

## Developing assessment policy and evaluating practice: a case study of the introduction of a new marking scheme

**Dr Fiona J. L. Handley, Southampton Solent University, East Park Terrace, Southampton, Hampshire, SO14 0YN, 0238 2013000<sup>1</sup>**

Dr Handley was a Quality Enhancement Officer at Southampton Solent University, and co-led the project management and evaluation of the Grade Marking Scheme.

**Dr Ann Read, Southampton Solent University, East Park Terrace, Southampton, Hampshire, SO14 0YN, 0238 2013000**

Dr Read is Dean of Academic Services at Southampton Solent University. Her background is in Occupational Psychology, and she previously worked at the University of Portsmouth.

### Abstract

In 2011 Southampton Solent University, a post-1992 university in southern England, introduced a new marking scheme with the aims of changing marking practice to achieve greater transparency and consistency in marking, and to ensure that the full range of marks was being awarded to students. This paper discusses the strategic background to the scheme's development, analyses the role of the working group and stakeholder involvement in developing the initiative, and presents a critical commentary on its success within the frame of the university as a "learning organization" (Senge, 2006).

Assessment, marking, spot-marking, working groups, staff development, policy, strategy, learning organization

### Introduction

University assessment policies and practices have come under increasing scrutiny as a result of clearer sector-wide strategic impetuses around learning and teaching (Hannan and Silver 2000), the reappraisal of assessment as a key motivator for student learning (Sambell et al. 2012), and more explicit codes of practice around assurance and enhancement (QAA). Attention has turned to the appropriateness of traditional award structures (Burgess 2007) and marking schemes (Yorke 2008, Rust 2011). As a result, some institutions are moving away from percentage based marking systems, and Grade Point Averages are being explored as an alternative marking and award scheme (HEA 2013; Oxford Brookes 2012).

In October 2011 Southampton Solent University replaced its percentage based marking system for all taught courses with a new marking scheme called Grade Marking, constructed from 17 marks ranging from F4 to A1 (Table 1). This was a major institutional initiative that drew on a wide variety of internal and external information to determine its form and dissemination. The strategic aim was to create a transparent marking scheme that justly rewarded students for their work. This paper presents the development of the initiative and critically analyses the results of

---

<sup>1</sup> Now at the University of Brighton [f.handley@brighton.ac.uk](mailto:f.handley@brighton.ac.uk)

the evaluation, and in particular the institutional commitment to being a learning organization incorporating bottom-up change and distributed leadership (Senge 2006) when faced with a major policy change involving the maintenance of academic standards. It is of interest to those managing and supporting assessment change initiatives, working groups, and stakeholder consultations.

Grade Mark	Numeric equivalent
A1	100
A2	92
A3	83
A4	72
B1	68
B2	65
B3	62
C1	58
C2	55
C3	52
D1	48
D2	45
D3	42
F1	38
F2	30
F3	20
F4	15

Table 1 The Grade Marking Scale as initially approved

### The development of the initiative

In 2009 the University created an Assessment Working Group with the remit to innovate assessment practice at the University in line with the University's Teaching and Learning Strategy. The Working Group was drawn from across the university with academic and professional services staff from all faculties, staff from central services, and a Students' Union representative. It used an action research approach (Zuber-Skerritt 1992) based on the examination of internal management data which had identified bunching of marks around classification thresholds, and feedback from external examiners who reported that marks in the first class range needed extending. This was developed through research into assessment best practice within the sector. One of the Group's first recommendations was the creation of Generic Grading Criteria for undergraduate students which was introduced in 2009, with Masters level criteria approved in the following year. The Grading Criteria made the rationale behind marking judgments clearer to students and encouraged staff to extend the marking range by giving them a clear basis on which to justify the marks.

To support this the Group also recommended a move away from marking on the 0-100 scale to point or spot marking, a system whereby the marker could only pick a mark from a much reduced scale of possible marks. This supported the use of Grading Criteria by having a clearer link between a mark and a criteria. The decision to reduce the number of marking points was backed up by research into inter- and intra- marker reliability, which suggested that fewer marking points were a better

reflection of how accurate markers could actually be (Newstead 2002; Newstead and Dennis 1994; Rust 2011; Yorke 2010). Fewer marking points also reduced the amount of time spent making marking decisions thus speeding up the time taken to return feedback and marks to students.

