# Chapter 2 Exploring the Relationship between Ipsative Assessment and Institutional Learning Gain

Gwyneth Hughes

Pre-publication draft from:

Hughes. G. (Ed.) (2017) *Ipsative Assessment and Personal Learning Gain: Exploring international case studies* Basingstoke: Palgrave Macmillan.

## Overview

This chapter explores in more detail how the concepts of ipsative assessment and personal learning gain are related to macro-level learning gain measurement, yet are largely absent from current educational discourse. While learning gain has currency as part of institutional monitoring, evaluation and ranking, the action of an individual learner in capitalising on information about their learning gain – for example through an ipsative assessment - is neither encouraged nor is it visible in competitive and selective assessment systems and schemes. As we have seen in the opening chapter, a fully ipsative approach which rewards progress or learning gain as much as achievement for all levels and backgrounds of learners would require considerable assessment reform. However, that is not to say that there are not countless small steps toward this vision being undertaken in educational institutions across the globe. This chapter will provide a conceptual and wider contextual basis for the book to support the professional evidence base for encouraging ipsative assessment in schools, colleges and universities.

The first section will examine why learning gain measurements for schools and universities are collated and how these provide a more just and equitable indication of teaching effectiveness than exit performance data alone. The limitations of learning gain measurement at scale will then be explored to include the difficulty of standardising tests and questions about data reliability alongside consideration of the desirability and value of national and global institutional performance rankings that can perpetuate elitism. Other related concepts that will be briefly covered including the discourse and practice of continuous improvement and use of learning gain data for research in learning and teaching.

The second section of the chapter will consider how not only similar but also different concerns confront the assessor of individual students. Using ipsative marking or measurements has an equalitarian purpose to motivate all learners, but that means it is problematic for credential-based systems of selection because it gives those with more to gain rather than top performers an advantage. Employers might also need convincing to accept ipsative credentials. The challenges of comparing two or more assessments will persist for learning gain at the individual level. Nevertheless, ipsative assessment might encourage high achievers to raise their game further so that the resistance from an educational elite can be tempered.

The range of purposes for gathering learning gain data will then mapped out for the different learning gain audiences along a micro-level to macro-level continuum. The chapter will end with an explanation of why, despite increasing recognition of learning gain as a concept, ipsative assessment is underused and largely invisible.

**Exploring the discourses and practices of large-scale learning gain measurement**

Exit performances of students are measured, averaged and widely used for comparing teaching effectiveness in schools or colleges. The data can also be used to produce rankings or comparative league tables of institutions and made publically available. In the UK, performance assessment arising from government-led inspections and end of course examinations is freely available online and influences school popularity and status. Meanwhile, international rankings of universities are gathered and published by a range of bodies such as QS World University Rankings and the Times Higher Education World University Rankings, which are also made public.

However, institutions which are selective or which have a greater proportional of academically and socially advantaged students will predictably outperform those which are less advantaged. Use of learning gain or value-added data may enable a more equitable comparison.

*Advantage of using learning gain rather than exit performances for institutional evaluation and ranking*

A teacher with engaged and capable, independently-working students is likely to obtain good outcomes unless the teaching is extremely inappropriate. Meanwhile, dedicated teachers are unlikely to get excellent results all round with more challenging students who struggle to spend time and effort on learning for whatever social and psychological reasons. Learning is complex business that cannot readily be captured by outcomes alone given that there are so many influences from students' broader social learning contexts: family, previous education, leisure pursuits etc. (Bloomer & Hodkinson, 2000). Exit performances do not necessarily give a good picture of teaching effectiveness.

Whether competition is global or local, institutional rankings tend to favour the elite. For example, the prominence of knowledge production and research in higher education rankings means that research intensive institutions invariably sit at the top of the tables, although the growing range of ranking categories, such as breaking down data into disciplines, means that many institutions can strategically highlight areas in which they do well (Marginson, 2009). The value of these rankings have been widely critiqued, yet rankings continue to be supported by those institutions which top the charts and less prestigious institutions aspire to improve their national and global ranking and reputation. University rankings are often given high national priorities (Hazelkorn, 2015).

