is multidimensional, and excellence in one dimension can compensate for poor performance in another'.

In a study where coursework essay markers in psychology were interviewed about the factors they considered important in essay marking, nine factors were mentioned by at least half of the tutors. These were structure, argument, answering the question, wide reading, content, clear expression of ideas, relevant information, understanding, and presentation (Norton, 1990, Table 13). Longhurst and Norton (1997) described five criteria that were considered by psychology tutors to encapsulate the essence of a deep approach to learning. These were: addresses the question throughout the essay, clear organisation with structure appropriate to the question, quality and relevance of argument, depth of understanding of underlying psychological issues, and evaluation of theoretical psychological issues.

A process of departmental discussion and consultation led to the selection of

seven aspects that included most of the main qualities specified in existing guides to good practice. They were: addressing the question, covering the area, understanding the material, evaluating the material, developing arguments, structure and organisation, and clarity of presentation.

The next step was to produce descriptions of the type of performance that was expected for each aspect at each grade. Draft descriptions were produced and a further round of meetings and consultation led to a set of criteria that had the broad support of the whole department. The criteria for Level 1 coursework essays and examination answers is given in Table 3.

We hoped that the criteria incorporated most of the features of good practice from the literature on teaching psychology in higher education and were consistent with more general principles about educational assessment. Quellmalz (1991), for example, recommended six essential characteristics for assessment criteria:

1. Significance (they should specify important components of performance).
2. Fidelity (they should represent standards that would apply appropriately under the conditions that performance takes place).
3. Generalisability (they should apply to a class or type of tasks, and markers should apply the criteria consistently within and between tasks).
4. Developmental appropriateness (they should specify a range of quality levels that are appropriate for the group being assessed).
5. Accessibility (they should communicate clearly and be able to be used by all the participants in the assessment process).
6. Utility (they should communicate information about quality with clear implications for decision-making and improvement).

The development of the criteria did not mean, however, that there was a universal consensus about the desirability of

Table 3. The assessment criteria for coursework essays and examination answers: Level 1.

CORE ASSESSMENT CRITERIA	1st (70–100%)	2:1 (60–69%)	2:2 (50–59%)	3rd (41–49%)	Pass (38–40%)	Fail (25–37%)	Fail (0–24%)
Addresses question asked	Directly, synthesising appropriate material and showing insight into the issues raised by the question	Directly, synthesising appropriate material to provide an answer to the question	Somewhat indirectly, by presenting relevant material and trying to 'link' it to the question; some synthesis of material	Partially, by presenting material to answer part of the question but not all of it; some synthesis of material	At a general level, by presenting material on the topic but not addressing the question; little synthesis of material	Presents some material which could be related to question, but the question is ignored; no synthesis of material	Not at all, answers a different question
Covers the area	Very well, providing a comprehensive account of the material based on extended reading – including current literature	Well, providing accurate accounts of relevant material – clear evidence of reading beyond lecture notes and core texts	Satisfactorily, but some errors and/or omissions in accounts of relevant material – largely based on lecture notes and core texts	Adequately, but some significant errors and/or omissions in accounts of relevant material – no evidence of reading beyond lecture notes and core texts	Superficially, with significant errors and/or omissions in accounts of relevant models, theories, etc. – evidence of bare minimum of required reading	Sketchily, providing partial descriptions of some of the material, but insufficient overall – little evidence of reading/lecture notes or core texts	Does not cover the material, presenting only own ideas and/or irrelevant material – no evidence of reading/lecture notes or core texts
Understanding of material	Depth of understanding of conceptual, theoretical and methodological issues	Good understanding across the breadth of the material and some depth	Good understanding of the core material and some depth	Reasonable understanding of core material, but no depth	A general understanding of the material at a basic level, but no depth	No clear understanding of core material and evident confusion	Basic misunderstanding of core material
Evaluates the material	Insightful critical evaluation of the material and elaboration of alternative perspectives and current controversies	Evaluation includes conceptual/methodological critique and an appreciation of alternative perspectives and current controversies	Some critical evaluation of material and an awareness of alternative perspectives and current controversies	Limited critical evaluation of material	Shows awareness of a critical perspective, but does not elaborate or discuss it	No evaluation of material	No evaluation of material or inappropriate criticisms rendered