Once approved in principle by Academic Board, the scheme was developed by the University's Academic Services with a period of desk-based research into similar schemes at other universities and the rationale for using shorter marking schemes. As a result, two further key decisions were made. The first was to move away from marking in numbers, because it was felt that for many staff the number was too bound up with conceptions of achievement that did not match the University's Grading Criteria, and research had demonstrated that using letter grades increased the range of marks used (Yorke et al. 2002). Several alternatives including using a lettering (e.g. B+, B, B-), or Grade Point Average (1.0-5.0) system, were considered, and a version based on the University of Warwick's on a scale from A1 to F4 was developed. The combination of letters and numbers was reflected in the scheme's name, Grade Marking.

The second key decision was that the shift away from numbers would only take place at the level of marking as there was no resource to change the underlying 'number based' IT system. This would have involved a major redevelopment of the algorithms for calculating awards and the student records system. This meant that the scheme only partially addressed criticisms of most numeric based marking schemes which use numeric data incorrectly, by using ordinal numbers (which represent a ranking) as if they were cardinal numbers. Only cardinal numbers represent a quantity and can therefore be meaningfully used in calculations (Dalziel 1998; Milton et al. 1986; Soh 2011; Yorke 2011).

A detailed implementation plan was drawn up, covering a consultation period, communication to students developed in partnership with Solent Students' Union, a programme of staff development, and evaluation.

The consultation was wide ranging, with the aim of fostering an institution-wide sense of involvement which would create a meaningful scheme which would be more easily accepted. The Students' Union supported the initiative because of its clear message of recognizing student achievement. Its particular concern was to ensure that the changes were clearly communicated to students. While few of the external examiners consulted had had direct experience of this kind of marking scheme, most knew and approved of the principles behind it. Consultation with University staff took place via workshops timed around exam boards during early summer 2011, allowing their feedback to inform the development of the final scheme. Feedback from administrative staff focused on the processes of the new scheme, for instance consistency in turning Grade Marks into the numeric equivalents.

The workshop materials were carefully developed to communicate the key messages of the initiative, particularly to academic staff. As the people with the most contact with students and who would be in the frontline in dealing with any problems, they had particular concerns about the uncertainties of the success of the roll out (Fullan 2007, 25). Research has shown that academic staff are wary of new

academic policy, especially if it interferes with their autonomy (Bryman 2007), and they look for hidden 'backstories' (Trowler 2008, 137). The feedback from staff generally followed these patterns. This was partially mitigated through presenting a strong academic rationale for the new marking scheme (Ramsden 1998) based on marking to points, spreading of marks and equity for students, and using staff developers with academic backgrounds. As a result, most staff immediately understood the benefits of the scheme to their working practices and student experience. However, concerns were raised about students becoming confused with different marking schemes (uncertainty), and the initiative being a backdoor way of hugely increasing marks (a hidden backstory).

The distribution of marks within the middle ranges was uncontroversial, as the three Grade Marks and their numeric equivalents of numbers ending 2, 5 and 8 were clearly understood as a way of pulling marks away from boundaries. Within the 'first' and fail ranges of marks however, the distribution of the Grade Marks was more contentious. The scheme had four Grade Marks covering 30% points in the first range, and another four Grade Marks covering the 39% points in the fail range, to allow the scale to tally with the Grading Criteria. The lowest mark that could be given by a marker for a credible attempt was F4 (15%), while the highest mark was put at A1 (100%).

Some staff were uncomfortable with the low first mark having a numeric equivalent of 74. For staff used to giving a borderline first of 70%, moving to 74% was challenging, particularly because it did not 'mirror' the highest fail mark of F1 (38%), or B1 (68%), by being 2 percentage points away from the threshold. Their feedback resulted in the lowest numeric equivalent in the 'first' range being reduced to 72.

The discussion around the marks in the fail range was less passionate. There were some staff who were concerned with the lack of 'symmetry' in not having a 0 when a mark of 100% was possible, or the marks not being distributed through the fail zone in the same way as marks were distributed in the first range, but overall staff were happy with the fairly arbitrary assignment of Grade Marks once it was established that the same pattern was used in other universities.

