Higher Education Pedagogies

ISSN: (Print) 2375-2696 (Online) Journal homepage: www.tandfonline.com/journals/rhep20

Making sense of learning gain in higher education

C. Evans, C. Kandiko Howson & A. Forsythe

To cite this article: C. Evans, C. Kandiko Howson & A. Forsythe (2018) Making sense of learning gain in higher education, Higher Education Pedagogies, 3:1, 1-45, DOI: 10.1080/23752696.2018.1508360

To link to this article: https://doi.org/10.1080/23752696.2018.1508360

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 03 Sep 2018.

Submit your article to this journal

Article views: 13392

View related articles

View Crossmark data

Citing articles: 11 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rhep20

HIGHER EDUCATION PEDAGOGIES
2018, VOL. 3, NO. 1, 1–45
https://doi.org/10.1080/23752696.2018.1508360

ARTICLE

OPEN ACCESS Check for updates

Making sense of learning gain in higher education

C. Evans[a], C. Kandiko Howson[b] and A. Forsythe[c]

[a]Southampton Education School, University of Southampton, Southampton, UK; [b]King's College London, University of London, London, UK; [c]School of Psychology, University of Liverpool, Liverpool, UK

ABSTRACT
Internationally, the political appetite for educational measurement capable of capturing a metric of value for money and effectiveness has momentum. While most would agree with the need to assess costs relevant to quality to help support better governmental policy decisions about public spending, poorly understood measurement comes with unintended consequences. This article provides a comprehensive overview of the development of measures of learning gain in higher education, exploring political contexts, methodological challenges, and the multiple purposes and potential of learning gain metrics for quality assurance, accountability and enhancement, and most importantly, we argue, the enhancement of learning and teaching. Learning gain approaches should be integral to curriculum design and delivery and not extraneous to it. Enhancing shared understandings of concepts, measures, and instruments, transparency in reporting and investment in developing pedagogical research literacy, including effective use of data are essential in the pursuit of meaningful approaches to measuring learning gain within higher education.

ARTICLE HISTORY
Received 20 May 2018
Revised 27 July 2018
Accepted 2 August 2018

KEYWORDS
Learning gain; higher education; post-secondary; pedagogy; research methodology

Introduction

The drive for transparency, accountability, equity, and value for money in higher education (HE) is now an international priority (Caspersen, Smeby, & Aamodt, 2017; Mountford-Zimdars, Sabri, Moore, Sanders, Jones, & Higham, 2015; Seifert, Gillig, Hanson & Pascarella, 2014). There have been several initiatives exploring the development of metrics that are thought to provide context for evaluating HE systems, institutions, and individual students. The need for better indicators to demonstrate excellence in teaching has been the main driver behind many learning gain developments and especially so in England, where the government declared measuring learning gain 'high priority work' (Johnson, 2017), with support and funding for improving indicators (Higher Education Funding Council for England/Office for Students (HEFCE/OfS, 2014, 2015).

Measuring learning gain is considered a policy panacea, a 'holy grail' (Willetts, 2017) and a concept to 'crack' (Havergal, 2017), but measurement is contentious due to the implied consequences of the outcomes through wider accountability regimes such as

CONTACT C. Evans c.a.evans@soton.ac.uk Southampton Education School, University of Southampton, University Road, Southampton SO17 1BJ, UK

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the U.K. Teaching Excellence and Student Outcomes Framework (TEF, Department for Education [DfE], 2017). Perverse incentives for higher education institutions (HEIs) are growing in scale and impact as universities jostle for improved positions in league tables leading to claims of 'gaming the system' through the redistribution of resources and/or investment in areas that ranking metrics emphasise (Edwards & Roy, 2017). Gaming serves to distort the HE landscape, encouraging a focus on outcome measures which are often superficial, transitory, and based on the subjective experience of impactful teaching rather than on meaningful learning (Dudas, 2015; McGrath et al., 2015, Guerin, Harte, Frearson, & Manville, 2015). Despite these risks, there remains a global appetite from government ministries and the media for a universal tool to measure learning outcomes at the institutional level to permit comparisons across HEIs in the U.K. and internationally (Caspersen et al., 2017). Whether meaningful learning and, by association, the quality of teaching can be measured in this way is questionable. In attributing gains to students' experiences of HE, caution is needed in that HEIs are not the only drivers of learning gain for students, and indeed, what students want from HE and are prepared to invest their efforts in are complex issues.

This article provides a comprehensive overview of the development of measures of learning gain in HE, exploring the complex interplay of political contexts, methodological challenges, and the multiple purposes and potential of learning gain metrics for quality assurance, accountability, and the enhancement of learning and teaching. Key imperatives in moving forward with learning gain initiatives are highlighted drawing on international research evidence, and findings from specific initiatives to include HEFCE/OfS-funded (2015–2018) pilot learning gain projects in England involving over 70 HEIs, selected examples of which are featured in the *Higher Education Pedagogies Learning Gain Special Issue 2018* (identified in italics throughout this article).

Our goal is to provide researchers with an overview of best practice with the purpose of fostering rigorous research which is both relevant to learning gain and ethical in nature. Important decisions are taken based on research findings: results influence government policy decisions, the targeting of scarce resources, educational strategy, technological advances, and decisions about practice. Understanding how research has been conducted and what was discovered supports better evidence-based decision-making. Transparency over the strengths and weaknesses of studies and how rigorously they have been conducted enables the selection of the best available evidence, improving the inferences policymakers can derive from findings, and thus improving accountability in decision-making. Such clarity gives data relevance in the future, supporting the next researcher to synthesise large bodies of research through methods such as meta-analysis (Dawson & Dawson, 2016).

To attend to the brief outlined above, this article is organised into three parts. In part one, we provide an overview of the development of learning gain, considering a range of conceptualisations, approaches to the measurement of it, and provide a background to the political and historical context in which the development of learning gain measures is taking place in U.K. HE and elsewhere. In part two, we offer a critique of the methods used to investigate measures of learning gain and suggest principles by which future research into the area can be conducted to ensure due rigour. We conclude in part three with a consideration of the pedagogical imperative of learning gain measurement and argue for the integration of learning gain approaches into the curriculum.

Part one: learning gain developments

What is learning gain?

Across different countries, discussions about learning outcomes, learning competencies, graduate attributes, and employability are often considered to be synonymous with learning gain. Debates about measuring learning gain are fuelled by the lack of consensus of what learning gain is, the varied purposes of learning gain measures, and the intended stakeholders: who it is for, what is measured, and how can it be done. This is confounded by the different terms used to describe learning gain that can mean the same and/or different things. Varied emphases in learning gain definitions revolve around several key dimensions, and combinations thereof, such as distance travelled, value added, specific and overall attainment.

This article explores the theoretical tensions and practical challenges of these dimensions, captured in Figure 1. These include: consideration of *what* to measure, and whether this is based on a holistic view of learning versus one focused on specific competencies, generalised gain versus attainment in specific tasks, acquisition of knowledge and understanding versus application of the same, academic versus broader social skills including preparation for employment; *how* to measure learning gain: methods and instruments, measurement between points in time as opposed to predictive measures relative to learning starting points; the role of individual differences and

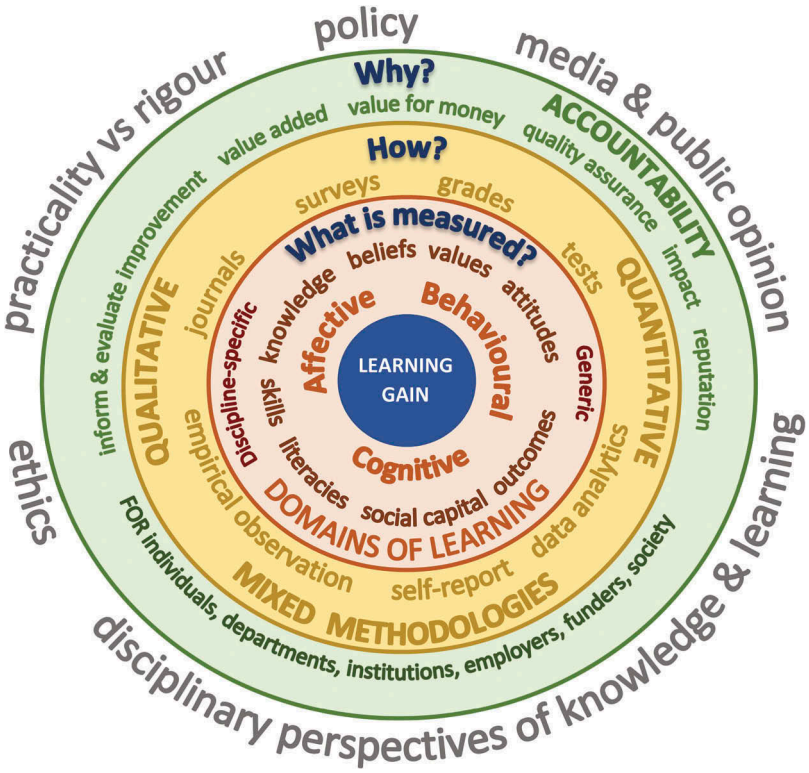

Figure 1. Considering learning gain. Courtesy of Corony Edwards.

contextual variables; process versus outcome measures; objective observations and performance versus student and lecturer perceptions; and the motivational and broader contextual factors that will influence the approaches adopted, including the purposes to which findings will be put, for whose benefit, and the broader political, social, ethical, and disciplinary contexts in which the measurements are to be made.

Learning gain has been described as 'the distance travelled' or 'the difference between the skills, competencies, content knowledge and personal development demonstrated by students at two points in time' (McGrath et al., 2015, p. xi). HEFCE/OfS offer a more holistic definition of learning gain as 'an attempt to measure the improvement in knowledge, skills, work-readiness and personal development made by students during their time spent in higher education' (HEFCE/OfS, 2015–2017[1]); this latter definition overcomes the potential issues associated with the limitations of two-point measurements designs, specifically response shift bias. Many learning gain measures incorporate value added, defined by McGrath et al. as 'the comparison between performance predicted at the outset of studies, and actual performance achieved' (2015, p. xi), although there are also considerable variations in how value added is measured in HE (see Kim & Lalancette, 2013 for review). In the United States (U.S.), learning gain is frequently discussed in terms of meaningful learning outcomes (U.S. Department of Education, 2006) and learning transfer, the latter being defined as the 'extent to which knowledge, skills and abilities learned in work-related training are generalised and maintained on the job' (Bates, Holton III & Hatala, 2012, p. 549). Transfer can be interpreted as the extent to which an individual can apply and adapt knowledge to a new context, and it also can measure the effectiveness of training in facilitating that learning.

Combining useful approaches from across the literature suggests that learning gain can be operationalised as a change in knowledge, skills, work-readiness, and personal development to include beliefs and values, and enhancement of specific practices and outcomes in defined disciplinary and institutional contexts (HEFCE, 2017a/OfS, 2018a). Beyond defining what learning gain is, philosophical and political debates continue about why learning gain should be measured, as well as methodological debates about how, or whether, it is possible.

Learning gain and the U.K. context

Learning gain stakeholders include students, academics, professional services staff, senior managers, the government, parents, employers, and the wider public. Internationally, changes to HE funding systems have raised interest in what students do during their time in HE and what they, and other stakeholders, gain from it. Given the additional costs associated with diversification of HE, increasing student numbers, and the consequent development of new funding systems, policymakers in the U.K. (and elsewhere) have identified the potential of student data for quality assurance, quality enhancement, and accountability. These shifts in focus, alongside measurement trends in the U.S., examining what students were gaining, or otherwise from their time in HE (Arum & Roksa, 2011), led to a joint venture between the U.K. Higher Education Academy (HEA), now known as Advance HE having merged with the Equality Challenge Unit and Leadership Foundation for HE; the government department

overseeing HE; the then-Department for Business, Innovation and Skills (BIS); and the then Higher Education Funding Council for England (HEFCE);, now known as Office for Students (OfS). This initiative commissioned the initial scoping study of learning gain by Rand Europe (McGrath et al., 2015) and led to a suite of 13 funded longitudinal pilot projects (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.

In parallel, the U.K. government began development of a Teaching Excellence and Student Outcomes Framework (TEF), which 'aims to recognise and reward excellence in teaching, learning and outcomes, and to help inform prospective student choice' (BIS, 2015). The TEF has been piloted with noted limitations of existing metrics as proxy measures of the assessment criteria. Some of the measures used in the initial iterations are student satisfaction, retention and completion, employment, and salary data. These are not necessary measures of high-quality student learning. Several measures are highly dependent on student characteristics such as socio-economic status. The intention is to 'incorporate new common metrics on engagement with study (including teaching intensity) and learning gain, once they are sufficiently robust' (BIS, 2015, p. 25). The potential inclusion of learning gain measures in a national accountability system has raised interest and concerns about the HEFCE/OfS pilot projects. This broader accountability agenda is part of a range of international efforts exploring outcomes of HE, which raise questions about why learning gain is being measured, what to measure, and what is possible to measure. Measures of learning gain could provide more robust metrics to assess areas of teaching and learning quality. They have the potential to contribute to a virtuous cycle, by holding institutions accountable while activities undertaken to raise outcomes could lead to improvements in teaching and learning and the student experience. There are arguments that it is too difficult to measure the complexity of student learning; although it may be challenging, the HEFCE/OfS learning gain pilot projects may nevertheless find better ways of capturing student learning outcomes than existing measures.

Approaches to learning gain: the international context

Interest in learning gain is not new. In the U.S., much emphasis has been placed on the measurement of higher level thinking skills such as the measurement of generic critical thinking skills via the Collegiate Learning Assessment Plus (CLA+) test developed by the Council for Aid to Education. The CLA+ test of critical thinking skills tests outcomes of the general education approach in the U.S. which differs from subject-specific degrees in England. The approach uses open-ended assessments which are focused on deep approaches to learning. The relevance and applicability to contexts outside of the U.S. and especially in relation to the U.K. context have been highlighted in some of the HEFCE/OfS projects (Kandiko Howson, 2018).

A prominent U.S. example of standardising assessment is that of the Association of American Colleges and Universities' Valid Assessment of Learning in Undergraduate Education (VALUE) project. In the VALUE project, rubrics are used to externally assess students' in-course assignments against nationally standardised learning outcomes (Drezek McConnell, & Rhodes, 2017; Rhodes, 2009). Samples of students' work are

scored using value rubrics for 16 domains of intellectual and practical skills, personal and social responsibility dimensions, and integrative and applied learning. Several principles underpin the rubrics including deep approaches to learning and student innovation and creativity. The approach has extensive institutional buy-in but it is resource and time-intensive. Also focused on outcomes, the Wabash National Study (2006–2012) led by the Center of Inquiry (2016) involving 49 institutions sought to explore practices and structures supporting liberal arts education and different methods of assessing liberal arts focusing on 12 outcome measures (e.g. critical thinking, moral reasoning, and openness to engaging new ideas and diverse people); the importance of using data collected to benefit student learning is emphasised as one of the key outcomes of the project (Blaich & Wise, 2011, Pascarella & Blaich, 2013).

Another approach focusing on student learning outcomes is the Voluntary System of Accountability (VSA) – a holistic accountability framework, which 'was created to provide greater accountability through accessible, transparent, and comparable information, and more recently, has been developed to support professional development opportunities to advance institutional data capacity' (VSA, n.d.). The VSA was introduced in 2007 by the U.S. Association of Public and Land-grant Universities (APLU) and the American Association of State Colleges and Universities based on the premise of offering straightforward, flexible, comparable information on the undergraduate experience, including student progress and learning outcomes. It provides a model for how multiple measures of student learning can be incorporated into a customisable portal. The 2017 VSA Vision focuses more directly on providing support to participating institutions to increase data, tools, and ability to develop and deliver exceptional evidence-based communications for a variety of stakeholders. An increasing emphasis on promoting access to HE and equity within it has seen the launch of the Center for Public University Transformation by the APLU involving 100 public research universities in the U.S. with the aim of developing, refining, and scaling innovative practices that also endeavour to close the achievement gap.

