DESIGNING CRITERION-REFERENCED ASSESSMENT

Kelley Burton
Faculty of Law
Queensland University of Technology, Australia
kj.burton@qut.edu.au

Abstract

Criterion-referenced assessment arguably results in greater reliability, validity and transparency than norm-referenced assessment. This article examines this assertion with reference to an example from a second year undergraduate law unit at the Queensland University of Technology, LWB236 Real Property A. When designing criterion-referenced assessment sheets for a course, an incremental approach should be taken to reflect that skills are progressively developed throughout the course. The incremental development and assessment of skills has been strongly supported by the literature as opposed to developing and assessing skills in a one-off manner. This article discusses how skills may be developed and assessed across three levels of a degree (or course). It builds on the existing research by recommending a model for taking an incremental approach to implementing criterion-referenced assessment across the three levels of a course. This recommended model is relevant to the designers of criterion-referenced assessment in all disciplines.

Norm-referenced and criterion-referenced assessment

The two main approaches to assessment are norm-referenced assessment and criterion-referenced assessment (Dunn, Morgan, O'Reilly & Parry, 2004, p. 22). Norm-referenced assessment "remains the dominant…[approach] within higher education and 'naturally' preferred by most markers" (Rust, Price & O'Donovan, 2003, p. 156). Biggs (1999) acknowledges that one of the main reasons for implementing norm-referenced assessment is for administrative convenience, but asserts that there is "no *educational* justification for grading on a curve" (p. 69).

Norm-referenced assessment ranks a student's performance against their peers in a particular cohort. A marker using a norm-referenced approach to assessment grades each student in a cohort "according to a preconceived notion of how the distribution of grades will turn out" (Dunn, Morgan, O'Reilly & Parry, 2004, p. 22). Fitting grades into such a pre-determined distribution is commonly referred to as a 'bell curve' (Centre for the Study of Higher Education, 2002). However, the pre-determined distribution for a unit is unlikely to represent a perfect bell as the assessment policy is unlikely to specify that a certain percentage of students must fail the unit.

Norm-referenced assessment has been criticised because it traditionally focussed on assessing content and the recent trend is to assess skills as well as content (Bond, 1996, p. 2). It suits units where there is an objective right or wrong answer (Dunn, Morgan, O'Reilly & Parry, 2004, p. 23). However, it is also an effective approach in assessing skills and content. In such a case, the assessor distributes the grades across a bell curve based on a "subjective judgement about performance that is backed by professional expertise rather than objectivity" (Dunn, Morgan, O'Reilly & Parry, 2004, p. 23).

In contrast to norm-referenced assessment, a criterion-referenced approach to assessment occurs when the assessor measures the performance of the students against a pre-set criteria (Le Brun & Johnstone, 1994, p. 185). The criteria serve the following purposes: "to describe, clarify, and

communicate requirements; to contextualise and fine-tune expectations; to facilitate the substantiation of judgments; to safeguard against subjectivity and bias; to ensure fairness; and to provide a defensible framework for assessing" (Scarino, 2005, p. 9).

Despite the distinct definitions for norm-referenced and criterion-referenced assessment above, the overlap between these two approaches to assessment is often overlooked. The purity of criterion-referenced assessment is diluted when markers using this approach to assessment are influenced by the performance of students from previous years and other students in the same cohort (Johnstone, Patterson & Rubenstein, 1998, p. 41). In such a case, criterion-referenced assessment is diluted because the assessor is influenced by the norm-referenced assessment approach.

It is recommended that an assessor using criterion-referenced assessment should monitor the spread of grades. This monitoring process is not to dilute criterion-referenced assessment, but it is done to gain a greater understanding of why criterion-referenced assessment led to a different outcome to norm-referenced assessment. For example, if criterion-referenced assessment resulted in the grades being bunched at the extremes. There may be several reasons that explain this such as there may have been a fault in the setting of the criteria or performance standards, the assessment task may not have had an appropriate degree of complexity, the markers may not have had a shared understanding of the criteria and performance standards with the students and the particular cohort may have been exceptionally better or worse than the cohorts in previous years. The assessor should reflect on the reasons and take them into account when setting assessment in subsequent years.

Similarly, an assessor using norm-referenced assessment should monitor and understand the reasons underlying the spread of the raw scores. The possible underlying reasons would be analogous to the ones stated above for criterion-referenced assessment.