The sessions demonstrated how some staff were keen to see marks fall equally above and below thresholds, and to have symmetry across marking ranges through having marks 'mirrored' in the first and fail range. This resonates with the way that staff traditionally mark to norms (Bloxham 2009; Yorke 2011). It may be that marking on a numeric scale encourages staff to search for patterns and symmetry because the numbers allow them to do this easily, and it is therefore presumed that this is one of the purposes of marking on a 0-100 scale.

In September 2011 the staff workshops shifted to having a more staff development focus rather than being consultative and were targeted at course leaders, for them to then disseminate to their teams. This is in line with fostering leadership at all levels and highlighting the importance of personal commitment to an institutional vision (Senge 2006). The sessions focused on using the Grade Marking scheme with the generic grading criteria supported with tools such as marking grids and rubrics. While the use of these tools would only go part of the way in supporting a change in marking practice (Sadler 2009), one of the purposes of the sessions was to

encourage course leaders to use these tools to engender local, disciplinary based understandings of marking practice.

### Evaluation

The scheme was launched at the beginning of the 2011-12 academic year, and evaluation began. As a result of initial feedback from staff further guidance materials on processes around marking portfolios were created, and the Students' Union requested that extra promotional material targeted at 3<sup>rd</sup> year students was created. Overall feedback from both staff and students at an evaluation event in spring 2012 was cautiously positive with a few issues raised (discussed below). By the time of the main summer exam boards there was a genuine sense of excitement as marks of 100% were approved, to the satisfaction of external examiners, academic and support staff, and students, and it was felt that a real institutional breakthrough had occurred.

After the referral boards in the autumn, the marks were analysed to assess the effects of the introduction on marking practice and student achievement. This showed that there was no dramatic increase in the numbers of marks in the high first range or low fail range given out, and this was reflected in the average marks which rose 1% (from 48.5% to 49.5%) between the two years, and the standard deviation which rose 1.2% (from 22.3% to 23.5%). This was triangulated through scrutinizing data from five courses where there had been an unexpected rise in good honours. Their marking patterns before and after the introduction of Grade Marking showed that, apart from one course, there had been no unexpected jumps in marks for the final year cohort who had been Grade Marked for the first time.

Feedback from the spring evaluation had suggested that some staff were not using the scheme in conjunction with the Generic Grading Criteria, and were marking in numbers, and then changing these to Grade Marks. This was problematic in the first and fail ranges, as some staff were still marking on the numeric scale where, in some subjects at least, a 'excellent' first could be a 74. They were then translating this to the nearest numeric equivalent of 72, rather than using the Generic Grading Criteria to identify what a 'good' first should be (an A3 with a numeric equivalent of 83) and giving A3 as the Grade Mark. Quantitative analysis of marks before and after the introduction of Grade Marking supported this, and showed that there had indeed been a redistribution of marks away from the A3 band to the A4. As a result, the University decided to change the Grade Mark numeric equivalent for A4 to 74, (in fact a return to the original suggested scheme), starting in autumn 2012, and supported this through further staff development work on using marking schemes.

In 2013 the University went through Institutional Review which confirmed that standards around assessment had been maintained, and in 2014 the first cohort of undergraduate students marked entirely in Grade Marks graduated, with management data since the introduction showing no unexpected changes in awards profiles. The introduction of the new marking scheme successfully changed marking practices and maintained standards through spreading marks and by increasing transparency and consistency. Student achievement was rewarded without radically changing awards profiles, and the introduction went ahead more smoothly than expected.

## Discussion

The evaluation highlighted some key challenges in leading initiatives involving assessment. The overall results show that the University has successfully transitioned to Grade Marking without compromising standards and with minimal disruption to staff and students. However, the evaluation showed that while staff were marking in Grade Marks, not all of them were marking as the University intended. Senge distinguishes between two types of learning in a learning organization, compliance and commitment; the former being a superficial response to an initiative and the latter a personal, values driven understanding. Evidently there had been successful compliance, but commitment had not been achieved. However, there is a balance here to be struck. For example, the Grade Marking scheme presented to staff showed the numeric equivalents, on the basis that being transparent about the workings of the scheme would enable staff to understand and therefore accept it faster. However, that knowledge also held back the culture change because it allowed old marking practices to survive. With hindsight, the University should have more seriously considered not sharing the numeric equivalents to academic staff during the roll out.

