Adopting new practice: using feedback to develop a university-wide moderation process that closes the assessment loop

Natalie Brown

Centre for the Advancement of Learning and Teaching, University of Tasmania, Natalie.Brown@utas.edu.au

In higher education, assessment practices are increasingly coming under scrutiny as a scholarship of assessment is being developed across the sector. Many institutions are promoting the use of criterion-referenced assessment, particularly in conjunction with the use of standards, yet despite this widespread support, the effective implementation of such a system can be problematic. Generating criteria that are clear and valid, developing a shared understanding of criteria and standards among students and assessors and arriving at final awards are key issues being addressed by scholars of assessment. At a meta-level, implementation of a fundamentally new assessment practice within an institution requires change management that takes account of a broader range of factors beyond issues of scholarship. This paper describes the procedure of articulating a moderation process in response to concerns about the removal of the requirement to 'grade on the curve'. The moderation process has three foci: the assessment design (pre-assessment focus), making judgments (point of assessment focus) and determining grading outputs (post-assessment focus). Importantly, the post-assessment focus explicitly builds in review and feed forward in relation to assessment design, thereby closing the loop of the assessment process.

Keywords: moderation; standards-based; criterion-referenced.

Introduction

David Boud tells us that "the impact of assessment on learning has been more negative than we normally care to admit" (2007, slide 8). When I display this statement on a PowerPoint slide in introductory assessment workshops I invariably look out on a sea of knowing faces, and if I probe for personal stories, the floodgates open. Perhaps this is not surprising. But when the audience is comprised of those who have arguably had the most success in assessment – academics with honours degrees and PhDs – it is a response of genuine concern. When, a little later in the session, I introduce the recent change of institutional policy to embrace standards-based criterion-referenced assessment, these same academics are almost unanimous in their support. It is not unusual during these workshops for confirmation to be sought that 'we are getting rid of the curve', followed by audible affirmations. Yet, embedding such a change of policy into institution-wide practice is not so straightforward. This paper gives a personal account of how one institution has approached this change of practice, drawing on current literature and practice across the sector to inform consultation and the development of resources.

There is no doubt that assessment in a higher education context is a vexed and challenging task (Barrie, Brew & McCulloch, 1999; Leathwood, 2005; Price, 2005; Tan & Prosser, 2004; Sadler, 2005). Perhaps adding to this complexity is the range of purposes to which assessment of student learning is applied: to guide meaningful student learning; to provide constructive feedback to students; to inform staff on the progress of students; to provide data to arrive at a final grade for a student; and to ensure academic quality and standards (Allen, Brown, Butler, Hannan, Meyers, Monkhouse & Osborne, 2007). It is therefore not surprising, given these important roles, that assessment of student learning is receiving significant attention from scholars of teaching and learning. This work, when combined with an agenda for strengthening accountability across the sector (Leathwood, 2005), is beginning to influence both policy and practice in higher education institutions.

One feature of the emerging discourse on assessment is the growing acceptance of a standards model of assessment as most applicable to higher education (Biggs & Tang, 2007; Leathwood, 2005; Price, 2005; Taylor, 1994). In such a model, student work is assessed against clearly articulated criteria, and decisions about the grades awarded to students are based on the attainment (or otherwise) of those criteria at stated performance standards (Allen et al., 2007; Biggs & Tang, 2007). A corollary to this is that the assessment practices are "moving assessment away from

a measurement model" (Falkichov & Thomson, 2007, p. 49). Indeed, Boud states that "most educational institutions have long moved from inappropriate norm-referenced assessment regimes" (2009, p. 30).

There are many accounts in the literature of issues associated with the use of a measurement model of assessment. Under this system, "proportions of students receiving the various grades are specified in advance as a matter of policy, and then applied to each class of students" (Sadler, 2005). Biggs and Tang (2007) write at length about the problematic assumptions underpinning a measurement model. Similarly, papers from Hornby (critiquing the 'spurious precision' inbuilt into measurement models, 2003, p. 440), Kassahun (2007) and Kulick and Wright (2008) all present powerful critiques of the underlying statistical assumptions of these models. In her work, Leathwood takes an equity perspective on measurement model as illustrated by this included quote from Spender:

> When only a few can get to the top, and when education is perceived as a distribution agency allocating people to their places, then many must be classified as failures. Through their barrage of marks, grades, tests, examinations, streaming, traditional educationalists promote the belief in the necessity of failure and the inevitability of inequality (Spender, 1986, cited in Leathwood, 2005, p. 311).