Presents and develops arguments	Originality in arguments developed; good use of theory and empirical evidence in debate to present a strong case	Develops own argument; uses theory and empirical evidence to debate the case	Develops mainly derivative arguments; presents supporting evidence	Presents some arguments with supporting evidence, but doesn't develop these arguments	Presents some arguments, but at a superficial level and makes little use of supporting evidence	Presents poor, unsupported arguments	Presents no arguments, or ones which are clearly erroneous
Structures answer and organises material	Clear structure, material organised around the question asked	Clear structure, material organised well	General structure clear, but organisation of material muddled in places	Overall structure unclear, but sufficient 'flow' to the material	Confusing structure, limited organisation of material	Answer unstructured and material disorganised	Material presented as unconnected points, use of note form
Shows clarity and coherence in presentation and expression	Clear expression of ideas and cogent argumentation	Material and arguments presented clearly and coherently	Some minor points of lack of clarity and coherence	Adequate for comprehension, but generally insufficient concern for clarity and coherence	Some material presented unclearly or arguments lacking coherence	Material presented unclearly and arguments not coherent	Material presented unclearly, incoherent/incomprehensible

Note to Table 3: The table sets out the core assessment criteria by degree classification as a guide for students and markers of coursework essays and examination essays at Level 1. It describes typical performance within each degree class, whilst allowing for variation of marks within that class. It is important to remember that assessment is multi-dimensional, so high performance on some dimensions may compensate for poorer performance on others.

representing the assessment criteria in this way. Among the concerns that were raised during the development process were:

- The possibility that using the criteria would increase the time taken by marking.
- The possibility that although the seven aspects were conceptually distinct, it would be impossible to assess one aspect independently of others.
- The fact that it was impossible to agree on a policy for weighting and combining the aspects to determine an overall mark for an essay or examination answer.
- The risk that specifying a rather formulaic approach to marking could potentially infringe academic freedoms in marking judgements.
- The possibility that specifying the criteria in such detail would encourage students to challenge the grades they had been awarded.

The assessment criteria were employed within the department in a number of ways.

1. They were published in course handbooks and course unit handbooks, and drawn to students' attention in course induction sessions, course unit revision sessions and meetings with personal tutors.
2. They were used in workshops and tutorials on key skills that were introduced to the Level 1 programme in a collaborative venture between the psychology department and the university's Learning Development Unit. The aim was to enhance generic and study skills that we hoped would support student achievement in the psychology programme as well as being important in their own right. The workshops and tutorials included sessions on essay writing and examinations, and in some cases there were exercises where students were asked to apply the criteria themselves to specimen essays.
3. They were issued to markers along with coursework assignments and examina-

- tion scripts. Markers were instructed to use them to guide their marking and remind them of the factors they should consider in awarding marks. Because of the difficulty with specifying formulaic methods of weighting and combining marks, markers were not asked to apply them rigidly. Instead they were encouraged to use their own judgements about how good performance on one aspect of the criteria might compensate for poor performance on another, depending on the type of question that had been posed. The criteria were also used to help resolve differences when first and second markers met to agree marks.
4. They were employed in research where a number of volunteers among the staff used the criteria in a much more formal way, rating each examination answer on each aspect of the assessment criteria at the same time as deciding on an overall mark for the answer.

Evaluation of the criteria

The assessment criteria were evaluated in several ways including formal evaluations of departmental practice associated with the criteria, and informal feedback from staff and students about their experiences of the criteria. The evidence is summarised here in relation to the issues that had led to the development of the criteria.

Firstly, by publishing the criteria in course and course unit handbooks, they became part of the documentation that was examined for the quality assurance subject review that took place shortly after the criteria were introduced. In the subject review, the department was commended for the course unit handbooks containing the assessment criteria, and those handbooks were described as models of good practice.

Secondly, student progression and achievement were monitored very closely during the period after the introduction of the criteria. The specific effects of the criteria have been difficult to assess because student

achievement is the focus of several departmental and university initiatives that are much wider in scope, and we cannot claim that the criteria led to quantifiable benefits in themselves. However, the criteria supported several of those wider initiatives and there was evidence that they contributed positively to several areas of teaching and learning.

For example, coursework feedback forms were developed to provide written feedback to students that focused on the criteria. These had several aims, all related to the criteria. Firstly to cue tutors to comment on aspects of the criteria when writing feedback. Secondly to remind students about the criteria and encourage them to use feedback to improve their work. Thirdly to save tutors having to write the same comments repeatedly, by providing frequently occurring comments that could be ticked, while leaving space for other comments about the criteria or any other aspects of the work (see Appendix 1).

