Evaluation of HEFCE's Learning Gain Pilot Projects Year 2

April 2018

Report to HEFCE by King's College London

Author: Dr Camille B. Kandiko Howson

Learning Gain Pilot Projects Evaluator

Senior Lecturer in Higher Education

King's College London

Table of Contents

List of Abbreviations

A&DC	Assessment and Development Consultants
AAC&U	Association of American Colleges and Universities
AASCU	American Association of State Colleges and Universities
AHELO	Assessment of Learning Outcomes in Higher Education
APLU	Association of Public and Land-grant Universities
ASSIST	Approaches and Study Skills Inventory for Students
BME	Black and minority ethnic
CAE	Council for Aid to Education
CALOHEE	Measuring and Comparing Achievements of Learning Outcomes in Higher Education in Europe
CLA+	Collegiate Learning Assessment
DHLE	Destinations of Leavers from Higher Education
ESE	Employability Self-Evaluation Test
GPA	Grade point average
HEFCE	Higher Education Funding Council for England
HELGA	Higher Education Learning Gain Analysis
HESA	Higher Education Statistics Agency
LEGACY	Learning and Employability Gain Assessment Community Project
NfC	Need for Cognition
NSS	National Student Survey
OECD	Organisation for Economic Co-operation and Development
SJT	Situational Judgement Test
SSA	Student self-assessment
STEM	Science, technology, engineering and mathematics
TEF	Teaching Excellence and Student Outcomes Framework
UKES	UK Engagement Survey
VALUE	Valid Assessment of Learning in Undergraduate Education
VSA	Voluntary System of Accountability

List of Tables

Notes

For this report, the projects are referred to by name of the lead institution but are indicative of the partnership group, but do not indicate institutional agreement or responsibility.

1   Executive Summary

Introduction

1.1   This is the second annual report of the evaluation of the Higher Education Funding Council for England (HEFCE)'s learning gain pilot projects programme. In 2015 HEFCE awarded over £4 million to 13 pilot projects involving over 70 higher education institutions, with the aim of testing and evaluating measures of learning gain in England for one to three years until 2018.

1.2   The pilot projects use a range of methods to explore questions about learning gain, including:

- what different approaches could be used to measure learning gain
- how robust and useful the data and other evidence arising from these approaches are, for example for supporting students and improving learning and teaching;
- and which methods and approaches have the potential to be scalable for use across the sector.

1.3   The multifaceted nature and purpose of higher education leads to a breadth of conceptions of learning gain. Combining useful approaches from across the projects leads towards understanding learning gain as **a change in knowledge, skills, work-readiness and personal development, as well as enhancement of specific practices and outcomes in defined disciplinary and institutional contexts.**

Policy implications

1.4   The development of measures of learning gain involves theoretical questions of what to measure, and turning these into practical measures that can be empirically developed and tested. This is in a broader political context of asking '**why' measure learning gain and, 'for what purpose'** with differing views from stakeholders. And as with all social science policy research, there are inevitable trade-offs between what is conceptually possible, methodologically rigorous and practically feasible. Policy drivers include: competition and value for money; quality assurance and quality enhancement; and social mobility.

1.5   There are questions of whether, and how, learning gain measures could be operationalised within the **Teaching Excellence and Student Outcomes Framework (TEF)**. From the projects, there is no simple, 'silver bullet' metric that accurately and effectively measures student learning gain comparatively across all subjects of study and institutional types. However, the pilot projects are developing tools and approaches that have the potential to offer valid and robust accounts of learning gain, at least within specific institutional, subject and pedagogical circumstances, that are contextualised for use at appropriate levels. Current challenges around scalability, engagement and comparability would need to be overcome for learning gain measures to be used in TEF or related activities.

1.6   Allowing for multiple approaches to measuring learning gain supports institutional diversification through evidence-based competition. A variety of data allows prospective students to make informed choices about what and where to study and

provides current students additional information to support their learning and develop their employability. It also offers employers greater opportunities to select diverse graduates and to target specific courses and institutions relevant to their needs.

Approaches to measuring learning gain

1.7   Three broad approaches to measuring learning gain have emerged, including:

- measures of general **cognitive gain** –what students think and know;
- measures of **soft skills development** –affective measures of attitudes and how students feel;
- **employability and career readiness** –largely behavioural measures of activities students have undertaken in preparation for the world of work. Evaluation work analyses the robustness and effectiveness of the projects through their validity, whether measures are accurately measuring what they are supposed to, and reliability, the consistency of measures. A few tools and approaches have already been amended or dropped.