However, lower performing institutions can demonstrate effectiveness through the value that they add for students. Recording learning gain has an advantage over using one-off exit performances for evaluation of teaching and learning because measuring a growth in performance can make allowances for teachers or schools which have high numbers of students who start from a lower base (Liu, 2011; Steedle, 2012).The baseline can be factored in so that a school or college with a low achieving intake could demonstrate a higher than expected improvement –indicative of successful teaching - without necessarily producing high outcomes. Such 'value-added' is a type of learning gain measurement made when predicted outcomes based on the first measurement are compared to actual outcomes. Improvements added by different teaching methods or schools are then comparable - although only schools at the extremes of high and low performance tend show meaningful effects (Goldstein et al. 2014).

Although learning gain and value added have been mainly applied to schools, the idea of measuring learning gain or learning growth as part of quality monitoring is taking root in higher education. There has been growing interest by governments, educational leaders and

educators in improving student engagement with learning at their institutions and in methods of benchmarking effective institutional teaching practice (Coates & McCormick, 2014). At present, indirect and questionable proxies are used as measures of teaching quality such as staff-student ratios, contact hours, completion rates or student evaluations of teaching, while learning gain might provide a more direct and meaningful measure of teaching quality (HEFCE, 2015). Nevertheless, the UK and many other countries do not yet use measurement of learning gain to compare different departments or institutions as part of an accountability system. Perhaps research-led elite institutions oppose more robust measurements of teaching quality which might shift rankings in favour of 'teaching' institutions. Such resistance could be exacerbated by a lack of standardisation of qualifications. In the UK for example, there are many ways of making decisions about boundaries between first, second or third class achievements, and so degree performances are difficult to compare across the higher education sector whereas at secondary level, where there are national examination systems, comparisons can be more robust.

*Measuring learning gain in higher education – a need for standardised tests*

Two systems from the USA that could be used to overcome the lack of standardisation in higher education are the grade point average (GPA) system and the Collegiate Learning Assessment (CLA).

A grade point average (GPA) system might provide more reliable comparison of learning gain between institutions. GPA is exactly what it sounds like – an average of all grades a student obtained on a programme of study. GPA is the method of classification of degrees in the US and is widely recognised internationally and is being explored as an eventual replacement for the degree classification system in the UK (Higher Education Academy,

2015). However, a GPA is not equivalent for different disciplines where different skills are being assessed and so would not be a valid measure for comparing institutions with a range of disciplines - this would require a generic standardised test.

There are standardised tests for undergraduates available which students can sit on entry and on exit of their course so that their learning gain can be measured. These tests are not discipline specific and instead measure a more general range of graduate attributes. Although exactly which skills a graduate should have can be readily disputed, there is broad agreement that all higher education aims to develop critical thinking as well as communication skills (speaking, reading and writing). The Collegiate Learning Assessment CLA was developed in the US and differs from other general ability tests like the Scholastic Aptitude Test (SAT) as it measures broad discipline independent competencies such as critical thinking and problem solving in real world contexts.

*What learning gain data can potentially reveal about student learning*

Use of learning gain data for quality assurance can backfire if there is the possibility of revealing large-scale inadequacies in an educational system. For example, Arum & Roska (2011) analysed large amounts of data from the CLA and, after making allowance for prior experience, used the data to claim that the majority of US students do not show significant learning gains in key graduate skills in the first two years of college, and are unlikely to gain much in the following two years. Students are they say "academically adrift":

> Many students come to college not only poorly prepared by prior schooling for highly demanding academic tasks that ideally lie in front of them, but - more troubling still

> they enter college with attitudes, values, norms and behaviours that are at odds with academic commitment. (p.3)

After taking account of school and social backgrounds, Arum & Roska (2011) used learning gain CLA data to suggest that time spent on study, for example, reading for more than 40 hours per week, predicts a high learning gain especially while working alone. They argue that many students are distracted by the social and networking aspects of college life, which may well have benefits and instil confidence, but which also take away time from study. Even group work, which has become a popular way of engaging learners in study, may give students social opportunities rather than provide academically demanding work. Nevertheless, the significance of these results may be overplayed. Being side-tracked from study is nothing new, meeting friends and partners has long been an important aspect of university life for full time students, and balancing family and work commitments has long posed challenges for part-time and professional learners.

Arum & Roska concede that there are wide variations in learning gain even within institutions demonstrating that the factors which enhance learning are highly complex. They argue that because a minority of students do develop the expected higher order skills this indicates that the CLA provides some useful comparative measurement. But 'before and after' standard tests have limitations: because of the many variables and possible unknown factors, all learning gain and value-added measurements are subject to unreliability.

