

**Improving Work Based Assessment:
Addressing grade inflation numerically or
pedagogically?**

Practitioner Research in Higher Education
Special Assessment Issue
Copyright © 2018
University of Cumbria
Vol 11(1) pages 80-86

Joy Robbins, Amanda Firth & Maria Evans
University of Bradford

Abstract

Work based assessment (WBA) is a common but contentious practice increasingly used to grade university students on professional degrees. A key issue in WBA is the potentially low assessment literacy of the assessors, which can lead to a host of unintended results, including grade inflation. We identified grade inflation in the WBA of the clinical module analysed for this study, and to address it we trialled two adjustments over a four-year period. The first and simpler adjustment, reducing the academic weighting of the WBA component of the module, appeared to lower grade inflation but actually had the inverse effect over time. The second adjustment, introducing a structured formative assessment, reduced the average WBA grade both initially and over time. In addition to this desired result, the second adjustment has brought ongoing benefits to the learning and teaching on the module as a whole.

Keywords

Work based assessment; grade inflation; assessment literacy; clinical education; formative assessment; workplace mentors.

Introduction

Work based assessment (WBA), or workplace or practice based assessment as it is variously known in some fields, is both a highly necessary and a growing practice across the Higher Education sector. In the UK, professional bodies in medical and related fields have increasingly required that the degrees they accredit contain significant observational assessment of students' performance in the clinical setting. At the same time the government agenda to raise apprenticeships to 3 million learners by 2020 (BIS, 2015) has driven the creation of new apprenticeship degrees and therefore new and more work-based assessment. While this form of assessment is on the rise, it nonetheless remains one of the more contentious approaches to measuring student ability.

Problems with WBA

The contention arises largely from the inherent schism of WBA: on the one hand, students are assessed at work by workplace mentors, usually on how well they have been observed to perform certain skills, and on the other hand, outside of the work environment, they are assessed by academics on how well they evidence broader learning outcomes that must incorporate their work performance. The validity of the latter is dependent on the former, and brings up a host of issues: the need for reliable assessment to be carried out by busy workplace professionals who may not have the time nor the educational training to assess students, the possibility of 'halo' effects or other bias when the workplace assessor and the student are essentially colleagues, and the need for objective measurement of performance in varied and potentially stressful real-life circumstances. Even in the field of medicine, arguably the forerunner of degree-level WBA, it has been (and perhaps still is) viewed with suspicion as "reductive 'tick-boxing' approaches to assess the complexities of

Citation

Robbins, J., Firth, A. and Evans, M. (2018) 'Improving Work Based Assessment: Addressing grade inflation numerically or pedagogically', *Practitioner Research in Higher Education Journal*, 11(1), pp. 80-86.

professional behaviour” (Academy of Medical Royal Colleges 2009). Studies have also found it to be highly problematic: too unreliable for summative assessment (Murphy et al., 2009), incorrectly grounded in quantitative psychometrics that cannot provide validity to what is inherently a socially-situated and therefore qualitative judgement (Govaerts and van der Vleuten, 2013), and sometimes surprisingly illogical in that the wrong people are asked to assess the wrong things (Crossley and Jolly 2012).

Within the field of nursing and midwifery, the focus of this practice evaluation, these WBA issues are well documented. For the past decade, the field’s professional regulatory body, the Nursing and Midwifery Council (NMC), has required student nurses and midwives to spend at least 50% of curriculum hours in clinical practice (NMC, 2009), wherein their performance must be summatively assessed by a registered nurse or midwife who has undertaken a recognised programme of training in assessment and mentorship (NMC, 2008). Yet despite this longevity, prominence, and required training for WBA, the above noted issues remain. Nursing and midwifery workplace mentors speak of difficulties in balancing the multi-faceted mentor/ assessor role in addition to their responsibilities in caring for patients and service users (Aston et al., 2014; McIntosh et al., 2014). Some even report that they view assessment as a potentially punitive part of their mentorship role (Bennett and McGowan, 2014) and a survey by Bray and Nettleton (2007) found that only 5% of clinical mentors named assessment as one of the most important parts of this role. All these findings point to an underlying issue across the disciplinary fields: we must use WBA but the assessors delivering it may not value it, trust it, or make sufficient time for it; such assessors are likely to have low assessment literacy.

