**Disrupting Learning Gain**

**Alex Forsythe1, Carol Evans2, Camille Kandiko Howson3, and Corony Edwards4**

1 University of Liverpool & Independent Performance Consultant

2 University of Southampton

3 King’s College London

4 Independent HE Consultant

Measurement promises to provide transparency and comparability, motivate staff, improve quality, and take me out for a beer on a Friday, except that measurement also makes empty promises and deceives. Measurement changes behaviour in perverse and cynical ways, and humans at work are adept at working to ‘the measure,’ for better and for worse. [Dina Gray](https://onlinelibrary.wiley.com/doi/book/10.1002/9781119960522) and colleagues provide a superb example of failed attempts to speed up movement through airports by incentivising baggage handlers to move luggage from planes faster. An athletic member of the handling team would grab the first bag from the hold, sprint to the reclaim, drop the bag onto the belt and the clock would stop. Everyone else’s bags were offloaded at leisure.

The same risks apply when measurement is applied to education. What can seem like a well-conceived measurement process may come with unintended consequences. The work of the French Psychologist Alfred Binet is possibly one of the most calamitous examples of how a prevailing climate can take something designed for the betterment of society and twist it. Binet was working with philanthropic motivations towards identifying and then helping underperforming students. He created the first practical intelligence test, which provided a mental age in comparison with chronological age that enabled a determination of how a child was performing. Binet’s intention for this scale was that it could help him advocate for the education of all children and to better help teachers understand their individual needs. However, Henry Goddard, the prominent American psychologist and eugenicist, discovered Binet’s work and popularised it with the intention of “*curtailing the reproduction of feeble-mindedness*". After all, measurement is not an objective fact, people decide what to measure, how to go about counting it, and what to do with that information.

That actual thinking had been almost completely suppressed with “the cancer of testology and testomania” was the historical and epistemological assessment of Leopold Szondi in the 1950s. Almost 70 years later the “distance travelled”, or the change in knowledge, skills, work-readiness, and personal development, conceptualised as “Learning Gain”, is the most recently proposed performance metric that students, academics, parents, employers, and other stakeholders hope will come about through investment in a university education.

Measuring learning gain is about trying to capture changes (gains) in students’ knowledge and skills through their learning experiences in higher education. While changes in certain attributes may be measurable, given the breadth of experiences both within and beyond higher education (HE), attributing cause and effect is more contentious. Furthermore, there is no one definition of what learning gain encompasses, with much depending on what one is valuing, and also how fine-grained a level of analysis is needed. Common to many definitions of learning gain, as identified above, is an emphasis on changes in students’ knowledge and skills; work-readiness, and employability; and personal development. The importance of including “value-added” as an integral part of measuring learning gain by comparing predicted and actual outcomes for students has also been advocated, building on the rich literature within the primary and secondary school system, and recognising concerns about differential learning outcomes within higher education.

Interest in learning gain within higher education in the UK is relatively new, although the concept has been widely applied in the US context, and variably across the globe. In the UK it has largely been driven by the government predicated around quality control (comparative data), value for money (employability and graduate earnings), and social justice issues around equity for all students moving into and through the HE system (differential learning outcomes). In the case of the latter, as Carol Evans, Camille Kandiko Howson, and Alex Forsythe note in *Making Sense of Learning Gain in Higher Education*, published in a special issue of [*Higher Education Pedagogies*](https://www.tandfonline.com/toc/rhep20/3/1?nav=tocList), fourteen years of primary and secondary schooling has not enabled the gap between low-income pupils’ attainment, and those from higher-income backgrounds to be narrowed, so the expectation that higher education institutions (HEIs) can do this in two or three years is problematic; this does not, however, mean we should not try.

The need for better indicators to demonstrate excellence in teaching has arguably been the main driver behind learning gain developments in the UK. In 2015 the Higher Education Funding Council for England (HEFCE), now Office for Students (OfS), funded [13 longitudinal pilot projects](https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/) (involving over 70 HEIs), and a National Mixed Methodology Learning Gain Project (with 10 HEIs) exploring methods of measuring learning gain and informing recommendations for scalability of different approaches within England. The projects investigate a range of approaches to, and dimensions of, learning gain, with emphasis on employability measures, and including development of new metrics on cognitive, affective, and behavioural, and metacognitive dimensions of learning, with some emphasising generic measures, and others focusing on discipline-specific approaches to learning.

There may be very strong and valid reasons why we should be investing in learning gain as a demonstration of ‘value-added’ and accountability. But by doing so we could also simply add to the data overload and pressure to perform to the metric that many students and academics already experience. Or worse, add further to student anxieties about their own personal learning journey. No more the road less travelled, rather, did they make the effort to travel far and fast enough?

