

The article “Designing criterion-referenced assessment” (Burton, 2006) appeared in the Journal of Learning Design Volume 1, Issue 2 in 2006. Nine years on, Associate Professor Burton reflects upon her original article.

Continuing my journey on designing and refining criterion-referenced assessment rubrics

Kelley Burton

USC Law School
University of the Sunshine Coast
Australia
kburton3@usc.edu.au

I am delighted that my 2006 article was in the top 10 most downloaded articles on the 10 year anniversary of the *Journal of Learning Design*. In 2006, when the article was written, I worked at the Queensland University of Technology (QUT). At that time, the development of criterion-referenced assessment rubrics was gaining momentum. In 2015, as I write this reflection, all 37 Australian universities that support a law school universally accept criterion-referenced assessment as opposed to norm-referenced assessment. This year, I started working at the University of the Sunshine Coast and I am proud to say that its assessment policy explicitly endorses the development of criterion-referenced assessment rubrics. I have published several journal articles on designing and implementing criterion-referenced assessment rubrics and it continues to be my quintessential research interest.

The main strength of my 2006 article is that it is based on fundamental principles of assessment, which remain relevant today. The discussion focussed on the dichotomy between criterion-referenced assessment and norm-referenced assessment; definitions for criteria and standards; strategies to enhance the reliability, validity and transparency of criterion-referenced assessment rubrics; recommendation that markers are provided with marked examples of student work; recommendation that students apply a criterion-referenced assessment rubric to formative assessment; and the implicit promotion of diverse and authentic assessment. These themes are multi-disciplinary, and appeal to a wide audience of teachers and students.

Designing criterion-referenced assessment rubrics based on a whole-of-curriculum approach

My 2006 journal article made a novel contribution in the area of criterion-referenced assessment by illustrating how criterion-referenced assessment rubrics could be designed to support the incremental and sequential assessment of a skill across three levels (introductory, intermediate and advanced) in a degree or program. In essence, the intermediate and advanced criterion-referenced assessment rubrics build onto the previous level’s criterion-referenced assessment rubric. Applying three rubrics that have the same foundation will arguably lead to an improved shared understanding of the performance standards between the teachers and students and, consequently, enhance reliability and transparency. This approach to designing criterion-referenced rubrics across a degree or program is consistent with the contemporary literature in legal education that supports a “whole-of-curriculum,” “integrated,” “incremental,” and “sequential” approach to teaching, curriculum renewal and development, and assessment (see, for example, Huggins, 2015; Johnstone, 2011).

Designing criterion-referenced assessment rubrics based on threshold learning outcomes (TLOs)

Criterion-referenced assessment rubrics should illustrate the alignment between performance standards and threshold learning outcomes (TLOs). In 2010, the Australian Learning and Teaching Council’s *Bachelor of Laws Learning and Teaching Academic Standards Statement* developed six TLOs for a Bachelor of Laws Program (Kift, Israel & Field, 2010). These signify what is expected of a law graduate and include: knowledge; ethics and professional responsibility; thinking skills; research skills; communication and collaboration; and self-management (Kift, et al., 2010, pp. 9-10). The TLOs offer a current framework against which performance standards can be mapped and the validity of criterion-referenced assessment rubrics can be strengthened.

Refining criterion-referenced assessment rubrics

Refining criterion-referenced assessment rubrics is an iterative process, and Figure 1 below supersedes Figure 2 in my original 2006 journal article (Figure 3 in 2015 reprinted article). A key improvement includes reversing the order the performance standards so that the best performance standard is on the left-hand side of the table. While I thought my original table showed a journey of progressive performance standards across the page, it appears that the consensus amongst teachers is to include the best performance standard first. Additionally, the performance standards should be mapped against the grading scheme for the university, for example, high distinction, distinction, credit, pass and fail; instead of solely linking them to excellent, very good, good, satisfactory and poor. In accordance with innovations in the higher education sector, Figure 1 refers to a whole-of-curriculum approach and creates a link between the performance standards and the TLOs.