Grade inflation within WBA

An unfortunate if unsurprising result of workplace mentors’ problematic relationship with WBA is grade inflation. It is well noted that clinical education students receive much higher marks in WBA than they do in their accompanying academic assessment. For example, within nursing, studies have found a four to one failure/referral rate for theory versus practice nationally (Hunt et al., 2012), and ongoing problems with ‘failure to fail’ student nurses in practice internationally (Hughes, Mitchell and Johnston, 2016).

There are many reasons why WBA is particularly susceptible to grade inflation (Duffy 2003; Scholes and Albarran 2005; Wilbur et al. 2017), and on top of this the ‘tripartite’ assessment relationship of student + workplace mentor + academic assessor adds multiple layers of complexity, from competing priorities between these different stakeholders (Norcini et al., 2011) to students’ manipulation of the different power structures at play (Hunt *et al.*, 2016).

In our own pre-registration midwifery degree at the University of Bradford, we found this same issue of grade inflation in the WBA component. An initial audit of year 1 (level 4) marks revealed that clinical placement grades were increasing every semester, from an average grade of 74.1% at the start of 2012 to an average grade of 80.2% by the end of 2013, while the concurrent academic grades did not show a similar increase. We should note here that this programme uses the typical mark range of UK degrees: 40-49% is a pass and considered a ‘3rd class’ degree mark, 50-59% is a ‘lower second class’ mark, 60-69% is an ‘upper second class’ mark which is awarded for work that is very good, and 70% and above is a ‘first class’ mark for work that is superlative. In other words, the average WBA grade in 2012 was, at 74.1%, already contentiously high, and the following year’s average grade of over 80% was a serious cause for concern.

Our evaluation of practice

We clearly needed to improve the WBA process, and we therefore implemented two consecutive adaptations, described below, in an effort to address the disparity between marks that students

achieved in clinical practice assessments and academic assessments. To evaluate the impact of each adjustment, grade data were collected and interrogated from the start of the audit to the present. Ethical approval was granted for the evaluation of clinical module grades. Student midwives on the programme completed 2 x 10 week clinical placements per year (named 1a and 1b in this paper) with a summative assessment in the final week of placement. Data was extracted from year 1 (level 4) of the programme, tracking the mean grades awarded per module across 5 different cohorts. Cohort sizes ranged from 35-43 across the 5 year data-period, with numbers varying due to allocated programme capacity per year, attrition due to illness, withdrawal from the programme or periods of intercalation. Student midwives in this sample were all undertaking clinical placement assessments in their 1st year of midwifery study and were inexperienced healthcare practitioners.

Throughout this evaluation period the Midwifery 1a and 1b modules comprised of 2 assessment elements: the grading of clinical practice by midwife workplace mentors (the WBA), and the grading of an eportfolio of development and reflection by midwife academics. In both cases, the students collected their reflections, evidence, and other relevant work in their individual eportfolios, and the grades were recorded electronically by the assessors also within the eportfolios. Students in clinical practice were graded using a criterion based rubric and awarded a grade out of 100% by workplace mentors. The same marking rubrics were used throughout the evaluation period and all clinical assessments took place either in a community midwifery or antenatal/postnatal ward setting. For the academic mark, students' portfolios of development and reflection from the clinical placement were assessed by academics and work was graded using an agreed level 4 portfolio marking rubric which was not altered during the entire evaluation period.