Another example of how the criteria contributed to teaching and learning was the Level 1 skills workshops and tutorials. These were evaluated by asking students to complete questionnaires nominating any parts or aspects of the sessions under three headings: 1. things that worked well and should be continued; 2. things that did not work well and should not be continued; and 3. things that were not included but would have been helpful. The comments 'essay writing', 'using sample essays', and 'preparing for exams' were the most frequently nominated items under the first heading. The assessment criteria themselves were mentioned less frequently under the first heading, but were never nominated as parts of the sessions that had not been useful. Comments like 'more information about style/structure for writing psychology essays', 'more detailed information about what markers look for in essays and exam answers', 'be more concrete about what is expected in exams/essays', and

'more focus needed on essay writing' were among the most frequent comments made under the third heading.

That feedback from students indicated that sessions on writing essays and examination answers were valued by students, but that greater use could be made of exercises linked to the assessment criteria. That conclusion was reinforced by focus group research conducted to investigate students' experiences of seminars in the department, which revealed very limited awareness of the assessment criteria. From a pedagogic perspective, therefore, the publication of the criteria appeared to help some students understand what aspects of their work contributed to the marks they were awarded, and how to improve those aspects, but the benefits were not sufficiently widespread among students.

Thirdly, the development and use of the criteria helped to inform the development of policy for double marking. A system of double marking just a sample of examination answers in each examination was adopted for one year on a trial basis, but after operating this system for one academic year, the department decided to return to double marking of all assessments that contribute to students' degree classifications. It was felt that we could not be sufficiently confident about the reliability of marking to abandon the system whereby every answer is looked at by two independent markers. Marking policy continues to be reviewed, and it is possible that sampling or moderation of scripts will be evaluated again in association with further developments of the criteria or other methods to support marking. The main concern about abandoning double marking, however, related to 'cognitive lapses' rather than poor judgement on the part of markers, and it is difficult to envisage criteria that would provide a safeguard against markers' lapses in concentration or attention.

Fourthly, the criteria provided an extremely helpful tool for research on the

psychology of marking judgements. The study involved seven markers across nine units of the psychology programme, who rated a total of 551 examination answers on each aspect of the criteria. The analysis examined the factor structure underlying aspect ratings, the relationships between aspect ratings and marks awarded for each answer, and differences between first and second markers in the aspect ratings that were associated with marks (Elander & Hardman, 2002).

Fifthly, the staff response to the introduction of the criteria was mixed, and many colleagues may have gone on marking in the way they had done previously. Many colleagues had been concerned that using the criteria would add significantly to the time taken for marking, and the initial experience of the markers taking part in the study appeared to support that view. Several markers also found it rather difficult to make separate ratings for each aspect of the criteria. However, the process became much quicker and easier with practice and increasing familiarity with the criteria. It is also possible, though there is no direct evidence, that making aspect ratings helped to guide markers' judgements about what marks to award. The subjective impression was that, having read an answer and been uncertain about the mark, rating specific aspects of the criteria helped to reach a decision, or at least helped markers feel more confident about the marks they awarded.

The existence of the criteria prompted one member of staff to develop a form for markers to record their views about the strengths and weaknesses of examination answers in terms of the criteria. This allows markers to review each answer in terms of the assessment criteria and reflect on the mark they award. It also provides a record of the reasons for the mark awarded, to guide markers when agreeing marks. Rather than having to read the answer again and try to remember why they awarded the mark they did, markers who have used a

form like this can quickly locate the reasons for their different marks and focus on those to agree a mark more quickly and more fairly (see Appendix 2).

Reflections on the psychology of marking

Previous articles about the psychology of marking have tended to adopt a psychometric perspective and focus mainly on evidence about the reliability of marking and the consistency of standards between institutions, disciplines, cohorts and genders of students (e.g. Newstead, 1996). Much less attention has been paid to the cognitive processes involved in marking and the cues that may influence markers' judgements. This may be partly because, with no objective criterion for the quality of essays or examination answers, it is difficult to separate factors that affect the quality of students' work from those that affect markers' judgements about it. There is nevertheless some evidence about the factors that influence markers' judgements and the ways that markers combine information about those factors to arrive at a grade. Judgement analysis provides a useful theoretical framework for those findings.

The marker of a coursework essay or examination answer has to make a global assessment that incorporates a number of more specific aspects, like those represented in assessment criteria. The requirements are similar to those of expert judgement in areas like medical diagnosis, personnel selection or commercial decision making. Einhorn (2000) identified a sequence of tasks that must be performed by the expert judge. They were firstly to identify information or cues from multidimensional stimuli, secondly to measure the amount of the cues, thirdly to cluster those cues into fewer dimensions, and finally to weight and combine the cues to arrive at an overall evaluation. Each of those tasks has a counterpart in marking.