Building on the work in the U.S. and related efforts across Australia and Europe, in 2012, the Organisation for Economic Co-operation and Development (OECD) undertook a feasibility study through the Assessment of Learning Outcomes in Higher Education project (AHELO) across multiple countries and subjects of study (Tremblay, Lalancette, & Roseveare, 2012). They faced challenges around what to measure, with international, cultural, and subject-level differences emerging. Due to concerns about data quality and use, and lack of buy-in from national governments, the project was not continued (Morgan, 2015; OECD, 2013a; 2013b).

Emphasis on the development of adequate measurements of achieved learning outcomes is reinforced in the Bologna Process in Europe and the European Qualifications Framework. The European Commission supports the Measuring and Comparing Achievements of Learning Outcomes in Higher Education in Europe (CALOHEE) project as part of the Tuning Framework (Calohee, 2018). This work focuses on aligning frameworks for course design rather than student outcomes. There are also national research projects in Germany (Blomeke et al., 2013), Brazil (Meguizo & Wainer, 2016), Italy (Cattani, Guidetti, & Pedrini, 2017), and Columbia (Shavelson et al., 2016) on student learning outcomes and the development of generic and discipline-specific tests to measure learning gain which have raised concerns about

student engagement, breadth of focus across sectors of HE, and practical challenges of such efforts. In pursuit of research into what works, there is increasing emphasis on embedded evaluation as an integral requirement of government funding (e.g. Australia, Germany, U.K., and U.S.). Drawing on the CLA+ and AHELO projects, the International Performance Assessment of Learning (iPAL) project (IPAL R & D, 2018) led by Shavelson, Solana-Flores, Tritschanskaia, and Marino aims to develop reliable and valid performance assessments of the twenty-first century ('generic') skills that can be used by HEIs nationally and cross-nationally to measure learning providing both formative and summative functions.

Student engagement has also been used as a proxy for learning gain (*Stoakes & Neves, 2018*) with the rhetoric suggesting that this is a better measure of the quality learning and teaching in HE than the use of standardised tests or student satisfaction surveys. Engagement surveys started in the U.S., and are now endemic within HE, spreading to Canada, Australia, Ireland, and many other countries and HE systems (Coates & McCormick, 2014). In the U.K., the Higher Education Policy Institute/ Higher Education Academy (HEPI/HEA/Advance HE) Student Academic Experience Survey has evolved to include questions on student well-being as part of student HE experiences. The national HEA/Advance HE's U.K. Engagement Survey provides nationally benchmarked data on students' engagement with their studies within and beyond the classroom. The U.K. National Student Survey (NSS) has evolved from its origins in 2006 to also include questions tapping into student voice and engagement in learning and teaching. Such surveys may have driven some improvements in teaching quality but they are not a strong measure of learning gain per se, rather their relative value is determined by how engagement is defined and measured. The link between engagement, student learning gain, satisfaction, and quality of teaching is complex and tenuous (Evans, Muijs, & Tomlinson, 2015), and the assessment and feedback dimension of the U.K. NSS has relatively weak predictive ability (Burgess, Senior, & Moores, 2018). With both engagement and student satisfaction surveys, individual student characteristics (e.g. learning orientation, discipline, conceptions of learning) impact both on how students self-select into studies and on how students respond to test items. Without an indefinite number of norm-reference groups to ensure that the sampled population is representative of the population of interest, these differences limit the extent to which reliable comparative analysis can take place and the extent to which measures can meaningfully influence policy and practice (Bennett & Kane, 2014).

Given the weight afforded to student satisfaction surveys in impacting the assessment of the quality of teaching provision in HEIs, the relative inadequacies of surveys such as the U.K. NSS (2005–2018) require urgent attention. Whilst the TEF in the U.K. has, in later iterations, reduced the value of NSS items in assessing institutional performance by half, this does not directly address the limitations of the survey which requires revision in terms of continuing fitness for purpose. The survey would benefit from increased evidence of its validity as a measure. Fundamentally, the NSS needs to be underpinned by a transformative model of HE experiences acknowledging the role of students in their learning as active contributors and not as passive recipients – vessels to be filled.

Furthermore, gaming of student surveys by HEIs and a consequent myopic focus on enhancing student satisfaction without the necessary attention afforded to research-

informed curriculum design can have negative consequences for the longer term enhancement of learning and teaching in HE (Burgess, 2018). There are very strong and valid reasons why we should be investing in learning gain initiatives, but this needs to be done mindfully and ethically.

The rationale for exploring learning gain

The worth of HE is under tremendous scrutiny globally given the increasing number of students in HE, the changing requirements of labour markets associated with the rise of artificial intelligence and the fourth industrial revolution (Edge Foundation, 2016), the increasing costs of education to the individual and the state, and the relatively poor rates of student completion in HE. In the U.K., 1 in 10 undergraduates drop out before their second year of study (HESA, 2018a); a third of students leave HE after 1 year in the U.S. (Fisherman, Ludgate, & Tutak, 2017). Completion rates vary considerably within Europe with figures of 81% in the U.K. and Denmark and 59% in Norway (Vossensteyn et al., 2015). Completion rates in Australia vary from 51% to 88% (AEN, 2018). In this HE landscape, the development of learning gain initiatives is seen as important in increasing the accountability of HE in terms of demonstrating value added leading to a 'learning outcomes race' (Douglas, Thomson, & Zhao, 2012) and the associated development of a quasi-commercial market where proponents of different approaches advocate their own wares (Caspersen et al., 2017). In this context, we need to be extremely judicious about how resources are used and for what purposes.

Widening participation and social mobility agendas provide an important rationale for the measurement of learning gain (Mountford-Zimdars et al., 2015). Social inequalities are perpetuated through quality judgements based on institutional reputation, a key sorting and selection criterion for many employers. Concerns about a lack of diversity in the workforce have led to a desire for more information to differentiate the quality of graduates beyond measures highly correlated with prior high socio-economic status (MDV Consulting, 2017). In response, many employers now design in-house recruitment mechanisms. There are large numbers of meta-analytical and other studies which suggest that measures such as cognitive ability and personality can predict performance across a broad range of jobs. Correlations remain modest with predictive validity co-efficients ranging from .10 to .40 and averaging around .30 (Schmitt, 2014). These psychometric tests are financially burdensome for employers and in-house alternatives are often methodologically flawed and create high inefficiencies for employers and graduates (Keep & James, 2010). This situation has led to a desire for metrics which permit the identification of students, courses, and institutions that demonstrate the knowledge, skills, and attributes that employers are looking for and that the economy needs.

Conceptual debates about what to measure lead to broader questions about the point of HE. HE has multiple purposes with different values placed on these by different stakeholders and not easily sated by the use of single measures. Perceptions and values of the different functions of education and potential purposes of education, for example, Biesta's (2010) notions of qualification, socialisation, and subjectification, impact on what is measured (see *Powell, Gossman, & Neame, 2018*). What aspects are focused on depend on how one views the purpose(s) of HE, which vary spatially and temporally

across the sector (Edge Foundation, 2016; Haldane, 2015; Marshall, 2017). Figure 2 highlights the different ways that information can be used, and at different levels to support individual, institutional, and national agendas; the extent to which these align is a significant challenge for HEIs.

The measures that are used are dependent on institutional priorities, and at different levels of inquiry (institution-wide, faculty, discipline, module, individual), and in relation to what one thinks the main goals of higher education are, ranging from sustaining local and global economies, to addressing social justice/equity, to individual development. A focus on global competitiveness implies developing flexible knowledge workers for the twenty–first—century economy, and training professionals to participate in a functioning society, whereas an aim to foster democratic citizenship means enabling participation in and contribution to a more just, peaceful and sustainable global community (UN, 2012); (UNESCO, 2005). Another perspective may be to work for social justice and equity, with higher education creating an engine for social mobility; or to focus on individual growth, cultivating in students a passion for a subject and skills for lifelong learning, with investment in skills for the accrual of individual capital (Collini, 2012; Kandiko Howson, 2012, 2018).

Level of use	Purpose of use
Government	Accountability; regulation; quality assessment; comparison; market indicators. Recognising the value added by participation in HE.
Employers	Recruiting graduates; diversifying workforce; better meeting the needs of employers, business, industry, professions.
Prospective students	Informed choice on value for money.
Institutional (strategic)	Programme review; inform strategy; enhancement, evaluate programmes; staff reward and recognition.
Institutional (service delivery)	Enhance and tailor student services (e.g. careers services); scaling-up potential of initiatives; integrated use of data.
Cross-institutional	Benchmarking, comparisons. To explore differential learning outcomes and impact of module/programme design within and across disciplines.
Programme-module level	Course management; pedagogical enhancement: generic and discipline-specific; team development; holistic evaluation of all elements of the programme.
Lecturer level	Pedagogical enhancement, data for teaching staff to tailor information to students; enhanced understanding of student needs. Impact analysis to inform curriculum development
Student level	Support student self-regulation: Provide data for reflection, awareness raising. Support student engagement with curriculum.

Figure 2. Summary of levels of use of learning gain metrics [adapted from Kandiko Howson HEFCE Year 2 Evaluation Report, 2018].

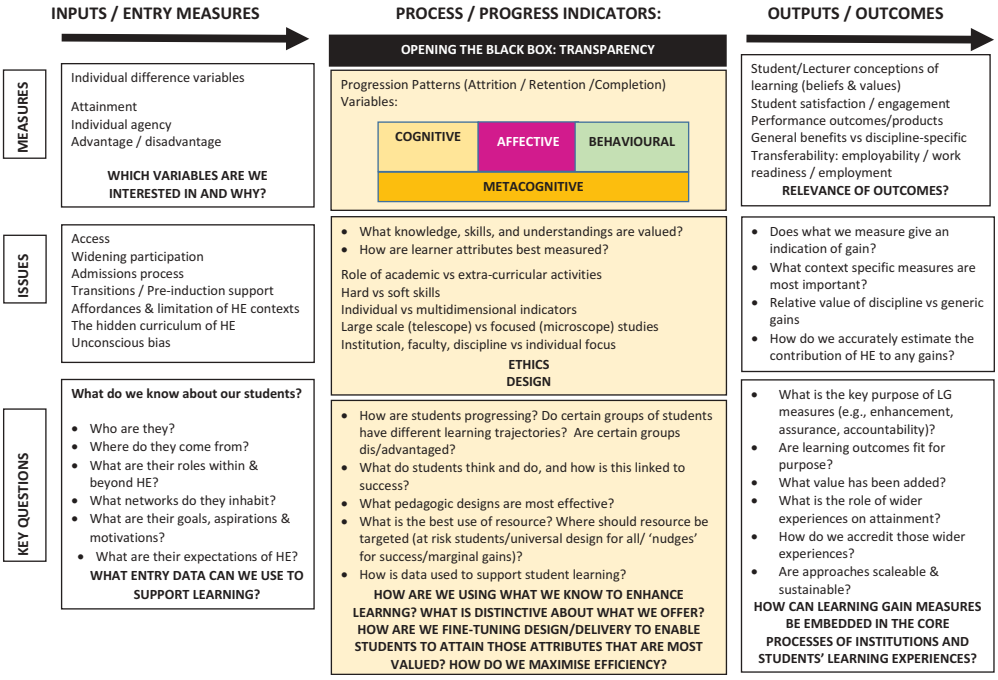

Figure 3. Inside the higher education black box.

In reviewing levels of use as outlined in Figure 2, we argue the importance of an integrated approach to supporting the implementation of research-informed practices to support learning gain initiative efficacy at the individual level and efficiency across the sector. This requires both top-down support and bottom-up approaches to impact policy development and practice.

A principal rationale for the measurement of learning gain relates to transparency and what happens inside the 'black box' of HE, asking what students should achieve and to what extent institutions are enabling this. The reality is, however, that a huge variety of students enter and exit HE with varying qualifications, skills, and social capital. Measuring learning gain of this eclectic mix requires some accounting for the inputs into HE, the student experience during HE, the subsequent outputs, and how these consolidate into meaningful learning. Figure 3 provides a framework for input, process, and outcome variables within HE, discussed in detail below. Key questions revolve around the decisions underpinning the collection, use, and management of data, and how this information can be used most effectively to enhance learning and teaching attuned to the development journey of a student, and groups of students as they navigate the many learning transitions into, through, and beyond HE.

Inputs

HE systems and institutions can be viewed with students seen as sets of inputs, extensively explored through research on access, admissions, and widening participation (*see Cameron, 2018*). In terms of measuring learning gain, input and entry measures are key, because not all students enter HE with the same knowledge, skills, and attributes, and

thus lack a fair and equal chance of success. Evidence suggests such differentials are not remediated throughout students' time within HE, with students with certain characteristics (e.g. low socio-economic status), not achieving as well as those from higher socio-economic backgrounds (Mountford-Zimdars, et al., 2015). The efforts of HE to remediate such issues are confounded by the fact that in the British education system, 14 years of primary and secondary schooling have not enabled the gap between low-income pupils' attainment and those from higher-income backgrounds to be narrowed (Field, 2010; Shaw, Baars, Menzies, Parameshwaran, & Allen, 2017).

Some students argue that there is a lack of useful information to help them make informed study choices and that the synthesis of league tables often provides an overwhelming challenge; this is also affected by differences in students' individual characteristics such as economic and social–cultural capital (BIS, 2011). Students are data rich, but insight poor. Bowes et al. (2015), and more recent studies by the Nuffield Foundation (2017) into how students use the data already available have shown that most do not know how to make good use of it, and instead may make decisions based on other considerations (media reputation, proximity to home, where their friends are going, etc.) (Diamond, Vorley, Roberts, & Jones, 2012; Nuffield Foundation, 2017). The development of robust, contextualised learning gain metrics could potentially support students from a variety of backgrounds to make better-informed choices about what and where to study, and for institutions to make better-informed decisions about which students to select. They could also, however, simply add to the data overload that many students already experience.

Process and progress indicators

In developing process and progress indicators, one of the challenges of using existing data is the large gap in student and staff expectations around assessment and feedback (Evans, 2013, 2016; *Forsythe & Johnson, 2017*). The multiple approaches being tested to measure learning gain globally and within the pilot HEFCE/OfS projects include:

- The development of new metrics on cognitive learning gains, both discipline-specific and generic development (see *Callaghan & Aloisi, 2018*; *Ilie, Vermunt, & Vignoles, 2018*; *Sands, Parker, Hedgeland, Jordan, & Galloway, 2018*; *Ylonen, Arico, Gillespie, Lancaster, & Ward, 2018)*;
- Higher level cognitive application skills such as problem-solving and critical reflection (see *Speight, Crawford & Hadelsey, 2018*); critical thinking skills (see *Callaghan & Aloisi, 2018*); conceptual understanding, particularly in Science, Technology, Engineering and Mathematics subjects (see *Sands et al., 2018*; *Scalise, Douskey, & Stacy, 2018*; *Ylonen et al., 2018*); and critical writing (*Ilie et al., 2018*);
- Affective measures such as self-efficacy (see *Forsythe & Jellicoe, 2018*; *Pampaka et al., 2018*); active and collaborative pedagogical approaches (see *Scalise et al., 2018*; *Turner, Gray, Sutton, Muneer, & Swain, 2018*);
- Behavioural measures of student engagement and employability, engagement in work preparation, work-readiness, and graduate employment (see *Callaghan & Aloisi, 2018*; *Cameron, 2018*; *Speight, Crawford, & Hadelsey, 2018)*; non-academic

skills development; civic activities, time on non-study activities (see *Stoakes & Neves, 2018*); increasing emphasis on soft skills development (Haldane, 2015);
- High-impact pedagogies – drawing on Kuh (2008) and Kuh, O'Donnell, and Schneider (2017), and the development of student research skills (see *Turner et al., 2018*).
- To attend to the domains highlighted, a raft of measures tap into students' affective, behavioural, and cognitive dispositions. Affective dimensions frequently include specific measures such as self-efficacy, and broader constructs including resilience, well-being, confidence, and satisfaction. Behavioural measures often include students' use of resources and opportunities, and increasingly the use of data analytics to map students' academic and co-curricular engagement (attendance, use of virtual learning environments, work-based learning, skills' assessments). Cognitive measures include general cognitive learning gains, discipline-specific cognitive learning gains, critical reasoning/thinking skills, situational judgement, and research methods competence. Many projects and studies focus on individual dimensions, while others aim to explore students' metacognitive and self-regulatory abilities through the integration of affective, behavioural, and cognitive domains.
- Whether we are tapping into those areas of direct relevance remains open to question.