Newble and Cannon (1989) suggest that implementing an assessment regime more oriented towards criterion-referenced assessment improves the validity of assessment (p. 99).

Validity

The validity of an assessment task is the extent to which it accurately measures the "desired learning outcomes" (Queensland University of Technology, 2003). In the context of a unit, these desired learning outcomes may also be referred to as unit objectives. Assessment is valid when it "measures what it is supposed to measure" (Dunn, Morgan, O'Reilly & Parry, 2004, p. 32).

The validity of an assessment task using a norm-referenced approach to assessment cannot be determined by analysing the pre-determined distribution of marks because it is possible that the student who received the top score did not achieve the unit objectives. The raw scores need to be analysed. The validity for norm-referenced assessment depends on how the marker allocates the marks to calculate the raw scores, for example, on the basis of prescriptive marking guidelines where there is little room for professional discretion and judgment or on the basis of professional discretion and judgment where the marking guidelines are not specific. The literature suggests that the allocation of marks in norm-referenced assessment is determined by how well it discriminates among students (Bond, 1996, p. 2). The same comment may be said about criterion-referenced assessment because it requires the assessor, when setting the criteria, to anticipate the strengths and weaknesses in student attempts at an item of assessment. However, criterion-referenced assessment is different to norm-referenced assessment because it specifically indicates the alignment between the assessment criteria and the unit objectives. Thus, criterion-referenced assessment arguably achieves greater validity.

An example of the alignment between the assessment criteria and the unit objectives on a criterion-referenced assessment sheet appears in Table 1 (below). This is an extract from the *LWB236 Real Property A* criterion-referenced assessment sheet, which was designed by the teaching team for a drafting exercise, file note and letter to a client. Assessment criteria are presented in the first column. A "criterion" is a "distinguishing property or characteristic of any

thing, by which its quality can be judged or estimated, or by which a decision or classification may be made" (Scarino, 2005, p. 9). The criteria set out in Table 1 involve instrument drafting skills, which correlates with unit objective 10. The unit objectives are set out in the unit's study guide. Unit objective 10 for this unit is:

> (10) draft specified instruments under the *Land Title Act 1994* (Qld) using appropriate drafting techniques supported by research and a written explanation to effectively communicate the legal and practical requirements.

For this unit, there are other criterions on the criterion-referenced assessment sheet, which are not extracted below in Table 1 that also relate to unit objective 10.

The second to fourth column in Table 1 are performance standards. A "standard" is defined as "a definite level of excellence or attainment, or a definite degree of any quality viewed as a prescribed object of endeavour or as the recgonised measure of what is adequate for some purpose, so established by authority, custom, or consensus" (Scarino, 2005, p. 9).

The total weighting for the drafting exercise, file note and letter to client is 20 per cent of the unit. Instrument drafting skills are weighted at five per cent of the unit. The weightings attached to the criteria depend on the importance of the unit objectives as well as the degree of professional marking judgment required, for example, in this unit, the content of the client letter is more important and requires less professional marking judgment than the form of the client letter.

Criteria	Excellent	Good	Satisfactory	Poor	Marks
Instrument Drafting Skills – Unit objective 10				**Max 5**	
• Understanding of forms' content and purpose, • Ability to transcribe information correctly • Compliance with the relevant law	Drafting exhibits all of the following - • an excellent understanding of forms' content and purpose, • no obvious technical drafting errors or omissions; • complies with relevant law **4.5-5**	Drafting exhibits all of the following - • a good understanding of forms' content and purpose; • at least one relatively minor technical drafting error or omission • complies with relevant law; **3.5-4**	Drafting exhibits all of the following - • genuine attempt to understand forms' content and purpose • contains one or more significant technical error or omission • generally complies with law **3-2.5**	Drafting exhibits one or more of the following - • limited or no demonstrated understanding of forms' content and purpose; • a number of significant technical drafting errors or omissions • fails to comply with the relevant law **<2.5**	/5

Table 1: Extract from the *LWB236 Real Property A* – Drafting exercise, file note and client letter assessment criteria and feedback sheet

Students are advised whether they have met the unit objectives by a tick in the appropriate performance standard for each criterion, individual feedback on the assessment item and additional personalised comments at the bottom of the criterion-referenced assessment sheet. This personalised feedback is also supplemented with meaningful generic feedback on the online teaching site. The quantity and quality of this feedback goes beyond simply providing a mark or

grade. Simply providing a mark or grade has been described as "cheating students" and "unprofessional teaching behaviour" (Ramsden, 1992, p. 193).