Firstly then, what evidence is there about markers' ability to identify the relevant cues in students' work? Norton (1990) conducted detailed interviews with coursework essay markers in psychology about the things they looked for when marking and what they considered important. Markers nominated 18 different factors between them and there were 'quite wide variations in what criteria tutors thought were important' (Norton, 1990, p.427).

The actual content of essays and the knowledge demonstrated by students was not mentioned by any of the markers in Norton's (1990) study, whereas in a survey of students reported in the same paper, content/knowledge was rated as the second most important factor. Markers' professional expertise could affect what content or knowledge they look for in students' work. 'Envisage that each examiner reminds himself from time to time of what he should be looking for in the answers he is reading, interpolating, as it were, a notional model answer into the sequence of real answers being assessed... Now the pair of examiners assigned to mark the same [answers] will nevertheless have different areas of professional expertise... and will interpolate different model answers as a basis for their judgements.' (Laming, 1990, p.247).

Markers may also detect cues that are not specified in the assessment criteria and are represented only indirectly in the students' work. In one study, markers of psychology coursework essays rated the students' effort, ability and motivation. All three ratings were highly correlated with the grades given to the essays, and multiple regression showed that both effort and ability, as perceived by markers, were significant independent predictors of grades (Norton *et al.*, 1999). Dennis *et al.* (1996) used structural equation modelling to analyse the marks awarded to student projects by supervisors and second markers. They found that some of the variance in the supervisors' marks had sources that did not

appear to have influenced the second marker and may have been related to the supervisors' personal knowledge of the students.

Secondly, how well are markers able to measure the cues or aspects that have been specified as important for assessment? This is a difficult question to answer because of the lack of evidence about how markers' ratings for specific aspects of essays or examination answers compare with objective criteria for those aspects. However, the limited evidence from studies using ratings of specific aspects does not support the reliability of those ratings. In Longhurst and Norton's (1997) research, both tutors and students were asked to rate coursework essays for five criteria. The tutors' and students' ratings for depth of understanding were significantly correlated, but for addressing the question, clear organisation and structure, quality of argument, and evaluation of concepts and evidence, there was no significant correlation between students' and tutors' ratings.

Disagreement between tutors and students does not mean that the tutors' ratings were inaccurate, of course, but professional markers also disagree about specific aspects of students' work. In Newstead and Dennis' (1994) research, 14 external examiners rated psychology examination answers for quality of argument; knowledge displayed; level of understanding; insight, originality and critical evaluation; and answering the question (almost exactly the same aspects as the five employed by Longhurst & Norton, 1997). In analysis of variance of the ratings, there was no significant interaction between scripts and aspects, 'suggesting that markers do not have a common view of where the strengths and weaknesses of each script lie' (Newstead & Dennis, 1994, p.218).

It is also questionable whether aspects that have been identified in advance as distinct attributes of students' work can be measured independently of one another.

Most of the research where markers have rated students' work for specific aspects has found that the ratings were closely correlated with one another. Newstead and Dennis (1994), for example, found quite high correlations between examiners' ratings of five aspects of examination answers. Elander and Hardman (2002) used principal components analysis to examine the ratings made by examination markers for the seven aspects of the criteria described in this article. Aspect ratings made by each of the markers were examined individually, and for five of the seven markers there was just one main component that accounted for up to 88 per cent of the variance.

One implication of this is that markers may overstate the number of separate attributes they are able to take into account when, like the markers interviewed in Norton (1990), they report large numbers of different factors that they consider important in their marking. Alternatively, it is possible that markers are able in principle to make independent assessments of specific aspects of students' work, but that those aspects tend in fact to be closely correlated in the essays themselves. Perhaps only experimental manipulation of specific aspects in artificially prepared essays would reveal whether markers could really assess aspects independently of one another.

Thirdly, can specific aspects be clustered into a smaller number of dimensions? This can certainly be achieved in committee, as it were, when longer lists of candidate aspects are distilled to a smaller number in the way that the criteria described in this article were developed. The five very similar aspects identified by Newstead and Dennis (1994) and by Longhurst and Norton (1997) show that deliberate selection of that kind can converge on the same set of aspects. It is much less clear how easy this is for individual markers to do in a spontaneous way. In Elander and Hardman's (2002) analyses, there was just one marker whose

aspect ratings produced a coherent, interpretable structure with more than one component.