For example, many programmes of study spend time developing critical thinking skills, while training students in the function of incubation and insight are largely overlooked. Problem-solving requires students to be able to remember critical aspects from other problems and be able to reproduce that knowledge in a new situation and for that to happen '*memory must be tickled*' (Halpern, 2014, p. 526). Incubation and insight are the thinking processes that are key to how humans depart from the principles of probability and take mental shortcuts known as heuristics to solve problems. The solutions to those problems are often, personally goal-centred, and we give favour to our own preferred conclusions. This rational thinking and problem-solving behaviour and attitudes are driven by one's goals and beliefs commensurate with the available evidence and it is not the same as thinking critically. Stanovich, West, and Toplak (2000) argue that more needs to be done to develop tests that explore rational thinking tendencies as well as errors of judgement and decision-making. A group of students may be excellent at critically evaluating the options for organisational investment, but if not one of them plans to decide and act, then the debate was a pointless one.

Output and outcome measures

The final stage of learning gain measurement is the development of output and outcome measures (for a detailed account of outcome measures consult OECD, 2013b report). These measures address what cognitive learning gains (e.g. increases in knowledge – thinking skills), employability skills, and other attributes students have achieved.

Summary

Many current outcome measures are poor proxy measures of student learning. This leads to a system of targets and key performance indicators that drive institutions to

compete on metrics that may detract from the student learning experience. For example, student satisfaction surveys can drive institutional behaviour towards keeping students happy at the cost of academic challenge and rigour; salary and job classification data are strongly correlated with socio-economic status, perpetuating inequalities in society, and are linked with institutional reputation; progression and completion of targets can lead to a lowering of academic standards or grade inflation (Higher Education Statistics Agency, 2018b). This has led to claims of a lack of accountability to the government or to the public.

More broadly, public and employer perceptions of quality have a strong foundation in institutional reputation, drawing heavily on institutional age, research performance, student selection, and graduate salaries. Reputation may be only partially, or even inversely, linked with the quality of the student learning experience. On an individual level, the quality of a student's degree has been signalled in England through the degree classification system. However, because of the limited four-point approach (first, upper second, lower second, and third) and through steady, long-term grade inflation, there is little differentiation amongst graduates, which has led to a rise in interest in the adoption of a U.S.-style Grade Point Average system in many U.K. HEIs (Advance HE, 2018).

These questions of quality have led to the desire for better metrics to account for what students have gained from their time in HE, and what added value do institutions provide as outlined in The Burgess Group Final Report (HEA, 2015; Universities UK, 2007). This would support sector-wide accountability for how much students have learned during their time in HE. Greater transparency of what happens inside the 'black box' of HE through the development of robust, contextualised metrics of learning gain could help address the challenges noted above. Within institutions, greater alignment of quality assurance and pedagogical enhancement, as staff, academic, and professional services, would be held to account for activities that directly led to student learning.

Part two: measuring learning gains: methodological considerations

Measuring learning gain is complex, involving philosophical questions of what to measure, and scientific questions on how to measure, with inevitable trade-offs between what is methodologically robust and practically deliverable, all of which are framed within the broader political debate about 'why' measure learning gain in the first place, and prerogatives of elite institutional and national resistance to challenge what has gone before. Arguments that measurement overemphasis, measurement limitations, and measurement misunderstandings distract from the main purposes of HE are all well-rehearsed within the literature (Peseta, Barrie, & McLean, 2017; *Ylonen et al., 2018*).

Presuming that the traits we are interested in for learning gain purposes are measurable, there are strengths and weaknesses inherent in different approaches to learning gain measurement as typified in the HEFCE/OfS projects and evidenced in the literature more widely (HEFCE/OfS 2018; Kandiko Howson, 2017; 2018; McGrath et al., 2015, pp. 38–42). Measurement of learning gain in the U.K. is relatively new with most of the available research on the issue emanating from the U.S., with a predominant focus on the development of generic as opposed to discipline-specific skills (McGrath

et al., 2015). To be at the vanguard of learning gain development, integrated approaches are required that marry the best in measurement, with research-informed approaches to pedagogy (Evans, Waring, & Christoudolou, 2017). This requires strong collaborative working across academic (research and teaching), professional services, and data teams.

Guiding principles underpinning the use of learning gain measures include: the importance of relevance to local context; potential for generalisability; transparency in how measures have been developed and implemented; validity and reliability of approaches; sustainability in terms of efficient use of time and money (Coates, 2016) and fundamental ethical considerations concerning student and staff engagement, use of data, and equity. However, less than 15% of the literature on learning gain is focused on the specific U.K. context (McGrath et al., 2015). There are several fundamental issues impacting what we measure, how we measure, and what can be deduced from our findings as outlined in the following section.

Measurement questions

Fundamentally, student achievement should in some way be measurable as learning gain(s), and measures of learning gain(s) should be able to predict some valued outcome. Significant responsibility comes with the design and measurement of learning gain(s) as poorly conceived measures change behaviours, triggering cynical and sometimes perverse actions in academia, and there are almost always unintended consequences (Edwards & Roy, 2017). If we are interested in measurement, we need to start from the core principles of good measurement design (Furr, 2013).

Time needs to be spent scrutinising and being explicit about the measurement process and the nature of the data to ensure that we are applying best measurement principles. It is essential that those at the forefront of academic measurement expertise are working closely with the teams that are implementing and running learning gain initiatives on the ground to ensure quality and academic rigour in research, greater confidence in findings, and potential to replicate approaches in different contexts. From an extensive review of the literature, there are several measurement issues to consider in the development of learning gain approaches; an overview of key concerns is outlined briefly in the following section.

Issue 1: assumptions concerning monotonicity: we should be able to see learning gains throughout a student's learning trajectory within HE

'Distance travelled' is wanting as a reasonable operationalisation of the construct of learning gain given that it is possible to define order but not distance. As noted by Thurstone (1925), the closer together attributes are, the more inconsistency in results there will be. Learning gains are most easily quantified when the magnitude of difference is large, because the closer together facets of learning are, the more inconsistency there will be between students.

A monotonic relationship is expected, in that over time, we should expect students' outcomes to get better and better. However, learning is complex, and may not represent something that is gainful in the linear sense. As observed by *Ylonen et al. (2018)*, students must attain a series of progressively higher academic standards but those standards are not singular in nature, there is not just one quantifiable variable, and

other factors invariably play a role. This is what Cattell (1944) identified when he highlighted the role of the '*interactive*' in measurement, which is the raw relationship between the environment and an individual's performance: family and personal circumstances, finances, self-efficacy, health, and well-being and so forth, all of which predict the probability of success and thus make up the dimensionality of gainful learning. *Rogaten and Rienties (2018)*, for example, explore these issues in their big data analysis of the individual student characteristics which may influence students' learning gain (i.e. gender, ethnicity, socio-economic status, and prior educational achievement) reporting that racial background may be the only demographic variable of statistical significance. Socio-economic status had limited influence on student performance, suggesting that the Open University widening access policy was supporting the widening of access. White students from non-low socio-economic SES who had A-levels or equivalent prior to the start of their degree showed high attainments and highest grade increase in comparison to non-white students and students from non-traditional backgrounds.

Issue 2: alignment of why, what and how in measurement

Normative measurement (the measurement of an individual's performance relative to a population or group), rather than ipsative testing (an individual's marks relative to previous results and performance on other measures), is the best way to evaluate how people or groups of people compare to one another. This recommendation is addressed by *Sands et al. (2018)* in their concept inventory evaluation and by *Callaghan and Aloisi (2018)* in their application of the CLA+. Such tests are complex to construct, however, and to be truly effective they require advanced analytical competencies such as Rasch modelling, which is possibly the only genuine technique available for constructing measures in human sciences that equate with measurement in the physical sciences, providing interval-scale measurement, that if performed correctly will remain invariant across use (Bond & Fox, 2015). Such approaches would avoid the considerable validity and reliability weaknesses found elsewhere (*Callaghan & Aloisi, 2018*), but because they are designed to compare cohorts, the data are often difficult to act upon.

Measurement whereby current individual preferences or performance is compared against prior preferences/performance often includes results between two or three individual assessments and commonly by using Likert forced choice measurement. Examples from the HEFCE/OfS projects include the measurement of academic self-efficacy, mindset, student self-assessment, confidence, academic career skills, and situational judgement, (see, for example, *Forsythe & Jellicoe; Neves & Stoakes; Speight et al.; Ylonen et al., 2018*). The aim of such scales is to show the relative position of someone on the scales relative to another, but some items are just easier to say *yes* to, more-or-less valued, best or better, and *Forsythe and Jellicoe (2018)* demonstrate this challenge in practice with their analysis of concurrent and ipsative mindsets. Ipsative measurement, while capturing students' stronger characteristics and weaknesses, says very little about how someone will compare against someone else with the same results. One student might have critical thinking as their strongest attribute, but still perform much worse than a student whose strength is analytical competence. Any measure of learning gain based on the nuances of student learning and experience in any given university would tell us very little about how students would stack up against others in their cohort, nor

would it tell us much about the performance of one university against another. Where such scales are useful, however, is that they are measures that can be acted upon, giving academics tailored advice about how to help students.

Issue 3: the need for transparency

Transparency is critical in educational research because it underpins the assumptions of our most important analytical tests: explicitness is not a principle that can be applied flexibly. The countless measurement tools available in the potential diagnosis and measurement of learning gain risks the creation of a naive expectation that somehow, we are applying the best measurement principles. Not enough time is often spent scrutinising and being explicit about the measurement process and the nature of the data. Bond and Fox (2015) argue that the Social Sciences have adopted a kind of 'pragmatic sanction' to explicitness in their data because it leads to more fruitful results. Several issues are evident within the literature:

Internal consistency (reliability). This measure is the unseen property which estimates the precision of scale scores. This is the true score that a researcher would detect if the scale of interest was perfectly precise and unaffected by measurement error. Measurement error inhibits the researcher's ability to obtain accurate measurements of participants' true scores. A test cannot be 'reliable' or 'unreliable' because it is the test scores for a specific population that determine reliability and not the test itself, one test administration could result in strong reliability and another may not. For example, the most widely reported measure of measurement reliability is to inspect the relationship between each test item and all the other items on a scale using the item-level internal consistency approach, Cronbach's alpha (Cronbach, 1957), where a high alpha supposedly translates as a reliable scale. Cronbach's alpha is widely used as *the* measure of reliability, but it only captures one component of reliability, there are many others, and it depends on the sources of variance that are considered to be relevant (Cortina, 1993). With alpha, longer scales are known to be more reliable but they may also have higher error rates because participants may become disengaged and fail to complete the test properly. With a shorter test, accuracy may be improved, but the internal consistency estimates may be lower. Therefore, investigators should not rely on published alpha estimates and should measure alpha each time the test is administered.

Again, with alpha, it does not automatically follow that because all test items are functioning well together (indicated by a high alpha), that the test is, in fact, measuring a unidimensional construct and that the trait is somehow valid. This is a question of *validity*, not r*eliability*, but often this is taken for granted, even when the soundness of the construct has not been considered at all (Sijtsma, 2009). Nor does a low alpha necessarily indicate a problematic test. For example, as a broad construct, learning gain will encompass growth, change, and sometimes failure. In this case, lower internal consistency is better for *validity* because it maximises the breadth of the domain that is being measured. Learning gain is likely to be a broad construct; therefore, it does not follow that high internal consistency makes a good test of learning gains.

Tau equivalence is worth a consideration to learning gains because tau gets to the heart of whether we define learning gains as a broad, narrow, open-ended, or closed construct. Tau equivalence means that each item on a scale contributes *equally* to the

total scale score. This is difficult to achieve in empirical research because some items will more strongly relate to a given construct while others are more weakly related. Some researchers will get around this issue by removing test items that are operating to reduce the overall alpha value. This step is perhaps not necessary when measuring learning gain as a broad construct but leaving them *in situ* violates the assumptions of alpha, and researchers seek to report a high alpha and thus demonstrate the reliability of their study. Cattell (1972) was a passionate advocate for the argument that measurement tools should be created to capture a breadth of examples that could tap into a construct of interest and was critical of the fact that through an obsession with reliability and internal consistency, researchers were removing test items unnecessarily. 'The fact that although the obsession of early psychometrists with internal consistency, under the impression that it was reliability, has long passed out of well-informed discussion, it dies hard as a superstition' (Cattell, 1972). Low within-scale correlations may, he argued, in fact, be a virtue. This is a critical point when trying to measure learning gains.

Sampling size requirements. Under- and over-estimated sampling sizes increase the risk of Type 1 and 11 errors, respectively (see Wolf, Harrington, & Clark, 2013). When planning the test, sample precision is critical. The widespread belief that large sample sizes are ideal for analysis is a fallacy. Larger than needed, samples leave studies vulnerable to over-sampling bias and an increase in false positives. Sample size varies with the number of variables, covariates, and the statistical analysis and error probability planned. For example, a study with 4 independent variables, 2 moderating variables, and 2 dependent variables, with a MANOVA (which remains the most commonly used multivariate test) potentially needs as few as 80 students to establish an effect. Testing more than required is not only unethical regarding the misuse of human and financial resources, but with a large sample, even a small inconsequential difference runs the risk of being flagged as significant. Appropriate sampling is efficient, and the data generated are reliable. We recommend that sample calculations are provided, and p values with effect sizes are reported.

Reporting bias. The selective inclusion of significant positive results and omission of non-significant or negative results can lead meta-analysts and the general population to have misleading understandings of the efficacy of an intervention (Dawson & Dawson, 2016, p. 1). This publication bias, also known as the file-drawer problem (Rosenthal, 1979), requires the calculation of a fail-safe number which estimates the number of unpublished studies in meta-analyses required to bring the meta-analytic mean effect down to a statistically insignificant level. Outside of psychology, psychiatry, and medicine, however, this correction would seem to be rarely applied (Heene, 2010).

Issue 4: research design: measurement points

The idea of a pre-and post-intervention study appears quite straightforward. For example, to evaluate the impact and outcomes of an educational programme or intervention students are asked the same, or similar sets of questions before and after the intervention, the difference is known as the response shift. Problems arise, however, because it is assumed that the students' frame of reference for the metric, or their

awareness of a variable, will not change between the two data collection points. This problem is known as the response-bias shift which is the contamination of measures that result from inaccurate pretest measures. A simplistic example could be an intervention designed to improve student team working with pretest questions probing team working experiences: 'When I am in a team, I rarely listen to what other people say', would probably elicit a high (socially acceptable) 'strongly disagree' response. Team exercises designed to encourage listening may trigger a realisation that sometimes the student forgets to listen to what others say, so they will re-evaluate their initial position. Post-testing will then present a lower score which reflects a re-evaluation of an earlier position, rather than improved listening skills from the intervention. For further reading and resolutions, see Howard (1980).