Primarily, the feedback serves to inform the students. However, the assessors should also use this feedback to inform their future teaching and learning approaches in the unit. Effective feedback should identify the strengths and weaknesses of an individual student, indicate ways of improving, be constructive, enhance student motivation and be timely (Crooks, 1988, p. 15). If several markers are involved in marking the assessment in a unit, the markers must provide consistent feedback to students to confirm that reliability is not compromised.

Reliability

The notion of reliability loosely equates to consistency in marking. An assessment task is unreliable if different markers award different grades to the same student attempt at the assessment or if one marker awards a different grade to the same student attempt at the assessment at a later point in time (Le Brun & Johnstone, 1994, p. 184).

The reliability of norm-referenced assessment depends on how the raw scores are calculated, that is, on the basis of using a prescriptive marking guidelines where there is little room for professional discretion and judgment or on the basis of professional discretion and judgment where the marking guidelines are not specific. It is unreliable for the purposes of comparing cohorts in different years because it assumes the knowledge and skills of cohorts from year to year are consistent. This means it "disguises absolute performance" (Dunn, Morgan, O'Reilly & Parry, 2004, p. 23). It does not acknowledge that a cohort in one year may be better than the cohort in another year because it spreads the raw scores across a bell curve based on predetermined cutoffs for the grades. For example, the top five percent of the students may receive a high distinction irrespective of the quality of their attempts at the assessment. Using the norm-referenced approach to assessment means that a particular student may pass in one year, but fail in another year. The Centre for the Study of Higher Education recognises that norm-referenced assessment is likely to be unfairer to smaller cohorts because it exaggerates the difference between the students, but alternatively norm-referenced assessment "might artificially compress the range of difference that actually exists" (Centre for the Study of Higher Education, 2002).

In contrast, criterion-referenced assessment establishes performance standards for each criterion. It is very prescriptive in nature. In the exemplar in Table 1 (above), the performance standards are presented across the page, that is, "excellent", "good", "satisfactory" and "poor". Within the School of Law at the Queensland University of Technology it is common to use the term "excellent" to mean a mark within the range of 85-100 per cent. The term "good" usually equates to a mark within the range of 65-84 per cent. The term "satisfactory" usually equates to a mark of 50-64 per cent. The term "poor" equates to a mark less than 50 per cent. There is no right or wrong answer on the number of performance standards to provide and it may be an odd or even number (Mueller, 2003, pp. 4-5). It depends on "the nature of the task assigned, the criteria being evaluated, the students involved and your purposes and preferences" (Mueller, 2003, p. 5). The Queensland University of Technology uses a seven point grading scale. However, the teaching team in *LWB236 Real Property A* provided four performance standards on the criterion-referenced assessment sheet rather than seven performance standards because the process of delineating the boundaries of the performance standards becomes more complicated as the number of performance standards increases.

In the example, in Table 1, the criteria has been weighted at five percent. The five percent is allocated across the four performance standards. Allocating a narrow range of marks or a single mark to each performance standard will lead to greater reliability because the marker has less discretion. Most of the performance standards in Table 1 offer a range within half a mark. This approach may be criticised for artificially compressing the marks. However, to rebut this it can be argued that this artificial compression is minimised when there are several criteria upon which students may perform at any standard. To overcome this difficulty in awarding numerical marks, the Teaching and Educational Development Institute suggests that the names of performance

standards awarded could be profiled according to their importance to arrive at an overall performance standard for the assessment, as opposed to a numerical mark (Teaching and Educational Development Institute, 1999, p. 15). For example, an excellent on a criterion, a good on another criterion and a satisfactory for another criterion may amount to an overall performance standard of good on the assessment.

Designers of criterion-referenced assessment sheets will find defining each performance standard the most difficult part of the process. The key is to anticipate the strengths and weaknesses in the student attempts at the assessment task. These strengths and weaknesses need to be articulated so that there is a clear limit between each performance standard. As mentioned above, this process becomes more complicated as the number of performance standards increase. When drafting the "excellent" performance standards, designers should avoid using descriptors that are almost impossible to achieve, for example, "All relevant issues considered". Designers should also make sure that the descriptors appropriately reflect the level of the performance standard, for example, "superficial analysis" would be inappropriate for the "satisfactory" performance standard and is better suited to the "poor" performance standard. The clarity of the performance standards will be refined over time in light of experience (Carlson, MacDonald, Gorely, Hanrahan, Burgess-Limerick,, 2000, p. 110).