During the introduction of most university initiatives a situation like this could have been dealt with through further staff development to encourage a committed approach, with a longer period of time before any decisions about the success of the initiative had to be made. However, the introduction of a new marking scheme is very high risk. Compared to other standards issues covered in the QAA's UK Quality Code for Higher Education (QAA 2014), such as credit frameworks, programme monitoring, and programme approval, assessment is one area where a major institutional change might be expected to affect almost all students simultaneously. While piloting or introducing the scheme year by year was considered it was decided that this was even higher risk as it may have increased confusion to have two marking schemes in operation at once, and there was a strong institutional desire to move forward with the initiative to address External Examiner concerns.

The Grade Marking initiative had a very narrow target of changing everyone's marking practice, and modifying it only slightly to extend the mark range. This is an example of "very clear, high stakes outcomes" (Trowler 2008, 151), a situation Trowler cites as directly opposed to one that encourages local adoption. In contrast initiatives that are indicative of a learning organization are open ended, encourage creative responses, risk taking and experimentation (Senge 2006, 191-215) – a scenario that runs counter to the introduction of a marking scheme.

The issue of maintaining standards also meant that there was little time for the initiative to bed in before its success had to be evaluated. One year is a very short amount of time to successfully embed organisational culture change, but is a critical amount of time in terms of maintaining standards relating to marks. The identification of a blip in marking practice meant that this had to be resolved through a policy change, which is a direct way of maintaining standards, but not a very efficient way of generating culture change and exactly the response to cause resistance in academic staff. The longer reporting cycle, over several years, shows that there was little change in award profiles, and it may well be that this would have been the case if the second policy change had not been put in place. It may also be that if feedback had been collected from staff two years into the initiative, low marking in the first

range would have corrected itself through culture change, and would not have been picked up either qualitatively or quantitatively.

Feedback from staff during the consultation had identified deep rooted ideas about 'correct' marking practice such as appropriate marks for low firsts, and the staff development sessions had therefore emphasised the use of Generic Grading Criteria and marking grids as a way of supporting the use of the new marking scheme. However, on reflection greater support could have been given to course leaders in preparing them to take a 'committed' rather than 'compliance' approach in disseminating this within their teams. Given the high risk scenario, more emphasis was placed on making sure that marking practice changed, with 'committed' approaches, where those occurred, considered as an added bonus. And while the University recognized the role of local discussion about marking practice as a successful output of the initiative, it didn't necessarily conceptualise this as an opportunity to counter the negative effects of a necessarily top-down approach.

Some of these issues may have been countered by having a different remit for the Assessment Working Group. In practice, the Working Group consisted of a core of staff who did the research and developed the recommendations, and others whose role was to consider the recommendations' impact on their colleagues. Early on, the Working Group identified key areas of development that effectively set a trajectory of activities, which meant that there was little leeway later in the project for developing new ideas. The role of the Working Group was to advise senior management on what to recommend, and once this was approved via the committee structure, it was disbanded. This was perhaps symptomatic of the institutional approach which saw the key challenge as the approval and implementation of the initiative, rather than the initiative requiring continued pedagogic input. In particular, after approval planning then focused on project management and especially identifying and managing risk. A further remit for the working group may therefore have been to model the opportunities presented by a successful introduction, for example a focus on how to engender institution wide discussions about marking practices in general. In line with Senge's work, this would have created a strong vision for the institution to work towards.

This reveals a lot about the University's understanding of how a learning organization should behave. Its approach, as in all of its initiatives, focused on stakeholder involvement, for example through proactively responding to staff feedback and through a commitment to transparency, both of which the analysis above shows may have had unexpected consequences. But by prioritizing this it overlooked the more managerial aspects of Senge's model in particular the importance of creating a strong vision to engender change which would have had a more profound impact on improving compliance.