That three-component structure may indicate how the criteria could be simplified in the future. One component comprised the aspects understanding the material, evaluating the material, and developing arguments. The second comprised addressing the question and covering the area. The third comprised structure and organisation and clarity in presentation. Elander and Hardman argued that the first component represented deep learning, the second represented surface learning, and the third represented presentation as opposed to content. It is possible that markers' ratings for those three 'composite' aspects, if the criteria were reorganised in that way, would have greater reliability and validity than more numerous and specific aspects.

Finally, there is the question of how specific aspects are combined and how much weight is attached to each aspect. In analyses of expert judgement in other areas, the combination of cues is regarded as the most difficult part of the process. 'People are good at picking out the right predictor variables and coding them in such a way that they have a conditionally monotone relationship with the criterion. People are bad at integrating information' (Dawes, 1982, p.395). In judgement analysis, the integration of specific information is examined using multiple regression to 'capture' the implicit 'policies' of individual judges by identifying the specific factors that influence their judgements (Cooksey, 1996). Newstead and Dennis (1994) conducted an analysis like that when they used examiners' ratings for five aspects of examination answers as predictor variables in multiple regression with the grade awarded as the dependent variable. In that analysis, all of the aspects except for level of understanding were significantly associated with grades. In Newstead and Dennis's analysis, however, data from 14 examiners were

combined, and it is possible that individual markers employ different policies. Indeed, the technique of 'policy capturing' is designed to provide insights into the judgements of individual experts (or groups of experts making judgements together).

Elander and Harman (2002) conducted separate 'policy capturing' analyses for each individual marker and for groups of markers acting as first and second markers. The rationale for comparing first and second markers was that the first marker is usually the person who taught the material and set the question, and the second marker is usually someone with more general expertise in the area of the assessment. That difference in perspective could lead to different aspects of the work being attended to or different weight being attached to aspects of the work.

These analyses revealed considerable individual differences between markers in the extent to which their overall marks reflected specific aspect ratings. The number of aspects that were independently associated with marks ranged from two to seven, and the proportion of variance in marks accounted for by aspect ratings ranged from 66 per cent to 96 per cent. There were more systematic differences between markers acting as first and second markers. Marks awarded by first markers were predicted by more of the aspect ratings, which accounted for 91 per cent of the variance, whereas those awarded by second markers were predicted largely by the aspect 'covers the area', and aspect ratings accounted for 71 per cent of the variance in marks (Elander & Hardman, 2002). This appeared to support the view that first markers, having taught the material and set the question, were in a better position to award marks that reflected the range of aspects specified in the assessment criteria.

Implications for practice

Developing the assessment criteria supported several other initiatives, like the

coursework feedback form, the skills workshops, and the examination marking form. The fact that the criteria led so readily to other initiatives reflects the key role that assessment plays in education. However, the impact of the criteria and the related initiatives was probably much less that it could have been. Publishing the assessment criteria and incorporating them in generic skills teaching did not impact very widely on students. That was disappointing but was consistent with previous research on students' perceptions of what counts in the marking of their work (e.g. Norton *et al.*, 1996a; 1996b). There was a comparable response among the staff, many of whom were reluctant to change their approach to marking.

The department has recently committed itself to efforts to increase student participation in subject-specific seminars and small group work. We have decided to devote more of the time spent in small group work within the psychology programme to helping students prepare for assessments, and to provide more guidance on what examination and coursework essay questions actually ask students to do. This could provide further opportunities to reinforce and consolidate students' awareness and understanding of the assessment criteria.

The coursework feedback forms link feedback directly to the assessment criteria, provide space for tutors to comment on each aspect of students' work, and provide frequently occurring comments to be ticked where they apply. Like the criteria themselves, however, their impact so far has been limited. The forms were not used uniformly because some markers preferred to write all their comments in the margins of students' work and others simply forgot to use the new form or could not make time to complete one for each essay. New procedures are not quickly or easily established as routine practice. The forms were distributed to markers for them to attach to

students' essays, and a better procedure might be to issue them in advance for students to attach to their work. This would increase students' exposure to the assessment criteria and enable them to use the frequently occurring comments as a brief checklist in the preparation of their work. Those were intended in the first place to save staff time, but their appearance as a standard feature on feedback forms could also help students understand the meaning of the criteria, in line with Norton's (1990) recommendations. The examination marking form is at an earlier stage of development and evaluation, but it could potentially help to focus markers' attention on the criteria and facilitate the process of agreeing marks.

The marking judgement study (Elander & Hardman, 2002) showed that first markers were more likely than second markers to award marks that reflected the range of aspects specified in the criteria, suggesting that greater weight should be given to first markers than second markers. Double marking was retained by the department, but the results of the research may have important implications for how markers go about agreeing marks. The findings could lead to greater awareness among markers of the ways that the perspective of the marker can affect their judgement about the quality of examination answers, and the examination marking form could contribute to a more reflective approach to agreeing marks.