With such powerful arguments against the measurement model it is difficult, at face value, to understand why a movement to a new assessment paradigm has been (and in some cases continues to be) so problematic in higher education. The reasons essentially lie in two realms. The first is the need to move from an established and relatively easily applied assessment model (measurement, with the use of 'grading on the curve'), despite its acknowledged shortcomings (Biggs & Tang, 2007; Price, 2005). The second is the sector's lack of confidence in adopting a new model that is not without problems of its own. This latter point, bound up with the need for teachers in higher education to engage in professional learning and dialogue about writing and interpreting criteria and standards, has been comprehensively addressed by assessment scholars (Barrie, et al., 1999; Hammer, 2007; Price, 2005; Sadler, 2005).

Context

The context for this paper is a medium-sized university with a long history of teaching and research. In response to student and staff feedback, and cognisant of a growing literature on assessment in higher education, a review of assessment across the institution was conducted in 2006 at the behest of its academic senate. The review was conducted through the formation of a working group of assessment fellows representing each of the institution's faculties. In September 2007, the fellows made a presentation to academic senate recommending an institution-wide change to standards-based criterion-referenced assessment (CRA). The adoption of CRA was supported by the senate, with the new assessment policy endorsed and a three-year implementation plan ratified the following year.

The implementation plan advocated an approach to change management that was outcomes-based (representing a corporate position) and collegial (representing shared values). To this end, a network of school-based champions was established to connect the implementation team, based in the academic development unit, to the schools (Cordiner & Brown, 2009). An important role of the network was to ensure the implementation process was monitored against progress goals and any issues of concern were raised.

Although never articulated in university-wide policy, the institution had a tradition, stronger in some faculties than others, of 'grading on the curve'. Implicit in the new policy was the abandonment of this practice. However, at the end of the first 18 months of implementation, there remained considerable feedback from academics in schools and school champions, that 'grading on the curve' continued to be enforced in some schools. It therefore became apparent that it was insufficient to introduce a new policy without explicitly prohibiting the use of any pre-determined grading curve.

Methodology

This paper describes the process that led to the development of a moderation document for use across the institution. It takes the perspective of an academic developer with leadership responsibility for the implementation of CRA across the university. It should be noted that the development of a moderation process and the

development of associated documentation were not the intended outcomes at the beginning of this initiative. The original aim of the author was to have high level support for explicit prohibition of the use of pre-determined grading curves. As a result of pursuing this goal, and interacting with colleagues and other stakeholders, the moderation process emerged as a useful tool to mediate the institution-wide change to a standards model of assessment that was no longer overlaid by a measurement model.

The methodology underpinning the current study is that of a particularistic case study focusing on a distinct situation, bound by time and space (Merriam, 1998; Stake, 1981) and "arising from everyday practice" (Merriam, 1998, p. 29). The presentation as a case aims to make the experience relevant to the reader through "resonating with their own experience" (Stake, 1981, p. 36). Consistent with the case study genre, my own role as "participant as observer" (Adler & Adler, 1994) needs to be declared, recognising also that the observations extend to reflection on my own effectiveness in bringing about change.

In this study, data was sourced predominantly from two domains. The first was data gathered from the field in which the work was taking place – that is, from the university staff the CRA implementation team was interacting with. This data was predominantly captured to inform the success, or otherwise, of the CRA implementation plan and was predominantly qualitative in nature. Specific data that informed the first phase of this project was gleaned from short, structured telephone interviews (five to 10 minutes) with school champions, undertaken at the end of the first full year of the project. The CRA implementation team also routinely took notes from workshop sessions and recorded responses to papers presented in various university fora such as school/faculty meetings and teaching and learning committee meetings. The second form of data was documentary in nature – the assessment literature and assessment policies and procedures of other higher education institutions. This data was used to inform the process and interactions of the author with her colleagues and the development of the moderation document itself.

Results and discussion

The results are presented as a description of the process undertaken by the author as she seeks to address the issue of continued use of a grade distribution curve in sections of the university. To reflect the nature of the process and the iterative engagement with primary and secondary data sources, the discussion has been integrated into this description.