1.8   **Cognitive gain** is the primary focus of the University of Cambridge strand of the University of Warwick's Learning and Employability Gain Assessment CommunitY (LEGACY) project through the development of a non-disciplinary survey tool, along with meta-cognitive, affective and socio-communicative components. Birmingham City University is piloting the US-based Collegiate Learning Assessment Plus (CLA+), a broad-based cognitive skills standardised test. Some projects explore to what extent grades are an effective measure of learning gain, including the Open University, the University of East Anglia, and the University of Reading which is also piloting the CLA+ test.

1.9   Measurements of **soft skills development** explore students' academic dispositions, resilience, graduate capital and identity, self-theories, motivation and confidence. Surveys are being developed across projects at the University of Manchester, the University of Portsmouth and the University of East London that involve testing scales and items from existing instruments as well as piloting new measures. Behaviour-based soft-skills, including student engagement, are being measured across half of the projects through the UK Engagement Survey (UKES). Additional contextualised behavioural measures include a survey on attitudes about and confidence in research methods at Plymouth University; a skills survey at the Manchester College tailored for Higher Education in Further Education students; self-efficacy at the University of East Anglia; and well-being surveys at the University of Lincoln and the University of Reading.

1.10   Four projects explore behavioural measures of activities focusing **employability and career readiness** through surveys, self-assessments and students' participation in work experience activities. The Careers Group is using career-readiness questions embedded into enrolment systems; Ravensbourne analysed the impact of work placement activities; three strands of work through the LEGACY project explore students' strengths, career adaptability and mobility; and the University of Lincoln developed a student self-assessment (SSA) of employment competencies.

Uses and scalability of learning gain measures

1.11   After two years, the strands of activity are now well placed to be categorised by the different uses and audiences for their approaches and resulting data, with examples or proposed uses. However, data used for classroom enhancements may not be scalable for comparative use across the sector and *vice versa*.

Level of use	Purpose of use
Prospective students	Informed choice on value for money
Current students	Provide data for reflection, awareness raising
Classroom enhancement	Pedagogical enhancement, data for teaching staff to tailor information to students
Course management	Management, pedagogical enhancement
Institutional (service delivery)	Enhance and tailor student services (e.g. careers services)
Institutional (strategic)	Programme review, inform strategy, enhancement, evaluate programmes, staff reward and recognition
Cross-institutional	Benchmarking, comparisons
Employers	Recruiting graduates, diversifying workforce
Government	Accountability, regulation, quality assessment, market indicators

Table 1. Summary of levels of use of learning gain metrics

1.12   Scalability explores the appropriateness, practicality and value for money of extending the use of measures of learning gain. Initial evaluation work indicates two categories; some approaches have potential for scalability at subject level across institutions, while others are not necessarily scalable as they need local embedding but there is potential use for benchmarking within institutions. Analysing the suitability and scalability of different measures and approaches will be the focus of evaluation work over the final year.

1.13   The projects have highlighted a number of challenges to measuring learning gain including: lack of student engagement; data protection, data sharing and research ethics; and financial costs of staff time and opportunity costs for students and institutions. Challenges to evaluating learning gain include: lack of clear purpose; breadth of the definition of learning gain; methodological complexity; multiplicity of uses and stakeholders; and institutional and political application.

1.14   Ten of the 13 funded projects continue until October 2018. A full evaluation of the learning gain pilots' programme of work will be reported in January 2019.

2   Introduction

2.1   This is the second annual report of the evaluation of the Higher Education Funding Council for England (HEFCE)'s learning gain pilot projects programme[1]. Following a call for expressions of interest issued in March 2015, HEFCE awarded over £4 million to 13 pilot projects involving over 70 higher education institutions, with the aim of testing and evaluating measures of learning gain in England. HEFCE funding is for one to three years until 2018, although some projects will be extending their work using internal funding.

2.2   In addition to the pilot projects, additional complementary activities support the learning gain programme. These include:

- the National Mixed Methodology Learning Gain Project, a HEFCE-administered multi-institutional longitudinal study combining a critical thinking and problem-solving test with self-reflective questions exploring academic motivation, attitudes to literacy and diversity, and dimensions of student engagement; and
- the Higher Education Learning Gain Analysis (HELGA) project, an assessment of the potential application of national datasets to learning gain issues, capacity building and networking events.

2.3   Information on learning gain was gathered in an independent scoping study carried out by RAND Europe[2]. Drawing on the RAND report, 'learning gain' is broadly considered by HEFCE to relate to **the improvement in knowledge, skills, work-readiness and personal development made by students during their time spent in higher education.**

2.4   The pilot projects use a range of methods to explore questions about learning gain, including:

- what different approaches could be used to measure learning gain;
- how robust and useful the data and other evidence arising from these approaches are, for example for supporting students and improving learning and teaching; and
- which methods and approaches have the potential to be scalable for use across the sector.