The assessment criteria themselves are the subject of ongoing development. In addition to regular Boards of Studies and Standard Boards where student learning and methods of assessment are reviewed, the department has begun holding annual Teaching and Learning Day events, where staff have the opportunity to discuss issues like the assessment criteria at greater length and in a wider context. For example, we are presently developing proposals for a more categorical marking system. At present each

piece of work is marked out of 100, and the assessment criteria provide little guidance to markers about distinctions within broad bands. The fail grades cover 37 percentage points and the first class grade covers 30 percentage points, and one of the main reasons for considering a categorical system is the difficulty of deciding marks at the lower and upper end of the percentage scale. A categorical marking system would require assessment criteria that described standards for a larger number of narrower grade bands than are given at present, and might involve combining aspects that presently are described independently of one another.

One question that was raised by the study of marking judgements concerned the validity of the aspects of the criteria. Just as there is no gold standard or external criterion for overall marks awarded for an essay or examination answer, there is similarly no external criterion for ratings of specific aspects such as addresses the question or develops arguments. This makes it possible to argue that there are in reality fewer than seven independent aspects of students' essays that can be accurately rated, or that aspect ratings merely reflect the mark awarded. There is a risk of circularity if markers make aspect ratings in line with the mark awarded and those ratings are then found to predict marks, although in Elander and Hardman's study, aspect ratings made by one marker were also used to predict marks awarded by another, independent marker.

The aspects of the criteria could be said to have face validity and content validity, but further research will be needed to establish the criterion validity and construct validity of markers' ratings of specific aspects of the criteria. One approach would be to conduct a sentence-by-sentence content analysis of essays or examination answers, and to relate measures derived from that analysis to aspect ratings made by markers.

Research like that could help to inform the further development of the criteria, by indicating the aspects for which markers can make valid as well as reliable assessments. It might also help to guide the simplification of the criteria by indicating ways in which the aspects should be combined. For example, it could be used to confirm whether the seven aspects could be reduced to the three 'composite' aspects described by Elander and Hardman (2002) that represented deep learning, shallow learning and presentation. It might also be used to investigate 'deep marking' and 'shallow marking' on the part of markers, by identifying ways in which marks awarded and markers' ratings of specific aspects are influenced by more significant and more superficial aspects of students' work.

The possibility of combining certain aspects of the criteria raises an issue about how the purposes of assessment criteria may sometimes conflict. To increase the reliability of marking there may be little value in specifying more than a few key aspects for markers to assess, whereas from a pedagogic point of view there may be advantages in differentiating more aspects of assessment. Markers may not benefit from making separate assessments of several specific aspects of deep learning, for example, but students may benefit from separate explanations of the meanings of aspects such as showing understanding, evaluation, and development of argument.

One implication of judgement analysis in other contexts is that mechanical combinations of specific information have been shown to outperform expert judgement. This means that once the specific cues that experts incorporate in their judgements have been identified, a statistical method for combining those measures was shown to be a better predictor of outcomes than the expert's global judgement. This has been shown to be true for medical diagnosis, parole board decisions, prediction of business failure and

student selection. In all of those areas, research has shown that the judgements made by trained experts are less accurate than statistical combinations of information about specific factors (e.g. Einhorn, 2000; Dawes & Corrigan, 1974; Dawes, 1982; 1994). Findings like that were sometimes disturbing for the experts concerned, but are not really surprising because what the statistical model does is capture the policy of the judge and apply it in a consistent way, eliminating random noise, idiosyncratic influences, or more systematic biases and limitations of judgement.

It is possible that this approach could be applied to marking. If specific aspects of students' work that the examination or assignment set out to assess could be identified and accurately rated, a mechanical combination of those aspects might provide a more accurate measure of the students' performance than a global judgement by an expert marker. There was some evidence in Elander and Hardman's (2002) research that this might be the case, at least for second markers. In that study, a simple model consisting of the sum of the seven aspect ratings made by second markers added significantly to prediction of the first markers' marks.

There is a great deal of work still to be done, however, before we would be in a position to introduce a statistical model to take over from human markers the function of combining specific information to arrive at a mark. Among the requirements would be very precise specification of the aspects of students' work that should contribute to their marks, high confidence in the measurement of those aspects, and an independent criterion of the quality of students' work (independent, that is, of the markers' judgements and the specific aspects). Those developments may still be some way off, but the application of judgement analysis to marking could potentially open the door to methods of assessment that do not require markers to make global judgements about

the quality of students' work. Instead they would assess specific aspects, and those specific measures would be combined using a formula to arrive at an overall mark for the piece of work. This would prevent the contamination of grades by factors that are not part of the assessment criteria, and avoid biases caused by the difficulty of combining different aspects of assessment in an overall, global evaluation.