*Reliability and validity of learning gain measurements*

Learning gain measurement for high stakes purposes such as evaluating and comparing institutional performance needs sophisticated statistical methods and models because there

are many variables that might influence the results. For example, the growth rate of those starting from a low base is not necessarily equivalent to the growth rate of those starting from higher up: it could be that starting low means that there is more to gain. Alternatively, those who are already advantaged from the start may progress more than those who are disadvantaged. Previous teaching and learning experience may influence measured learning growth as might class size, the selectivity of the institution or the resources available to learners. Although statistical methods can compensate to some extent for these variables using covariate and multivariate statistical models (Anderman et. al, 2015; Lui, 2011), the uncertainty over reliability suggests that learning gain measurement should be used with caution as different statistical methods of measuring learning gain can produce very different results. To improve reliability, it is generally better to have many cycles of gain rather than one wave. Combining statistical models with other qualitative methods such as self-assessment against planned learning outcomes, standardised tests, skills audits, personal development portfolios, student evaluations and graduation or persistence rates also helps (Anderman et. al, 2015; McGrath et al., 2015; Steedle 2012).

Others question the validity of CLA measurements and other generic systems and argue that the range of skills that are being tested is limited and not necessarily helpful. Critics have argued that the CLA test is not related enough to the specific knowledge taught in degree courses, is too generic and in any case measures prior learning rather than learning gains at university (Lodge & Bonsanquet, 2014). The generic higher order skills may be discipline dependent and so the CLA is not necessarily a good proxy for overall learning. For example, those taking a philosophy course might develop critical thinking more readily than applied sciences students irrespective of teaching quality. As Hazelkorn (2015) has identified there is a trade-off between measuring what is easy and measuring what is meaningful.

It could be that learning gain is more valuable as an aspiration rather than a reliable measure of teaching quality. Such an aim to enhance or improve performance is also captured in discourses of continuous improvement and quality enhancement in which institutions self-monitor their performance. There are some useful points to be made next by locating learning gain within the concept of continuous improvement as well as some more caveats.

*Learning gain as continuous improvement – raising the bar too quickly?*

Learning gain measurement is part of a wider discourse of continuous improvement. Continuous improvement started in the 1930s and uses data on outputs, originally mainly statistical, to identify how a system is functioning and how the performance of the system can be improved (Dew & Nearning, 2004). Continuous improvement can apply to a range of organisations from manufacturing to health and education. For example in education, performances of departments from student evaluations can be recorded over time with the aim of demonstrating year on year improvement in performance. Such continuous quality improvement or quality enhancement can take place at the micro-level as small changes, or these changes can have a multiplication effect so that large-scale institutional shifts occur.

One example of continuous improvement from a micro-level practitioner perspective is action research. It consists of a cycle of planning for change, taking action, observation and reflection leading to another cycle of improvement (McNiff & Whitehead, 2006). Historically it has been practitioners who led action research with a dual aim to improve practice and contribute to knowledge (Norton, 2008). Educational action research is also manifest in terms and approaches such as ‘teacher-as-researcher’ and ‘reflective practitioner’ (Schön, 1991).

However, we might argue that maintaining the status quo is enough of an ambition for an educational organisation and that continuous improvement whether through quality outcomes measures or action research may be unachievable especially if external conditions are unfavourable. For example, my institution has recently undergone a merger with a larger university and a reversal or halt in continuous improvement trends is expected during a period of instability. Emphasising continuous improvement for teachers or even individual students could be equally problematic and for me produces a certain amount of unease.

Learning and improvement for an individual seem on the surface to be tautologous. If there is no improvement how has learning taken place? Nevertheless, applying the concept of continuous improvement to an individual may not always be desirable or possible. Always aiming for a new personal best in a particular skill or learning outcome may be unrealistic and learning may sometimes be better viewed as keeping up a level of achievement as circumstances shift or even deteriorate. For example, an elderly person may experience reduced physical or mental capacity and maintaining independence may require considerable effort and learning. A student may encounter a limit to time and resources available for learning as a consequence of illness or increased demands from the workplace or family. For such a student temporarily preserving the status quo may be a challenging enough goal (Hughes, 2014). But, it is implied in commonly used terms such as ‘continuing professional development’ and ‘teaching excellence’ that teachers should strive to develop their teaching throughout their careers in response to external changes. For example, emergence of new technologies alongside a continual pressure for greater efficiency has fuelled the mantra ‘do more with less’. Continuous improvement as an aspiration may motivate both staff and students, but at times simply standing still may be enough, and institutional leaders need to apply brakes to unrealistic upward trajectories of achievement.