Evaluation approach

2.5   The aims of the evaluation are to:

i.   evaluate the success of the learning gain projects against the aims of the scheme;

ii.   evaluate the progress, outputs and outcomes of each pilot project funded against their individual aims and success criteria;

---

[1] Evaluation of HEFCE's learning gain pilot projects: Year 1 report
www.hefce.ac.uk/pubs/rereports/year/2017/lgeval/
[2] Rand Europe report on Learning Gain:
http://www.hefce.ac.uk/pubs/rereports/Year/2015/learninggain/

iii. analyse the success, feasibility and challenges of the different methods and approaches for learning gain in England based on evidence gathered from the learning gain projects;

iv. oversee the progress of the pilot projects to identify emerging themes and particular issues as they arise;

v. identify knowledge gaps across the pilot project portfolio for which further investigation is required;

vi. disseminate findings from the evaluation work among the learning gain projects and wider external audience; and

vii. use the outcomes of the evaluation to make recommendations to inform HEFCE's advice to Government on future learning gain policy.

2.6 Given the disparate nature of the projects, the evaluation operates at two primary levels. The first is against each project's unique success criteria and the second is against an overall evaluation framework. These two approaches operative iteratively. The evaluation framework has four key areas of focus: development of measures of learning gain; robustness and effectiveness; suitability; and scalability. Information on the individual projects can be found on HEFCE's website[3].

---

[3] http://www.hefce.ac.uk/lt/lg/

3   Evaluation of Learning Gain Pilot Projects Year 2

3.1   The evaluation of the projects operates at two levels: (1) across the suite of projects, and (2) evaluating each project against the evaluation framework and its own unique success criteria. At the macro level the evaluation draws out broad themes and findings across the 13 projects, covering the development of measures of learning gain; robustness and effectiveness of different measures and approaches; the suitability of measures for different purposes; and the potential scalability of measures and approaches.

Evaluation Framework

3.2   <u>Development of measures of learning gain</u>. The first stage of the evaluation framework focuses on the theoretical and practical components of measuring learning gain. Theoretically, what to measure is based on philosophical questions of what one thinks higher education is for, what the purpose of higher education actually is and the motivation for wanting to measure learning gain. The projects address these questions through focusing on different areas: cognitive gain; soft skills development; and employability and career readiness.

3.3   The next component involves how theoretical concepts are operationalised into practical measures that can be empirically developed and tested. It also involves the context in which projects are developed, targeting specific student groups, subject areas, regions or institutional types. The development of measures of learning gain was the focus of Year 1 of the evaluation of the pilot projects[4].

3.4   <u>Robustness and effectiveness</u>. The second stage of evaluation builds on the rationale of what is being measured and explores *how* it is being measured through evaluating robustness. This element has two interrelated components: reliability and validity.

3.5   Reliability explores the consistency and accuracy of a measure, covering technical aspects of the approach. Validity has theoretical, practical and technical aspects. Theoretically, validity explores whether a measure is conceptually measuring what is intended, often through qualitative research with students and other stakeholders. Practically, validity explores whether the metric is *actually* measuring what is intended, involving a combination of qualitative and quantitative approaches. Finally, there are also technical aspects of evaluating validity, particularly for developing items and scales.

3.6   As with developing learning gain measures, context is important for notions of validity and reliability: to what extent are measures being tested with specific student groups, subject areas, regions or institutional types? Further considerations are the representativeness of sample populations and respondents. Additionally, evaluating validity depends on the purpose of a measure, for example, assignment grades may be useful for differentiating across student gains in a module but not valid for comparison across institutions. Robustness and effectiveness have been major

---

[4] www.hefce.ac.uk/pubs/rereports/year/2017/lgeval/

areas of focus for the Year 2 evaluation of the pilot projects and will continue into Year 3.

3.7 <u>Suitability</u>. The third stage of evaluation, suitability, explores contextualised validity through feasibility and usability. Feasibility explores how practical the measure was to obtain, and brings in questions of value for money related to the first stage of the evaluation on the motivation for wanting to measure learning gain. Some projects tested instruments with hundreds of items, which may result in robust data but very few students would complete it.

