

Maintaining alignment in management education: The potential for drift in assessment

Ian Herbert, Loughborough University, John Joyce and Trevor Hassall, Sheffield Hallam University

DOI:10.3794/ijme.72.215

Received: May 2007

Revised: December 2007; April 2008

Accepted: April 2008

Abstract

The paper explores the issues involved in maintaining operational alignment between curriculum aims, teaching and the assessment of student learning. Whilst various conceptual frameworks can help to shape learning outcomes that reflect a constructivist approach across an aligned scheme of education there are, nonetheless, opportunities for misalignment to occur with the potential to significantly dilute the aims of the curriculum. A particular focus of the paper is the use of verbs to articulate learning outcomes and how these follow through into assessment mechanisms. The paper argues that drift is likely to occur in all forms of education, although empirical evidence is usually difficult to access due to the confidential nature of assessment processes. In order to illustrate the conjectured concerns, a case study drawing on the published syllabi, examinations and marking schemes from the examinations of a professional accounting body is presented.

Keywords: assessment; constructive alignment; learning objectives; marking processes

Introduction

In extended learning programmes, such as university degrees or those qualifications leading to membership of a professional body, the core aims of the curriculum will be both qualitative and complex, in that they will go beyond simple acquisition and recall of knowledge and will likely be context dependent. For example, the core aims stated in the syllabus of the Chartered Institute of Management Accountants [CIMA] (2005, p. 4) refer to:

Assuring society that those admitted to membership are competent to act as management accountants for entities.... have adequate knowledge, understanding and mastery of the stated body of knowledge and skills... and complement the practical experience and skills development programme.

Underpinning these core aims will be a series of more detailed subject-based learning outcomes that feed into an overall programme of learning and assessment. The extent to which student learning achieves the core aims is dependent on two things. First, the extent to which the design of the learning and assessment programme enables students to internalise the outcomes and construct their own meaning of the knowledge content, what Biggs (1996) refers to as 'constructive alignment'. Second, the extent to which the actual learning activities and assessment tasks reflect the intended learning outcomes, what is referred to here as 'operational alignment'. The significance of the first aspect is that a poorly aligned programme will not achieve its learning aims: students will be taught and tested on areas different to those in the curriculum. In the case of the second aspect, a properly aligned programme might still fail to achieve its aims if alignment is not maintained in the operational mechanism, that is between the designed scheme and the actual learning activities and assessment.

In practice, there is likely to be a natural reduction in alignment due to four primary causes:

Trevor Hassall is Professor of Accounting Education at Sheffield Hallam University. He is the leader of the Accounting and Business Education Research Centre at Sheffield Hallam University. Specific research interests include the development of vocational skills in undergraduate and professional accounting students and the pedagogic use of case studies.

John Joyce is Professor of Management Accounting Education at Sheffield Hallam University. He is an active member of the Accounting and Business Education Research Centre and is the leader of the Professional Portfolio at the University. His research interests include the development of vocational skills in undergraduate and professional accounting students.

Ian Herbert is Lecturer in Accounting and Financial Management at Loughborough University Business School. He is the Academic Co-ordinator for the Loughborough MBA. His research interests include graduate employability, accounting education and the role of finance in shared services.

1. Drift occurring through the actions of individual actors. As each person interprets the intentions of the preceding stages of the education process: core curriculum aims, to syllabus learning outcomes, to teaching schemes, to assessment tasks, to model solutions, to marking schemes and finally, to the marking of students' scripts.
2. Drift occurring over time. As the intentions articulated in the curriculum become forgotten, or else are reinterpreted by new members of faculty. A particular issue in business and management subjects is the need for continual updating of technical and contextual elements as the external environment changes, whilst the master curriculum is updated only periodically.
3. Drift in assessment style. For example, learning outcomes that relate to the affective domain, such as 'demonstrate awareness', might be eroded during the assessment process as tasks and measurements based on the cognitive domain provide for more straightforward test construction, marking expediency and result reliability (Krathwohl, 1956).
4. Drift as a result of operational constraints and managerial imperatives such as 'modernising' staffing and delivery methods. In UK universities, the demands of rising student numbers together with various initiatives such as 'casualisation', privatisation, virtual learning environments, greater flexibility, team teaching, summer semesters etc., might lead to compromises in alignment (e.g. Biggs, 1996). For example, if tutors are assigned to a module just before teaching commences there might be a tendency for them to teach 'what they normally do' rather than tailor their teaching and assessment to the learning outcomes of the specific module in question.

However, it is difficult to access and present evidence of actual drift in any systematic way due to the confidential and sensitive nature of assessment processes, especially in universities. Thus we draw on data in the public domain to illustrate how, in just one aspect of the education process - assessment - and in one dimension of the construction of learning outcomes - between syllabus learning verbs and task requirement verbs - alignment is not always clear cut and an element of drift is apparent.

The paper is structured in six sections:

1. The need for alignment between curriculum aims, learning outcomes, teaching and assessment practices is introduced, followed by the role and nature of conceptual frameworks in achieving alignment.
2. The potential for drift across the various stages is explored within the context of one mechanism of alignment: learning outcome verbs to task requirement verbs.
3. The emerging issues are discussed in the context of professional bodies.
4. A case study based upon a set of examinations, solutions and marking schemes in the public domain is presented to illustrate the potential for drift in practice.
5. A discussion of the findings and some parallels with university degree programmes are noted.
6. Concluding remarks with some suggestions for further research.