Phase 1: Addressing the need to explicitly prohibit the use of school- or faculty-imposed grade distributions

Prompt: School champions and other academic staff members attending assessment workshops were continuing to report persistent use of grade distribution curves in some schools. This was confirmed through analysis of school champion interview data where three champions (from two faculties) indicated that grading on the curve continued to impact on assessment in some way. In addition, anecdotal evidence suggested there was a perception among many staff that a university-wide policy insisting on the use of post-assessment grade distribution remained in existence, regardless of whether it was actively promoted within the school or faculty.

Response: A paper was prepared for the university's teaching and learning committee in November 2009, outlining the issue and providing a basis for discussion. The paper's perspective was that standards-based criterion-referenced assessment was identified as good practice across the higher education sector and, when implemented, removed the need to apply any predetermined grading scheme. It also highlighted the inconsistencies of such an approach with the current policy. During the in-committee discussion, the underpinning (incorrect) assumptions of the effectiveness of the measurement model, as identified in the literature, were raised, particularly erroneous statistical assumptions (Biggs & Tang, 2007; Hornby, 2003; Kulick & Wright, 2008), separation of assessment from quality of teaching (Biggs & Tang, 2007; Sadler, 2005), lack of transparency (Price, 2005) and questions of equity and justice (Leathwood, 2005; Stowell, 2004).

In making the case for removing reference to grade distributions, it was acknowledged that setting standards 'up front' with the inability to 'scale' students may cause some anxiety for staff in ensuring they have appropriately described standards to reflect each of the assigned grades. This was addressed in the university's teaching and learning committee discussion paper as below.

> As academics who offer units as part of a degree course, a number of which are subject to accreditation by outside bodies, it is reasonable to assume that we can set minimum acceptable standards for our students to pass a unit. Indeed, it could be argued that most academics clearly have this in their minds when they are writing units, learning outcomes and assessment tasks, and may have always done so. These standards should be independent of student cohort and should stand up to scrutiny when benchmarked with similar units within the University or in comparable units in other institutions. It is recognized, however, that this assumption does require that as academics working in particular schools we have access to processes that allow us to become familiar with standards across the school and faculty and there are robust and regular conversations regarding standards across the school and faculty, including opportunities for external benchmarking (N. Brown & G. J. Hannan, personal communication, November 13, 2009).

It was also noted in the discussion paper that there was an implementation team, including a dedicated lecturer in assessment, to support staff in writing criteria and standards.

Outcome: A robust discussion ensued in the teaching and learning committee. Concerns raised by members included the critical requirement for high quality criteria and standards to be written and understood by assessors and students. There was also a question of whether a 'safety net' in the form of a grade distribution was needed for cases where criteria and standards were deficient. Specific instances of assessment problems were also raised in connection with assessment by sessional or clinical staff who were not well versed in university expectations. The ability to deal with assessment 'aberrations' in a systematic but time efficient way was seen as a vital component of any new policy regime and needed to be addressed.

There was no resolution with respect to the recommendations in the discussion paper to issue a high level statement directing abandonment of grade distribution. Further discussion was deemed to be necessary and scheduled for a future meeting.

Phase 2: Addressing the need to reassure university staff that abandonment of school- or faculty-imposed grade distributions would not lead to significant errors in awarding grades to students, or to a pattern of grade inflation

Prompt: In the transition to CRA, there was genuine concern among some senior members of academic staff that the inability to apply post-assessment scaling (with reference to a grade distribution curve) would cause significant issues in arriving at final results. As a result, the use of the distribution curve remained, at least as a reference point, in some faculties.

Response: It was the assertion of the author that genuine implementation of CRA would not be possible while the normal distribution curve was given such pre-eminence in discussions about assessment, regardless of whether or not it was *actually* being used post-assessment to amend student results. Nevertheless, it was necessary to acknowledge academics' concerns in order to move forward on this issue. In doing so, it was important to ensure that good practice in assessment was encouraged without imposing a great deal of extra work on academics, heads of school and associate deans. The approach adopted in phase one, critiquing the measurement model (and consequent grading according to pre-determined schemes) was clearly insufficient to result in a shift away from historical assessment practices. In this second phase, perceived weaknesses or questions about the standards-based criterion-referenced system needed to be addressed to raise confidence levels in this new assessment model.