3.8 Usability explores how measures are practically put into practice and what they theoretically could be used for within and beyond institutions. This relates to testing validity, as it depends on what level and for what purpose measures of learning gain are being used as to whether they are valid. As these are institution-based projects, the context and embeddedness of the measures impact on the ability to evaluate the suitability of the metrics. Suitability and usability have been major areas of focus for the Year 2 evaluation of the pilot projects and will continue into Year 3.

3.9 <u>Scalability</u>. The fourth stage, evaluating scalability, is a synthesis of the main outcomes of the previous stages. Drawing on the first stage (the motivation for measuring learning gain), scalability explores what could be done with the metrics, and at what level. Evaluation of scalability also involves the robustness and effectiveness of a measure for a given purpose factoring in practicality, motivations, and costs. On one project, researchers visited classrooms for every single course to get students to complete surveys; this would be a very costly and time-intensive activity to conduct across the higher education sector.

3.10 The importance of contextualisation is brought into questions of scalability: to what extent can the approach being piloted be scaled up across student characteristics, subjects and institutions? Finally, broader considerations of research and data ethics, data protection, data sharing and wider institutional, sector and public buy-in all factor into evaluating scalability. Scalability has been explored in the Year 2 evaluation and will be the major focus of Year 3.

3.11 Challenges to evaluating learning gain are reflected throughout the report. Firstly, there was no clear purpose for measuring learning gain established at the outset, rather this has evolved over the duration of the pilot projects. The difficulties of project administration and challenges of student engagement have impacted on the ability of some projects to collect sufficient data to be able to judge the efficacy of the approach. Regarding scalability, all the projects are conducted in specific contexts so it is not possible to fully evaluate wider scalability of any measure without further piloting. And finally, the philosophical orientation of those developing measures may vary from that of the end-user, such as Government or prospective students, and particularly for potentially high-stakes accountability purposes.

> Learning gain data is part of evidence-informed decision making and the ambition for impact and evaluation of educational processes. The purpose being explored and the audience drive which aspects of learning gain data are used and what would be most appropriate.

Summary of measures of learning gain

3.12   As the pilot projects have developed, three broad approaches to measuring learning gain have emerged. These include measures of: general cognitive gain; soft skills development; and employability and career readiness. Some projects focus on a single area and others integrate multiple approaches or operate several strands of activity.

3.13   Cognitive gain. There are two primary methods used across projects and strands of work exploring measures of cognitive gain. The first includes developing non-subject specific questionnaires or performance-based standardised assessment to explore general cognitive gain; the second focuses on subject-specific learning through an analysis of grades trajectories.

3.14   The Cambridge strand of the University of Warwick's Learning and Employability Gain Assessment CommunitY (LEGACY) project is piloting and testing an instrument to measure cognitive, meta-cognitive, affective, and socio-communicative gain, assessing its validity, transferability and scalability. This is done through students' perspectives, questionnaires and test-like instruments across selected disciplines.

3.15   Two other projects incorporate piloting an existing instrument, the Collegiate Learning Assessment (CLA+), a performance-based critical-thinking assessment based on the general education model in the US. It measures students' performance in analysis and problem solving, scientific and quantitative reasoning, critical reading and evaluation, and critiquing an argument, in addition to writing mechanics and effectiveness. This is led by Birmingham City University as part of a project consortium of institutions as well as in a project by the University of Reading. The projects are exploring the feasibility of administering the instrument as well as its applicability in the English higher education context, which differs from the US in offering subject-specific degree courses.

3.16   The second approach to measuring cognitive gain is through analysis of existing institutional data on trajectories of grades students receive through their course of study. Strands of work across the University of East Anglia and the University of Reading explore grade trajectories across modules, subjects and years of study. Analysis done by a partnership of three institutions led by the Open University also explores grade trajectory patterns across institutions.

3.17   A strand of activity through the University of East Anglia project explored embedding a disciplinary-based concept inventory into classroom teaching practices.

Cognitive gain	New data	Existing data
General cognitive gain	Questionnaires; surveys	
Subject-specific cognitive gain	Tests; concept inventories	Module marks; Grade Point Average (GPA); grade trajectories; Degree classification

Table 2. Measures of cognitive learning gain

3.18   Soft skills development. Soft skills incorporate non-cognitive skill development including affective measures, such as attitudes, interests and values which capture how students think and feel. There is an aim to understand how students feel about the experience of university through completing surveys and questionnaires at various points in time and seeing how these measures impact on their progression and attainment.

3.19   The project led by the University of Portsmouth is developing, testing, and validating psychometric measures of non-cognitive skills. Through longitudinal self-reported data, they are developing new scales on resilience, graduate capital and identity, and self-theories of self-esteem, self-efficacy and false uniqueness along with data on demographics, attainment and outcomes. Similarly, the University of Manchester is running a project exploring the distance travelled by students in terms of knowledge, work readiness, critical thinking skills and personal development, mapped onto students' trajectories and experiences of higher education.