Yet at the same time students or teachers who are underperforming for whatever reason do need to be identified and supported. Thus, there is a distinction between recording an absence of continuous improvement that is to be expected because of shifts in external circumstances, and recording lack of improvement which is not expected and warrants intervention or further action. Viewing learning gain as an aspiration for individual or collective continuous improvement illustrates further that the interpretation of learning gain measurements is highly context dependent and this point is also well illustrated by the use of learning gain data in pedagogic research.

*Educational research and learning gain*

A final and rather different use of learning gain data is for educational research. Some educational practitioners and researchers use learning gain measurements to compare the effectiveness of teaching methods. For example, learning gain measurements in medical education in Australia were used to compare a small group who completed a task together as a unit and a group where the task was divided into sub-tasks with peer teaching of each sub-task to the remainder of the group. Results indicated that the latter produced a greater and more lasting learning gain (Kooloos, et al, 2011).

Another study in Hong Kong used self-reported learning gain measurements to demonstrate that higher education students showed learning gain not only in subject knowledge, cognitive and intellectual skills, but also in personal, social and cultural matters (Tam, 2004). Results indicated that those who engaged most with peers and teachers demonstrated most learning gain which is not surprising for developing social and personal skills. But, these findings do not complement those of Arum & Roska (2011) mentioned earlier who argued that social contact alone is not sufficient for development of high-level intellectual skills and that

independent study and time spent on task are the key factors. However, given that the contexts of these studies are different we might expect different findings. There is also a question here about the reliability of self-reporting.

Because pedagogic studies like this are often localised, small scale and highly context dependent they may have a limited effect on enhancing teaching and learning more generally. Furthermore, many busy teachers may not have time to gather and scrutinise their own learning gain data and adjust their teaching accordingly. Learners have even less access than their teachers to quality personal learning gain data as part of assessment processes as we shall see next.

**Ipsative assessment as personal learning gain**

So far the discussion has been mainly of learning gain measurements in the hands of governments, accountability bodies and managers who monitor school and university and teacher performances to identify underachievement so that it can be rectified, or to showcase and benchmark excellence. While learning gain data may have some advantages over learning outcome measurements in helping governments monitor education to assure that it delivers value for money and is fit for purpose, there is no guarantee that any data gathering exercise will improve teacher performance and/or school effectiveness. Critics have questioned how far quality control over education is desirable and empowering for teachers who are caught up in the machinery of accountability and performativity which at times may seem divorced from educational goals and the professionalism of teachers (Ball, 2013). But could learning gain be judged at the individual level?

*Learning gain monitoring for teachers and student: from large scale data-sets to personal learning gain*

Measurement of learning gain for comparative and quality monitoring purposes synthesises and averages out data from often very large numbers of learners. Such methodologies using data analysed using statistical modelling is not accessible to learners and may not provide sufficient granularity to help individual teachers or students. That the large-scale methods of measuring learning gain do not usually provide personal data for learners that can motivate and help them plan their learning is I suggest a missed opportunity.

But, learning gain information does not have to be linked to the management of education if teachers and students themselves can identify what I term 'personal learning gain' through comparing individual marks or achievement of learning outcomes over time as part of an assessment process. There are other forms of qualitative personal learning gain information such as ipsative feedback on progress for self-improvement that individuals could also use. However, this kind of grass-roots personal learning gain or ipsative activity is not usually recorded in educational reports or institutional data. We shall explore later why personal learning gain is hidden. First, we need some discussion about the acceptability of ipsative summative assessment using personal learning gain information.

*Ipsative marking – is it feasible?*

Ipsative summative assessment occurs when a high stakes measurement of personal learning gain is recorded. It is more controversial than measuring learning outcomes and is at present unlikely to be used at any level of education where the main purpose is to measure attainment for a qualification. This is because professional and government agencies who manage

qualifications require standards and criteria that are absolute; in this way assessment can be selective and competitive (Broadfoot, 1996). If the learner's journey or progress were to be the basis of a qualification there would likely be concern that those who progress from a lower base have the advantage over high fliers because they have 'more to prove' and this could upset selection based on the ideal of meritocracy. Professions such as medicine also have non-negotiable requirements, for example that patient safety standards are met. In addition the problems of standardisation of learning gain measurements identified earlier will apply to any assessments made at two or more points in time.