When implementing the criterion-referenced assessment sheet, the assessment will be more reliable when each marker has a consistent understanding of the words used in the performance standards. For example, on the *LWB236 Real Property A* criterion-referenced assessment sheet, some of the ambiguous phrases include, "sophisticated and intellectual level of analysis", "high, but not comprehensive level of analysis", "lack of analysis" and "superficial or no analysis". Arguably, the extract in Table 1 is not a best practice model because it contains ambiguous terms, for example, "geuine attempt". Ambiguous phrases are open to interpretation by the markers. To overcome this problem, some criterions could be expressed more objectively, for example, the phrase "footnotes predominantly conform with the style guide" could be replaced with "more than 60% of the footnotes conform with the style guide". However, this may require the marker to count the number of footnotes and then count the number of times the footnotes conform with the style guide. This process is time consuming and tedious for the marker. Consequently, it is submitted that it is more efficient to use ambiguous terms in the criterion-referenced assessment sheet, but necessary to employ strategies that ensure there is a consistent understanding of the criteria and performance standards between the markers.

One strategy that can be used is to encourage the markers to provide feedback on the criterion-referenced assessment sheet before it is released to students. This will give the markers a sense of ownership over the criterion-referenced assessment sheet and generate interest in it (Burton & Cuffe, 2005, p. 170). Another strategy is to provide the markers with marked examples of assessment using the criterion-referenced assessment sheet. This will give the markers a greater understanding of how to apply the criterion-referenced assessment sheet and illustrate the types of comments to be provided to students. It will also guide the markers on where to place the ticks within the boxes for the performance standard descriptors, for example, in the middle of the box, or more towards the left or right. The placement of the ticks may seem trivial but some markers have agonised over this (Sumsion & Goodfellow, 2004, p. 338). In addition to the markers having a shared understanding of the criteria and performance standards, the students must also have a consistent understanding with the markers. This is better achieved under criterion-referenced assessment as opposed to norm-referenced assessment because it is more transparent.

Transparency

"Within Higher Education there is an increasing acceptance of the need for a greater transparency in assessment processes" (Rust, Price & O'Donovan, 2003, p. 147). The transparency of an assessment task measures whether the students understand what they are required to do in order to get a particular mark.

Norm-referenced assessment does not clearly indicate to students what they need to do to be awarded a certain mark because they are marked against their peers. As a result, norm-referenced assessment forces students to be more competitive because "students perceive they can achieve best results by pulling others back" (Jackson, 2004). Competition has been referred to as a side effect of assessment, for example, if only a certain percentage of students can receive the highest grade and the cohort is exceptional compared to previous cohorts, "there will not be enough rewards to go around" (Dunn, Morgan, O'Reilly & Parry, 2004, p. 12).

On the other hand, criterion-referenced assessment does not have arbitrary cutoffs. It clearly articulates to the students the criteria and performance standards (if the descriptors are well-written). It encourages the students to focus on the unit objectives because it shows the alignment between the assessment criteria and unit objectives. Criterion-referenced assessment compels students to devote time and effort on the important aspects of a task and not to waste time on things they are not required to do (Johnstone, Patterson & Rubenstein, 1998, pp. 30-1). In theory, if criterion-referenced assessment is used, there are enough rewards to go around when the cohort is exceptional.

If designers use ambiguous terms in criterion-referenced assessment sheets (as discussed above this may be necessary), they should explain such terms to the students. Devoting class time to discussing the criteria and performance standards is important given "the pivotal role of assessment in teaching and learning and the difficulties students have in understanding exactly what is required in concrete assessment tasks" (Johnstone, Patterson & Rubenstein, 1998, p. 37). Further strategies to increase transparency include providing students with examples of marked assessment using criterion-referenced assessment and asking the students to apply the criteria and performance standards to a piece of assessment (Burton & Cuffe, 2005, p. 170).

Criterion-referenced assessment arguably achieves greater validity, reliability and transparency. However, criterion-referenced assessment sheets should not be implemented randomly in a course. Designers should use criterion-referenced assessment to reinforce the incremental assessment of skills across the units in the course.