3.20   Several projects highlighted the importance of contextualisation when considering measuring learning gain. A project consortium led by the University of East London is measuring personal development, cognition and confidence in learning situations by testing different measures, analysing their variability with time across a wide range of learners, and exploring trajectories and interrelations with inequality. The project focuses on the importance of student characteristics on patterns of attainment and progression.

3.21   Two projects explored the learning context of college-based and further education. Plymouth University is leading a project evaluating research methods learning across seven institutions and multiple disciplines, exploring learning motivation, research orientation, research self-efficacy and research environment. Focusing on the further education context, The Manchester College is leading a longitudinal project to develop metrics for personal skills development, learner engagement and academic development.

3.22   A strand of work from the project by the University of East Anglia explored the development of self-efficacy and confidence through peer learning. This approach was embedded in the pedagogy of selected courses and involved regular data collection throughout a module mapped onto attainment data.

Soft skills development	Student characteristics	Classroom context	Institutional context
General skills development	Surveys; student records		Surveys
Subject-specific skills development	Surveys	Innovative pedagogies	Surveys

Table 3. Measures of soft skills development

3.23   Employability and career readiness. Four projects explore behaviour-based measures of activities focusing on employability and career readiness through surveys, self-assessments and students' participation in work experience activities. Across a broad range of institutions, the Careers Group project assesses whether a

small number of self-reported questions asked during student enrolment can track development of student employability during their time in higher education can predict employment outcomes and evaluate the effectiveness of employability strategies and interventions.

3.24 Three strands of work from the LEGACY project focus on specific aspects of employability, including the distance travelled in development of self-awareness, self-efficacy, confidence, career readiness and ability to secure graduate employment or further study; the distance travelled of concern, control, curiosity and confidence of students with lower career readiness compared to their peers; and a qualitative project on aims, motivation and self-perceived employability of students who have studied or worked abroad compared to on-campus international experience.

3.25 Drawing on institutional contexts and missions, contextualised aspects of employability are explored. The University of Lincoln is running a project combining reflective student self-assessments (SSAs) with data on academic achievement, attendance, engagement in extra-curricular activities, and work experience. Ravensbourne completed a project last year which measured gains from work-based learning and work preparation activities learning in the creative industries and the effects of these activities on employability.

Employability and career readiness	Embedded within course	Institutional support services
Student self-assessment	Surveys	Surveys
Engagement with activities	Surveys; student records	Surveys; qualitative measures

Table 4. Measures of employability and career readiness

Key considerations: Data source

Using existing data
'Big data' and learner analytics draw on existing data from students' experiences at their institutions, including prior educational attainment and data such as participation, engagement and experiences, progression and grades. These can be useful for uncovering patterns in progress and attainment across student groups or courses of study. Challenges include separating meaningful differences from module or subject-specific idiosyncrasies and that when interesting findings are made, the data often indicates correlational relationships but does not explain why, thus requiring further qualitative analysis. Such approaches may be useful for programme and institutional improvement but may be less effective as scalable, evaluative measures of learning gain.

Collecting new data
The second approach focuses on collecting new data using existing instruments or developing new general, institution-specific or subject-specific tools. Methods include tests, surveys and qualitative measures. The challenge of these is the appropriateness of using existing instruments and the robustness of new instruments as well as gathering sufficient data to be generalisable across student characteristics, subjects and institution type. A major hurdle has been getting students to complete additional tests and surveys. By far the most successful projects have embedded

surveys, tests and questionnaires in the curriculum and standard institutional processes, such as registration.

4   Development and characterisation of measures of learning gain

4.1   Conceptual models or frameworks for measuring learning gain help contextualise the definitions, offer a rationale for approaches and provide scope for inclusion of additional measures. Models can also help clarify how the different elements being studied relate to one another, such as the relationship between affective and behavioural measures. These are grounded in philosophical conceptions of the purposes of higher education, such as creating new knowledge; benefiting the public good; envisioning learning and teaching as ends in themselves; employability and career development; and immediate employment. The purposes, values and aims that underpin the projects can be broadly grouped by the three approaches to measuring learning gain.

Cognitive gain

4.2   Measures of general cognitive gain explore the variation in non-subject-specific learning through academic years. Thus far, these approaches concentrate on the developmental value of the metrics for individual students, rather than institutional comparison.