There are other dangers, too: the incentives and reward structures in academia are a study in how not to encourage excellence and ethical high standards, with existing quantitative performance measures already driving perverse teaching behaviours such as teaching to the test, competing for the fastest feedback turn-around times at the expense of quality, emphasis on short-term learning, and class size increase and grade inflation (see Mark Edwards and Siddartha Roy’s, 2015 review of *Academic Research in the 21st Century:* [*Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition*](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5206685/)). There is no reason to believe that the introduction of learning gain measures will do anything to alleviate this situation.

Data collected about students’ experiences are already used well beyond their intended purposes, leading to a plethora of metrics that are not aligned with people and processes. Module-level evaluation data is used to fire individual teaching staff—even if the issues students are not happy with are outside their control. Satisfaction has been the dominant focus of measuring teaching in the UK over the past decade—even though studies show it is inversely related to learning. It turns out being happy does not really mean you have learned anything, but if you really push students to work hard, you’ll plummet in the league tables. How would learning gain measures fit with this paradigm?

The inadequacy of existing measures is a further problem. For example, the National Student Survey (NSS) has demonstrated inadequacies to such an extent that its value has now been reduced by 50% for the Teaching Excellence and Student Outcomes Framework (TEF) in England. Reducing the contribution of an invalid measure to 50% of a performance evaluation, does not, however, suddenly make it a more valid, fit for purpose measure. The myopic focus of the NSS is counterproductive, gameable, and encourages academics to focus on approval at the expense of real learning, and we know that achieving meaningful learning is often an unsatisfying, disruptive struggle that doesn’t necessarily equate with popularity, at least in the short timescales involved.

If applied in a critical way as an integral part of curriculum design and delivery, and through utilizing robust research designs, learning gain as a concept has huge potential in being able to offer valuable insights into the learning process and progress of all students, and hence to inform pedagogic design. However, the quest for a universal measure of learning gain is a futile one. Standardised measurement does not enable valid comparisons between institutions because those institutions need to be able to adapt and contextualise learning gain relative to the students whom they serve. At a practical level and given the increasing flexibility inherent in students’ degree programmes, the precise nature of what is taught and how, and the specifics of assessment mapped to the progressive development of module and programme level outcomes, makes the relevance and value of standardised assessment options limited.

Investment in measures of precisely targeted, empirically-tested learning interventions can help students to optimize their potential, but it does not necessarily follow that those interventions are appropriate for all programmes of study or for all institutions. In some instances, large-scale measurement may be poorly attuned to contextual idiosyncrasies. Similarly, small-scale measures may be of little relevance outside a specific discipline.

We might also argue that if assessment were fit for practice in the first place, there would be no need for additional learning gain measures. We cannot expect one overarching assessment to capture all we need it to do, but we ought to expect assessment to do a better job of capturing improvements in key areas of knowledge, skills, and understanding, both discipline-specific and generic.

Students differ from one another in their learning motivations and behaviours, which also vary over time. Therefore, student development and learning cannot be monotonic; i.e., it does not always increase, and never decreases. This means that the operationalisation of learning gain as “distance travelled” is problematic as it confounds distance travelled with order (coming first, second, or last). Secondly, the closer together attributes are, say for example, the difference between a first and second year student, or, the differences between final-year students across all universities, the more inconsistency there will be. These inconsistencies mean that when researchers try to measure differences between students, their measurements will be inaccurate because they introduce large levels of error variance between students; the spurious noise that researchers pick up on when trying to measure differences between individuals or variables of interest that can result in false positive and false negative results.

Measuring the learning gained by a student at a specific point in time (for example, at the end of a programme of study) may seem straightforward, but will tell us little or nothing about that student’s learning journey, or their progress relative to their peers’, and their potential for future development. Therefore, it may be theoretically possible to define the rank order between students over long time periods, but it doesn’t follow that this will provide a measure of distance travelled, or, that such a measure could be used to compare the performance of universities or students in their progression. And that is before even beginning to consider how to account for different starting points, which include a plethora of different entry qualifications, previous educational opportunities, and personal circumstances.

Meaningful learning gain measurement ought to capture the value-added component of a university experience. For this to be achieved attention needs to be directed away from a narrow measurement of summative achievement towards developing an understanding of the factors contributing to learning. The focus would be more productively orientated towards more meaningful measures of learning and teaching, especially in the pursuit of student equality of opportunity, the approaches to which may not be generalizable to large populations. When our aim becomes developing and evaluating early pedagogical interventions that impact student learning outcomes, we will target support where it matters most, with the result that we level the playing field and support all learners to maximise their potential.

The educational, social and biological sciences have gained ground in this area, having developed theoretical foundations that support high impact pedagogies (HIPs). HIPs research in the [US](https://www.tandfonline.com/doi/abs/10.1080/00091383.2017.1366805?journalCode=vchn20) and the [UK](https://www.heacademy.ac.uk/knowledge-hub/engaged-student-learning-high-impact-strategies-enhance-student-achievement) promotes deep engagement with, and ownership of, learning commonly identified within inquiry-based learning contexts, involving students actively in research, and with students co-authoring their assessment experiences.