Employers are also unlikely to be convinced that ipsative assessment will help them identify suitable candidates for jobs and they tend to seek out those who have already attained qualifications rather than those with learning potential. I have argued that this attitude perpetuates inequality because the disadvantaged can rarely catch up even if they make huge strides in the right direction. Giving at least some recognition for distance travelled could go some way towards equitable forms of assessment and might provide useful information for educators and employers alike. For example, professional knowledge needs to be continually updated and what matters is an employee's ability and willingness to do this alongside realistic goals.

In the opening chapter I hinted that it might be possible to satisfy both the personal learning and measurement goals of assessment. I have previously suggested that a dual system that combines an ipsative regime with a standards and outcomes based regime might be feasible (Hughes, 2014). This already happens when students undertake supervision such as for a doctorate. Here there is a developmental phase where progress is monitored and recorded that is separate from a final submission phase – the viva voce - where a conventional summative grade or thesis pass is given.

Exploring the relationship between ipsative assessment and learning gain

From the above discussion we can see that there is a real possibility of including assessment of an individual's progress as part of an assessment regime. Learning gain information at the personal level – both qualitative and quantitative - can be made available to learners to form the basis of ipsative assessment and feedback. But such a view of learning gain has different methods and audiences from learning gain conducted at scale and a summary will be presented next.

*Mapping the relationship between ipsative assessment and learning gain measurement*

Table 1 summarises the different purposes of ipsative assessments and learning gain measurements with examples, the different methodologies and the different audiences.

*Table 1* The purposes of measuring or capturing learning gain

| **Personal Learning Gain** ⟷ | | | **Learning Gain at Scale** |
|---|---|---|---|
| Ipsative feedback about progress to enhance learning | Ipsative assessment as a measure of personal learning gain | Learning gain measurement used in practitioner research or for continuous improvement | Measurement of learning gain for comparison or benchmarking between classes, cohorts, programmes or institutions |

| | | | |
|---|---|---|---|
| e.g. teacher, peer or self-assessment of progress | e.g. individual progress scores | e.g. comparing the effectiveness of different methods of group work | e.g. measures of school or university effectiveness |
| Qualitative | Qualitative or quantitative | Quantitative sometimes combined with qualitative data | Quantitative but could be combined with qualitative data |
| Student audience | Student and teacher audiences | Researchers and teachers as main audiences | Audience is educational managers, policy makers and accountability monitors |

**Explaining the invisibility of ipsative assessment and personal learning gain**

The table above might imply that all things are equal along the micro-level to macro-level learning gain continuum, but that is far from the current reality. There are a number of reasons why ipsative assessment is underused and personal learning gain measurements are largely invisible to learners and teachers alike with only a few exceptions. Firstly, ipsative assessment based on learning gain is not usually part of the formal assessment of students, although there are some possible options. Secondly, although formative assessment is widely practised, it tends to be carried out on a short-term basis and a longer-term ipsative approach is not easy to

establish. Thirdly, ipsative assessment may be undertaken verbally in informal tutorial and classroom settings and is therefore unrecorded. Invisibility means missed opportunities for assessors and learners alike and it is worth exploring each of these points in a little more detail.

Formal assessments and examinations usually measure achievement of pre-defined criteria with standards judged using a marking scheme or rubric. The assessment stands alone and is considered independently of previous work and credit is not given for progress. Any deterioration in standards is likewise not visible. As maximising the objectivity of marking is the aim, information on the learner's past history may be viewed as creating preconceptions of learners and consequently a bias to the assessment process. But in an attempt to minimise marking bias, the opportunities are lost for using progress to motivate and assist planning, or for using lack of progress as a warning.

One well established exception to this is the recording progress in portfolio assessments. In a portfolio a learner collects evidence of their practice and provides a narrative to demonstrate learning that has taken place. Sometimes a learner may be asked to showcase their 'best' work such as in creative disciplines while at other times a learner may be asked to capture a developmental journey, for example, in teaching or other professional practice. There may be ipsative processes going on in portfolio construction, but that does not necessarily mean that the overall assessment recognises process and the final mark may only be for the quality of the content. In such cases, ipsative assessment may be side-lined in comparison to criteria and standards-based assessment and largely invisible unless there are clear assessment criteria that refer to development work and a learner's progress.