4.3   Analysis of existing data on students' academic performance allows for measuring across various points in time and across disciplines and institutions. Such modelling approaches capture the global achievement and improvement of the university, and the relative between- and within-programme differences within the same cohort, as well as between-cohort differences. This analysis supports academic practice, as well as teaching and learning. However, it is important to balance analysis of grades with flexible policies which can ensure standardisation without suppressing naturally-occurring performance fluctuations. Otherwise, there is a risk that accountability outcomes become increasingly detached from actual student learning, and that examination results progressively lose their validity as measures of (and for) learning.

4.4   Measures of cognitive abilities include critical thinking, analytic reasoning ability, problem solving, deep processing; meta-cognitive processes, including self-regulation and time management; socio-communicative attributes including written communication skills and conceptions of academic writing; and attainment measured through grades.

4.5   Developing rigorous new cognitive tests and surveys is very time consuming and challenging to ensure high quality of psychometric properties, validity and reliability. Most projects use existing instruments, and many are creating bespoke surveys using selected scales and items from existing surveys. The Cambridge strand of the LEGACY project is testing a Learning Gain Measurement Tool with cognitive, meta-cognitive, affective and socio-communicative components at three time-points across two years. The US-based CLA+ from the College for Aid to Education (CAE) is a broad-based cognitive skills standardised test. This is being run through the projects at Birmingham City University and the University of Reading.

4.6   The University of Lincoln also purchased access to an existing instrument, the Graduate Dilemmas Situational Judgement Test (SJT) run by Assessment and

Development Consultants. However, due to lack of student motivation and engagement with the test it has been dropped from the project and replaced with a transitions survey. The University of East Anglia is using pre-developed simple concept inventories in chemistry, biology and pharmacy with normalised gain measured as the difference between two sittings.

4.7   Some projects explore to what extent grades are an effective measure of learning gain (directly with the Open University, the University of Reading, the University of East Anglia, and indirectly with others). Conclusions show that grades are subjective measures, with variations across individual markers, modules, courses, disciplines and institutions. Different trends on grading patterns are linked with local practices rather than students' performance. Furthermore, there are external pressures to increase or inflate grades, such as quality assurance processes, league table rankings, student satisfaction, and retention and completion metrics.

4.8   The University of East Anglia analysed the use of grades in their project, concluding that the nature of subjects give different marking profiles, with mathematical subjects producing a different (bimodal) distribution of marks when compared to essay-based subjects which tend to be more clustered. The nature of the assessment design varies from course to course with some students having to produce different numbers of assessments for modules of the same credit size. They found an acceptance of the subjectivity of the marking process in some subjects, especially when it came to small differences (for example 2 per cent) in marks awarded. While a generic marking scale is applied across the university, several subject areas have developed more subject-based marking rubrics. The opportunities to discuss marking and assessment approaches between schools are limited.

Soft skills development

4.9   Measurements of soft skills draw out data that is useful for individual students as well as courses and institutions. Data on students' academic dispositions, resilience, graduate capital and identity, self-theories and assessment of students' critical skills can be linked with grades and entry and exit measures to conduct longitudinal and cross-sectional value-added models of learning gain. This allows for a transferrable methodology for effectively evaluating how students learn in their time in higher education through easily administered self-reported skills-based survey instruments and institutional data on grades.

4.10   Analysis of soft skills can also be sensitive to groups of students, local contexts and specific types of higher education provision. For institutions, comparable measurements of learning gain in diverse higher education environments and disciplines cannot be a valid measurement without an element of contextualisation. Measures of learning gain that focus on generic interpretations of skills and competencies and personal development will undoubtedly provide a widely applicable and scalable option for collection and analysis. To this end, several projects attempt to amalgamate this perspective with a contextualised appreciation of institutional missions, subject areas and pedagogical approaches that underpin students' learning experiences.

4.11 Measures of soft skills, or non-cognitive development, include: learner resilience and well-being, graduate capital and identity, and self-theories such as confidence-based academic self-efficacy, self-esteem, and false uniqueness. Further measures cover confidence with maths; behaviour confidence; disposition to complete chosen course, disposition towards critical thinking attitudinal/affective measures; openness, confidence and attitude towards research ; need for cognition; motivation; competence in applying critical reasoning skills (through self-reporting questionnaires). Behavioural student engagement measures include: how students spend their time; their levels of engagement; interpersonal skills; contemporary knowledge skills; level of academic challenge; reflective and integrative learning; and module and overall satisfaction. Also covered are engagement with subject learning and effective pedagogical approaches. Most measures look at the developmental trajectories of non-cognitive skills, and various predictive factors.