| Whole-of-Curriculum Approach to Designing Criterion-referenced Assessment Rubrics | | | | | | |
|---|-----------------------------|-----------------------------|--------------------|---------------|-------------|-------------|
| 1st Level Introductory TLO | | High Distinction | Distinction | Credit | Pass | Fail |
| 2nd Level Intermediate TLO | | High Distinction | Distinction | Credit | Pass | Fail |
| 3rd Level Advanced TLO | High Distinction | Distinction | Credit | Pass | Fail | |

Figure 1: Whole-of-Curriculum approach to designing criterion-referenced assessment rubrics for TLOs

It is hoped that this reflection will help build a community of practice to disseminate exemplars of contemporary criterion-referenced assessment rubrics and further develop the notion of a whole-of-curriculum approach to designing criterion-referenced assessment rubrics for TLOs. Please feel free to email me at kburton3@usc.edu.au, if you are interested in being a part of this community of practice. I am looking forward to continuing my journey on designing and refining criterion-referenced assessment rubrics; and wonder what they will look like in another nine years’ time.

The following presents the article as it appeared in 2006. With the author's approval, minor edits have been made, the referencing and other conventions updated to APA 6.0 and the numbering of figures altered to reflect the new publication.

Designing criterion-referenced assessment

Kelley Burton

Faculty of Law
Queensland University of Technology
Australia

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 A¹. 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.

Keywords

Criterion-referenced assessment, norm-referenced assessment, validity, reliability, transparency, criteria, performance standards, legal education, assessment, skills

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). 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) acknowledged 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

¹ While the University currently offers a comparable unit of study in Property Law, course design updates have meant that the unit code and name have changed.

“according to a preconceived notion of how the distribution of grades will turn out” (Dunn et al., 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). It suits units where there is an objective right or wrong answer (Dunn et al., 2004). 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 et al., 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 pre-set criteria (Le Brun & Johnstone, 1994). A “criterion” is a “distinguishing property or characteristic of anything, by which its quality can be judged or estimated, or by which a decision or classification may be made” (Scarino, 2005, p. 9). Assessment 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 & Rubenstein, 1998). 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 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: (a) there may have been a fault in the setting of the criteria or performance standards; (b) the assessment task may not have had an appropriate degree of complexity; (c) the markers may not have had a shared understanding of the criteria and performance standards with the students; and (d) the particular cohort may have been exceptionally better or worse than the cohorts in previous years. The assessor should reflect on possible 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) suggested 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, n.d.). In the context of a unit, here a semester-long period of study, 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 et al., 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). The same comment may be made 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 Figure 2. This is an extract from the *LWB236 Real Property A* criterion-referenced assessment sheet designed by the teaching team for a drafting exercise, file note and letter to a client. The assessment criteria adopted were: (a) understanding of forms’ content and purpose, (b) ability to transcribe information correctly, and (c) compliance with the relevant law. The criteria correlate with Unit Objective 10 which stated that students will:

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.

The four columns in Figure 2 represent 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 recognised measure of what is adequate for some purpose, so established by authority, custom, or consensus” (Scarino, 2005, p. 9).

| Excellent | Good | Satisfactory | Poor |
|--|---|--|--|
| Drafting exhibits all of the following: | Drafting exhibits all of the following: | Drafting exhibits all of the following: | Drafting exhibits one or more of the following: |
| <ul style="list-style-type: none"> • an excellent understanding of forms' content and purpose. • no obvious technical drafting errors or omissions. • complies with relevant law. | <ul style="list-style-type: none"> • a good understanding of forms' content and purpose. • at least one relatively minor technical drafting error or omission. • complies with relevant law. | <ul style="list-style-type: none"> • genuine attempt to understand forms' content and purpose. • contains one or more significant technical error or omission. • generally complies with law. | <ul style="list-style-type: none"> • 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. |
| 4.5-5 | 3.5-4 | 3-2.5 | <2.5 |

Figure 2: Extract from the *LWB236 Real Property A – Drafting exercise, file note and client letter assessment criteria and feedback sheet*

The total weighting for the drafting exercise, file note and letter to client was 20 per cent of the unit's final grade while instrument drafting skills made up five per cent. Weightings attached to the criteria depend on the importance of the unit objectives as well as the degree of professional marking judgement required.