Secondly, we have already encountered the idea in the opening chapter that developmental or formative assessment is potentially ipsative because it helps learners establish goals and next

steps based on current levels of work. However, such goals are often short-term – addressing the next piece of work - and thus do not address longer term goals or review the distance travelled as part of a learning trajectory. Short-term goal setting processes for monitoring achievement may stunt growth. For example, learners may respond to feedback with an action plan for improvements or correct their errors, but be unaware of whether or not they have implemented recommendations. Given that students have choice over when and how to respond to feedback, it is difficult to isolate and measure student responses to feedback (Price et al., 2010). Learners may repeat the same errors or ignore the more challenging aspects of intended feedback, especially if this is peer feedback (Walker, 2015). But over time, judgements of progress in response to feedback, whether from self or others, could be very helpful, perhaps showing when feedback has had negligible impact.

Students need time and support to become self-regulating and learn how to manage feedback (Nicol & Macfarlane-Dick, 2006). But while a longitudinal or programme level approach to assessment could enable students to identify repeated unhelpful behaviour, this is not easy with modularised curricula that are not very coherent (Hughes, et al. 2015). Without a visible and systematic approach, formative assessment which is potentially developmental in the longer-term may go unnoticed by students and teachers. Formative assessment might therefore benefit from defining and establishing ipsative approaches to tracking the longitudinal development – or personal learning gain - of learners and we shall see some good examples later in the book.

A third reason for the invisibility of ipsative assessment is its association with verbal feedback, perhaps used as a motivational device. Progress may be discussed informally in the classroom or in a tutorial, but when such feedback is spoken there is usually no permanent record and agreed learning goals may be easily forgotten. Even when discussion of progress is captured for a formal progress review, the outcome is likely to be very general – for example in the

ipsative phase of doctoral supervision a tutorial or review might record overall progress on a thesis as either satisfactory or unsatisfactory. A more nuanced perspective in which personal learning gain can be compared for a range of skills and attributes is then missed. This begs the question of how the clarity and precision with which standards are articulated, in supposed 'good assessment practice' can be replicated for measurement of personal learning gain. But for reliability to even be an issue, ipsative assessment must first be given more status and recognition.

**Conclusion and summary**

Learning gain defined simply is the difference between two (or possibly more) measurements of achievement, and is usually collated for large numbers of students; ipsative assessment occurs when a student's present and past (and possibly future) performances are compared over time as a personal learning gain. The chapter has briefly reviewed the burgeoning use of learning gain data at the macro-level to monitor and compare institutional and teacher performance and introduced some of the debates that are relevant to the collection. Commentators on the use of large scale data-sets for comparing institutional performances tend to view such data gathering as valuable if only the data analysis quality could be improved. Yet performance monitoring, however thorough, is not necessarily a desirable or helpful activity for teachers and learners.

This book is not directly concerned with the wider politics of educational evaluation and government control; it explores instead examples of teachers and students using ipsative assessment and personal learning gain information voluntarily as part of their practice with educational and emancipatory aims. But, despite the growing world-wide practice of measuring learning gain or value added for quality monitoring at scale, ipsative qualitative

judgment and measurement of personal learning gain are not routinely used as part of the assessment strategy for individual students to guide and assist their learning. If such practices exist then they lack visibility. Giving weight to personal learning gain is very likely to be controversial because it may mean unsettling the status of top performers - a theme that will reoccur in this book.

To begin to rectify this situation, the collection offers case studies of practice which explore the benefits and challenges of different methods of ipsative assessment and individual learning gain measurement in a variety of educational contexts.

**References**

Arum, R & Roksa, J. (2011) *Academically adrift: limited learning on college campuses* (Chicago, London: University of Chicago Press).

Anderman, E.M., Gimbert B, O'Connell A.A. & Riegel, L. (2015) Approaches to academic growth assessment, *British Journal of Educational Psychology*, 85:138-153.

Ball, S.J. (2013) *The Education Debate* (2nd ed. Bristol: Policy Press).