Incubation effect. Furthermore, the instrumental testing of students, for example, on critical thinking tests is problematic because we know that training with similarly modelled questions fosters performance on later assessments (and test manufacturers have built an industry around this). Those students who have incubated information from previous assessments are more likely to show insight; Duncker's (1945) radiation problem is a classic undergraduate teaching example of how to prime individuals to transfer knowledge. As such, students who stay the course on such assessments are less likely to be representative of the entire student population, self-selecting into the assessment for personal motivations and interests, and generally feeling more confident than their peers who either avoid the assessment altogether or fail to complete it. It would be difficult to conclude that the latter group of students are behaving and thinking irrationally, as we all have a limit on the resources we have available (Friedlander et al., 2011), and priority often must be given to tasks that we are evaluated against, rather than investing in collecting poorly conceived institutional data. Academics are no less prone to this kind of behaviour than students, with our decisions on where to focus our time and effort depending on what the incentives are, and how we apply those values to ourselves (Edwards & Roy, 2017).

The inclusion of multiple measurement points is also important to address the lack of linear development of learning gains and indeed 'learning loss', identified in a number of the U.K. (HEFCE/OfS 2015–2018) learning gain projects (Kandiko Howson, 2018).

Issue 5: standards for reporting

It is commonly found that articles that struggle with clarity around the reporting of research design and analysis also suffer from lack of understanding of the current best practices regarding the methods they have adopted (Academy of Management Learning and Education, (AMLE, 2017), and this is no less true of studies examining learning gain. Several systematic behaviours were identified in our review of the learning gain literature and associated projects, problems which have resonated elsewhere in the educational literature (Academy of Management Learning and Education (AMLE), 2017), to include key design issues as discussed below.

Common method variance (CMV) is the spurious error that becomes attached to variables through the measurement process, particularly in cross-sectional designs. If researchers use the same method to collect data for their criterion and their predictor

variables, (for example, through a large questionnaire with different scales) then CMV is a risk. In other words, the method will influence the variables being measured by increasing the error variance between the variables being measured and thus the statistical analysis (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). The trend towards 'super surveying' (facilitated by survey software such as Qualtrics and Survey Monkey) is increasing this problem and rarely do researchers make even the most rudimentary checks in this regard.

The individuals who decide to participate in such surveys are also less likely to be representative of the entire population. This will lead to two types of sampling error. These self-selecting participants are drawn to a survey for some personal reasons or interests, and because they are more likely to have good feelings about the purpose of the study, they are more likely to complete. The individuals who take part are different from the students who drop out or do not engage, and this results in a sampling bias known as the non-response error and the coverage error (because there are some students that you will never be able to reach).

Using data from studies that have not been purposefully planned (e.g. gathering data from usual teaching efforts) are problematic for a whole host of reasons and especially from an ethical perspective. In such examples, the motivation underpinning such studies has often been identified after the learning activities have taken place. The context is then often poorly planned; the research questions posed are often not well informed by, or anchored to, current theory and research. That is not to say that there is no place for exploration and serendipitous discovery in educational research, but that this type of work should be reported transparently, recognising its inductive nature and with a more tentative interpretation of its results.

Detailed standards for reporting are now commonplace in fields that education science draws upon, (see, for example, APA, 2008; Appelbaum et al., 2018; Levitt et al., 2018) but those standards are not necessarily so transparent in pedagogical research. Drawing on recommendations from the American Psychological Association and the AMLE (Köhler, Landis, & Cortina, 2017) (see Table 1 on the reporting of experimental studies), typical reporting standards should include:

- The rationale and problem formulation supporting the project, clear research questions or hypotheses that clearly inform the management and testing of the data, a method section which provides sufficient detail to permit accurate replication by another researcher, including the design and logic of the study, the variables tested and/or sources of evidence, the measurement process, including details of the materials used, ethical processes and permissions, and an analysis which includes interrogation of the underlying assumptions of the data, before statistical inference and a reporting of all relevant results, including those that do not support the original hypotheses (see also Dawson & Dawson, 2016).
- Research within this field, regardless of paradigm needs to be able: to report findings clearly and should demonstrate pedagogical clarity (what are the key elements of the design); methodological transparency (explication of methods, samples, context, process), methodological congruence (alignment between aims, methodology, methods of data collection and analysis; suitability of tools), be evidence-based (consideration of reliability, validity, etc.); implications and

Table 1. Standards for experimental studies adapted from the APA and AMLE.

Introduction	
	• A clear statement regarding the importance of the problem, the theoretical and practical applications. • Clarity about how this work adds to or develops current knowledge. • Clarity about how theory has generated questions or hypotheses to be examined.
Methodology	
Sampling	• Methods used to determine sample size. • Intended and actual sample size. • Changes in participant numbers during the study. • Inclusion and exclusion criteria. • The nature of sampling including details of systematic sampling, probability or non-probability sampling, percentage approached, self-selection, snowballing, settings and locations for data collection, inducements, rewards, and payments. • Detailed demographic characteristics should be provided as well as any topic-specific details such as stage of educational achievement, etc.
Procedure	• How was the data collected? • What instruments were used? How were those validated?
Design	• Were the conditions naturally occurring, or manipulated in some way, randomised or quasi-experimental? • Operational statements of variables being examined.
Ethical approval	• Ethical issues addressed ethical board approval reference.
Results	
Statistical assumptions	• Descriptive statistics for variables, means, standard deviations, and other relevant descriptive statistics. • Assumptions of the data and the distributions • Explanations for missing data, methods for addressing missing data.
Significance	• Direction, effect sizes, *p* values, and exact *p* value if no effect is detected. • For regression, multivariate analysis of variances, structural equation modelling, and hierarchical linear model correlations should be included and relevant tests for collinearity. • Exploratory analyses.
• **Discussion**	
	• Brief reminder of the rationale for the study followed by a statement of support or non-support for questions and/or hypotheses posed. • Any *post hoc* explanations. • Findings framed and evaluated considering the work of others. • Evaluation of potential bias, validity issues, or other study weaknesses. • Generalisability of the findings. • Discussion of implications for future research and practice. • Conclusions

recommendations should be made; the process of implementation and context should be detailed to enable replication of the approach in other settings to be robust, and demonstrate clarity around the pedagogy/approach being implemented (See Evans et al., 2015).

Issue 6: holistic research design planning and evaluation

Learning gain initiatives represent complex interventions with many important design and delivery considerations to consider from the outset. Failure to have tight research-informed designs impacts the credibility of findings and is resource-wasteful. In framing learning gain measurement approaches, considerable lead in time is needed to align all dimensions of projects. Figure 4 identifies key considerations when undertaking complex interventions drawing on *Medical Research Council* (MRC) research (MRC, 2008, p. 8; Moore et al., 2014) and the associated literature (NESTA, 2013, 2016; WHO,

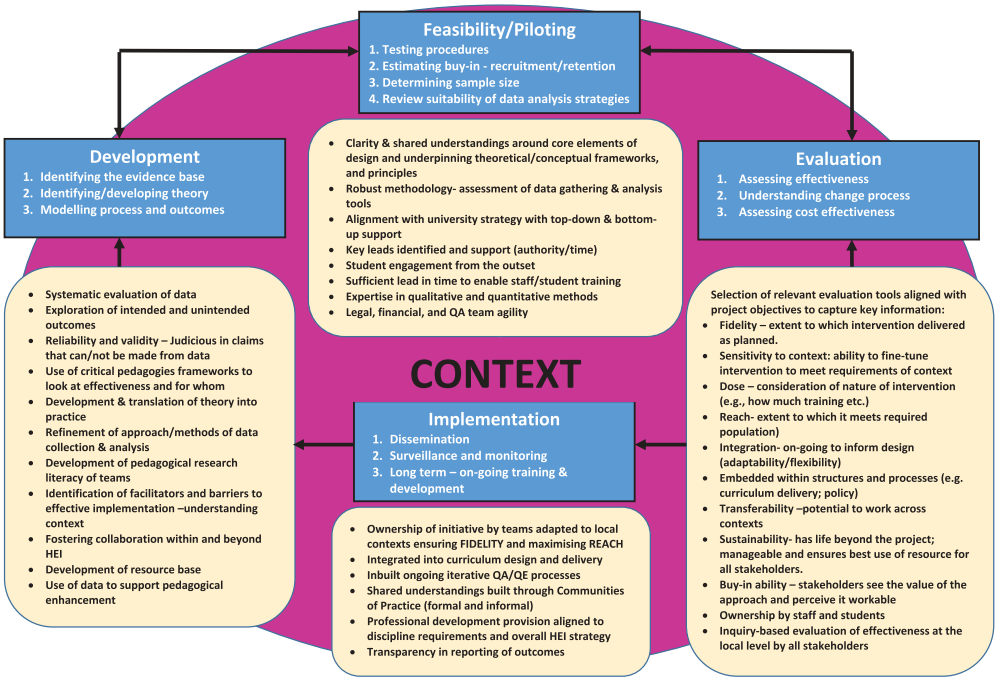

Figure 4. Considerations in Managing Complex Interventions (adapted from MRC, 2008, Moore et al., 2014; Evans & Zhu, 2018)

2008, 2010, 2011); useful guidance on process evaluation can be found in Moore et al., (2015), p. 4). The core themes emerging from the data and reviews of the literature to support effective learning gain initiatives include: (i) ensuring ownership of learning gain approaches at all levels; (ii) employing integrated approaches that build in support from all stakeholders; (iii) embedding learning gain approaches within curriculum design and delivery; (iv) training to support shared understandings of initiatives; (v) using data effectively to support enhancements in pedagogy requiring nimble data mining and analysis; (vi) effective dissemination of research to support pedagogical enhancement.

In addressing such concerns already highlighted, the importance of ongoing integrated evaluation of interventions cannot be underestimated. Key elements of the development and evaluation process and the relationships between them are shown in Figure 4 adapted from MRC guidance. These elements include: feasibility and piloting, evaluation, implementation, and development; the nature of such processes may or may not be linear or cyclical; there are many possible permutations of the process. The importance of context cannot be underestimated; interventions need to be sufficiently flexible to mould to local contexts, and in turn, will be shaped by how they play out in naturalistic settings. Several key constructs highlighted by Moore et al. (2015) in the MRC guidance have applicability within HE and the learning gain context, to include: (i) *fidelity* – the degree to which an intervention is delivered as intended; (ii) *dose* – the quantity of the intervention implemented; (iii) *reach* – whether the

intervention reaches and impacts the intended population. Linked to fidelity, dose, and reach is the importance of *sensitivity to context* especially if aiming to embed ideas, and to support sustainability agendas. The effects of an intervention may be consistent or vary across contexts; learning gain projects need to be able to capture the minutiae if we are to understand what works well, and under what conditions, and with whom. Crucially, and with the U.K. learning gain projects in mind, time should be factored in to ensure sufficient evaluation of the feasibility and piloting stage of interventions prior to wider scale implementation. Even with high-quality designs, time is needed to build shared understanding of the implementation process, the principles underpinning the research, the mechanisms to support the process, and training for academics and students in all stages of the process.

In choosing an appropriate approach to measure learning gain, there are limitations with large- and small-scale measures; the former not being attuned to contextual idiosyncrasies (Ifenthaler, 2017), and the latter may have little relevance to other contexts. To address this issue, Caspersen et al. (2017, p. 28) argue that:

> Perhaps the middle road is the most productive: the systematic development of different indicators, with systematic comparisons of results, will probably provide the best way forward in terms of costs and benefits. Also, systematic meta-reviews of the results and experiences from a wide range of assessment projects may be useful.

Systematic meta-reviews have value, but we must not ignore the rich bank of qualitative data that exist within HE research. Much emphasis has been placed on the use of experimental designs and randomised control trials within the learning gain arena, but this should not detract from the importance and credibility of good qualitative research. We must be careful that we do not privilege quantitative over qualitative evidence and vice versa (NESTA, 2016 quoting Stephen Morris). As part of this work, greater sophistication in research designs is needed to explore variation at the individual level as well as at the group level (Asikainen & Gijbels, 2017), and in exploring the causal mechanisms that underlie differential learning outcomes (Mountford Zimdars et al., 2015).

For all types of analyses, a key limitation is how constructs are defined and measures are operationalised, rendering it difficult to make comparisons across studies or to replicate approaches (Duckworth & Yeager, 2015). To address the different interpretations of constructs, there are calls across the sector for greater standardisation of the definition of concept, measures, and of measurement instruments (Van Der Zanden, Denessen, Cillessen, & Meijer, 2018) derived from clearly specified process models: 'Investment in precisely targeted, theoretically based, interventions could help students optimize their potential and would provide empirical tests of proposed process models of tertiary achievement' (*Richardson, Abraham, & Bond, 2012,* p. 376).

The critical questions to ask are: to what extent are these measurement principles and statistics relevant in this context? As Bond and Fox (2015, p. 5) opine, 'we ought to spend more time investigating our scales, than investigating with our scales', and it may be, that the answer to our questions is that our endeavours are not valid in psychometric terms, but that they do have valid usefulness.

Part three: the pedagogical imperative of learning gain policy

> . . .the true value of learning gain data is to be found not in its potential to enable a national comparison between universities, but in its ability to influence and improve teaching, learning, and student development (Haddelsey, Speight, & Brumhead, 2017; HEFCE, 2017a/Office for Students, 2018a)

In this section, we argue that the key aim of learning gain measures should be to inform pedagogy and not merely to provide proof of quality (*Sands et al., 2018*). The emphasis should not be about promoting homogeneity, it should be concerned with maximising learning and teaching effectiveness in specific contexts which may or may not be generalisable to wider populations (*Powell et al., 2018*). As noted by *Turner et al. (2018)* 'It is unlikely that there will be a single solution, and institutions will need to adapt and contextualise any learning gain measure that they employ'. Effective pedagogical approaches to measuring learning gain require a comprehensive understanding of local contexts, the idiosyncratic nature, composition, and needs of the student population, the core requirements of the discipline, and how best to teach and assess those requirements.

Learning gain initiatives need to be embedded within the curriculum although there are currently many ways in which learning gain approaches can be used to support enhancements in pedagogy, with HEIs often combining several approaches. Table 2 identifies seven ways in which learning gain approaches have been used with variable success to inform pedagogy, the rationale for such choices, and indications of the efficacy of specific approaches to include: (i) use of specific measures to test student understanding (e.g. critical thinking); (ii) exploring student satisfaction; (iii) use of learning analytics; (iv) using metrics to explore students (individual and group) learning trajectories; (v) exploring the role of individual difference variables on learning; (vi) considering the impact of specific learner behaviours on student learning outcomes; and (vii) exploring the impact of pedagogical interventions that may combine a variety of approaches. As identified in Table 2, in choosing an approach, there is an inevitable compromise between methodological rigour, high-quality pedagogy, and what is feasible within naturalistic settings. It is notable that most of the HEFCE/OfS (2015–2018) learning gain projects are not embedded within the curriculum/institutional processes, and this limits both sustainability and scalability.

A key elephant in the 'HE room' is that of student engagement which is impacted by variable levels of student attendance; the latter of which is a key factor in impacting learning outcomes (Schneider & Preckel, 2017). How HEIs address the attendance issue along with autonomy and agency concerns is fundamental. Student surveys have their place but as part of a suite of evaluative measures to better capture the potential offered by the student voice (Darwin, 2017). Big data also have massive potential but whether HE systems are sufficiently refined and have the capacity to handle such complex analyses and especially across institutions is debatable. A key question is how can HEIs work together to identify systems and processes that are fit for purpose, and how can we train colleagues (staff and students) most effectively in the ethical use of data. It is becoming increasingly possible to triangulate findings from learning gain projects although the use of different measures to measure the same and different things

Table 2. Pedagogical focus of learning gain approaches.