The need for alignment

Constructive alignment is a concept developed by Biggs (1996) to represent 'a marriage' between constructivist learning theory, in which the learner is encouraged to construct their own meaning of knowledge from learning activities, and the alignment of instructional activities across curriculum design, teaching and assessment. A key feature of such an approach is the use of learning objectives/outcomes (e.g. Eraut, 1989). In management education especially, outcomes are likely to be expressed in qualitative terms, and the mix of learning and assessment activities will likely become more complex as a programme of study progresses. For example, students might be expected to analyse situated contexts and then suggest a course of action to deal with a particular problem set. Thus care and creativity is required when planning an appropriate mix of tutor-led and student-centred learning activities to ensure that students achieve, and are assessed against, the learning outcomes in the curriculum.

In addition to upfront programme design, operational alignment of the processes of teaching and assessment is necessary to ensure that students actually learn, and are tested on, what they are supposed to learn. The extent to which individual aspects of the instructional and assessment activities are interdependent is highlighted by Ramsden (1992) who argues that past assessments can define the 'actual' curriculum, thus shaping what students actually study. Rather than working from the stated learning outcomes, students often focus their studies on questions and answers from previous diets in an attempt to anticipate what is really likely to be asked (question spotting), and in what manner, in future examinations. Biggs (1999, p. 141) calls this "the backwash effect". Put simply, students will tend to study what they think they will be tested on, although this is only 'bad' learning if the assessment task itself is bad. If the task is good, and reflects a constructively aligned learning scheme, then 'backwash' can be positive. Indeed, assessment tasks properly aligned with the curriculum will help to moderate any drift that might occur in the teaching phase.

A constructively aligned scheme aims to foster deep learning by students, but no matter whether some choose to take a more surface, or strategic, approach, the design of aligned assessment tasks, together with

clear model solutions and marking schemes (for students and markers), is critical if student learning is to reflect the learning outcomes. In team teaching situations especially, a shared understanding of the purpose of assessment and a common conceptual framework of learning is fundamental to maintaining alignment.

Conceptual frameworks

Various conceptual frameworks have been proposed to inform the processes of curriculum design and delivery. These fall into three broad types:

1. Taxonomies which seek to classify the generic elements of learning outcomes as a series of levels but across one dimension (e.g. Bloom *et al.*, 1954).
2. Multidimensional grids often based upon the refinement, or combination, of one or more of the one-dimensional taxonomies (e.g. Krathwohl *et al.*, 2002).
3. Approaches that seek to express learning as a series of generic developmental processes, e.g. the SOLO Taxonomy (Biggs & Collis, 1982).

For overviews of the development of the analysis and classification of learning objectives see Carter, 1985; Imrie, 1995; De Landsheere, 1989; Eraut, 1989; DeMong *et al.*, 1994.

Whatever the framework of choice, a shared basis for articulating learning outcomes at the design stage and linkage of those outcomes to student development and testing across tutors, students and assessors, is essential for the design of an aligned learning structure. Also necessary is a modus operandi that provides for: operational alignment across teaching and assessment process; and consistent recognition of student achievement between students and between cohorts over time. Imrie (1995) notes that a mismatch between curriculum outcomes and the actual behaviour of students could occur without a systematic framework.

The best known and widely used framework is “The Taxonomy of Learning Objectives in the Cognitive Domain” (Bloom *et al.*, 1956) in which learning progresses through six stages: knowledge, comprehension, application, analysis, synthesis and evaluation. These six stages are further analysed into component parts. For example, ‘evaluation’ is comprised of ‘evaluate’ internal data and ‘judge’ external data. The Taxonomy encourages learning outcomes to be expressed as a verb phrase (the cognitive process) and noun phrase (the subject content), for example, ‘explain the convention of prudence in accounting’. The project team led by Bloom also considered two further taxonomic frameworks: the affective domain covering desired behaviours (see Krathwohl *et al.*, 1954) and the psychomotor domain it is the cognitive taxonomy that has become the basis of many educational schemes over time (Imrie, 1995). However, its present ubiquity should not imply that it is without criticism. Indeed, David Krathwohl, a member of the original Taxonomy Working Party, later argued that the cognitive taxonomy had a number of problems, not least the inclusion of ‘comprehension’ rather than ‘understanding’ (Krathwohl, 2002). An amended scheme (Anderson & Krathwohl, 2001) depicts knowledge across four categories which now comprise one dimension of a Taxonomy Table with the other dimension comprising the cognitive processes now labelled, ‘remember, understand, apply, analyze, evaluate, and create’.