The analysis of these findings demonstrates that rolling out institutional change can be effectively project managed, but the challenges of 're-culturing' an organization are consistently underestimated (Fullan 2007, 25), The policy change and the supporting staff development were very successful in terms of a transition to the new marking scheme and minimalizing confusion for students, as well as protecting standards. However, a narrowly defined institution-wide target, implemented through policy and interfering directly with everyday working practices of academic staff runs

counter to the principles of local adaptation and disseminated leadership. Attempts to counter this through an approach to creating a learning organization based mainly on stakeholder involvement were not sufficient. It may have been better to have had a clear vision of a committed organisation and to have continued to develop the initiative towards this.

### References cited

Bloxham, S. 2009. "Marking and moderation in the UK: False Assumptions and Wasted Resources." *Assessment & Evaluation in Higher Education* 34 (2): 209-220.

Bryman, A. 2007. *Effective Leadership in Higher Education*. London: Leadership Foundation for Higher Education.

Burgess, R. 2007. *Beyond the Honours Degree Classification (The Burgess Group Final Report)*. London: Universities UK.

Dalziel, J. 1998. "Using Marks to Assess Student Performance, Some Problems and Alternatives." *Assessment & Evaluation in Higher Education* 23 (4): 351-366.

Fullan, M. 2007. *The New Meaning of Educational Change*. 4th ed. London: Routledge.

Hannan, A. and H. Silver. 2000. *Innovating in Higher Education: Teaching, Learning and Institutional Culture*. Buckingham: Open University Press.

HEA. 2013. Grade Point Average. <https://www.heacademy.ac.uk/workstreams-research/themes/assessment-and-feedback/grade-point-average-gpa>

Milton, O., Pollio, H. R., and J. A. Eison. 1986. *Making Sense of College Grades. Why The Grading System Does Not Work and What Can Be Done About It*. San Francisco, California: Jossey-Bass.

Newstead, S. 2002. "Examining the Examiners: Why Are We So Bad at Assessing Students?." *Psychology Learning and Teaching* 2 (2): 70-75.

Newstead, S. and I. Dennis. 1994. "Examiners Examined." *The Psychologist* 7 (5): 216-219.

Oxford Brookes University. 2012. *Recording Student Achievement: A White Paper for Oxford Brookes*. Oxford: Oxford Brookes University.

QAA. 2014. *The UK Quality Code for Higher Education*. <http://www.qaa.ac.uk/assuring-standards-and-quality/the-quality-code>

Ramsden, P. 1998. *Learning to Lead in Higher Education*. London, Routledge.

Rust, C. 2011. "The Unscholarly Use of Numbers in Our Assessment Practices: What Will Make Us Change?." *International Journal for the Scholarship of Teaching and Learning* 5 (1) <http://digitalcommons.georgiasouthern.edu/cgi/viewcontent.cgi?article=1254&context=ij-sotl>

Sadler, D. R. 2009. "Indeterminacy in the Use of Pre-set Criteria for Assessment and Grading." *Assessment & Evaluation in Higher Education* 34 (2): 159-179.

Sambell, K., McDowell, L., and C. Montgomery. 2012. *Assessment for Learning in Higher Education*. Abingdon: Routledge.

Senge, P. M. 2006. *The Fifth Discipline: The Art and Practice of the Learning Organisation*. London: Random House.

Soh, C. K. 2011. "Grade Point Average: What's Wrong and What's the Alternative?." *Journal of Higher Education Policy and Management* 33 (1) 27-36.

Trowler, P. 2008. *Cultures and Change in Higher Education: Theories and practice*. Basingstoke: Palgrave Macmillan.

Yorke, M. 2011. "Summative Assessment: Dealing with the 'Measurement Fallacy'." *Studies in Higher Education* 36 (3) 251-273.

Yorke, M., Barnett, G., Bridges, P., Evanson, P., Haines, C., Jenkins, D, Knight, P., Scurry, D., Stowell M., and H. Woolf. 2002. "Does Grading Method Influence Honours Degree Classification?." *Assessment & Evaluation in Higher Education* 27 (3): 269-279.

Yorke, M. 2008 *Grading Student Achievement in Higher Education: Signals and Shortcomings*. Abingdon: Routledge.

Yorke, M. 2010. "How Finely Grained Does Summative Assessment Need to Be?." *Studies in Higher Education* 35 (6) 677-689.

Zuber-Skerritt, O. 1992. *Action Research In Higher Education: Examples and Reflections*. London: Kogan Page