Three levels of embedding and assessing skills

Students enrol in a course with diverse backgrounds and varying skills. They are not a homogenous group, but the literature suggests that they have a common view that their course "will better enable them to succeed in professional employment, assist them to make career changes, strengthen their potential for a more personally fulfilling life, or some combination of these" (Australian Technology Network, 2000). To meet this student demand in a law school context, law schools have rigorously overhauled their curriculum to embed lawyering and generic skills, and to assess them in an authentic and learner-centred manner. Lawyering skills are those skills that are essential to practice law, for example, drafting skills and legal research skills (Kift, 1997, pp. 50-52) Generic skills are those skills that may be transferred to other contexts, for example, communication skills and teamwork skills. The literature suggests that skills should not be learned or assessed in a "one shot or inoculation model of teaching, which is commonly characterised by having one skills unit at the beginning of the course and a "booster" unit/shot at the end" of the course (Christensen & Kift, 1997, p. 220). Students should have the opportunity to incrementally develop their skills as they progress through the substantive law units in the law course.

Nathanson (1987) labels the incremental development of skills from lower to more complex levels, a "vertical transfer" (p. 191). Christensen and Kift (1997) effectively applied the notion of a vertical transfer when they unpacked the development of skills into three levels (p. 219). At level 1, students are "instructed on the theoretical framework and application of the skill, usually at a generic level. This skill may be practised under guidance and feedback provided. Assessment will usually include a critique of the skill as practised" (Christensen & Kift, 1997, p. 219). Level 1 is notionally the equivalent to the first year undergraduate core units in the law course. Level 2 builds on level 1 and is notionally the equivalent of the second year undergraduate core units. It

requires "a degree of independence…This may involve some additional guidance at an advanced level of the skill, an environment in which to practise the skill in a real world legal scenario, and feedback to students on their progress. Students will be encouraged to reflect on their performance and on ways to improve. At this level, individually or within a group, a student should be able to complete a task utilising a range of skills in relation to a simple legal matter" (Christensen & Kift, 1997, p. 219). Level 3 builds on level 2 and is the equivalent of the third and fourth year undergraduate core units. It requires students to "draw on their previous instruction and transfer the use of the skill to a variety of different circumstances and contexts without guidance. Students should be able to adapt and be creative in the ways they approach the context and use particular skills. Reflection on performance will be a key aspect. At this level, individually or within a group, a student should be able to complete a task utilising a range of skills in a complex legal matter for a knowledgeable and critical audience" (Christensen & Kift, 1997, p. 219).

Since 2000, the QUT School of Law has been overhauling its units to embed and assess lawyering and generic skills across the three levels of the law course. Since the beginning of 2004, its challenge has been to shift its assessment practices more strongly towards criterion-referenced assessment. The plan is to design criterion-referenced assessment sheets for all items of assessment in all law units by the end of 2007 (Queensland University of Technology Teaching and Learning Committee, 2003). In meeting this challenge, the need for an incremental approach to criterion-referenced assessment across the law course has emerged.

The incremental approach to criterion-referenced assessment used in *LWB236 Real Property A*

To take an incremental approach to assessing a particular skill using criterion-referenced assessment, the designer of the criterion-referenced assessment sheet must identify how the skill has been assessed in previous units and how it is assessed in later units in the course. This identification process was simplified in the Queensland University of Technology School of Law because examples of criterion-referenced assessment sheets across the three levels were readily available to assessors on an online teaching site. For example, the design of the *LWB236 Real Property A* criterion-referenced assessment sheet was informed by *LWB143 Legal Research and Writing*. In particular, *LWB143 Legal Research and Writing* developed legal research skills, legal analysis skills, written communication skills and document management skills at level 1 and *LWB236 Real Property A* builds these skills at level 2. These skills are further developed at level 3 in *LWB434 Advanced Research and Legal Reasoning*. However, as at semester 2 2005, *LWB434 Advanced Research and Legal Reasoning* had not introduced criterion-referenced assessment with descriptors for the performance standards. Drafting skills are embedded and assessed for the first time in the law course in *LWB236 Real Property A*. This means that a second year law unit assesses drafting skills at level 1. *LWB237 Real Property B* had previously assessed drafting skills using criterion-referenced assessment and this informed the criterion-referenced assessment sheet in *LWB236 Real Property A*.