Adjustment 1 – Changing of Assessment Weighting

In 2012, the beginning of this evaluation period, the Midwifery 1a and 1b overall module grade was calculated using a weighting of 70% assessment of clinical practice by a workplace mentor and 30% academic assessment of eportfolio. Figure 1 demonstrates that during this period the average module grade rose steadily from 72.8% in 2012 to 80.2% in 2013. Adjustment 1 was implemented in an attempt to counteract this grade inflation and consisted of a change in the weighting of the assessments from 70%/ 30% to an equal 50%/ 50%, with a greater amount of the module grade now being attributed to the student's academic work in this practice module. Figure 1 shows that although the mean module grade initially reduced, within a year the mean grade had risen to 78.8%, almost as high as when the decision to adapt the WBA weighting had been taken. As discussed below, we realised this numeric approach to fixing the grade inflation was overly simplistic: the base issue of the workplace mentors' engagement with assessment and assessment literacy remained, and the impact of this on grades quickly returned. We therefore decided to implement a more involved change to the assessment process, whilst also maintaining the 50%/ 50% assessment weighting already established.

Adjustment 2 – Introduction of Structured Formative Feedback

Qualitative student placement evaluations revealed that many students only received a cursory intermediate interview (a mandatory formative feedback discussion at the mid-point of the clinical placement), and some students disclosed a lack of trust in this feedback opportunity. This distrust was justifiable: at this point (in 2014) the eportfolio prompt for the intermediate interview feedback was an open textbox which asked mentors to consider the learning outcomes and assessment targets for the student and comment on their progress. Some mentors did not complete this and some mentors gave vague feedback that was neither constructive nor helpful to the student, in all probability because they did not adequately understand the assessment process. Literature supports this, stating that mentors frequently give generic formative feedback to conceal their own difficulty in understanding learning outcomes and educational jargon in assessment documentation (Scholes

and Albarran 2005). Our design in this showed the common WBA mistake of asking the wrong person to assess the wrong thing (Crossley and Jolly, 2012).

To address this, adjustment 2 was implemented in 2014 and consisted of the introduction of a criterion based formative feedback rubric, which was identical to the summative feedback rubric but with the removal of any numerical scores. Workplace mentors were asked to indicate which column the students were currently working within for each area of competency, with numerical grades in columns replaced by adjectives such as ‘unsafe, safe, good’ rather than numbers. The aim of this formative use of the summative marking scheme was to both improve the feedback given to the student and to increase the workplace mentor’s familiarity with the learning outcomes and competencies being assessed at the summative point. The adjustment was designed to give repeated exposure and practice in using the grading tool in an effort to enhance the reliability of workplace mentors’ grades (Heaslip and Scammel 2012), while at the same time avoiding any additional grading for the students lest they become grade-centric and reduce or increase their future efforts depending on their contentment with the grade, as per Lefroy et al. (2015). (It should be noted these same authors offer an alternative perspective that students who don’t agree with a grade awarded at the formative assessment, or who are deemed ‘borderline’, are challenged to improve, needing to work through that dissonance prior to the summative assessment point).