Level	Definition	Sample illustrative verbs
Knowledge	Memorize information presented	Define, describe, identify, match, memorize, name, order, recognize, recall
Comprehension	Able to restate in own words	Classify, convert, distinguish, estimate, express, extend, generalize, give examples, infer, predict, recognize, rewrite, restate, translate
Application	Applying knowledge to different or new situations	Apply, change, choose, compute, discover, employ, interpret, manipulate, modify, operate, relate, schedule, show, solve, use, write
Analysis	Breaking a larger problem into its smaller component and noting relationships	Analyze, break down, calculate, categorize, compare, contrast, criticize, differentiate, examine, experiment, identify, infer, model, question, relate, reorganize, revise, set up, summarize, tell, write
Synthesis	Rearranging component ideas into a new whole	Arrange, assemble, collect, combine, construct, create, design, develop, devise, formulate, generate, integrate, manage, organize, plan, propose, rearrange, reconstruct, relate, reorganize, revise, set up, summarize, synthesis, tell, write
Evaluation	Making decisions based on the whole situation	Appraise, argue, assess, choose, compare, conclude, contrast, defend, discriminate, estimate, evaluate, explain, judge, justify, interpret, relate, predict, rate, select, summarize, support, value

Table 1: Bloom’s Cognitive Taxonomy and illustrative requirement verbs
Source: Hall and Johnson, 1987, as cited in Imrie, 1995, p. 178.

Whilst a taxonomic approach to classifying the common elements of learning objectives is intuitively appealing, in practice there is significant potential for overlap between the categories, and the assignment of appropriate requirement verbs to individual learning levels can be problematic to say the least. For example, Table 1 shows the Bloom Taxonomy together with a set of indicative verbs against each stage accredited to Hall and Johnson (1987, as cited in Imrie, 1995). "The Handbook of the Cognitive Domain", in which Bloom *et al.* (1956) set out the Taxonomy and the process of its development, does not in itself have a list of verbs, although it does provide extensive examples of tasks against each level. In the Hall and Johnson scheme, 'explain' appears only in the highest level category - evaluation. The Association of Chartered Certified Accountants [ACCA] do not provide a list of verbs against learning levels although each subject syllabus sets out learning objectives. In categorising the intellectual challenge of individual subjects to learning levels, ACCA choose to group the six levels of Bloom into three pairs. By contrast, CIMA list the six learning levels of Bloom and then explicate a set of verbs against each level, although they prefer their own arrangement of verbs. For example, in the CIMA scheme 'explain' appears under comprehension (CIMA, 2005, p. 5).

A further issue is that a taxonomic approach can be criticised as mechanical and atomistic, leading to an overly reductionist approach to assessment (Hyland, 1994), although he also notes that the use of the Bloom Taxonomy is at least systematic and consistent. To expand on the example of the verb 'explain', this might also be presented as a capstone to other task requirement verbs. For example:

With reference to (a given scenario), *explain* to the Managing Director how the *application* of the prudence convention might affect the valuation of inventory.

Whilst 'explain' is still the primary requirement verb, the student now needs to analyse a business scenario and *apply* subject knowledge to be in a position to explain. The requirement could be further extended to the highest level of the Bloom Taxonomy, evaluation as follows:

With reference to (a given scenario), *explain* to the Managing Director the likely impact on the company's share price of applying the prudence convention to the company's inventory valuation.

This now requires internal *evaluation* and external *judgement*. Note that it is assumed to be good practice for individual task requirements to contain a single requirement verb, although this can lead to a tendency to combine lower verbs into a single higher one.

A further issue is that taxonomies are usually presented as a series of progressive levels and, whilst this may be quite appropriate in terms of the structure of an individual curriculum, things become complicated when a student moves across different programmes of study. For example, the use of 'explain' in secondary education may not be comparable with its use on the final year of a degree, in which case signals from the wider context of the curriculum and the learning situation will need to be assimilated to clarify the intention of the learning outcomes.

Additionally, learning outcomes can harbour other educational dimensions. For example, subject content (knowledge) versus transferable skills, cognitive versus affective skills. Whilst the use of learning verbs and content nouns is helpful, the style of teaching and assessment will also reflect the relative importance educators attach to individual aspects of a range of competing educational dimensions within learning outcomes. Thus a further disconnection between intention and practice may manifest. For example, 'explain something to a hypothetical MD' might also require evidence of specified communication skills, such as a written report/essay/briefing note or, alternatively, a verbal presentation. Within these individual task styles, evidence of further attributes such as: use of English, persuasiveness, reasoning, presentation or structure might be sought. Whilst a good marking scheme will make explicit the weighting of marks attached to each attribute, practical application to student scripts still involves a considerable degree of subjectivity. Notwithstanding these wider tensions, the focus for the empirical exploration of drift in this paper is restricted to the use of learning verbs in the cognitive domain, and specifically the Bloom Taxonomy, as this forms the basis of the educational structure in the case study.

The purpose and practice of assessment

Lines and Gammie (2004) argue that there are three purposes of assessment: to support and thus enhance learning; to provide certification of progress; as a form of accountability to stakeholders in the educational process. More specifically, it provides feedback on performance to individual students, enabling them to monitor their progress and reflect on strengths and weaknesses. Such feedback can be either quantitative (grading) or qualitative (providing guidance on content and omissions). Student achievement feeds forward into the maintenance and improvement of the quality of teaching and learning within a faculty, and outside through inter-institutional benchmarking. The achievement of individual learning targets also provides a source of motivation to individual students and allows them to make comparisons within their peer group (Lambert & Lines, 2000).