For the present, one of the most important effects of investigating marking from the perspective of judgement analysis may be to promote reflection among individual members of staff about their roles as markers. Staff often place a great deal of confidence in the judgements they make about awarding marks, sometimes pointing to the length of their experience and their appreciation of the subtle interplay between qualities of the answer they are reading as evidence that their judgements should not be challenged. Applying judgement analysis to marking may increase markers' awareness and appreciation of the ways their marking could be affected by the same kinds of limitations and biases that have been shown to affect expert judges in other areas.

Acknowledgements

This article reflects the work of many individuals among the staff and students in the department of psychology. Many thanks especially to Robin Iwanek and Liz Charman for the roles they played in the development of the assessment criteria, to Karin Marson for the examination marking form, and to the members of staff who took part in the marking judgement study. Thanks also to the editor and the reviewers for their comments on an earlier draft.

References

- Bradley, C. (1984). Sex bias in the evaluation of students. *British Journal of Social Psychology*, 23, 147–163.
- British Psychological Society (in conjunction with the Association of Heads of Psychology Departments). (1989). *Guidelines for external examiners on undergraduate psychology degrees*. Leicester: British Psychological Society.
- Caryl, P.G. (1999). Psychology examiners re-examined: A five-year perspective. *Studies in Higher Education*, 24, 61–74.
- Cooksey, R.W. (1996). *Judgement analysis: Theory, methods and applications*. London: Academic Press.
- Dawes, R.M. (1994). *House of cards: Psychology and psychotherapy built on myth*. London: The Free Press.
- Dawes, R.M. (1982). The robust beauty of improper linear models in decision making. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press, pp.331–407.
- Dawes, R.M. & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95–106.
- Dennis, I., Newstead, S.E. & Wright, D.E. (1993). A new approach to exploring biases in educational assessment. *British Journal of Psychology*, 87, 515–534.
- Dracup, C. (1997). The reliability of marking on a psychology degree. *British Journal of Psychology*, 88, 691–708.
- Einhorn, H.J. (2000). Expert judgement: Some necessary conditions and an example. In T. Connolly, H.R. Arkes & K.R. Hammond (Eds.), *Judgement and decision making: An interdisciplinary reader* (2nd ed.), pp.324–335.
- Elander, J. & Hardman, D. (2002). An application of judgement analysis to examination marking in psychology. *British Journal of Psychology*, 93, 303–328.
- Hall, J., May, S. & Shaw, J. (2001). Widening participation – what causes students to succeed or fail? *Educational Developments*, 2, 5–7.
- Hartog, P. & Rhodes, E.C. (1935). *An examination of examinations*. London: MacMillan.

- Husbands, C.T. (1976). Ideological bias in the marking of examinations: A method of testing for its presence and its implications. *Research in Education*, 15, 17–38.
- Laming, D. (1990). The reliability of a certain university examination compared with the precision of absolute judgements. *Quarterly Journal of Experimental Psychology*, 42A, 239–254.
- Longhurst, N. & Norton, L.S. (1997). Self-assessment in coursework essays. *Studies in Educational Evaluation*, 23, 319–330.
- Marton, F. & Saljo, R. (1976). On qualitative differences in learning. 1. Outcome and process. *British Journal of Educational Psychology*, 46, 4–11.
- Miller, A.H., Imrie, B.W. & Cox, K. (1998). *Student assessment in higher education: A handbook for assessing performance*. London: Cogan Page.
- Murphy, R.L.J. (1978). Reliability of marking in eight GCE examinations. *British Journal of Educational Psychology*, 48, 196–200.
- Murphy, R.L.J. (1982). A further report of investigations into the reliability of marking of GCE examinations. *British Journal of Educational Psychology*, 52, 58–63.
- Myron-Wilson, R. & Smith, P.K. (1998). A matter of degrees. *The Psychologist*, 11, 535–538.
- Newstead, S.E. (1996). The psychology of student assessment. *The Psychologist*, 12, 543–547.
- Newstead, S.E. & Dennis, I. (1990). Blind marking and sex bias in student assessment. *Assessment and Evaluation in Higher Education*, 15, 132–139.
- Newstead, S.E. & Dennis, I. (1994). Examiners examined: the reliability of exam marking in psychology. *The Psychologist*, 7, 216–219.
- Norton, L.S. (1990). Essay writing: What really counts? *Higher Education*, 20, 411–442.
- Norton, L., Brunas-Wagstaff, J. & Lockley, S. (1999). Learning outcomes in the traditional coursework essay: Do students and tutors agree? In C. Rust (Ed.), *Improving student learning: Improving student learning outcomes* (pp.240–248). Oxford: The Oxford Centre for Staff and Learning Development.
- Norton, L.S., Dickins, T.E. & McLaughlin Cook, A.N. (1996a). Coursework assessment: what are tutors really looking for? In G. Gibbs (Ed.), *Improving student learning: Using research to improve student learning* (pp.155–166). Oxford: The Oxford Centre for Staff Development.
- Norton, L.S., Dickins, T.E. & McLaughlin Cook, A.N. (1996b). Rules of the Game in essay writing. *Psychology Teaching Review*, 5, 1–14.
- Partington, J. (1994). Double marking students' work. *Assessment and Evaluation in Higher Education*, 19, 57–60.
- Quellmalz, E.S. (1991). Developing criteria for performance assessments: The missing link. *Applied Measurement in Education*, 4, 319–331.
- Smith, P.K. (1990). The distribution of psychology degree classes in the UK. *The Psychologist*, 3, 147–152.