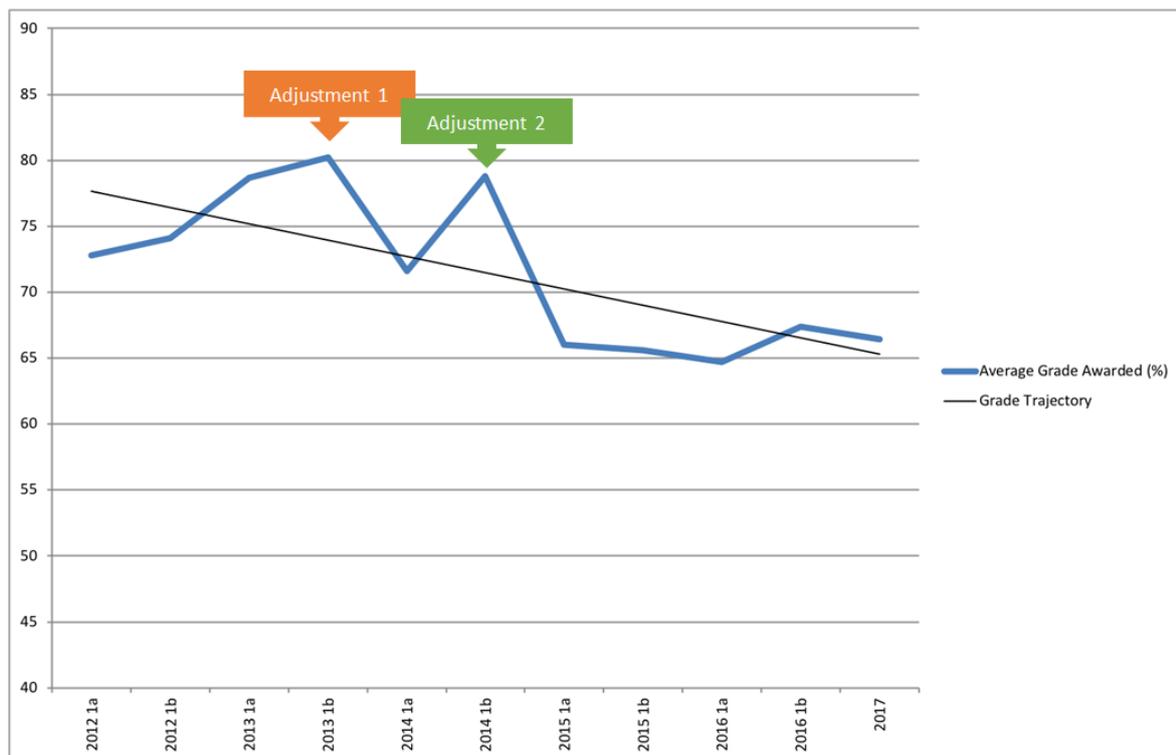


Figure 1. Effects of assessment adjustments on summative grades.

Figure 1 demonstrates that since this adjustment to the WBA process, the year 1 grade average has reduced significantly and is now consistently in the mid-60% range, which is where we want good UK university marks to be. The combination of these two adjustments to the WBA process – though particularly the second – has therefore improved the clinical assessment, successfully lowering grade inflation.

Discussion

Adjustment one, the altering of the assessment weighting to increase the importance of the academic proportion of the work, did reduce the average grade as hoped but its success was short-lived, affecting grades for only one academic year and quickly proving to be an unreliable approach in isolation. Upon reflection, our attempt to fix the grade inflation through numeric adjustment demonstrates a level of naivety in trying to enhance assessment validity by attending to only the assessment tool, ignoring the people involved. While the academic team and the students knew about the assessment change, we did not adequately involve the workplace mentors in discussions about grade inflation. Wu et al. (2015) recognise this as a common issue in the dichotomy of academic assessment versus work based assessment, stating that academics may design the assessment tools but that it is mentors in practice who implement the assessment. Bindall et al. (2011) supports this, stating that WBA tools are often misused, with both assessors and trainees viewing them as hurdles to be jumped. An important consideration is the effect of the workplace mentor/ student relationship. As discussed previously, there is subjectivity in the assessment process, with mentors potentially influenced by student's expectations of a particular grade, mitigation for known personal issues in the student's life or 'giving the benefit of doubt' to a weaker student (Duffy 2003). By making the academic grade more important than it had previously been to students' overall summative mark, we may have incentivised some students to focus even more on what they felt they could influence – their mentor's subjective assessment of their workplace performance – essentially making the WBA more pressured than it had been before. Used on its own, a numeric approach to addressing WBA grade inflation would appear to be a bad bet; it certainly backfired for us.

Adjustment 2, the formative use of the final assessment rubric but without any grades and within an already-required scaffolded conversation ('intermediate interview'), appears on the whole to be much more successful. Because this adjustment required a change to the eportfolio page for the formative WBA point – replacing an open textbox prompt for formative feedback with the ungraded rubric – we had to reach out to the workplace mentors to advise them of the change, and this prompted useful discussions around why we were changing the assessment and what the assessment was for. In hindsight, these discussions would have been a good place to start. The formative rubric has also increased tripartite conversation between academics, students and workplace mentors, resulting in weaker students being identified sooner and receiving criterion-based feedback and clear developmental action plans for the remainder of their placement.