Students were advised whether they had 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 was also supplemented with meaningful generic feedback on the online teaching site. The quantity and quality of this feedback went beyond simply providing a mark or grade which has been described as "cheating students" and "unprofessional teaching behaviour" (Ramsden, 1992, p. 193).

Primarily, the feedback provided served to inform the students. However, 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). 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's attempt at the assessment or if one marker awards a different grade to the same student's attempt at the assessment at a later point in time (Le Brun & Johnstone, 1994).

The reliability of norm-referenced assessment depends on how raw scores are calculated, that is, on (a) the basis of using prescriptive marking guidelines where there is little room for professional discretion and judgment; or (b) on the basis of professional discretion and judgment where the

marking guidelines are not specific. Norm-referenced assessment 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 et al., 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 per cent 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 recognised that norm-referenced assessment is likely to be unfairer to smaller cohorts because it exaggerates the difference between the students and may “artificially compress the range of difference that actually exists” (Centre for the Study of Higher Education, 2002, para. 5).

In contrast, criterion-referenced assessment establishes performance standards for each criterion. It is very prescriptive in nature. In the exemplar in Figure 2, performance standards are presented across the page, namely, “excellent,” “good,” “satisfactory” and “poor.” Within the School of Law at the Queensland University of Technology [at time of writing], it was common to use the term “excellent” to mean a mark within the range of 85-100 per cent; “good” usually equated to a mark within the range of 65-84 per cent; while, “satisfactory” usually equated to a mark of 50-64 per cent. The term “poor” equated 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). They depend 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 Figure 2, the criteria have been weighted at five per cent. The five per cent 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 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 (TEDI) suggested 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 (TEDI, 1999). 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 students' 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).

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 Figure 2 is not a best practice model because it contains ambiguous terms, for example, “genuine attempt.” Ambiguous phrases are open to interpretation by the markers. To overcome this problem, some criteria 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 contended 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 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). 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). 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

Rust et al. (2003) noted that “within Higher Education there is an increasing acceptance of the need for a greater transparency in assessment processes” (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 et al., 2004, p. 12).

On the other hand, criterion-referenced assessment does not have arbitrary cutoffs. It clearly articulates to students the criteria and performance standards (if the descriptors are well-written). It encourages 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 et al., 1998). 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), 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 et al., 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).

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 the context of Law, 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). 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). Students should have the opportunity to incrementally develop their skills as they progress through the substantive law units in the law course.

Nathanson (1987) labelled the incremental development of skills from lower to more complex levels as a “vertical transfer” (p. 191). Similarly, Christensen and Kift (1997) effectively applied the notion of a vertical transfer when they unpacked the development of skills into three levels. 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)

From 2000 [to 2006, at time of writing], the QUT School of Law overhauled its units to embed and assess lawyering and generic skills across the three levels of the law course. Since the beginning of 2004 [to 2006, at time of writing], its challenge was to shift its assessment practices more strongly towards criterion-referenced assessment. The plan was 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 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 were 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 were embedded and assessed for the first time in the law course in *LWB236 Real Property A*. This meant that a second year law unit assessed 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 used 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 gave them a greater sense of ownership and arguably increased their willingness to embrace change (Burton & Cuffe, 2005). 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 assessed 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 an unreasonable and unrealistic expectation of students at Level 1. If all references were 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 Figure 3 achieves this progression. 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.