4.12 Affective-based soft skills are captured through several different instruments. These include:

- an institution-based well-being survey at the University of Reading;
- questionnaires on disposition, transition, perceptions of support and pedagogic practices at the University of Manchester;
- self-efficacy through self-reported confidence in formative assessment performance, self-assessment skills through the statistical association of student confidence and student performance, and confidence gains through the proportion of correct responses and confident responses to questions before and after peer instruction at the University of East Anglia;
- existing questionnaires (Approaches and Study Skills Inventory for Students (ASSIST) and Dweck's Implicit Theories of Intelligence), as well as developing new psychometric tools to assess the development of non-cognitive skills at the University of Portsmouth;
- incorporating scales from a few existing instruments: Academic Behavioural Confidence ; Need for Cognition (NfC); Predicted Grades; Time Spent, Personal Development and Skills Development subscales of the UK Engagement Survey (UKES); and new Multiple Motivation and Critical Employability scales in Year 3 at the University of East London.

4.13 Behavioural soft-skills are captured through several instruments. The main facet being explored is student engagement, through use of the UKES survey. UKES is being used fully, partially or in an amended form by six projects: The Manchester College, Plymouth University, the University of Portsmouth, the University of Reading, the University of East London, and Birmingham City University. Additional behavioural measures include a self-report survey developed for reporting on attitudes about and confidence in research methods at Plymouth University, and a self-reported skills survey at The Manchester College.

Employability and career readiness

4.14 Employability measures can support the development of individual students, institutional strategies and support services and cross-institutional benchmarking. Combining employability measures with progression and attainment data allows for

better understanding which factors most influence the extent and pace of gains. Such analysis can uncover common characteristics of students who are most likely to achieve notable gains during their undergraduate studies, and how that improved understanding might be used to shape teaching practices and non-academic service provision.

4.15 Employability measures overlap with several soft skills measures including self-awareness, self-efficacy, resilience, motivation, concern, control, curiosity and confidence. Specific competencies explored include: global citizenship, agility, commercial awareness, influencing, leadership, and emotional intelligence. Career readiness is linked with its effectiveness as a predictor of employment outcomes for graduates; the extent to which students are engaged in employability enhancing activities and the impact that this has on their career readiness; and the effectiveness of employability strategies and interventions.

4.16 Employability is explored through several instruments, most using or amending existing instruments. The Careers Group is using two core and several optional career readiness self-reported questions developed by the University of Leeds, predominately embedded into compulsory enrolment systems. Ravensbourne ran the Solent Capital Compass model using the Employability Self Evaluation Test (ESE), providing a before and after snapshot of work-placement activities. Ravensbourne also ran a telephone survey based on the Destinations of Leavers from Higher Education (DHLE) model, following up after three years. From the LEGACY project, the University of Warwick employability strand ran the R2 Strengths Questionnaire and the University of Nottingham strand is administering the Career AdaptAbility Scale (on-line and a UK version of the existing scale). The University of Lincoln developed an student self-assessment (SSA) that includes nine questions focusing on seven core employment competencies.

4.17 Several projects combine cognitive, soft skills and employability approaches to broadly trial measures of the improvement in knowledge, skills and personal development made by students during their time spent in higher education.

Key considerations: Generic or subject-specific learning gain

Characterisations of learning gain
Measures of learning gain can be divided into three psycho-social constructs: *cognitive* measures capturing students' intellectual gains including critical thinking, problem solving skills and subject knowledge; soft skills, or *affective* measures of how students feel, their attitudes and approaches to their learning; and employability and career readiness, or *behavioural* measures covering what students do and how they engage with their learning.

Generic or subject-specific gain
Some measures are generic, for example problem solving tests. However, most measures have more variation across subjects than across institutions. Other measures are subject specific, for example exploring critical thinking in engineering differently than in education. There are also many different definitions and meanings of employability across students, subjects, institutions and government.

Entry and outcome measures

4.18   Most projects link new survey or test data with existing student records data, providing information on student background characteristics and entry qualifications. Based on good research practice, it is ideal to be able to engage students before or at their arrival point to ensure that measurements capture the full impact of the higher education experience. Due to the timing of starting the projects in 2015, little new baseline data was collected on students to be able to follow their progress over three years. However, nine of the 13 projects collected new data on first year students in the first year of the projects.

4.19   Entry level was not specifically accounted for in a few projects, but may be added in final project models. Several of the employability projects use baseline measures when students first enrol or 'before' questionnaires prior to interventions. Half of the projects use tariff points, however, several projects noted specific challenges of using standardised base line data because, due to the nature of the institutions, the majority of the students enter with either non-formal qualifications or no qualifications at all. Some projects found that various measures that could be used were not uniformly collected across partner institutions.