In addition to drawing on the criterion-referenced assessment sheets from units before and after the one in question, the teaching team discussed the criteria, the weightings of the criteria and performance standards at face-to-face meetings and via email. Asking the teaching team for their input on the criterion-referenced assessment gives them a greater sense of owernership and arguably increases their willingness to embrace change (Burton & Cuffe, 2005, p. 170). Even though *LWB236 Real Property A* attempted to incrementally assess skills via criterion-referenced assessment by drawing on the criterion-referenced assessment sheets used in earlier and later units, using the recommended model discussed below will result in greater consistency and efficiency across the three levels of a course.

Recommended model for approaching criterion-referenced assessment across the three levels of a course

When designing criterion-referenced assessment sheets, it is important that the performance standards reflect an appropriate expectation of skill development. For example, the "excellent" performance standard used in *LWB143 Legal Research and Writing*, which assesses legal citation at level 1 is, "All references correct and conform with style guide". The word "all" suggests that something slightly less than perfect would not be excellent, which is unreasonable and an unrealistic expectation of students at level 1. If all references are correct at level 1, there is no scope for the students to incrementally develop citation skills at levels 2 and 3. There is also no scope for the designers of criterion-referenced assessment sheets at levels 2 and 3 to incrementally expect more of the students. The criterion-referenced assessment sheets implemented in level 2 and 3 units cannot simply repeat the same performance standards implemented in level 1 units. Each level should build onto the previous level to demonstrate the logical incremental progression of the assessment of skills.

The recommended model presented in Table 2 achieves this. At each level, there is an increased expectation of the skill development. For example, "excellent" at level 1 is only worth "good" at level 2 and is only worth "satisfactory" at level 3. Further, each unit in all three levels uses the same number of and name for the performance standards.

The recommended model will be more efficient for criterion-referenced assessment designers who assess students at level 2 because they will be able to copy the descriptors for "excellent", "good" and "satisfactory" from level 1 and paste them into the "good", "satisfactory" and "poor" descriptors at level 2. The designers at level 2 will only need to design a descriptor for "excellent" at level 2. This will obviate the need for the level 2 designers to draft all of the descriptors for the performance standards at level 2 because they are building on the level 1 descriptors. Similarly, the level 3 designers can build on the work done by level 1 and 2 designers and merely need to design a descriptor for "excellent" at level 3. Academics who are inspired to implement the recommended model for approaching criterion-referenced assessment should consider its impact on the workloads of staff across the three levels. In particular, the designers of criterion-referenced assessment should keep a record of their time spent on setting criteria and performance standards, explaining criteria and performance standards to students, supervising other markers to ensure there is a shared understanding of the criteria and performance standards, collecting, marking, grading, processing marks or grades and providing feedback to students. The hours spent on these tasks should be compared with the number of contact hours in the course (Andresen, Nightingale, Boud & Magain, 1992, p. 8).

In addition to being more efficient for criterion-referenced assessment designers, it is submitted that the recommended model improves the understanding of the criteria and performance standards (and expectations) by the markers and students who progress through the three levels of the course because it repeats the criteria and performance standards in the manner illustrated in Table 2 and thus reinforces the meanings attributed to them.

The recommended model is also more pedagogically sound than simply repeating the level 1 descriptors at level 2 and 3 because it advocates incremental assessment across the three levels of the course and applies the notion of a vertical transfer, which is discussed above.

To facilitate this incremental approach to criterion-referenced assessment, it is recommended that assessors across the three levels meet to discuss how levels 2 and 3 build onto levels 1 and 2. A further initiative is to place all criterion-referenced assessment sheets on a shared drive so that all assessors can access them and readily copy and paste the relevant descriptors. This will be useful for assessors who, for example, need to create a criterion-referenced assessment sheet for a new item of assessment. Periodic meetings should be scheduled for assessors across the three levels of the course to review the skills embedded in the course, the assessment criteria and descriptors for the performance standards.

		Level 3	Poor	Satisfactory	Good	Excellent
	Level 2	Poor	Satisfactory	Good	Excellent	
Level 1	Poor	Satisfactory	Good	Excellent		

Table 2: **Recommended** model for approaching criterion-referenced assessment across three levels of a course

Conclusion

The criterion-referenced assessment of skills should be incremental across the three levels of the course. Designers of criterion-referenced assessment sheets should take a consistent approach by using the same number of performance standards and using the same terminology across the units in the course. This will enhance the shared understanding of the criteria and performance standards by the markers and students. Designers should also use the recommended model in Table 2 to ensure that the expectation of skill development increases over the course. This incremental approach to criterion-referenced assessment will better meet the demands of students by preparing them for the real world.