Focus	HEFCE/OfS learning gain projects	Rationale/Issues	Findings
(1) Using **specific measures to test student understanding** within and beyond existing curriculum offer to include individual and multi- dimensional frameworks versus one-dimensional tests (e.g. standardised tests – generic concept inventories, generic critical thinking tests e.g. CLA+, AHELO; student self-assessment of competencies; conceptual frameworks measuring range of variables e.g. cognitive, metacognitive, affective, socio-communicative; epistemological beliefs, attitudes; Discipline-specific concept inventories)	*Callaghan and Aloisi (2018)*; *Ilie et al. (2018); Pampaka et al. (2018)*; *Speight et al. (2018)*; *Ylonen et al. (2018)*	**Rationale** • Standardised assessments have potential to be used across institutions to enable comparisons. • Efficient use of time and allows for comparison across HEIs or does it? • Ability to tap into large populations to test validity and reliability of constructs. **Issues** • Applicability of U.S. models such as CLA to U.K. context. • Assumes pedagogical content covered is similar across HEIs. • Can learning gain be independent of subject-specific content? • Concept inventories may be more appropriate to specific disciplines/groups of disciplines such as STEM. • Administration/financial costs versus value.	Students who performed worse in their first tests exhibited larger absolute improvements in their conceptual understanding than their better performing peers (n = 73) (*Ylonen et al., 2018*) Role of different disciplines and interaction of individual differences identified (*Speight et al., 2018*). MCQ tests more reliable than open-ended questions-based tests such as CLA+ (Caspersen et al., 2017). Scores from the CLA+ test only meaningful if aggregated at the course or institution level; not reliable measures to use for student-level decisions. A ceiling effect was found with >50% of participants had little room for improvement. (*Callaghan & Aloisi, 2018*). CLA+ scores appear discipline-dependent (Kandiko Howson, 2018).
(2) **Exploring student satisfaction with the intent of improving learning and teaching delivery. Identifying strengths and areas of relative weakness across disciplines. Exploring trajectories over time.** Use of student satisfaction and engagement surveys (e.g. National Student Surveys, Engagement Surveys); module and programme evaluation data.	*Stoakes and Neves (2018)*	**Rationale** • Student satisfaction/engagement can be used as proxies for the quality of teaching. • Relatively easy to administer. • Such surveys provide one strand of evidence to explore issues at depth. **Issues** • At best, they are poor proxies for the quality of teaching. • NSS-type surveys have led to metrics chasing in HEIs rather than focusing on improving fundamental aspects of learning and teaching? • To what extent do they measure learning and teaching elements that matter?	Assessment remains the area that students appear least satisfied with internationally. The assessment feedback gap has been much debated and linked to different conceptions/ beliefs lecturers and students have about assessment (Evans, 2013). Tension between accountability and improvement motivations impact use of student survey data (Darwin, 2017)

(*Continued*)

Table 2. (Continued).

Focus	HEFCE/OfS learning gain projects	Rationale/Issues	Findings
(3) Use of learning analytics **'Big Data' and Artificial Intelligence**	*Baum (2018)*; *Rogaten and Rienties (2018)*	**Rationale** • To make better use of existing data sets to identify patterns rather than seeking to exploit new ones – efficiency arguments. • As a tool for quality assurance and improvement. • To boost retention rates. • As a tool for assessing and acting upon differential outcomes among students. • To support personalised/adaptive learning. Dashboards of individual student data can enable continuous feedback of engagement and academic performance as a means to 'nudge' behaviour. • To support students in using data well. **Issues** • Data does not give us the WHYs? It does provide us with a starting point. • Ethics in management and use of data. • HEIs are being drowned in data. • Need for training in collection and use of data. • Capability of systems to capture relevant data sets. • How being used to inform pedagogy?	Potential identified in numerous reports but not yet realised. First, national deployment of learning analytics in world taking place in U.K. with JISC working with 50 HEIs. Evidence to suggest that learning analytics can help to improve outcomes. Issue of investment in systems and people. (See Ifentahler, 2017; Sclater et al., 2016; Shacklock, 2016; VitalSource learning matters quoting Hughes and King, 2018). https://www.jisc.ac.uk/sites/default/fil es/learning-analytics-in-he-v2_0.pdf http://www.policyconnect.org.uk/hec/sites/site_hec/files/report/419/fieldre portdownload/frombrickstoclicks- hecreportforweb.pdf Lack of student interest in exploring data unless embedded in course design. Lack of use of data for enhancement purposes, also linked to senior management engagement in initiatives. (Kandiko Howson, 2018)

(*Continued*)

Table 2. (Continued).

Focus	HEFCE/OfS learning gain projects	Rationale/Issues	Findings
(4) Using metrics to explore student learning trajectories – in order to better understand the process of learning and specific individual/group dynamics Supporting learner transitions – identifying patterns of development for cohorts, for specific groups of learners to use at the individual/group level to measure progress (e.g. focus on widening participation). Use of grades to measure learning gain	*Callaghan and Aloisi (2018)*; *Cameron (2018)*; *Ilie, Vermunt, Vignoles (2018)*; *Scalise et al. (2018)*; *Ylonen et al. (2018)*	**Rationale** • Emphasis on how best to support learner transitions from point of entry to support students to achieve their academic potential. • To understand the characteristics of students/groups of students and how these may impact progress. • To address assumptions that staff may have about students and their learning. • To identify at risk students. • To develop holistic profiles of students based on academic and extra-curricular activity to explore interrelationships at individual and group level. • To explore the trajectories of different groups of students to look at patterns of performance and identify trends with the aims of addressing inequalities/enhancing design of teaching and supporting learners. To explore whether any specific groups are being disadvantaged by the curriculum. • To be able to use data at individual level to raise awareness and to enable student reflection. • **Issues** To understand how learners develop transferable skills. • Trajectories impacted by individual difference and contextual variables. Nature of research design matters. • Need to be able to explore individual patterns of development. • Changes between point 1 and point 2 may mask variations within the year. • Scores are dependent on activities – e.g. assessment deadlines.	Importance of conceptual support early on in students' development (Peeters et al. (2016). Students' variable use of metacognitive monitoring and strategy use is evident early on in their HE journey and emerged through exposure to course content and not easily detectable through prior achievement measures. Higher achieving students set more specific goals, made strategy changes that were specific to their performance, and selected more effective study strategies than lower achieving students (Di Francescca, Nierfeld & Cao, 2016). Greater value of differentiated support at an early stage in a students' HE career (*Scalise et al., 2018*). Richardson et al. (2012) highlight the importance of interventions early in a students' HE career given that performance self-efficacy and goal orientation are likely to be more fluid during the early stages of skill development. Multifaceted interventions may be more effective in supporting students, but interventions targeting specific cognitive changes may be more cost-effective (Richardson et al., 2012). The stability or volatility of student' approaches to learning are associated with the initial subgroup they belong to (Asikainen & Gijbels, 2017). Students' learning patterns are associated with their attributions of academic success and self-efficacy and their perceptions of the learning environment, which may also be dependent on their current learning pattern (Vermunt & Donche, 2017). Limited value of grades as an effective measure of learning gains (*Callaghan & Aloisi, 2018*; *Ylonen et al., 2018*).

(*Continued*)

Table 2. (Continued).

Focus	HEFCE/OfS learning gain projects	Rationale/Issues	Findings
(5) Predictive: Exploring the **role of individual difference variables** on students' learning (Importance of psychological, social, and human capital). Consideration of student characteristics and impact on attainment (innate and contextual to include intelligence measures; entry tariffs; socio-economic class; gender; ethnicity, protected characteristics; etc.). Utilisation of existing measures, new measures, and hybrid versions Much emphasis on affective measures (self-efficacy; self-esteem; grit; resilience; motivation; confidence)	*Cameron (2018)*; *Forsythe and Jellicoe (2018)*; *Pampaka et al. (2018)*; *Parker (2018) Rogaten and Rienties (2018)*; *Ylonen et al. (2018)*	**Rationale** • To identify which variables have most impact on learning and which ones can be addressed as part of curriculum design and delivery. • To reduce differential learning outcomes. • To promote inclusion through identifying the needs of specific groups and identify where best to place efforts. • To ensure curriculum is accessible to all students and that no student is inadvertently disadvantaged. • To tailor provision to the needs of specific students; identify practices that are good for all students. • Potential to address 'unconscious bias regarding certain groups of students'. **Issues** • Being able to minimise the 'noise' in naturalistic learning contexts – accounting for confounding variables. • Relative stability and mobility of individual difference constructs. • Ability to develop adaptive rather than adapted learning environments as part of inclusive approaches. • Professional development requirements. • Competing priorities on the part of the academic, and in relation to promotional criteria.	Significant differences between student groups identified (n = 5103). Variations in learning gains mainly due to module/programme design characteristics with only a small amount accounted for by individual difference variables (*Rogaten & Rientes*, 2018). Psycho-social demographic variables are at best small predictors of performance although when considering marginal gains, they can make a difference (Owens & Tibley, 2014). However students' degree of academic and social integration impacts student retention (Bluic et al., 2011; Van der Zanden et al., 2018). Prior achievement and intelligence show the strongest relation with achievement of all variables (based on 3330 effect sizes) (Schneider & Preckel, 2017). For first years, previous academic success and intrinsic motivation related to learning outcomes (Van der Zanden et al., 2018). Key factors impacting student performance include academic/performance self-efficacy, goal orientation, and effort regulation (Richardson et al., 2012). Highlights the importance of goal-setting interventions. Emphasis on affective and behavioural (employability characteristics) with relatively little emphasis on cognitive measures in U.K. context.

(*Continued*)

Table 2. (Continued).

Focus	HEFCE/OfS learning gain projects	Rationale/Issues	Findings
(6) Considering the **impact of specific learner behaviours on outcomes** (e.g. **Involvement/ participation**: Student engagement with VLE; attendance; **Nature of engagement**: immersion: time on task; distribution of time on task; **Interaction**: participation in dialogue; questioning; generation of questions; quality of questions; online presence; contribution to teams' peer collaboration; **Engagement with technology;** **Interpersonal/intrapersonal**: interaction; collaboration; team skills; cooperation; practical application; self-management of tasks.)	*Forsythe & Jellicoe (2018)*	**Rationale** • Identify behaviours linked to positive outcomes to enhance retention and learning outcomes. • Share with learners what we know about productive 'behaviours' to support positive changes in approaches. • Can be used to model personalised learning systems. **Issues** • Learners variably disposed to make good use of information provided. • Comparative data may be detrimental to some students depending on their dispositions. • Needs to acknowledge 'outliers' those who do not perform to type.	Attendance key factor impacting student learning outcomes (Schneider & Preckel, 2017).Studies have highlighted the potential value of student engagement in activities beyond the taught curriculum, the positive impact of independent study, and the negative impact of high levels of paid work on learning outcomes (Blackman, 2018). Mastery goal orientation, challenging interventions from feedback, and motivational intentions are most essential personal constructs linked to behaviour change (*Forsythe & Jellicoe, 2018*).

(*Continued*)

Table 2. (Continued).

Focus	HEFCE/OfS learning gain projects	Rationale/Issues	Findings
(7) Explore the **impact of pedagogical interventions** (e.g. use of formative feedback; changes in assessment approach; development of student self- assessment skills; delivery model; teaching hours; value of placements in skills development/authentic assessment; engagement activities; inclusive curriculum; critical reflection/thinking; service delivery models; Student self- assessments; etc.)	*Cameron (2018)*; *Sands et al. (2018)*; *Scalise et al. (2018)* *Speight et al. (2018)*; *Turner et al. (2018)*; *Ylonen et al. (2018)*.	**Rationale** • Explore the relative effectiveness of different curriculum approaches and impact on different groups of students with intention of reducing differential learning outcomes • To enhance curriculum design and delivery. To support students to self-assess and enable lecturers to gauge student understanding of a particular pedagogy. • Enhance student self-efficacy through focus on development of self-assessment skills. • Focus on formative assessment to support student progression with emphasis on developing conceptual understanding. • Value of undergraduate research in supporting learning. **Issues** • Identifying the key elements of the pedagogic design that make the difference. • Variable impact of approaches due to individual differences including volition of students and staff. • Need for integrated approaches across modules/programmes to be able to embed key ideas. • Competing demands on time. • Need to know the precise context in order to be able to reproduce/translate the idea to new contexts. • Knowledge of the discipline and the programme. • Student attendance.	Peer instruction combined with self-assessment was successful in supporting overall class improvement (n >= 200). Students with higher academic self-efficacy scored higher in formative assessment. The approach enhanced performance of low performers and also fostered high self-efficacy beliefs in high performers. Highlights the importance of an integrated pedagogical approach within the disciplines to support skills development and especially embedding self-efficacy in design. *Ylonen et al. (2018)*. Using formative assessment support – 43% reduction in students at risk (n = 943). The timing and nature of support mattered. The group that received additional support with conceptual understanding early on in the programme continued to outperform the original comparison group even though the latter group had subsequently been given the treatment. Early support had sustained impact over time and especially for certain groups. *Scalise et al. (2018)* Value of Inquiry: Importance of developing research methods provision with UG populations but also dependent on perceived relevance within specific disciplines and how framed within the discipline. Greater impact on students with lower prior achievement. Research focus benefitted all students but there were variations in the extent of impact related to individual difference variables and context (discipline) (n = 5027). *Parker (2018)*

is not helpful. We need to be building a compendium of reliable and valid learning gain tools and measures for the HE fields to enable more effective comparisons; the HEFCE/OfS learning gain projects can make a valuable contribution to this agenda. What is evident is the power of early interventions to impact student learning outcomes, coupled with an awareness of the needs of specific groups to enable targeted support where it matters most, to support student strategy development both in the immediate and longer terms.

In such ways, learning gain as a concept has huge potential in being able to offer valuable insights into the learning process of all students if applied in a critical way as an integral part of curriculum design and delivery, and through utilising robust research design. Learning gain approaches have the potential to open the lid on the learning process through exploring what students know, how they come to know, and in what ways (*Scalise et al., 2018*). Through such approaches, we can explore the patterns of understanding students' exhibit as they move towards mastery, and how our pedagogies can enable students to manage their learning more effectively. In doing so, academics and students can be supported in coming to know 'what works, why, when, and for whom', to make informed decisions about learning and teaching and thereby attend to individual agency in learning and workplace contexts. More informed use of data can support enhancements in individual and organisational learning (c.f. recent work on learning analytics, e.g. Sclater, Peasgood, & Mullan, JISC report, 2016). Learning gain approaches should not be metrics chasing tools. Researching practice in a rigorous way should be agentic in enabling individuals and organisations to challenge those approaches not informed by a strong evidence base or not sufficiently nuanced to the requirements of the specific context/discipline in question.

Importantly, in analysing individual and contextual variables impacting learning at institutional and individual levels, there is considerable potential to address social justice and equity issues to reduce differential student learning outcomes (Mountford Zimdars et al., 2015). Exploring how individual and contextual factors impact learners' transitions in HE can inform decision-making about teaching design imperatives. Investigation of the impact of specific pedagogies and organisation of curricula on students' access to, and engagement with, learning and consequent outcomes can be interrogated along with the relative effectiveness of different approaches to learning and the likelihood of a range of outcomes through the different choices that are made. Crucially, by providing insights into those variables that matter most, it can allow resources to be allocated more effectively by attending to those areas that have the potential to make the most difference.