Analysis of issues raised in the teaching and learning committee discussion highlighted as key concerns the setting of appropriate standards; the interpretation of criteria and standards by markers; and comparability between units at the same level (introductory, intermediate and advanced). These were wholly consistent with findings from the literature critiquing the standards model of assessment. This suggests that we can also look to the literature for ways of overcoming these issues in practice. One of the most important of these is engaging in dialogue regarding assessment tasks, criteria and standards to develop shared understandings (Barrie, et al., 1999; Hammer, 2007; Price, 2005; Tan & Prosser, 2004).

Reviewing the literature highlighted the value of formalising this dialogue regarding assessment through a process of moderation (Hammer, 2007; Hughes, 2008). Indeed, moderation is a key element of the assessment process in a number of universities (for example, Macquarie University, 2010; Queensland University of Technology, 2010; Royal Melbourne Institute of Technology, 2010; and the University of South Australia, 2009). Perhaps unsurprisingly, the interpretations of the process of moderation differ in these documents but would generally fit under the definition offered by Sadler (1998, cited in Hughes, 2008), who defined moderation as a process for developing consistency and comparability of assessment judgments which could apply within classes, units, courses, faculties and institutions.

Concentrating on developing and articulating a process of moderation that would address concerns appeared to offer a way forward to the impasse from phase one. What was particularly attractive about this approach was that moderation strategies were already routinely employed at the university. In this way, what was being proposed was not new or additional, but bound in current (good) practice.

In designing a tri-focal moderation process, moving from an emphasis on scrutinising distribution of grades towards the stages of assessment design and making judgments was intentional. This was seen as being consistent with a standards-based, criterion-referenced approach in which assessment tasks need to be aligned with learning outcomes to ensure validity, and decisions about standards that are appropriate for study at the level of the unit concerned are made up front (Biggs & Tang, 2008; Hammer, 2007; Stowell, 2004). Moderation at the assessment design stage can also address the critique of the standards model that argues that an individual lecturer can make assessment tasks either too easy or too difficult for the standard and level of the course, so that a "disproportionate number of students may achieve very high or very low grades" (Kassahun, 2007, p. 36). Employing a moderation process at the point of assessment, when judgments about standards are being made provides an opportunity to "build shared understandings among teaching team members" (Hammer, 2007, p. 55). Price (2005) notes that with increasing student numbers the need for whole teaching teams to be versed in an understanding of common standards is gaining importance. She also contends that in sharing standards, the articulation of "tacit knowledge" (Price, 2005, p. 215) is essential, citing the use of exemplars and marked work as being powerful in helping making this explicit.

Arguably, effective moderation at these first two points should render moderation post-assessment as unnecessary. However, this focus remains as a final check of assessment design and the making of judgments. Indeed, given the strength of concern expressed by senior university colleagues, this focus was deemed necessary, at least while the shift to CRA was being embedded. The use of the term 'triggers for review' was chosen to indicate the possibility of an issue of concern and provide a safety net for examiners should the moderation process fail at the first two points. In providing this safety net two points have been emphasised. Firstly, a trigger for review does not mean that there is necessarily a problem with the assessment, and secondly, when an assessment review takes place this should be of the whole unit – not be dealt with through the application of scaling, or the alteration of students' marks to fit a pre-determined grade distribution. When such a review takes place, feed forward into the assessment design phase of the unit in the next offering would be an expectation. This effectively closes the loop and provides data to inform the ongoing review of the effectiveness of the teaching, learning and assessment design of the unit. In order to present the moderation process in a way that emphasised its genesis as being from current (good) practice, an outline of how suggested strategies relate to policy and practice was created (Table 1).

Outcome: The moderation process document was accepted in principle, subject to minor alterations as recommended by associate deans. This process replaces the use of grade distributions for monitoring grade outputs.