References

Andresen, L., Nightingale, P., Boud, D., & Magin, D. (1992). *Strategies for assessing students*. Kensington: Professional Development Centre, The University of New South Wales.

Australian Technology Network. (2000). ATN report on generic capabilities. Retrieved October 25, 2005, from http://www.clt.uts.edu.au/TheProject.htm

Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education Research and Development,* 18(1), 57.

Bond, L. (1996). Practical assessment, research & evaluation. Retrieved October 25, 2005, from http://pareonline.net/getvn.asp?v=5&n=2

Burton, K., & Cuffe, N. (2005). The design and implementation of Criterion-referenced assessment in a first year undergraduate core law unit. *Legal Education Review*, 15, 159.

Centre for the Study of Higher Education. (2002). A comparison of norm-referencing and criterion-referencing methods for determining student grades in higher education. Retrieved October 25, 2005, from http://www.cshe.unimelb.edu.au/assessinglearning/05/normvcrit.html

Carlson, T., MacDonald, D., Gorely, T., Hanrahan, S., & Burgess-Limerick, R. (2000). Implementing criterion-referenced assessment within a multi-disciplinary university department. *Higher Education Research & Development*, 19, 103.

Christensen, S., & Kift, S. (1997). Graduate attributes and legal skills. *LegalEducation Review*, 11, 207.

Crooks, T. (1988). *Assessing student performance.* Kensington: Higher Education Research and Development Society of Australasia.

Dunn, L., Morgan, C., O'Reilly, M., & Parry, S. (2004). *The Student Assessment Handbook.* London: Routledge Falmer.

Jackson, S. (2004). A project to facilitate the implementation of criterion-referenced assessment in the school of law. Retrieved October 25, 2005, from https://olt.qut.edu.au/udf/FELLOW09/gen/index.cfm?fa=getFile&rNum=1638031&nc=1

Johnstone, R., Patterson, J., & Rubenstein, K. (1998). *Improving Criteria and Feedback in Student Assessment in Law.* Sydney: Cavendish Publishing.

Kift, S. (1997). Lawyering Skills: Finding their Place in Legal Education. *Legal Education Review,* 8, 43.

Le Brun, M., & Johnstone, R. (1994). *The Quiet (R)evolution Improving Student Learning in Law.* Sydney: The Law Book Company.

Mueller, J. (2003). Authentic assessment toolbox – rubrics. Retrieved October 25, 2005, from http://jonathan.mueller.faculty.noctrl.edu/toolbox/rubrics.htm

Nathanson, S. (1987). Putting skills and transactions together in professional legal training. *Journal of Professional Legal Education,* 5, 187.

Newble, D., & Cannon, R. (1989). *A handbook for teachers in universities and colleges: A guide to improving teaching methods,* New York: Kogan Page.

Queensland University of Technology. (2003). Manual of Policies and Procedures. Retrieved October 25, 2005, from http://www.qut.edu.au/admin/mopp/C/C_09_01.html

Queensland University of Technology Teaching and Learning Committee. (2003). CRA implementation plan for assessment policy. Retrieved October 25, 2005, from http://www.library.qut.edu.au/academics/cra-infolit.jsp

Ramsden, P. (1992). *Learning to teach in higher education.* London: Routledge.

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education,* 28(2), 147.

Scarino, A. (2005). Going beyond criteria for judging performance. *Barbel: Journal of the Australian Federation of Modern Language Teachers' Association*, 39(3), 8.

Sumsion, J., & Goodfellow, J. (2004). Identifying generic skills through curriculum mapping: A critical evaluation, *Higher Education Research & Development,* 23(3), 329.

Teaching and Educational Development Institute. (1999). Assessment criteria, standards and marking schemes. Retrieved October 25, 2005, from http://www.tedi.uq.edu.au/teaching/assessment/criteria.html

Please cite as: Burton, K.. (2006). Designing criterion-referenced assessment. *Journal of Learning Design*, 1(2), 73-82. http://www.jld.qut.edu.au/