By moving the emphasis from assessment *of* learning to assessment *for* learning, adjustment 2 not only raised the assessment literacy of the workplace mentors, it also helped all stakeholders negotiate a more shared vocabulary around assessment. We found for example that mentors were able to better explain and qualify what had previously been more intuitive judgements: instead of saying things like 'she's not a midwife but I can't explain why not', they could point to clear areas where there were issues. This move towards shared clarity of assessment standards achieved through dialogue, with a focus on feedback over grades, has long been called for in the pedagogic literature (Gibbs, 2006; Hattie and Timperley, 2007; Price et al. ,2008) and has a clear common-sense appeal to anyone with an educational background. Nonetheless, in the time-pressed circumstances of WBA, often replete with professional-body stipulated tick boxes and other above-noted barriers to meaningful assessment, it can be tempting to take shortcuts. We note here for the benefit of other practitioners' line managers that our 'shortcut' adjustment 1 actually caused more harm than good, and that our more pedagogically sound adjustment 2 was not that difficult: we simply swapped out a textbox asking for feedback about learning outcomes with a formative rubric. The resulting dialogue is still ongoing, as are the resulting enhancements to workplace mentors' assessment literacy. There is no denying all this dialogue takes time, but it has also saved us some assessment surprises, some exam board appeals, and some untold stress.

Conclusion

Overall, the adjustments to WBA we trialled have successfully lowered grade inflation on our programme, and it is hoped other practitioners will be able to apply our lessons learnt to their own WBA. Our approach confirms previous findings about the importance of reviewing grading documentation and ascertaining a common understanding of assessment language amongst all parties (Cassidy et al. 2012), and of working in partnership with workplace mentors. A numeric approach to lowering WBA grade inflation is discouraged, unless used in conjunction with a pedagogic approach that increases structured formative assessment. The ensuing collaborative conversations about assessment and grading clinical practice that result from this not only effectively reduce grade inflation, but also put the focus squarely back where it should be in WBA: learning and development.

Limitations and Recommendations

The scope of this audit and evaluation only considers module grades for 1 cohort per year, for 1 programme over a 5 year period. Full consideration of all 3 cohorts over 5 years may have presented different data and this is acknowledged as a limitation. It is recommended that further research is undertaken on the value of using criterion-based formative feedback which mirrors the summative assessment in terms of language and format. This paper agrees with the work of Heaslip and Scammell (2012) who suggest that reliability and validity of WBA is improved when assessors have repeated exposure and practice in using grading tools.