Table 1: The proposed moderation process

Focus		Possible strategies	Comments
Pre-assessment	**Assessment design**	Use of university's good assessment guidelines (e.g. alignment with learning outcomes, range of assessment tasks, opportunity for early feedback, not too many or too few tasks) Clearly articulated criteria and standards for major assignments	These are becoming standard practice under CRA. The concept of 'constructive alignment' is introduced to all staff in foundations' programs and embedded in the unit outline template.
		Peer review of units	This has been introduced as recommended practice in the CRA implementation plan.
		Benchmarking between units at the same level	This occurs in some schools as standard practice; it is also carried out by some course coordinators in specialist degrees.
		Ensuring progression of complexity in units at successive levels	This is carried out in some degree courses, and has occurred to some extent with the CRA project.
		Benchmarking against other institutions	This is on the national agenda. Australian Learning and Teaching Council discipline scholars will begin to explore how this can be facilitated.
Point of assessment	**Making judgments**	Heads of school, or their delegates, should ensure that all staff involved in marking (including casual academic staff) are well prepared As a minimum, marking guidelines (in addition to a criteria sheet) and representative work samples when possible, should be provided	This is standard practice in many schools. Many schools ensure that the unit coordinator is available for reference during marking.
		Group marking exercise to agree on standards, particularly for large numbers of markers, markers from different areas, or inexperienced markers	Assessment workshops can be organised through the academic development unit for sessional staff. This does occur, to some degree, in many large units.
		Selection of sample to double mark (e.g. random) or borderline/failures/high distinctions	Does occur in some schools – many have operational guidelines for this This is carried out by some unit coordinators, but the practice is not systematic, particularly with regards to documentation.

Focus		Possible strategies	Comments
		Use of triggers for review of grades awarded in individual assessment tasks prior to returning the work to the student	This is also an opportunity for assessors to review how students are tracking against the learning outcomes to inform teaching.
		These triggers may be:	It is important that the findings, from any investigation of an assessment review trigger, feed forward into future assessment design and/or practice.
		• discrepancies between grade allocations of individual markers	
		• high numbers of failures or high distinctions	Schools need to have clearly documented procedures about how discrepancies are dealt with (e.g. Who does the remarking? How many papers are re-marked? Who settles disputes?). These exist in some schools, is there an opportunity for us to share practice?
		• clustering of marks	
		• discrepancies between grades allocated to individual students in successive assessment tasks	
Post-assessment	**Grading outputs**	*Triggers*[1] for review of assessment include:	Any queries about assessment should be handled prior and then reported to examiners' meetings for consideration.
		• disproportionate allocation of marks according to *historical* data (in a large first year course this may be a version of a normal or bimodal distribution; in specialist units it may be something quite different)	If a trigger for review of assessment is identified a review of assessment *in the unit* should occur. Procedures should be consistent with those used at point of assessment, clearly documented and available to students.
		• large numbers of failures among students who have *participated* in the unit	Review, any correction or other course of action should, where possible, occur within the current assessment timeframe, as is consistent with current practice. In deciding upon any course of action, no student should be disadvantaged.
		• large numbers of students who have received the same grade	
		• discrepancies between grades allocated to individual students in different units	
		• substantially late submission of results	When an assessment review identifies an issue related to assessment, the examiners should determine how the findings will feed forward into future assessment design or practice.

Conclusion

An academic developer with responsibility to facilitate teaching and learning policy implementation and to enhance practice walks a fine line. Respect for context and current practice is a critical starting point for any desired change in practice. The challenge then becomes how to bridge the gaps between current practice and better practice in a way that is accessible and practical given the many competing demands on academics' time.

In this instance, a tool to mediate change (the moderation process) emerged from a study of the literature and current good practice. Much of what is incorporated is not new: it is embodied in current policies and guidelines, and is the practice of some academics. However, the formal articulation of these practices, putting a focus on dialogue and the importance of assessment design, has enabled the university community to move into a new assessment paradigm with greater confidence that it will enhance students' experiences of assessment. The moderation process is being progressively introduced in 2010 and feedback on its use is being collected as a component of the systematic review of the Assessment policy.

---

[1]   A trigger for review of assessment is not a sign that the assessment is inaccurate. The assessment review should be applied to the whole student cohort (e.g. through random sampling), not individual students. These triggers may also provide helpful information about teaching and assessment of the unit more generally.

Acknowledgements

References

Adler, P.A., & Adler, P. (1994). Observational techniques. In N. K. Denzin & Y.S. Lincoln (Eds.), Handbook of qualitative research (pp. 377-392). Thousand Oaks: Sage.

Allen, P., Brown, N., Butler, L., Hannan, G., Meyers, N., Monkhouse, H., & Osborne, J. (2007). Guidelines for good assessment practice. Hobart: University of Tasmania.

Barrie, S., Brew, A., & McCulloch, M. (1999). Qualitatively different conceptions of criteria used to assess student learning. Proceedings of the AARE-NZARE Conference, Melbourne. Retrieved January 22, 2010, from http://www.aare.edu.au/99pap/bre99209.htm.

Biggs, J., & Tang, C. (2007). Teaching for quality learning at university (3rd ed.). Maidenhead: McGraw-Hill/Society for Research into Higher Education and Open University Press.