In business and management programmes, case studies are often used to create linkages to real-world scenarios. However, achieving an appropriate balance between practical versus theoretical aspects, and subject content (knowledge) versus transferable skills, is challenging. Additionally, there may be tensions between setting tasks that produce higher test reliability, such as objective tests, and tasks that enable assessors to take a more holistic view of the student’s work, thus providing for higher test validity. The latter is more desirable as learning levels progress, and assessment scenarios and requirement tasks become more complex, context dependent, unstructured, uncertain and with problem ambiguity, as is the case in the final stages of degree programmes and professional examinations.

According to Lines and Gammie (2004, p. 4), a good assessment task “will be *valid* in that it will test what it sets out to test, and *reliable* [italics added] if the result will be exactly the same across all occasions, tasks, observations and settings”. Whilst these twin aims are in essence complementary, in practice they can oppose (Lambert & Lines, 2000). Some of the dimensions typically involved in assessment of management education are represented schematically in Figure 1 which shows how typical assessment task dimensions might be depicted across the continuum of reliability versus validity. The levels of the Bloom Taxonomy are also overlaid schematically.

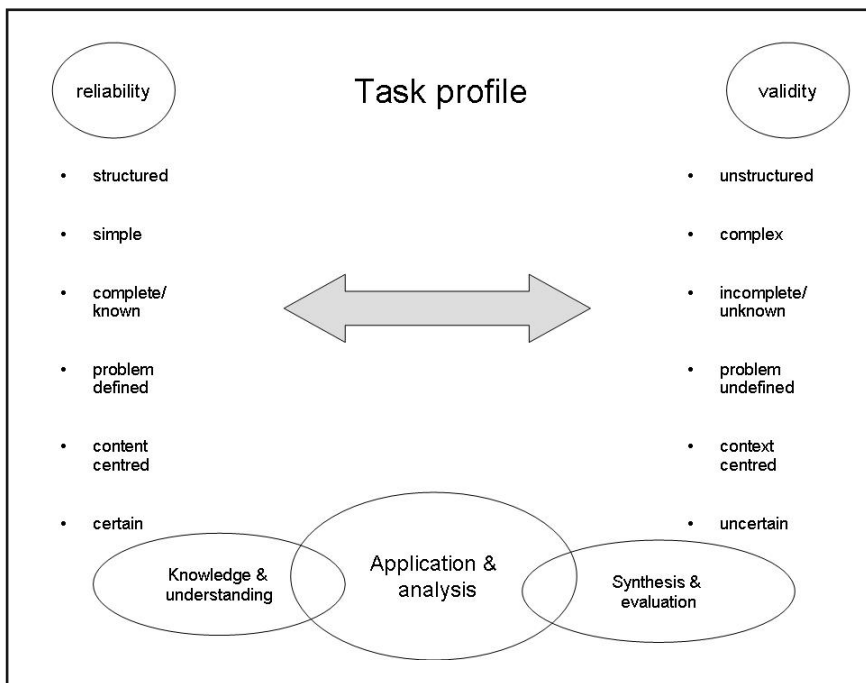


Figure 1: Continuum of requirement tasks

The potential for drift

In the Introduction, four causes of drift were identified, and whilst each of these might apply across the gamut of teaching and assessment activities, it is the manner in which drift might manifest between the intentions and actions of the personnel involved across the various stages in the assessment cycle that this paper is now concerned.

Table 2 shows an indicative scheme for professional education. The assessment cycle includes stages 7 to 15. Lambert and Lines (2000) explain the process of setting and marking external secondary education examinations such as ‘A’ Levels and GNVQs.

Such a professional scheme has parallels with degree programmes in universities, although professional bodies tend to be more formalised in terms of publishing model answers and marking guidelines. Indeed, some universities actively seek to reflect the syllabi and examination styles of professional bodies in their degree programmes, both to facilitate exemptions of professional examinations for their students and to use past professional assessments within their learning materials. This may be seen as an interesting instance of a backwash effect across programmes.

	Education stage	Interpretation by:
1	Survey of stakeholders (essentially employers) - subject knowledge, skills, attitudes	Executive officers, curriculum design team, approved by Council of Members
2	Core aims - notion of a qualified professional person, role within society	Subject curriculum design team, external tutors and students
3	Subject syllabus aims - revised periodically typically, 5-10 years	Subject curriculum design team, external tutors and students
4	Syllabus content by subjects with detailed learning outcomes together with generic and topical knowledge required. Note: might be revised formally each year (e.g. tax or emergent business context - e.g. ENRON)	Subject examiners, external tutors and students
5	Pilot papers - further guidance on content and assessment emphasis	3rd party authors and subsequent new examiners
6	Study systems - where a body publishes or endorses learning materials these will be seen as further guidance on subject content and assessment style/emphasis	External tutors and students
7	Examination paper with task requirements - by diet - typically every six months	Examiner(s) to produce answers, (checked by cold paper sitters), exam panel, external tutors and students
8	Detailed marking scheme - including guidance on alternative approaches	Markers
9	Candidate script - first marking	Markers
10	Second marking by sample	Examiner and marking monitor
11	Statistical analysis	Exam board
12	Exam review panel	Executives and educational experts
13	Examiner's published model answer	External tutors and students
14	Published marking scheme - broad guidance on mark allocation	External tutors and students
15	Post exam guidance to students	External tutors and students
16	Certificate of competence to candidate	Employers to evidence competence of needs in #1 above

Table 2: Typical stages in curriculum design and assessment processes

Ensuring operational alignment across assessment processes over time is likely to be problematic, especially if a number of different individuals are involved, or if there are changes in personnel. When designing an assessment, an examiner may not interpret the learning outcomes exactly as intended by the syllabus design team and subsequently markers may not interpret or apply the examiner's marking scheme exactly as the examiner intended. For example, the potential confusion over the otherwise innocuous verb 'explain'. Whilst any process that involves interaction between human actors is likely to produce differences between intention, interpretation and action, we argue that in the case of assessment such differences will tend to be both progressive and cumulative, hence the term 'drift'.