Interpreting this work into valid and practical learning gain measurement requires robust research designs, supported by transparent reporting of information, that substantiate findings and permit replication. Including students as partners, learning gain becomes then a pedagogical philosophy that drives programme design and evaluation. Such an approach can advance shared understandings of concepts, measurement, instrumentation and transparency.

The political craving for simple measures requires objective scrutiny with more time spent investigating measurement, than investigating with measurement. Not enough time is spent scrutinising, and being explicit about the measurement process and the nature of the data, with all of the negative consequences described. This has been justified as the pragmatic sanction adopted by the social sciences because it is publication-fruitful (another perverse behaviour), and satisfies external, politically-driven demands, but it does not follow that the data is plausible. Across international efforts to measure learning gain there is often a disconnect between research design, and actual learning and teaching in institutional contexts. All too frequently, learning gain data is used in a reductionist way to gather data from students on their attitudes towards learning, and to identify competencies in specific, isolated dimensions (e.g., thinking skills) not attuned to the sensitivities of the taught programme or contextualised in complex practice. When divorced from a student’s teaching and learning experience, there is need to question the nature of data collected, and especially whether it has any direct relevance to how students are learning as part of their taught delivery. Often transparency is poor, and reliability and validity questionable. Non-specialists, and even specialists, rarely challenge the plausibility of data and even if they do feel the data is plausible it doesn’t follow that the information it generates should be accepted.

Unsurprisingly, when proposals to measure learning gain nationally across England were first discussed, initial reactions focused on how to game the metrics—savvy institutions would tell students to dumb down answers on the first test, to allow more room for “gain” by the end. There were also concerns that elite institutions would be unfairly disadvantaged because their students were already so clever (or well-coached) when they started. This reaction indicates a fear that some universities may not be doing much more than recruiting the best and the brightest and churning them out a few years later—without adding much value along the way.

Carol Evans, Camille Kandiko Howson, and Alex Forsythe in their detailed evaluation and critique: *Making Sense of Learning Gains in Higher Education*; as part of a forthcoming Higher Education Pedagogies journal special issue featuring nineteen articles based on work in the UK and overseas, identify core themes in the current learning gains debate. Principally, that ‘the sum of the parts does not make a whole’ is evident in the lack of an integrated approach within higher education to exploring learning gains. Integrated design requires researchers, practitioners, policymaker and professional services colleagues to work together from the outset, and leadership to facilitate the integration of approaches across an institution. Ideally, ‘integrated academics’ are needed who can take the best of research, appraise it critically, apply it through implementing contextually appropriate pedagogies, and through good design, use outcomes from practice to inform research.

HE has multiple purposes with different values placed on those purposes by different stakeholders. This requires a common value of ensuring ownership of learning gain approaches at all levels. Knitting together relationships among knowledge, concepts and skills, then embedding evidence-driven learning gain approaches within curriculum design and delivery can result in learning gain measures that inform and reflect pedagogy; not merely a metric-chasing tool presented as ‘a proof’ of quality. Integral learning gain approaches have the potential to lift the lid on learning processes through exploring what students know, and in what ways; what works well, why, when and for whom, personalising learning, personalising challenge.

The question of ethics looms large; this is not just about informed consent and the ethical use of data, it is about the very question of why and how data is being collected. The validity question is paramount. Asking students generic questions that may bear little direct relevance to their actual studies is likely to result in ‘garbage in, and garbage out’ in terms of what can be inferred. In the case of reliability we are obsessed with ‘more data is better’, but that is simply not true; it is about the appropriate numbers to satisfy the requirements of the test. All this comes back to questions as to how we want lecturers and students to use their time while in higher education. What knowledge, understanding, and skills are we valuing most, and how best can we develop and measure these?

Training is needed to support shared understandings of initiatives. How can we work together to identify systems and processes that are fit for purpose and train students and staff effectively in the ethical use of data? How can we use data effectively to support enhancements in pedagogy requiring nimble data mining and analysis? For example, there is potentially a wealth of data that can be used to evaluate gainful learning in students. Linking multiple qualitative and quantitative data sets is becoming increasingly possible, albeit with major General Data Protection Reporting (GDPR) implications that HEIs are grappling with, but the lack of consistency in the nature of, and application of tools to measure constructs (e.g., self-efficacy) is not helpful.

Assessment criteria vary across programmes, institutions and nations; they reflect different structures and values, making comparisons difficult. This can only be resolved through a compendium of reliable and valid learning tools and measures for the higher education fields, embedded within the curriculum and compared at the appropriate level. On the other hand, over-standardising programmes drives maladaptive behaviours. It will not support the drive for increasingly flexible programmes of study or the development of agile competencies for a changing workforce. Keeping that value at the forefront of the learning gain question helps avoid confusion between the symptoms of problems and their causes; it will support the future planning of education in the face of change, and then we can be assured that our collective actions will contribute to our field, leaving it a better place.

Word count (2957)