References

- Academy of Medical Royal Colleges (2009) *Improving Assessment*. Available at: http://www.aomrc.org.uk/doc_view/49-improving-assessment (Accessed: 10 April 2018).
- Aston, E., Aston, L. and Hallam, P. (2014) *Successful mentoring in nursing*. Exeter: Learning Matters.
- Bennett, M. and McGowan, B. (2014) 'Assessment matters—mentors need support in their role', *British journal of nursing*, 23(9), pp. 454-458.
- Bindal, T., Wall, D. and Goodyear, H. M. (2011) 'Trainee doctors' views on workplace-based assessments: Are they just a tick box exercise?' *Medical teacher*, 33(11), pp. 919-927.
- BIS HM Government (2015) *English Apprenticeships: Our 2020 Vision*, London: Department for innovation business and skills.
- Bray, L. and Nettleton, P. (2007) 'Assessor or mentor? Role confusion in professional education', *Nurse education today*, 27(8), pp. 48-855.
- Cassidy, I., Butler, M. P., Quillinan, B., Egan, G., Mc Namara, M. C., Tuohy, D., Bradshaw, C., Fahy, A., O'Connor, M. and Tierney, C. (2012) 'Preceptors' views of assessing nursing students using a competency based approach', *Nurse Education in Practice*, 12(6), pp.346-351.
- Crossley, J. and Jolly, B. (2012) 'Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people', *Medical education*, 46(1), pp. 28-37.
- Duffy, K. (2003) *Failing students: a qualitative study of factors that influence the decisions regarding assessment of students' competence in practice*. Glasgow: Caledonian Nursing and Midwifery Research Centre.
- Gibbs, G (2006) How assessment frames student learning, in: Bryan, C., Klegg, K. (eds.) *Innovative Assessment in Higher Education*. London: Routledge, pp. 23–36.
- Govaerts, M. and Vleuten, C. P. M. (2013) 'Validity in work-based assessment: expanding our horizons', *Medical education*, 47(12), pp. 1164-1174.
- Hattie, J. and Timperley, H. (2007) 'The power of feedback', *Review of educational research*, 77(1), pp. 81-112.
- Heaslip, V. and Scammell, J. M. E. (2012) 'Failing underperforming students: The role of grading in practice assessment', *Nurse Education in practice*, 12(2), pp. 95-100.
- Hughes, L. J., Mitchell, M. and Johnston, A. N. B. (2016) "'Failure to fail" in nursing – A catch phrase

- or a real issue? A systematic integrative literature review', *Nurse Education in Practice*, 20, pp. 54-63. doi: 10.1016/j.nepr.2016.06.009.
- Hunt, L. A. *et al.* (2012) 'Assessment of student nurses in practice: A comparison of theoretical and practical assessment results in England', *YNEDT*, 32, pp. 351-355. doi: 10.1016/j.nedt.2011.05.010.
- Hunt, L. A., McGee, P., Gutteridge, R., and Hughes, M.. (2016) 'Manipulating mentors' assessment decisions: Do underperforming student nurses use coercive strategies to influence mentors' practical assessment decisions?' *Nurse Education in Practice*, 20, pp. 154-162. doi: 10.1016/j.nepr.2016.08.007.
- Lefroy, J., Hawarden, A., Gay, S. P., McKinley, R. K. and Cleland, J. (2015) 'Grades in formative workplace-based assessment: a study of what works for whom and why', *Medical education* 49(3), pp. 307-320.
- Murphy, D.J., Bruce, D.A., Mercer, S.W., and Eva, K.W. (2009) 'The reliability of work-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom', *Advances in Health Sciences Education*, 14(2), pp. 219-232.
- McIntosh, A., Gidman, J. and Smith, D. (2014) 'Mentors' perceptions and experiences of supporting student nurses in practice', *International journal of nursing practice*, 20(4), pp. 360-365.
- NMC (2008) *Standards to support learning and assessment in practice*. London, Nursing and Midwifery Council.
- NMC (2009) *Standards for pre-registration midwifery education*. London, Nursing and Midwifery Council.
- Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., Galbraith, R., Hays, R., Kent, A. and Perrott, V. (2011) 'Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference', *Medical teacher*, 33(3), pp. 206-214.
- Price, M., O'Donovan, B., Rust, C., and Carroll, J. (2008) 'Assessment Standards: A Manifesto for Change', *Brookes eJournal of Learning and Teaching*, 2(3). Available at: http://bejlt.brookes.ac.uk/paper/assessment_standards_a_manifesto_for_change-2/. (Accessed: 10 April 2018).
- Scholes, J. and Albarran, J. (2005) 'Failure to fail: facing the consequences of inaction', *Nursing in Critical Care*, 10(3), pp. 113-115.
- Wilbur, K., Hassaballa, N., Mahmood, O. S. and Black, E. K. (2017) 'Describing student performance: a comparison among clinical preceptors across cultural contexts', *Medical Education*, 51(4), pp. 411-422.
- Wu, X. V., Enskär, K., Lee, C. C. S. and Wang, W. (2015) 'A systematic review of clinical assessment for undergraduate nursing students', *Nurse Education Today*, 35(2), pp. 347-359.