Bloomer, M. & Hodkinson, P. (2000) Learning Careers: continuity and change in young people's dispositions to learning, *British Educational Research Journal*, 26(5): 583- 597.

Boud, D. & Molloy, E. (2013) 'What is the problem with feedback?' in D. Boud & E. Molloy (eds.) *Feedback in Higher and Professional Education: Understanding it and doing it well* (London: Routledge) pp. 1-10.

Coates, H. & McCormick, A.C. (2014) 'Introduction: Student Engagement-A Window into Undergraduate Education' in H. Coates & A. C. McCormick (eds.) *Engaging University Students: International Insights from System-Wide Studies*. (Dordrecht: Springer) pp. 1-12.

Dew, J. R. & Nearing, M.M. (2004) *Continuous Quality Improvement in Higher Education* (Westport: American Council on Education/Praeger)

Goldstein, H., Huiqi, P., Rath, T. & Hill, N. (2000) *The Use of Value Added Information in Judging School Performance* (London: Institute of Education, University of London, published online http://www.bristol.ac.uk/cmm/team/hg/value-added-school-performance.html (Accessed July 2015).

Hazelkorn, E. (2015) *Rankings and the Reshaping of Higher Education*
*The Battle for World-Class Excellence* (2nd ed. Basingstoke: Palgrave Macmillan).

Higher Education Academy (2015) *Grade point average: Report of the GPA pilot project 2013-14* (York: Higher Education Academy).

Hughes, G. (2014) *Ipsative Assessment: Motivation through marking progress* (Basingstoke: Palgrave Macmillan).

Hughes, G., Smith, H. & Creese, B. (2015) Not seeing the wood for the trees: developing a feedback analysis tool to explore feed forward in modularised programmes, *Assessment & Evaluation in Higher Education*, 40(8): 1079-1094.

Kooloos, J., Klaassen, T., Vereijken, M., Van Kuppeveld, S., Bolhuis, S. & Vorstenbosch, M. (2011) Collaborative group work: Effects of group size and assignment structure on learning gain, student satisfaction and perceived participation, *Medical Teacher*, 33 (12): 983–988.

Laurillard, D. (2012) *Teaching as a design science: building pedagogical patterns for learning* (London: Routledge).

Liu L. (2011) Value-added assessment in higher education: a comparison of two methods, *Higher Education*, 61(4): 445-461.

Lodge, J. M. & Bonsanquet, A. (2014) Evaluating quality learning in higher education: re-examining the evidence, *Quality in Higher Education*, 20 (1): 3-23.

Marginson, S. (2009) The Knowledge Economy and Higher Education: Rankings and Classifications, Research Metrics and Learning Outcomes Measures as a System for Regulating the Value of Knowledge, *Higher Education Management and Policy*, 21(1): 31-46

McGrath, C.M., Guerin, B., Harte, E., Frearson, M. & Manville, C. (2015) *Learning gain in higher education* (Cambridge: Rand Corporation), published online

http://www.rand.org/content/dam/rand/pubs/research_reports/RR900/RR996/RAND_RR996.pdf (Accessed July 2015).

McNiff, J. & Whitehead, J. (2006) *All you need to know about Action Research* (London: Sage).

Nicol, D. & Macfarlane - Dick, D. (2006) Formative assessment and self-regulated learning: a model and seven principles of good feedback practice, *Studies in Higher Education*, 31( 2): 199-218.

Norton, L.S. (2008) *Action Research in Teaching and Learning: a practical guide to conducting pedagogical research in universities* (London: Routledge).

Price. M., Handley, K., Millar, J. & O'Donovan, B. (2010) Feedback: all that effort, but what is the effect?, *Assessment & Evaluation in Higher Education*, 35(3): 277-289.

Schön, D. (1991) *The Reflective Practitioner: How Professionals Think in Action*. (2nd ed. Aldershot: Arena Ashgate).

Steedle, J. T. (2012) Selecting value-added models for postsecondary institutional assessment, *Assessment & Evaluation in Higher Education*, 37 (6): 637- 652.

Tam, M. (2004) Using students' self-reported gains as a measure of value-added, *Quality in Higher Education*, 10 (3): 253-260.

Walker, M. (2015) The quality of written peer feedback on undergraduates' draft answers to an assignment, and the use made of the feedback, *Assessment & Evaluation in Higher Education*, 40 (2): 232–247.