Not only can such approaches provide insights into those factors supporting high levels of learning gain for different populations (*Callaghan & Aloisi, 2018*; *Rogaten & Rienties, 2018*), through the combination of learning gain and learning analytic approaches, there is the potential to personalise learning (Sclater et al., 2016). However, the use of data analytics requires care in terms of data ownership and security along with considerable resource. For example, investment is needed in flexible data mining tools, new statistical methods including machine learning algorithms, visualisations that provide all relevant stakeholders with an overview of relevant information, and recruitment and development of specialised staff (Ifenthaler, 2017). Training is also

essential in the uses of such tools and techniques along with developing enhanced understanding and contextualisation of moderating impacts on learning gain within specific contexts at the individual, module, programme, faculty, and university levels.

Considerable investment has been placed globally in developing learning gain measures in search of a pedagogical El Dorado: the meaningful assessment of high-quality teaching. However, the relationship between teaching and learning is not a linear one, with many contextual and individual difference variables impacting results. To reveal those practices that support high-quality pedagogies, we need robust research designs supported by transparent reporting of information to substantiate claims and enable replication of approaches across contexts. There is a significant body of literature on high-quality learning and teaching practices underpinned by sound theoretical frameworks that can offer guidance in this endeavour. For example, Dinsmore's (2017) work on effective strategy use; Schneider & Preckle's (2017) meta-analysis on variables associated with achievement in HE; high-impact pedagogies' analyses (Evans et al., 2015; Kuh, O'Donnell, & Schneider, 2017; Strang, Bélanger, Manville, & Meads, 2016); differential student performance (DiFrancesca, Nietfeld, & Cao, 2016; J. T. E. Richardson, 2015; M. Richardson et al., 2012; Seifert et al., 2014; Mountford Zimdars et al., 2015); cognitive and educational psychology insights (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Kozhevnikov, Evans, & Kosling 2014; Waring & Evans, 2015), and neuroscientific applications (Dubinksy, Roehrig, & Varma, 2013).

Assessment of student achievement is at the heart of learning gain approaches. It could be argued that if assessment was fit for purpose, there would be no need for additional learning gain measures. Data generated through student assessment should be usable for learning gain purposes (*Boud, 2018*), although the inherent inadequacies of assessment in HE make the use of grades as an accurate outcome measure of students' performance questionable (Bloxham & Boyd, 2012). However, as *Sands et al. (2018)* note, we cannot expect one overarching assessment to capture all we need it to, but we can expect assessment to do a better job of capturing improvements in key areas of knowledge, skills, and understanding through identifying progressive points in key study modules, and, or across an entire programme. Our ability to extract useful information from student assessment data is fraught with difficulty given that assessment criteria and scaling vary across programmes, institutions, and nations and reflect different knowledge structures, so their value for comparison purposes is limited (Caspersen et al., 2017). It is also difficult to ascertain whether any increases in marks are the result of learning gain or different norms and cultures *(Ylonen et al., 2018)*.

To address this problem, *Boud (2018)* argues for a common set of standards and criteria which could be achieved through a programme-level approach. However, given the drive to provide increasingly flexible programmes of study for students, programme-level assessment approaches are difficult to implement where students are studying across a range of disciplines and faculties. At a fundamental level, questions have been raised as to whether assessment actually measures learning at all, or whether it simply measures an individual's ability to successfully pass assessments. In being able to use assessment data, we need to consider the authenticity, relevance, and validity of assessments; the extent to which assessments accurately measure intended learning outcomes, and whether the proposed meaning of test outcomes or uses of a test are warranted by its qualities and justified within its context (Messick, 1989).

Placing pedagogy at the centre of learning gain initiatives: key considerations

In sum, the following principles are central to the implementation of learning gain approaches focused on supporting enhancements in learning and teaching:

- A holistic and integrated approach to curriculum including assessment design
- A research-informed approach from pedagogical and methodological perspectives
- Embedding approaches within disciplines – discipline aligned
- Ensuring learning gain ownership

Learning gain approaches should be holistic and integrated into the curriculum including assessment design

Learning gain measures and approaches should be embedded into the fabric of how students learn and are assessed (*Ylonen et al., 2018*), and not extraneous to them. Leaning gain needs to consider the process of student learning itself (*Varsavsky Matthews, & Hodgson, 2014*). For this to be successful, explicit integration of learning gain measures into the curriculum is essential, along with the opportunities for students to engage in a dialogue around their growing knowledge, skills, and experience in the area the measure is centred. This also requires training for students and staff in the use and application of information.

Curriculum design impacts authentic and meaningful measurement of learning. Bringing isolated goals and learning objectives together into a coherent whole is essential in supporting student and staff mastery and independence in learning (Evans, 2016). 'The knitting together, or relationships among the knowledge and concepts [and skills] are very important to the development of student [and staff] understanding' (*Scalise et al., 2018*). An integrative pedagogical approach is essential in order that we can consider the whole learning journey of the student and how progressive learning opportunities are to be built into the curriculum and signalled to students (e.g. *Evans' EAT, (2016)* integrated Assessment and Feedback Framework drawing on Evans (2013)).

Clear signalling of what the core concepts are, how they will progressively develop, what activities will support the development of such conceptual understandings, and how they will be assessed is needed. The authenticity of assessment and whether it actually measures what students need to know, and what is valuable for them to know, and to be able to do is being questioned with the increasing emphasis on the importance of non-cognitive skills (self-confidence, self-esteem, relationship-building, negotiation skills, and empathy) over cognitive attributes in enhancing well-being and future employability in an increasingly artificial intelligence world (Edge Foundation, 2016; Haldane, 2015; Heckman & Masterov, 2007). We need to consider how we are progressively measuring those learning outcomes that matter in a meaningful way over the duration of a student's academic career within HE. Marking and grading need to be fit for the key purpose of indicating what a student has achieved (*Boud, 2018*). A programme-level approach to assessment to ensure assessments are made in relation to programme-level outcomes using standards and criteria that are consistent across outcomes, and that the same programme-level outcomes are assessed at various points

in the students' journey is advocated (Rust, 2017). At the same time, it needs to be acknowledged that there are many different approaches to programme-level assessment attuned to the requirements of the discipline and context.

Learning gain approaches should be research-informed from pedagogical and methodological perspectives

A rigorous, research-informed, and critical approach is needed to consider those variables identified as most important in impacting student learning outcomes. From a teaching perspective, this includes an emphasis on developing student social interaction skills, the promotion of meaningful learning including the use of conceptually demanding learning tasks, clear learning goals, and the development of student self- and peer assessment (Evans, 2013; Schneider & Preckel, 2017; *Ylonen et al., 2018*). Supporting students to self-manage and self-evaluate their learning is well rehearsed in the literature (Boud & Molloy, 2013), as part of an increasing emphasis on the promotion of self-regulation to support the quality of student learning (Evans, 2016, 2018; Peeters, De Baker, Kindenkens, Triquet, & Lombaerts, 2016). In attending to cognitive, metacognitive, and affective dimensions of learning, the importance of student self-efficacy in impacting learning outcomes is evident (Komarraju & Nadler, 2013; Richardson et al., 2012). Students' approaches to learning are associated with the initial subgroups they belong to (Asikainen & Gijbels, 2017), suggesting the importance of identifying the needs of such 'tribes' from the outset and in addressing the relational dimension of learning along with cognitive dimensions. Similarly, DiFrancesca et al. (2016) identify the importance of context along with prior academic student achievement in impacting student learning outcomes in HE, with context impacting the adoption of learning patterns early in students' HE careers.

Forsythe and Jellicoe (2018) advocate pedagogical interventions that promote goal-setting in order to support students to develop higher level motivational intentions (e.g. mastery of a task rather than a specific performance outcome) and the importance of targeted feedback that challenges students to consider how to address areas of relative weakness. Strategies and approaches to learning need to be explicitly modelled as it cannot be assumed that all students will intuitively select the most appropriate approach given their varied entry points, backgrounds, experiences, and demands on their time (Gibson, 2015). Learning gain initiatives need to consider students' beliefs, attitudes, conceptions of learning, learning orientations, and social interactions within their programmes, given their potential to impact students' approaches to learning and learning outcomes (Bennett & Kane, 2014).

The way that different variables interact to impact student learning outcomes is context related, and at the micro-level can be highly individualistic, so what applies for one student may not apply for another; more research is needed on the impact of the learning environment on student success. Caution is needed in the interpretation of results given the role of moderator variables (e.g. socio-economic status; assessment design) in impacting student learning outcomes (*Ilie et al., 2018*). In being able to compare results across studies, the need for transparency in reporting is paramount given the number of potentially confounding moderator variables (e.g. timing of interventions; different or same tools used to measure similar constructs; nature of tools used; process in how approaches have been applied; the composition of

populations involved; disciplinary factors; clarity about whether data sets can be aggregated; variance within or between modules, subjects, and institutions). To address questions like the ones proposed above, differentiating is required between the general (equivalent in magnitude and direction for all students) and conditional effects (where certain subgroups in the sample are affected to a greater or lesser extent by the programme/experience/intervention), Seifert et al. (2014, p. 534).

Learning gain initiatives need to be authentic and aligned with the requirements of the discipline

The potential effectiveness of approaches to learning gain depends on how they are implemented at the micro-level. Module design, delivery, and especially the characteristics of assessment can 'straitjacket' students' learning (*Rogaten & Rienties, 2018*), given the known relatively high impact of module-specific characteristics on students' learning gains in their first year compared to individual difference variables (Nguyen, Rienties, Toetenel, Ferguson, & Whitelock, 2017). In considering generic learning gain approaches, Schneider and Preckel's (2017) review of meta-analyses signals that improving students' strategies within the context of their academic discipline is more effective than training them in extracurricular settings with artificially created problems. Generic measures (tests/inventories) not closely related to the module context can result in spurious findings (Asikainen & Gijbels, 2017). In supporting the argument for authentic integration, Ruge and McCormack (2017) found students' skills for employability were facilitated through discipline-based curriculum design linking university and industry skills expectations, where clear interweaving of learning contexts and assessments to enable students to identify academic and professional learning dimensions were most effective.

There is an assumption that discipline-specific methods are important. They may be so less on grounds of construct validity than of face validity (i.e. they need to fit with the ontological and epistemological orientations of academic staff who are 'fully paid up' members of their disciplines). Epistemological and ontological assumptions about the nature of knowledge, and how you know within disciplines, can be crucial in the development and application of learning gain measures. For example, it is notable that the development of concept inventories is much stronger in STEM disciplines. In designing potential learning gain measures, disciplinary nuances must be attended to. For example, Turner et al. (2018) found 'The differential framing of research methods in disciplines shape[d] students' pedagogical engagement with research methods, … their research orientations, learning motivation, and sense of self-efficacy'.

Learning gain ownership

The engagement of students in learning gain initiatives is an issue (in terms of actual numbers, representativeness and retention) (Kandiko Howson, 2017). It is imperative that students invest in the process as part of an ipsative approach (Burke, 2017) to explore all dimensions of their learning (cognitive, affective, and metacognitive) as part of a self-regulatory approach (Evans, 2016, 2018); this requires the approaches to be relevant and integral to a student's learning experience. Upskilling is needed for all in the 'how and what to' measure, and in the interpretation of such findings. A 'third person' perspective (Rosenfeld & Rosenfeld, 2011) is advocated whereby students are

not seen as vessels to fill and from which to pluck data but are fully briefed about the purposes of research and are clear on how all data collected is being used and to what ends; active participation by students in the research process is crucial. Students should have opportunities to be centrally involved in research design, data collection, and analysis, and be debriefed on the outcomes of any data gathering exercise; these activities should also not be solely the preserve of final-year students. Students and staff collaborating effectively as partners in learning and teaching is arguably one of the most important issues being faced in HE in the twenty-first century (Healey, Flint, & Harrington, 2014). The student as an active collaborator and co-producer has more potential for meaningful transformation (Dunne, Zandstra, Brown, & Nurser, 2011). More effort needs to be expended in developing meaningful partnerships to address the pervading concern that full partnership is rarely achieved (Evans et al., 2015; Healey et al., 2014).

Recommendations for policy and practice

We have argued that there is no one single magic bullet that can solve the learning gain issue. There are many different definitions of learning gain aligned to how one sees the main purposes of HE. The quest for a universal measure of learning gain is a futile one given the implicit and explicit differences in context. Standardisation of learning outcomes across HE will not enable valid comparisons of learning across specific contexts. A single solution is unlikely; it is important that institutions adapt and contextualise any learning gain measures to suit specific requirements. From an ethical perspective, while also looking for more robust ways to measure learning gain, 'we should continually consider whether our collective actions will leave our field in a state that is better or worse than when we entered it' (Edwards & Roy, 2017, p. 56). Learning gain initiatives should be focusing on developing more meaningful measures of learning and teaching, especially in the pursuit of equity in learning for all students, and addressing the current deficits in existing measures that are currently poor proxies for learning gain and the measurement of effective learning and teaching.

We have already noted the limitations of the definition of learning gain 'as distance travelled', and Caspersen et al. (2017) add to this debate in cautioning against mixing the measurement of learning growth with the measurement of knowledge at a given point in time. They argue that growth does not tell us about proficiency. Unless there is a value-added component, any measures of knowledge do not actually tell us about learning, and grades do not tell us directly about the quality of learning or the quality of teaching for that matter. Similarly, while serious questions have been raised regarding the value of students' perceptions of their own learning over more objective measures, students can accurately assess their experiences of the learning environment, and with focused attention on developing their self-regulatory capacity can make accurate estimates of their learning (Boud & Molloy, 2013; Evans, 2016, 2018; Waring & Evans, 2015).

In opening up the black box of HE, greater emphasis needs to be paid to the factors contributing to students' learning rather than focusing narrowly on outcome measures (e.g. students' acquired competence) (Caspersen et al., 2017; Strang et al., 2016).

If our institutions are to live up to their potential in developing and supporting equitable educational opportunities for all students, our understanding of student learning and development must become more nuanced, with educators using this knowledge to modify and tailor programs, practices, and policies accordingly (Seifert et al., 2014, pp. 560–561).

Learning gain initiatives will work most effectively if embedded within the formal curriculum (Kandiko Howson, 2017), the success of which requires intentional design, theoretical grounding, a clearly defined purpose, and nuanced to the requirements of the context (Kuh et al., 2017). A multidimensional view of student success is required if we are to gain better understandings of those factors impacting student learning outcomes (Van Der Zanden et al., 2018). There is currently a paradox that needs resolving. On the one hand, research from several perspectives (e.g. neuroscientific, educational, and cognitive psychology inquiry; high-impact pedagogies research (Kuh, O'Donnell, & Scheider, 2017) has given us considerable insights into effective practices; however, few studies excepting Finlay and Brown McNair (2013) have examined whether all students benefit from good practices in similar ways (Seifert et al., 2014). This is especially important given that students with certain characteristics (e.g. lower socio-economic groups; first-generation HE students; black and minority ethnic; disability, etc.) have consistently lower attainment and progression outcomes (Mountford Zimdars et al., 2015).

High-quality implementation research is required on teaching quality and learning gain, especially in the U.K. context. The HEFCE/OfS learning gain projects make a valuable contribution to this endeavour, but more is required, with greater sophistication and rigour in research designs as a starting point. The extent to which HEIs collect and use empirical evidence well to inform policy, practice, and pedagogy has been noted (Seifert et al., 2014). To avoid '*garbage in and garbage out',* rigorous research design and reporting is essential (see Table 1). Moving from 'islands of innovation' to a focus on engaged learning practices (Kuh et al., 2017) through enhanced collaboration within and across HEIs is essential if we are to pool the wealth of information available to enhance understanding of learning gain. HEI machinery needs to be agiler in facilitating more effective collaborations across institutions (see HEFCE, 2017b/OfS 2018b; learning gain and addressing barriers to student success projects within in the U.K.); however, accountability and competition work against transparency in the sector.