Price and Rust (1999) found that whilst an assessment grid based upon a common educational framework should hold considerable benefits, their study conducted within a UK business school found that achieving consistent shared understanding across staff was extremely difficult in practice. Literature on the efficacy of assessment procedures is sparse in comparison to other aspects of education, but in an experimental exercise that involved remarking of portfolios in universities, Baume *et al.* (2004) found that differences between the first and second markings occurred. Whilst various factors might be at play in the experiment they suggested that:

What the [task] requirement actually meant was not as transparent as the course designers thought it was. (p. 456)

In a similar vein, Hornby (2003) concluded from a survey of staff in a business school that:

the standards to be applied to various pieces of work were not acquired in a systematic way or more generally through training. (p. 16)

Removing the need for all but the most straightforward judgements by markers can lead to meaningless assessment. Moreover, in attempting to increase reliability in the marking process through the imposition of more tightly specified marking schemes whilst reducing the scope for variation between markers through

atomising the task requirements within tests, the potential for drift might, paradoxically, be exacerbated. This is because a reductionist approach can lead to a defensive culture whereby markers award marks merely for evidence of knowledge (such as key words and phrases) and thus dilute curriculum aims that might seek to develop higher level cognitive skills. Moreover, removing the contextual information might allow candidates to demonstrate that they can apply the higher level skills expected of proficient practitioners rather than competent technicians, thereby impairing the validity of the assessment. Krathwohl *et al.* (1956) highlight the possibility of learning outcomes based on the affective domain (e.g. willingness, awareness) being eroded and being replaced by cognitive-based assessment tasks, which are generally easier to mark consistently.

A further instance of the potential for ambiguity, which is often highlighted in the review of marginal scripts, is the practice of positive marking. For example, one marker might argue that a student does not deserve to pass as there are weaknesses in their overall approach and thus insufficient evidence of subject mastery. Such a view might typically appeal to the core curriculum aims and the consequences for employers. Alternatively, another marker might argue that the same student has accumulated sufficient marks, based upon evidence of technical criteria against the marking scheme, and as such cannot be failed. A specific example of this dichotomy occurs in accountancy when a student includes, say, depreciation (a non-cash expense) as an outflow in a cash budget. Under a positive marking regime the error would simply be ignored, but it is easy to appreciate the concern of a future employer who might rely on the institutional certification of that student as evidence of competence in line with the core aims. In general, positive marking increases marking reliability (and operational expediency) as there is no scope for markers to express opinions as to the degree of 'wrongness' in student scripts, but it is difficult to claim that the practice improves validity if the notion of competence is embedded in the core aims. This depreciation/cash issue is a relatively straightforward exemplar compared to judgements about responses to task requirement verbs. For example, when does a list of bullet points become, or fail to become, an 'explanation', and should it be seen as critical to competence around the pass/fail mark?

Setting and marking assessments involves dealing with numerous opportunities for ambiguity, not least in respect of learning verbs. Quality assurance is usually based upon an assessor's qualified status and ability to have an inherent notion of what is an appropriate standard for a prospective professional person or university graduate (Eraut and Cole, 1993). The phrase 'I know a pass script when I see one' (Price, 2006) is common in many spheres of education, but it is a moot point whether one should look to existing staff to anchor alignment or new staff to refer back to the original syllabus when existing staff have drifted. Hornby (2003) suggests that new staff might be 'infected' by existing staff such that new staff adopt inappropriate (unaligned) marking practice. Barrie (2007) argues:

far from a shared understanding of such attributes as the core outcomes, academics hold a variety of disparate understandings of the nature of generic attributes and their place amongst the outcomes of a university education. (p. 439)

The context of professional bodies

The importance of constructing and maintaining alignment between the core aims and assessment is magnified when an institution is an examining body only. Professional bodies tend not to provide teaching and so tutors working for third party colleges are likely to reinforce rather than correct any misalignment that is manifest in published examination questions, answers and marking schemes, as tutors interpret the perceived 'real' requirements on behalf of their students (the backwash effect). This may result in students successfully learning the wrong things, at least in so far as their learning may not fully reflect the original intentions of the curriculum.

In the case of professional bodies, script marking is especially pressured due to a high number of students across a global reach and a limited time window in which sufficient experienced and qualified markers are available to mark literally thousands of scripts. As an example of the complexity, and thus the potential for drift that might arise, it is typical for a team of around 20 markers, from a range of institutional backgrounds, to meet for a whole day to achieve a common approach.