Boud, D. (2007). Great designs: what should assessment do? Keynote address at International Online Conference sponsored by the REAP Project: Assessment design for learner responsibility. Retrieved July 18, 2008, from http://www.reap.ac.uk/reap07/ConferenceSessions/Keynotesessions/ProfessorDavidBoudKeynotetheme2/tabid/269/Default.html.

Boud, D. (2009). How can practice reshape assessment? In G. Joughin (Ed.), Assessment, Learning and Judgement in Higher Education (pp. 29-43). Dordrecht: Springer.

Cordiner, M., & Brown, N. (2009). Using a distributive leadership strategy to improve the quality of assessment across a university: initial results of the project, Proceedings of ATN Assessment Conference: Assessment in Different Dimensions. Melbourne: Royal Melbourne Institute of Technology.

Falchikov, N., & Thomson, K. (2007). Assessment: What drives innovation? Journal of University Teaching and Learning Practice, 5(1), 49-60.

Hammer, S. (2007). Demonstrating quality outcomes in learning and teaching: examining 'best practice' in the use of criterion-referenced assessment. International Journal of Pedagogies and Learning, 3(1), 50-58.

Hornby, W. (2003). Assessing using grade-related criteria: a single currency for universities? Assessment & Evaluation in Higher Education, 28(4), 435-454.

Hughes, C. (2008). Moderation: A discussion paper prepared for the Assessment Subcommittee of the UQ Teaching and Learning Committee. Unpublished discussion paper. Retrieved January 22, 2010, from http://www.tedi.uq.edu.au/teaching/assessment/assessmentResources.html.

Kassahan, D. (2007). Standardization techniques for grade-inflation problems at higher educational institutions of Ethiopia: the case of Addis Ababa. Assessment & Evaluation in Higher Education, 33(1), 33-44.

Kulick, G., & Wright, R. (2008). The Impact of Grading on the Curve: A Simulation Analysis. International Journal for the Scholarship of Teaching and Learning, 2(2).

Leathwood, C. (2005). Assessment policy and practice in higher education: purpose, standards and equity, Assessment & Evaluation in Higher Education, 30(30), 307-332.

Macquarie University. (2010). Assessment policy. Sydney: Author. Retrieved October 29, 2010, from http://www.mq.edu.au/policy/docs/assessment/policy.html.

Merriam, S. (1998). Qualitative Research and Case Study Applications in Education. San Francisco: Jossey Bass.

Price, M. (2005). Assessment standards: the role of communities of practice and the scholarship of assessment. Assessment & Evaluation in Higher Education 30(3), 215-230.

Queensland University of Technology. (2010). Review and moderation of assessment. Brisbane: Author. Retrieved October 29, 2010, from http://www.mopp.qut.edu.au/C/C_05_01.jsp#C_05_01.07.mdoc.

Royal Melbourne Institute of Technology. (2010). Moderation and validation of assessment. Melbourne: Author. Retrieved October 29, 2010, from http://www.rmit.edu.au/browse;ID=jof1m3aqzf23.

Sadler, R. (2005). Interpretations of criteria-based assessment and grading in higher education, Assessment & Evaluation in Higher Education, 20(2), 175-194.

Spender, D. (1980). Educational institutions: where cooperation is called cheating. In D. Spender & E. Sarah (Eds.), Learning to lose: sexism and education (pp. 39-48). London: The Women's Press.

Stake, R. E. (1981). Case study methodology: an epistemological advocacy. In W.W. Welsh (Ed.), Case study methodology in educational evaluation. Proceedings of the 1981 Minnesota Evaluation Conference. Minneapolis: Minnesota Research and Evaluation Centre.

Stowell, M. (2004). Equity, justice and standards: assessment decision making in higher education. Assessment & Evaluation in Higher Education, 29(4), 495-510.

Tan, H. K., & Prosser, M. (2004). Qualitatively different ways of differentiating student achievement: a phenomenographic study of academics' conceptions of grade descriptors. Assessment & Evaluation in Higher Education, 29(3), 267-282.

Taylor, C. (1994). Assessment for measurement or standards: the peril and promise of large scale assessment reform. American Educational Research Journal, 31, 231-262.

University of South Australia. (2009). Moderation. Sydney: Author. Retrieved October 29, 2010, from http://www.unisa.edu.au/ltu/staff/practice/assessment/moderation.asp.