Correspondence

James Elander

Department of Psychology
London Guildhall University
Calcutta House, Old Castle Street
London E1 7NT.

Tel: 020 7320 1073

E-mail: elander@lgu.ac.uk.

Appendix 1: Formative feedback form: Coursework essay.

This form is intended to help students improve the quality of their coursework. Comments that are ticked or written by the marker indicate areas you could pay attention to in future essays. (The number of circled or written comments is not necessarily related to the overall grade.)

1. Addressing the question

- Focus on the title or question throughout the essay
- Relate the material to the essay title more explicitly
- Use the introduction to explain how you will approach the question
- Good approach to the title or question

5. Presenting and developing arguments

- Try to link each point to what went before
- Try to use the material to develop arguments
- Make clear to the reader how the material relates to the argument you are presenting
- Clear, well developed line of argument

2. Covering the area

- Focus more on psychological theory and research
- Avoid giving too much detail about a limited part of the material
- Include a wider range of material
- Do more reading and studying on the topic
- Good evidence of reading and research

6. Structuring answer and organising material

- Break the material up into shorter paragraphs
- Explain in the introduction how the essay will be structured
- Include a concluding paragraph
- Follow the departmental guidelines for setting out references
- Good clear structure

3. Showing understanding of the material

- Try to make points using your own words
- Explain fully the points you make
- Don't include material that does not relate to the question
- Good understanding of the issues

7. Showing clarity and coherence

- Try to make your handwriting easier to read
- Use a larger font for word processed work
- Pay attention to spelling/grammar in your writing
- Good clear writing style

4. Evaluating the material

- Explain the reasons or the basis for the points you are making
- Make more evaluative points
- Use a conclusion to summarise the most important points
- Good use of critical evaluation

8. Overall or more generally

Appendix 2: Examination marking form.

Script Number:

Question: Mark:

	Very poor	Poor	Acceptable	Good	Excellent
Addressing the question					
Covering the area					
Showing understanding of the material					
Evaluating the material					
Presenting and developing arguments					
Structuring answer & organising material					
Clarity and coherence					

Negative:

Includes irrelevant information YES NO
 Includes wrong information YES NO

Positive:

Relevant, additional info (not covered in lectures) YES NO
 Relevant links to material from other lectures/units YES NO

Question: Mark:

	Very poor	Poor	Acceptable	Good	Excellent
Addressing the question					
Covering the area					
Showing understanding of the material					
Evaluating the material					
Presenting and developing arguments					
Structuring answer & organising material					
Clarity and coherence					

Negative:

Includes irrelevant information YES NO
 Includes wrong information YES NO

Positive:

Relevant, additional info (not covered in lectures) YES NO
 Relevant links to material from other lectures/units YES NO

Question: Mark:

	Very poor	Poor	Acceptable	Good	Excellent
Addressing the question					
Covering the area					
Showing understanding of the material					
Evaluating the material					
Presenting and developing arguments					
Structuring answer & organising material					
Clarity and coherence					

Negative:

Includes irrelevant information YES NO
 Includes wrong information YES NO

Positive:

Relevant, additional info (not covered in lectures) YES NO
 Relevant links to material from other lectures/units YES NO