Firstly, agreement is required on the meaning of marking scheme in terms of how marks are to be awarded for: different levels of technical achievement in line with the learning outcome verbs; other educational dimensions and the validity of alternative technical answers (beyond the examiner's model solution). This often involves debate concerning what it is reasonable to expect a student to produce in the time allowed and even what the examiner really intended to ask. Whilst examiners are usually present at such meetings, this does not always dissuade lengthy and sometimes heated debate.

Secondly, when a common understanding of the marking scheme has been achieved it is then necessary to evidence a level of marking consistency amongst the team. To achieve this before and during the markers'

meeting a number of sample scripts will be marked and then discussed. It is sometimes debatable whether the marking scheme produced after the markers meeting is more closely aligned with the original learning outcomes or, alternatively, the collective product of students' learning, as evidenced on the scripts. Inevitably, there have to be compromises in agreeing a workable range of alternatives for the sake of expediency; too much elaboration of the marking scheme and discretion for the markers (higher validity) may result in inconsistency in marking (lower reliability). As markers meetings occur towards the end of the educational process, they provide an opportunity to correct previous lapses in alignment, although this may disadvantage students whose learning has been tailored to the original learning objectives.

In order to increase the reliability of assessment processes various statistical analyses can be employed. However, such procedures can only seek to ensure that the process of marking has been consistent between members of the marking team. In other words within reason natural variability between markers can be accommodated, but structural drift that occurs when assessments have not been properly designed to test the learning outcomes in the syllabus is more problematic.

To summarise, a number of issues might result in drift in operational alignment between curriculum intentions and student assessment, and it has been suggested that within assessment processes these issues are particularly acute for professional bodies. The next section scrutinises the published assessment materials of a global professional body. Requirement verbs in examinations questions and marking schemes are compared to the stated learning outcomes of the curriculum to demonstrate how ambiguity, and thus the potential for drift, can occur in practice.

Case study

The case study focuses on the assessment package of the Association of Chartered Certified Accountants [ACCA]. The study was based on the 2005 Syllabus Guide (which gives details of the intended learning aims and subject content), examination papers for June 2005 and the associated answer papers including mark allocation guides (ACCA, 2005a). These documents are all in the public domain and indeed are purchased by students as study aids. The academic levels exhibited in this material were mapped by the use of Bloom's cognitive taxonomy (1956) and are shown in Table 1 along with sample illustrative verbs and definitions.

The ACCA assessment scheme contains 16 papers across three levels. Parts 1 and 2 contain three and six papers respectively. All of the nine papers in the first two levels are compulsory. Part 3 consists of seven papers of which five have to be taken: three are compulsory but the candidate then has the choice of two from the remaining four papers. In order to fully benchmark the academic levels, all of the seven papers in Part 3 were reviewed.

The Syllabus Guide has a separate introductory section for each Part. These state the objectives of the papers and the skills to be tested in each of the levels of the syllabus. These objectives and skills were taken as the starting point for the mapping of the academic levels in the overall assessment scheme for this case study.

Syllabus Review

The desired level in Part 1 is perhaps best indicated by the statement in the syllabus objectives: "The knowledge and skills will be tested separately by subject and will be limited to straight-forward examples of application" (ACCA, 2005b). This is consistent with the verbs used in the statement of the skills to be tested such as 'identify', 'retrieve', 'use' and 'apply'. Consequently, it appears that the intention of Part 1 is to focus on the first three levels of the taxonomy (knowledge, comprehension and application) and culminate in the ability to apply knowledge in different or new situations.

Part 2 develops incrementally from Part 1 by aiming to test "application of the theory in the context of recognisable problems and conceptual understanding" (ACCA, 2005c). It also indicates a desire to 'develop candidates' ability to criticise current practices'. The verbs now stated in the skills to be tested narrative are 'analyse', 'evaluate', 'apply', 'identify', 'define', 'rank', 'interpret' and 'criticise'. These verbs map predominantly on to the fourth and sixth levels of the taxonomy (analysis and evaluation).

In Part 3, the objectives are more contextual and include reference to professional competence. An indication of academic level is given by the statement in the Guide that: "Examinations at this stage will be set at a level equivalent to a UK masters degree" (ACCA, 2005d). The verbs used in the skills to be tested are 'integrate', 'analyse', 'interpret', 'diagnose', 'formulate', 'adapt' and 'communicate'. There is also a reference to "draw on knowledge across all earlier papers studied" and "exercise judgement drawing on technical political and commercial awareness in developing and evaluating alternatives and in proposing solutions" (ACCA, 2005d). These verbs map predominately onto the two highest levels of the taxonomy (synthesis and evaluation) and culminate in the ability to make decisions based on the whole situation.

The three parts of the examinations leading to the ACCA qualification therefore move candidates through Bloom's cognitive taxonomy, with each successive part promoting the development of candidates' abilities in incremental but overlapping steps.

Review of Examination Papers

The next stage of the case study was to review the exam papers for June 2005 for each of the 16 subjects. A detailed analysis of the verbs used and corresponding mark allocations in each one of the exam papers is shown in Appendix 1.