Investing in the development of the pedagogical research literacy of teams is essential in promoting meaningful approaches to learning gain. Greater efficiency in the mining and interpretation of data for pedagogical enhancement within institutions is a must. Researchers also need to be equipped with a new set of fundamental competencies in order to manage big data, use technology to personalise provision, and develop interdisciplinary learning and teaching approaches cognisant of cognitive, behavioural, social, and emotional perspectives on learning (Ifenthaler, 2017).

In sum, we need to ensure that learning gain initiatives do not detract us from the essential purposes of HE to support students' self-regulation and associated independence in learning, and thirst for lifelong learning. Learning gain initiatives need to be judicious and ethical regarding the use of data, best use of resources, methodologies employed, and inferences that can be reasonably made from such research. Students need to be clear about how their data are being used and ideally should be actively

involved in reviewing data with academics as co-researchers. Investment in training for staff and students in the use and application of data is vital. Learning gain approaches should be integral to curriculum design and delivery and not extraneous to it. Greater focus is needed on process rather than product if we are to have a better understanding of the impact of HE on students' learning, with more emphasis on early estimations of student progress to enable students and academics to use data most effectively. Enhancing shared understandings of concepts, measures, and instruments, transparency in reporting, and investment in developing pedagogical research literacy are all essential in supporting collaborative efforts in the pursuit of meaningful approaches to measuring learning gain within HE. It is hoped that HE reforms (OfS, 2017; REF 2018/02/02) that promote the value of impacts on teaching within/across disciplines and institutions as an important component of research outputs will lead to greater interest in evaluating effective educational practices.

If, as a sector, we can achieve all this, only then we will have at our disposal measures of learning gain that are truly fit for purpose. In acknowledging that learning gain is a messy business, we need to be clear about what is 'good enough' through marrying the best we can from research design and effective pedagogies with what is feasible and sustainable in specific contexts.

Acknowledgments

We would like to sincerely thank Corony Edwards for her insights, guidance, and full support in enabling us to develop this article.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

C. Evans http://orcid.org/0000-0001-9344-8952

References

Academy of Management Learning and Education (AMLE). (2017). Editorial: Establishing methodological rigor in quantitative management learning and education research: The role of design, statistical methods, and reporting standards. *Academy of Management Learning & Education*, *16*(2), 173–192.

Advance, H.E. (2018). Grade point average. Retrieved from https://www.heacademy.ac.uk/individuals/student-success/assessment/grade-point-average#

AEN. (2018). Batchelor degree completion rates. Retrived from http://www.universityrankings.com.au/degree-completion-rates.html

APA. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*(9), 839–851.

Appelbaum, M., Cooper, H., Kline, R., Mayo-Wilson, E., Nezu, A., & Rao, S. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist*, *73*(1), 3–25.

Arum, R., & Roska, J. (2011). *Academically adrift*. Chicago, USA: University of Chicago Press.

Asikainen, H., & Gijbels, D. (2017). Do students develop towards more deep approaches to learning during studies? A systematic review on the development of students' deep and surface approaches to learning in higher education. *Educational Psychology Review*, *29*, 205–234.

Bates, R., Holton, E.F., III, & Hatala, J.P. (2012). A revised learning transfer system inventory: Factorial replication and validation. *Human Resource Development International*, *15*(5), 549–569.

Baum, D. (2018). Towards a measure of learning gain. A journey. With obstacles. *Higher Education Pedagogies*. *This issue*.

Bennett, R., & Kane, S. (2014). Students' interpretations of the meanings of questionnaire items in the national student survey. *Quality in Higher Education*, *20*(2), 129–164.

Biesta, G. (2010). *Good education in an age of measurement*. Boulder, CO: Paradign Publishers.

BIS (Department for Business Innovation and Skills). (2011). *Higher Education: Students at the Heart of the System*. London: BIS.

BIS (Department for Business Innovation and Skills). (2015). *Fulfilling our potential: Teaching excellence, social mobility and student choice*. London: BIS. Retrived from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/474227/BIS-15-623-fulfilling-our-potential-teaching-excellence-social-mobility-and-student-choice.pdf

Blackman, T. (2018). What affects how much students learn? HEPI policy note 5. Higher Education Policy Institute. Retrieved from http://www.hepi.ac.uk/2018/01/08/5266/

Blaich, C., & Wise, K. (2011). From gathering to using assessment results: Lessons from the Wabash national study. *NILOA Occasional Paper No. 8*. Urbana, IL: University of Illinois and Indiana University, National Institute of Learning Outcomes Assessment.

Bliuc, A-M., Ellis, R. A., Goodyear, P., & Hendres, D. M. (2011). Understanding student learning in context: Relationships between university students' social identity, approaches to learning, and academic performance. *European Journal of Psychology Education*, *26*, 417–433.

Blömeke, S., Zlatkin-Troitschanaskaia, O., Kuhn, C., & Fege, J. (2013). *Modeling and measuring competencies in higher education*. Rotterdam: Sense Publishers. Retrieved from https://www.sensepublishers.com/media/1531-modeling-and-measuring-competencies-in-higher-education.pdf.

Bloxham, S., & Boyd, P. (2012). *Developing effective assessment in higher education: A practical guide*. Maidenhead, Berkshire: Open University Press.

Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch model, fundamental measurement in the human sciences* (3rd ed.). New York, NY: Routledge.

Boud, D. (2018). Assessment could demonstrate learning gains, but what is required for it to do so? *Higher Education Pedagogies*. *this issue*. doi:10.1080/23752696.2017.1413671

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment and Evaluation in Higher Education*, *38*(6), 698–712.

Bowes, L., Evans, J., Hughes, D., Olewe-Richards, S., Roberts, J., & Birkin, G. (2015) *UK review of information about higher education: Report on the review of the key information set and Unistats (Annex B: Research on information use by students and their advisers)*. A report to the UK higher education funding bodies by CFE Research. Retrieved from http://www.hefce.ac.uk/pubs/year/2015/201527/

Burgess, A. (2018). How to do well in the national student survey. Retrieved from https://wonkhe.com/blogs/how-to-do-well-in-the-national-student-survey-nss/?platform=hootsuite

Burgess, A., Senior, C., & Moores, E. (2018, February). A 10 year case study on the changing determinants of university student satisfaction in the UK. *PLOS ONE*, *22*, 2018.

Burke, G. (2017). *Ipsative assessment and personal learning gain*. London: Palgrave Macmillan.

Callaghan, A., & Aloisi, C. (2018). Threats to the validity of the collegiate learning assessment (CLA+) as a measure of critical thinking skills and implications for Learning Gain. *Higher Education Pedagogies*, *this issue*.

Calohee. (2018). Calohee. Measuring and comparing achievements of learning outcomes in HE in Europe. Retrieved from https://www.calohee.eu/

Cameron, A. (2018). An investigation into the comparative learning gain and 'value added' for students from widening participation and non-widening participation groups: A case study from sports degrees. *Higher Education Pedagogies*, *this issue*.

Caspersen, J., Smeby, J.-S., & Aamodt, P.O. (2017). Measuring learning outcomes. *European Journal of Education*, *52*, 20–30.

Cattani, L., Guidetti, G., & Pedrini, G. (2017). Overeducation among Italian graduates: Do different measures diverge? *Economia Politica*, 1–31. Retrieved from https://link.springer.com/article/10.1007/s40888-017-0075-1

Cattell, R.B. (1972). The 16 P.F. and basic personality structure: A reply to Eysenck. *Journal of Behavioural Science*, *1*(4), 169–187.

Center of Inquiry. (2016). Wabash national study 2006–2012. Centre of inquiry. Retrieved from http://www.liberalarts.wabash.edu/study-research/

Coates, H. (2016). *The market for learning: Leading transparent higher education*. Melbourne, Australia: Springer.

Coates, H., & McCormick, A. (eds.). (2014). *Engaging university students: International insights from system-wide studies*. London: Springer.

Collini, S. (2012). *What are universities for?* London: Penguin.

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104.

Cronbach, L.J. (1957). The two disciplines of scientific psychology [1]. *American Psychologist*, *12*, 671–684.

Darwin, S. (2017). What contemporary work are student ratings actually doing in higher education. *Studies in Educational Evaluation*, *54*, 13–21.

Dawson, P., & Dawson, S.L. (2016). Sharing successes and hiding failures: 'Reporting bias' in learning and teaching research. *Studies in Higher Education*, 1–12. Retrieved from https://0-doi-org.lib.exeter.ac.uk/10.1080/03075079.2016.1258052

DfE. (2017). *Teaching excellence and student outcomes framework specification*. London, UK: Department for Education. Retrieved from. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/658490/Teaching_Excellence_and_Student_Outcomes_Framework_Specification.pdf

Diamond, A., Vorley, T., Roberts, J., & Jones, S. (2012). *Behavioural approaches to understanding student choice*. York: Higher Education Academy. Retrieved from https://www.heacademy.ac.uk/system/files/resources/student_choice.pdf

DiFrancesca, D., Nietfeld, J.L., & Cao, L. (2016). A comparsion of high and low achieving students on self-regulated learning variables. *Learning and Individual Differences*, *45*, 228–236.

Dinsmore, D.L. (2017). Toward a dynamic, multidimensional research framework for strategic processing. *Educational Psychology Review*, *29*, 235–268.

Douglass, J.A., Thomson, G., & Zhao, C.-M. (2012). The learning outcomes race: The value of self-reported gains in large research universities. *Higher Education*, *64*, 317–335.

Drezek McConnell, K., & Rhodes, T.L. (2017). *On solid ground, value report 2017*. Washington, DC: Association of American Colleges and Universities.

Dubinksy, J.M., Roehrig, G., & Varma, S. (2013). Infusing neuroscience into teacher professional development. *Educational Researcher*, *43*(6), 317–329.

Duckworth, A.L., & Yeager, D.S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251.

Dudas, B. (2015). What lies ahead for students and learning gain in the UK? National union of students. Presentation at learning gain in higher education – HEFCE/OfS Second National Conference 10 November, 2016 London, UK.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, *58*. doi:10.1037/h0093599

Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J., & Willingham, D.T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–5.

Dunne, E., Zandstra, R., Brown, T., & Nurser, T. (2011). *Students as change agents: New ways of engaging with learning and teaching in higher education*. ESCalate (Education Subject Centre of the Higher Education Academy; University of Exeter, corp creators: Retrieved from http://dera.ioe.ac.uk/id/eprint/14767.

Edge Foundation. (2016). *The digital revolution: The impact of the fourth industrial revolution on employment and education*. London, England. Retrieved from https://www.voced.edu.au/content/ngv%3A73764.

Edwards, M.A., & Roy, S. (2017). Academic research in the 21st-century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science*, *34*(1), 51–61.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, *83*(1), 70–120.

Evans, C. (2016, 2018). *Enhancing assessment feedback practice in higher education; The EAT Framework*. Southampton: University of Southampton.

Evans, C., Muijs, D., & Tomlinson, D. (2015). *Engaged student learning: High impact strategies to enhance student achievement*. York: Higher Education Academy.

Evans, C., Waring, M., & Christoudolou, A. (2017). Building teachers' research literacy: Integrating practice and research. *Research Papers in Education*, *32*(4), 303–423.

Evans, C., & Zhu, X. (2018). *Considerations in managing complex interventions in Evans et al., 2018 Supporting student agency and success in higher education and beyond through the development of assessment feedback skills (the ability to self-monitor and self-evaluate. Experimental innovation final report for higher education funding council for england/office for students (HEfCE/OfS)*. Southampton: University of Southampton.

Field, F. (2010). *The Foundation Years: Preventing poor children becoming adults*. London: HM Government UK. Retrieved from http://webarchive.nationalarchives.gov.uk/20110120090141/http://povertyreview.independent.gov.uk/media/20254/poverty-report.pdf

Finlay, A., & Brown McNair, T. (2013). *Assessing underserved students' engagement in high-impact practices*. Washington, DC: Association of American Colleges and Universities. Retrieved from https://www.aacu.org/assessinghips/report

Fisherman, T., Ludgate, A., & Tutak, J. (2017). Success by design, improving outcomes in American higher education. A Deloitte center for higher education excellence series on student success. Deloitte University Press. Retrieved from https://www2.deloitte.com/insights/us/en/focus/reimagining-higher-education.html.

Forsythe, A., & Jellicoe, M. (2018). Predicting gainful learning in higher education; a goal orientation approach. *Higher Education Pedagogies* (*this issue*).

Forsythe, A., & Johnson, M. (2017). Thanks, but no thanks for the feedback. *Assessment and Evaluation in Higher Education*, *42*(6), 850–859.

Friedlander, M.J., Linda Andrews, L., Armstrong, E.G., Aschenbrenner, C., Kass, J.S., Ogden, P., … Viggiano, T.R. (2011). What can medical education learn from the neurobiology of learning? *Academic Medicine*, *86*, 415–420.

Furr, R.M. (2013). *Psychometrics: An introduction* (3rd ed.). Thousand Oaks, California: Sage.

Gibson, S. (2015). When rights are not enough: What is? Moving towards new pedagogy for inclusive education within UK universities. *International Journal of Inclusive Education*, *19*(8), 875–886.

Haddelsey, S., Speight, L., & Brumhead, A. (2017). HEFCE/OfS, 2017 – Lincoln/Huddersfield project. Retrieved from HEFCE http://www.hefce.ac.uk/lt/lg/projects/uni-lincoln/

Haldane, A. (2015). Speech by Andrew G Haldane, chief economist, Bank of England, 12 November 2015. Retrieved from http://www.bankofengland.co.uk/publications/Documents/speeches/2015/speech864.pdf

Halpern, D. (2014). *Thought and knowledge, an introduction to critical thinking* (5th ed.). New York: Psychology Press.

Havergal, C. (2017). Cambridge looks to crack measurement of 'learning gain'. Times Higher Educational Supplement, December 7, 2017. Retrieved from https://www.timeshighereducation.com/news/cambridge-looks-crack-measurement-learning-gain

Healey, M., Flint, A., & Harrington, K. (2014). *Engagement through partnership: Students as partners in learning and teaching in higher education*. York: Higher Education Academy. Retrieved from https://www.heacademy.ac.uk/system/files/resources/engagement_through_partnership.pdf

Heckman, J.J., & Masterov, D.V. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics, American Agricultural Economics Association*, *29* (3), 446–493. Retrieved from http://www.nber.org/papers/w13016

Heene, M. (2010). A brief history of the fail-safe number in applied research. arXiv:1010.2326 [stat.OT]. Retrieve from https://arxiv.org/ftp/arxiv/papers/1010/1010.2326.pdf

HEFCE/OfS. (2014). *Listed in hefce/office for students (ofs) grant letters for 2014-15*. London/Bristol: Higher Education Funding Council for England

HEFCE/OfS. (2015). *Listed in hefce/office for students (ofs) grant letters for 2015-2016*. London/Bristol: Higher Education Funding Council for England

HEFCE. (2017a). Learning gain. Retrieved from http://www.hefce.ac.uk/lt/lg/projects/

HEFCE. (2017b). Addressing barriers to student success. Retrieved from http://www.hefce.ac.uk/sas/barriers/projects/southampton/#d.en.114689

Higher Education Academy. (2015). *Higher education achievement report*. York: HEA. Retrieved from. http://www.hear.ac.uk/about/history-development

Higher Education Statistics Agency HESA. (2018a). Non-continuation: UK performance indicators. Retrieved from https://www.hesa.ac.uk/news/08-03-2018/non-continuation-tables

Higher Education Statistics Agency HESA. (2018b). Higher education student statistics: UK, 2016/17 – Qualifications achieved. Retrieved from https://www.hesa.ac.uk/news/11-01-2018/sfr247-higher-education-student-statistics/qualifications

Howard, G.S. (1980). Response-shift bias; A problem in evaluating interventions with pre/post self-reports. *Education Review*, *4*, 93–106.