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

Figure 3: Recommended model for approaching criterion-referenced assessment across three levels of a course

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).

In addition to being more efficient for criterion-referenced assessment designers, it is suggested 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 Figure 3 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 previously discussed notion of a vertical transfer.

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.

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 Figure 3 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, Australia: Professional Development Centre, The University of New South Wales.
- Australian Technology Network. (2000). *Generic capabilities of ATN university graduates*. Retrieved from <http://www.worldcat.org/title/generic-capabilities-of-atn-university-graduates/oclc/223603532>
- Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education Research and Development*, 18(1), 57-75.
- Bond, L. (1996). *Practical assessment, research & evaluation*. Retrieved from <http://pareonline.net/getvn.asp?v=5&n=2>
- Burton, K. (2006). Designing criterion-referenced assessment. *Journal of Learning Design*, 1(2), 73-82. doi:10.5204/jld.v1i2.19
- 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, 159-175.
- 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, 103-116.
- 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 from <http://www.cshe.unimelb.edu.au/assessinglearning/06/normvcrit6.html>
- Christensen, S., & Kift, S. (1997). Graduate attributes and legal skills. *Legal Education Review*, 11, 207-237.
- Crooks, T. (1988). *Assessing student performance*. Kensington, Australia: Higher Education Research and Development Society of Australasia (HERDSA).
- Dunn, L., Morgan, C., O'Reilly, M., & Parry, S. (2004). *The student assessment handbook*. London: Routledge Falmer.
- Huggins, A. (2015). Incremental and inevitable: Contextualising the threshold learning outcomes for law. *University of New South Wales Law Journal*, 38(1), 264-287.
- Jackson, S. (2004). *A project to facilitate the implementation of criterion-referenced assessment in the School of Law*. Unpublished paper, Queensland University of Technology, Brisbane, Australia.

- Johnstone, R. (2011). Whole-of-Curriculum Design in Law. In S. Kift, M. Sanson, J. Cowley & P. Watson (Eds.), *Excellence and Innovation in Legal Education* (pp. 1-29). Chatswood: LexisNexis.
- Johnstone, R., Patterson, J., & Rubenstein, K. (1998). *Improving criteria and feedback in student assessment in law*. Sydney, Australia: Cavendish Publishing.
- Kift, S. (1997). Lawyering skills: Finding their place in legal education. *Legal Education Review*, 8(43), 67-71.
- Kift, S., Israel, M., & Field, R. (2010). *Learning and Teaching Academic Standards Project: Bachelor of Laws Learning & Teaching Academic Standards Statement December 2010*, Australian Learning & Teaching Council. Retrieved from http://www.utas.edu.au/__data/assets/pdf_file/0007/456829/altc_standards_LAW.pdf
- Le Brun, M., & Johnstone, R. (1994). *The quiet (r)evolution improving student learning in law*. Sydney, Australia: The Law Book Company.
- Mueller, J. (2003). *Authentic assessment toolbox – Rubrics*. Retrieved 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(2), 187-200.
- 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 Teaching and Learning Committee. (2003). *CRA implementation plan for assessment policy*. Unpublished report, Queensland University of Technology, Brisbane, Australia.
- Queensland University of Technology. (n.d.). *Manual of policies and procedures*. Retrieved from http://www.mopp.qut.edu.au/C/C_05_01.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), 147-164. doi: 10.1080/0260293032000045509
- Scarino, A. (2005). Going beyond criteria for judging performance. *Barbel: Journal of the Australian Federation of Modern Language Teachers' Association*, 39(3), 8-16.
- Sumsion, J., & Goodfellow, J. (2004). Identifying generic skills through curriculum mapping: A critical evaluation, *Higher Education Research & Development*, 23(3), 329-346.
- Teaching and Educational Development Institute (TEDI). (1999). *Assessment criteria, standards and marking schemes*. Retrieved from <http://www.tedi.uq.edu.au/teaching/assessment/criteria.html>