Part 1

The syllabus for Part 1 is primarily concerned with the Bloom Taxonomy's categories of knowledge, comprehension and application. The 'aim' stated for each of the syllabi for the three individual papers that form Part 1 begins with the phrase: 'To develop a knowledge and understanding of the application'. This is again consistent with the rationale for Part 1. Also, the individual syllabi each have stated objectives. The verbs used in these objectives are broadly consistent with the desired academic level. However there is a mismatch in that two of the verbs used ('explain' and 'appraise') appear in the highest category of the taxonomy (i.e. 'evaluation').

The analysis revealed that there appears to be overall alignment between the stated objectives of this level within the assessment structure of ACCA, the individual syllabi, the question papers, the answers and the marking schemes. However, there is a lack of clarity about the use of several verbs (explain, advise and distinguish) which, whilst indicative of the highest cognitive level - evaluation - are being used in this context (confirmed by the answers and marking scheme) to search out basic knowledge, comprehension and application. As noted earlier, CIMA categorise 'explain' under 'comprehension': much lower in the Bloom Taxonomy than ACCA (CIMA, 2005, p. 5).

Part 2

The stated objective of Part 2 is the 'application of the theory in the context of recognisable problems and conceptual understanding' and to 'develop candidates' ability to criticise current practices'. However, the verbs used in the question papers (see Appendix 1) appear to indicate that the cognitive level has not risen to that indicated in the stated objectives for Part 2 of the assessment structure. The use of the verb 'explain' is again problematical. Also, because of the differing requirements of questions in optional sections of papers, students can choose to minimise the cognitive level they attempt.

Part 3

The stated objective of Part 3 was indicated by the verbs used in the skills to be tested. These were 'integrate', 'analyse', 'interpret', 'diagnose', 'formulate', 'adapt' and 'communicate'. There is little direct evidence of the verbs in the learning objectives being used in the requirements of the questions on the June 2005 papers at this level (see Appendix 1). There is a range across the Taxonomy, with illustrative verbs at all levels but perhaps predominately at the levels of application and analysis. There is ambiguity in the use of certain verbs, especially 'explain'. Moreover, certain verbs such as 'explain' are used throughout all levels of the educational programme. One noteworthy complication is the co-existence, within the final level, of subjects that have progressed throughout the course of the programme to culminate in a more abstract requirements at a higher cognitive level, such as management accounting, and subjects appearing for the first time in the later stages such as taxation, where even at the final level there is a significant requirement to handle extended technical computations. Although these calculations are technically demanding in terms of knowledge required, they involve lower level cognitive skills.

Discussion

The case study has highlighted instances of ambiguity and apparent drift in the use of learning verbs across the various stages of an examination regime. Lines and Gammie (2004) depicted the relationship between validity and reliability on a matrix similar to Figure 2 and suggested that both validity and reliability could be increased through examiner and marker guidance. We further suggest that if such improvements in assessment design enable assessment systems to move, hypothetically from say point A to point B (as shown in Figure 2), then without a robust consensus of the learning outcomes across the educational team, rooted in an appropriate conceptual framework, the assessment tasks will tend to move towards an emphasis on atomised tasks at lower cognitive levels. This will result in drift towards the more limited aim of ensuring reliability, point C. Notwithstanding the quality of assessment design, drift to point D might also occur in the chain of marking procedures, as both the validity and reliability of the assessment is compromised by marking processes and the inclination of individual markers to 'play safe' by favouring evidence of concrete technical knowledge over the more abstract aspirations of the core aims and higher level learning outcomes.

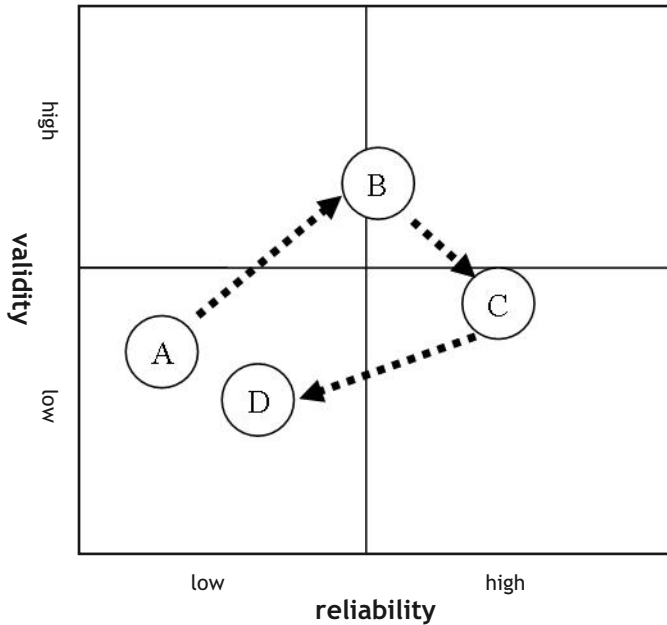


Figure 2: Potential for drift in assessment procedures
Adapted from Lines and Gammie (2004)

We suggest that not only is the potential for drift indicated in the case study likely to afflict university degree programmes but, as professional bodies are able to devote a much greater level of resources to setting and marking examination papers, such drift is likely to be greater in universities. Professional bodies specialise in assessment: examination papers take around a year to set and go through various stages of technical and holistic moderation, often including a ‘cold sit’ of the final paper by an independent person. In comparison, universities have better scope for constructive alignment when one person teaches and assesses a subject, but there are perhaps greater dangers of operational drift, especially if there are operational constraints which impact on staff training and preparation time. In either system, reliance on statistical measures is not effective if the drift has occurred in teaching or assessment before marking commences.