Ifenthaler, D. (2017). Are higher education institutions prepared for learning analytics? *Tech Trends*, *61*, 366–371.

Ilie, S., Vermunt, J., & Vignoles, A. (2018). Conceptualising and measuring learning gain in higher education. *Higher Education Pedagogies*, *this issue*.

iPAL. (2018). The iPal Project. International collaborative for performance assessment of learning in higher education – Research and development. *Ipal R & D*. Retrieved from http://www.ipal-rd.com/

Johnson, J. (2017). *Higher education funding for 2017–18*. London: Department for Education and Department for Business, Energy & Industrial Strategy. Retrieved from http://www.hefce.ac.uk/media/HEFCE,2014/Content/News/2017/Grant-letter-2017.pdf

Kandiko Howson, C.B. (2012). *Student expectations and perceptions of higher education*. London: King's College London.

Kandiko Howson, C.B. (2017). Evaluation of HEFCE's learning gain pilot projects. Year 1 report. HEFCE. Retrieved from http://www.hefce.ac.uk/pubs/rereports/Year/2015/learninggain/

Kandiko Howson, C.B. (2018). *Evaluation of HEFCE's learning gain pilot projects*. London, UK: The Office for Students. Year 2 report. Retrieved from https://www.officeforstudents.org.uk/media/1386/evaluation-of-hefce-s-learning-gain-pilot-projects-year-2.pdf

Keep, E., & SJames, S. (2010). Recruitment and selection – The great neglected topic. *SKOPE research paper 88*. Cardiff: Cardiff University, SKOPE.

Kim, H., & Lalancette, D. (2013). Literature review on the value-added measurement in higher education. OECD. Retrieved from http://www.oecd.org/edu/skills-beyond-school/Litterature%20Review%20VAM.pdf.

Köhler, T., Landis, R. S., & Cortina, J. M. (2017). From the Editors: Establishing methodological rigor in quantitative management learning and education research: The role of design, statistical methods, and reporting standards. *Academy of Management Learning & Education*, *16*(2), 173–192. Retrieved from https://journals.aom.org/doi/10.5465/amle.2017.0079

Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences*, *25*, 67–72.

Kozhevnikov, M., Evans, C., & Kosslyn, S. (2014). Cognitive style as environmentally sensitive individual differences in cognition: A modern synthesis and applications in education, business and management. *Psychological Science in the Public Interest*, *15*(1), 3–33.

Kuh, G. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Washington, DC: Association of American Colleges and Universities (AAC & U).

Kuh, G., O' Donnell, K., & Schneider, C.G. (2017). HIPs at ten. *Change: The Magazine of Higher Learning*, *49*(5), 8–16.

Levitt, H.M., Bamberg, M., Creswell, J.W., Frost, D.M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA publications and communications board task force report. *American Psychologist*, *73*(1), 26–46.

Marshall, S. (2017). *Rising to the challenges of tomorrow*. York: Higher Education Academy.

McGrath, C., Guerin, B., Harte, E., Frearson, M., & Manville, C. (2015). *Learning gain in higher education*. Santa Monica, California and Cambridge UK: RAND. Retrieved from https://www.rand.org/content/dam/rand/pubs/research_reports/RR900/RR996/RAND_RR996.pdf

MDV Consulting. (2017). Is your graduate selection differentiating the best from the rest? Retrieved from http://mdvconsulting.co/wp-content/uploads/Is-your-graduate-selection-differentiating-the-best-from-the-rest2.pdf

Meguizo, T., & Wainer, J. (2016). Toward a set of measures of student learning outcomes in higher education: Evidence from Brazil. *Higher Education*, *72*(3), 381–401. Retrieved from https://link.springer.com/article/10.1007/s10734-015-9963-x

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., … Baird, J. (2014). *Process evaluation of complex interventions: Medical research council guidance*. London, UK: MRC Population Health Science Research Network.

Moore, G. M., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015). *Process evaluation of complex interventions*: *Medical research council guidance*. BMJ 2015, 3350:h1258

Morgan, J. (2015). OECD's Ahelo will fail to launch, says education director times higher education supplement Sep 21, 2015. Retrieved from https://www.timeshighereducation.com/news/oecds-ahelo-will-fail-launch-says-education-director

Mountford Zimdars, A., Sabri, D., Moore, J., Sanders, J., Jones, S., & Higham, L. (2015). Causes of differences in student outcomes. London, UK: HEFCE. Report to HEFCE by King's College London, ARC Network and the University of Manchester. Retrieved from http://dera.ioe.ac.uk/23653/1/HEFCE2015_diffout.pdf

MRC. (2008). Developing and evaluating complex interventions: New guidance. *Medical Research Council*. Retrieved from www.mre.ac.uk/complexinteractionsguidance

NESTA. (2013). Understanding how innovation works. NESTA. Video retrieved from http://www.nesta.org.uk/resources/understand-how-innovation-works

NESTA. (2016). Using research evidence. A practice guide. NESTA, ESRC. Retrieved from http://www.nesta.org.uk/publications/using-research-evidence-practice-guide

Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, R., & Whitelock, D. (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior*, *76*, 703–714.

Nuffield Foundation. (2017). Improving social well-being through education, research and innovation. Retrieved from http://www.nuffieldfoundation.org/news/young-people%E2%80%99s-subject-choices-influences-and-impact

OECD. (2013a). Assessment of higher education learning outcomes. Feasibility study report. Volume 2: Data analysis and national experiences. OECD. Retrieved from http://www.oecd.org/education/skills-beyond-school/AHELO%20FS%20Report%20-%20Volume%202%20Executive%20Summary.pdf

OECD. (2013b). Assessment of higher education learning outcomes AHELO. Feasibility study report. Volume 3: Further insights. Retrieved from http://www.oecd.org/education/skills-beyond-school/AHELOFSReportVolume3.pdf

OfS. (2017). *REF 2021. Initial decisions on the Research Excellence Framework 2021*. London, UK: Author. Retrieved from. http://www.ref.ac.uk/publications/2017/initialdecisionsontheresearchexcellenceframework2021.html

Office for Students. (2018a). Learning gain. Retrieved from https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/

Office for Students. (2018b). Addressing barriers to student success. Retrieved from https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/addressing-barriers-to-student-success/what-are-the-projects

Owens, J., & Tibley, M. (2014). *Enhancing employability through enterprise education: Examples of good practice in higher education*. York: Higher Education Academy.

Parker, J. (2018). *Undergraduate research, learning gain and equity: The impact of final year research projects. Higher Education Pedagogies (this issue)* .

Pampaka, M., Swain, D., Jones, S., Williams, J., Edwards, M., & Wo, L. (2018). Validating constructs of learners' academic self-efficacy for measuring learning gain. *Higher Education Pedagogies this issue*.

Pascarella, E.T., & Blaich, C. (2013). Lessons from the Wabash national study of liberal arts education. *Change: the Magazine of Higher Learning*, *45*(2), 6–15.

Peeters, J., De Baker, F., Kindenkens, A., Triquet, K., & Lombaerts, K. (2016). Teacher differences in promoting students' self-regulated learning: Exploring the role of student characteristics. *Learning and Individual Differences*, *52*, 88–96.

Peseta, T., Barrie, S., & McLean, J. (2017). Academic life in the measured university: Pleasures, paradoxes and politics. *Higher Education Research and Development*, *36*(3), 453–457.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Powell, S., Gossman, P., & Neame, C. (2018). *Pain, gain, mission. Higher Education Pedagogies, this issue*.

Retrieved from https://www.nesta.org.uk/sites/default/files/standards_of_evidence.pdf

REF. (2018/02). *Consultation on the draft panel criteria and working methods*. Bristol, UK: the Higher Education Funding Council for Wales, the Scottish Funding Council, the Department for the Economy, NI, and Research England. Retrieved from. http://www.ref.ac.uk/media/ref,2021/downloads/Consultation%20on%20the%20draft%20panel%20criteria%20and%20working%20methods%20REF%202018_02.pdf

Rhodes, T. (2009). *Assessing outcomes and improving achievement: Tips and tools for using the rubrics*. Washington, DC: Association of American Colleges and Universities.

Richardson, J.T.E. (2015). The under-attainment of ethnic minority students in UK higher education: What we know and what we don't know. *Journal of Further and Higher Education*, *39*(2), 278–291.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387.

Rogaten, J., & Rienties, B. (2018). A case-study of multilevel growth modelling to understand learning gains: which first year students are making most learning gains in STEM subjects? Higher Education Pedagogies (*this issue*).

Rogaten, J., & Rienties, B. (2018). A case-study of multilevel growth modelling to understand learning gains: Which first year students are making most learning gains in STEM subjects? *Higher Education Pedagogies*, *This issue*.

Rosenfeld, M., & Rosenfeld, S. (2011). Illustrating a complementary paradigm for styles research: From a third-person to a second-person perspective. In S. Rayner & E. Cools (Eds.), *Style differences in cognition, learning, and management: Theory, research and practice* (pp. 143–159). New York, NY: Routledge.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*(3), 838–1641.

Ruge, G., & McCormack, C. (2017). Building and construction students' skills development for employability-reframing assessment for learning in discipline-specific contexts. *Architectural Engineering and Design Management*, *13*(5), 365–383.

Rust, C. (2017). Re-thinking assessment – A programme leader's guide. Retrieved from http://ocsld.brookesblogs.net/2017/12/22/re-thinking-assessment-a-programme-leaders-guide

Sands, D., Parker, M., Hedgeland, H., Jordan, S.E., & Galloway, R. (2018). Using concept inventories to measure understanding. *Higher Education Pedagogies*, *3*(1), 60–69.

Scalise, K., Douskey, M.C., & Stacy, A.M. (2018). Measuring learning gains and examining implications for student success in STEM. *Higher Education Pedagogies*, *this issue*.

Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*, 45–65.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education. A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600.

Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education. JISC. Retrieved fromhttps://www.jisc.ac.uk/sites/default/files/learning-analytics-in-he-v2_0.pdf.

Seifert, T.A., Gillig, B., Hanson, J.M., Pascarella, E.T., & Blaich, C.F. (2014). The conditional nature of high impact/good practices on student learning outcomes. *The Journal of Higher Education*, *85*(4), 531–564.

Shacklock, X. (2016). *From bricks to clicks – the potential of data and analytics in higher education. London*. Higher Education Commission. Retrieved from http://www.policyconnect.org.uk/hec/sites/site_hec/files/report/419/fieldreportdownload/frombrickstoclicks-hecreportforweb.pdf

Shavelson, R.J., Domingue, B.W., Mariño, J.P., Mantilla, A.M., Forero, A.M., & Wiley, E.E. (2016). On the practices and challenges of measuring higher education value added: The case of Colombia. *Assessment & Evaluation in Higher Education*, *41*(5), 695–720. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/02602938.2016.1168772?journalCode=caeh20

Shaw, B., Baars, S., Menzies, L., Parameshwaran, M., & Allen, R. (2017). *Low income pupils' progress at secondary school. LKMco and education datalab*. London, UK: Social Mobility and Child Poverty Commission. Retrieved from www.gov.uk/government/organisations/social-mobility-commission

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120.

Speight, L., Crawford, K., & Hadesley, S. (2018). *Towards measures of longitudinal learning gain in UK higher education: The challenge of meaningful engagement (this issue)*.

Stanovich, K.E., West, R.F., & Toplak, M.E. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioural and Brain Sciences.*, *23*, 645–726.

Stoakes, G., & Neves, J. (2018). UKES, Learning gain and how students spend their time. *Higher education pedagogies*, *this issue*.

Strang, L., Bélanger, J., Manville, C., & Meads, C. (2016). *Review of the research literature on defining and demonstrating quality teaching and impact in higher education*. York: Higher Education Academy. Higher Education Academy/RAND Europe. Retrieved from www.heacademy.ac.uk/knowledge-hub/review-research-literature-defining-and-demonstrating-quality-teaching-and-impact

Thurstone, L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, *16*, 433–451.

Tremblay, K., Lalancette, D., & Roseveare, D. (2012). Assessment of higher education learning outcomes (AHELO) feasibility study report, Volume 1. OECD. Retrieved from www.oecd.org/edu/ahelo.

Turner, R., Sutton, C., Muneer, R., Gray, C., Schaefer, N., & Swain, J. (2018). Exploring the potential of using undergraduates' knowledge, skills and experience in research methods as a proxy for capturing learning gain. *Higher Education Pedagogies*, *3*(1), 110–136

U.S. Department of Education. (2006). *The Spellings commission report. September. (2006)*. Washington DC,Washington: US Department of Education.

UNESCO. (2005). *The UN decade of education for sustainable development 2005–2014*. France: Author. Retrieved from. http://unesdoc.unesco.org/images/0014/001416/141629e.pdf

United Nations. (2012). Resolution adopted by the general assembly on 27 July 2012. 66/288. The future we want. UN General Assembly. Retrieved from http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/66/288&Lang=E

Universities UK. (2007). *Beyond the honours degree classification. The burgess group final report*. Retrieved from http://www.universitiesuk.ac.uk/

Van der Zanden, P.J.A.C., Denessen, E., Cillessen, A.H.N., & Meijer, P.C.M. (2018). Domains and predictors of first-year student success: A systematic review. *Educational Research Review*, *23*, 57–77.

Varsavsky, C., Matthews, K.E., & Hodgson, Y. (2014). Perceptions of science graduating students on their learning gain. *International Journal of Science Education*, *36*(6), 929–951.

Vermunt, J., & Donche, V. (2017). A learning patterns perspective on student learning in higher education: state of the art and moving forward. *Educational Psychology Review*, *29*, 269–299. doi:10.1007/s10648-017-9414-6

VitalSource learning matters (THE). (2018). Weighing up the benefits of learning. *Times Higher Education*. 34–35.

Voluntary System of Accountability. (n.d.). The new VSA. c/o APLU, 1307 New York Ave NW, Ste 400 Washington, DC 20005. Retrieved from http://www.voluntarysystem.org/.

Vossensteyn, H., Kottmann, A., Jongbloed, B., Kaiser, F., Cremonini, L., Stensaker, B., … Wollscheid, S. (2015). Dropout and completion in higher education in Europe. Main report. European Commission. Center for Higher Education Policy Studies and Nordic Institute for Studies in Innovation Research and Education. Retrieved from http://supporthere.org/sites/default/files/dropout-completion-he_en.pdf

Waring, M., & Evans, C. (2015). *Understanding pedagogy: Developing a critical approach to teaching and learning*. Abingdon, Oxford, United Kingdom: Routledge.

WHO. (2008). Scaling up health services: Challenges and choices. *Technical Brief*, *3*(3) June 12. World Health Organisation Retrieved from http://www.who.int/healthsystems/topics/delivery/technical_brief_scale-up_june12.pdf.

WHO. (2010). Nine steps for developing a scaling-up strategy. World Health Organisation. Retrieved from http://www.who.int/reproductivehealth/publications/strategic_approach/9789241500319/en/

WHO. (2011). Beginning with the end in mind. Planting pilot projects and other programmatic research for successful scaling up. World Health Organisation. Retrieved from http://www.expandnet.net/PDFs/ExpandNet-WHO%20-%20Beginning%20with%20the%20end%20in%20mind%20-%202011.pdf

Willetts, D. (2017). *A university education*. Oxford: Oxford University Press.

Wolf, E.J., Harrington, K.M., & Clark, S.L. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *76*(6), 913–934.

Ylonen, A., Arico, F., Gillespie, H., Lancaster, S., & Ward, N. (2018). Lessons in learning gain: Insights from a pilot project. *Higher Education Pedagogies*, *this issue*.