Conclusion

The research has questioned the ability of accounting education programmes to consistently design and operate aligned assessment regimes. The case study demonstrated actual instances of drift across learning outcomes, examinations and marking schemes. Educators need to think carefully about assessment design within a consistently applied conceptual framework and place more emphasis on examiner and marker guidance. Moreover, attempts to improve reliability, save time or save cost during the marking process will compromise the curriculum outcomes. We believe that issues arising in the setting and marking of the professional examinations in the case study will resonate with tutors in higher education who typically do not have the time and resources to dedicate to setting assessments in comparison to large professional bodies.

Whilst a number of issues have been suggested which might, in practice, create drift in assessment, it is acknowledged that the research is limited to one aspect of the overall process - comparing the usage of requirement verbs in learning outcomes to examination papers, answers and marking schemes. Further empirical research into other stages of the assessment process, particularly the manner in which markers interpret examiners’ answer schemes and script marking processes, would be useful. Other fruitful lines of enquiry might focus on the tensions between other dimensions of learning outcomes such as subject knowledge versus transferable skills or the cognitive versus affective domains.

Acknowledgements

The authors are grateful to Carol Newbold, Laurie McAulay and two anonymous referees for comments on earlier drafts of this paper.

References

ACCA (2005a). ACCA Professional Scheme Archive. Retrieved December 5, 2007, from http://www.accaglobal.com/students/study_exams/qualifications/acca_choose/acca_professional_scheme/

- ACCA (2005b). *Professional Scheme Part 1*. Retrieved December 5, 2007, from http://www.accaglobal.com/students/study_exams/qualifications/acca_choose/acca/professional_scheme/part1/
- ACCA (2005c). *Professional Scheme Part 2*. Retrieved December 5, 2007, from http://www.accaglobal.com/students/study_exams/qualifications/acca_choose/acca/professional_scheme/part2/
- ACCA (2005d). *Professional Scheme Part 3*. Retrieved December 5, 2007, from http://www.accaglobal.com/students/study_exams/qualifications/acca_choose/acca/professional_scheme/part3/
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Barrie, S. C. (2007). A conceptual framework for the teaching and learning of generic graduate attributes. *Studies in Higher Education*, 32, 439-458. doi:10.1080/03075070701476100
- Baume, D., Yorke, M., & Coffey, M. (2004). What is happening when we assess and how can we use our understanding of this to improve assessment? *Assessment & Evaluation in Higher Education*, 29, 451-477. doi:10.1080/02602930310001689037
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: the SOLO taxonomy (structure of the observed learning outcome)*. Educational Psychology. New York: Academic Press.
- Biggs, J. B. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364. doi:10.1007/BF00138871
- Biggs, J. B. (1999). *Teaching for quality learning at university: What the student does*. Buckingham, England: Society for Research into Higher Education.
- Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: David McKay.
- Carter, R. (1985). A taxonomy of objectives for professional education. *Studies in Higher Education*, 10, 135-149. doi:10.1080/03075078512331378559
- Chartered Institute of Management Accountants (2005). *Qualification and syllabus structure*. London: CIMA.
- De Landsheere, V. (1989). Taxonomies of objectives. In M. Eraut (Ed.), *The International encyclopedia of educational technology* (pp. 352-362). Oxford, England: Pergammon Press.
- DeMong, R. F., Lindgren, J. H. Jr., & Perry, S. E. (1994). Designing an assessment program for accounting. *Issues in Accounting Education*, 9, 11-27.
- Eraut, M. (Ed.) (1989). *The international encyclopedia of educational technology* (pp. 341-352). London: Pergammon Press.
- Eraut, M. & Cole, G. (1993). Assessment of competence in higher level occupations. *Competence and Assessment*, 21, 10-14.
- Hornby, W. (2003). Maintaining standards? An analysis of assessment practices in a business school. *The International Journal of Management Education*, 3(3), 3-17
- Hyland, T. (1994). *Competence, education and NVQs: Dissenting perspectives*. London: Cassell.
- Imrie, B. W. (1995). Assessment for Learning: quality and taxonomies. *Assessment & Evaluation in Higher Education*, 20, 175-189. doi:10.1080/02602939508565719
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41, 212-218. doi:10.1207/s15430421tip4104_2
- Lambert, D., & Lines, D. (2000). *Understanding assessment: Purposes, perceptions, practice*. Key issues in teaching and learning. London: Routledge/Falmer.
- Lines, D., & Gammie, E. (2004). *Assessment methods report*. New York: IFAC.
- Price, M., & Rust, C. (1999). The experience of introducing a common criteria assessment grid across an academic department. *Quality in Higher Education*, 5, 133-144. doi:10.1080/1353832990050204
- Price, M. (2006). *I know a 2.1 when I see it. Challenging perspectives in assessment*. Retrieved November 15, 2007, from <http://stadium.open.ac.uk/stadia/preview.php?s=39&whichevent=736>
- Ramsden, P. (1992). *Learning to teach in higher education*. London: Routledge.