4.20   Few projects specified outcome measures. Most outcome measures involve various aspects of what is being studied (affective, behavioural and cognitive gains), often put together under the umbrella of 'employability'. Most projects operationalise employability as a combination of various affective, behavioural and cognitive learning gain measures. Several projects use existing exit data (grades, degree outcomes, DHLE employment data) for outcomes measures. Following up on specific entry and outcomes measures will be focus of the Year 3 evaluation work.

Key considerations: Outcome measures

Outcome measures encompass *affective*, *behavioural* and *cognitive* measures, which have different meanings across different subjects and institutional types. Using contextualised multiple metrics allows for using different measures of learning gain for different purposes.

---

5   Robustness and effectiveness

5.1   Validity and reliability of instruments and measures are essential to ensure the effectiveness and appropriateness of measures of learning gain. Each project explores the technical aspects of validity and reliability in relation to their study design and instruments used. Analysis of findings is part of judging the validity of measures of learning gain. Additionally, contextualisation, taking into consideration the purpose and level of use, is a fundamental part of establishing validity and reliability.

5.2   Most projects are using existing instruments, which have been previously tested and trialled. Further verification of instruments is among the aims of the projects (e.g. testing the usefulness of the CLA+ in an English educational context). However, further research will need to be conducted by projects amending survey questions or combining items from different instruments.

5.3   The challenge of getting students to complete tests and surveys, and to have sufficient motivation when doing so, is a significant issue for testing the validity of approaches. In some cases, this has impacted on testing whole instruments, and across most projects there has not been sufficient engagement for validating disciplinary breadth or longitudinal analysis. Further piloting and analysis will be necessary for most measures to be used across all subjects and institution types.

Evaluating validity and reliability

5.4   All the projects are exploring the validity of different measures of learning gain as noted in their individual success criteria. For projects designing and testing new instruments this has been a major focus of their work so far. There are multiple approaches to judging the validity and reliability of measures of learning gain. These are conducted through statistical techniques, interviews with stakeholders such as staff and students, and analysis of findings.

5.5   <u>Content and face validity.</u> Content validity measures the extent to which a measure represents all of the elements of a construct. For example, a test with strong content validity represents the topics *actually* taught to students, rather than asking unrelated questions. Content validity is a statistical way of ensuring only meaningful elements are included. This is related to face validity, a qualitative judgement such as asking students if they think a test or survey is well designed and useful. Instruments being developed need to be clear about what they are trying to measure and test whether this is indeed the case.

5.6   <u>Concurrent and predictive validity.</u> Most projects are using multiple approaches to measure learning gain. This allows for analysis and triangulation across different approaches to judge the effectiveness of different measures. For example, if there is no relationship between students' scores on a test and their grades than the two approaches are not effectively measuring the same aspect of learning gain. Similarly, predictive validity explores correlations between new measures of learning gain and existing 'objective' measures.

5.7   External validity. Across the projects collecting new data from students, there is a danger that 'more engaged' students are participating and that this group of students may differ from students who choose not to participate. This factor can be somewhat mediated by statistical controls based on analysis of the whole population, as well as complemented by projects, or strands within projects, that encompass all students in the analysis. The mix of different subjects selected across projects also helps to support the external validity of different measures across the projects.

5.8   Internal validity. Whether the questions asked on a test or survey really explain the outcomes sought is explored through each of the individual instruments used across the projects. Much of this is done through statistical testing on findings, but some of this analysis needs to wait until the Year 3 data comes through to measure the effectiveness of measures for longitudinal analysis.

5.9   Reliability. Reliability is a measure of consistency. A measure has high reliability if it produces similar results under consistent conditions. Projects designing new instruments are conducting reliability tests, including factor analysis, Rasch measurement framework, test reliability theory, and modifying instruments between waves. Full reports on reliability of new instruments, particularly for longitudinal use, will be available when the projects have completed data collection and analysis. All the projects are conducting focus groups and interviews with students and other stakeholders, such as staff, parents and employers to explore the validity and reliability of instruments.

Validity and reliability findings and analysis

5.10   Initial data from several pilot measures, including grades and the R2 Strengths employability tool, found indications of 'learning loss'. Projects are exploring if this is due to challenges of the reliability of the measure or whether this is a true measure of students' experiences.

5.11   The location of variance is another emerging issue in the validity and reliability of instruments. This explores, for example, whether there are greater differences within or between groups of students. This has consequences for how the data is used. The University of Reading pilot found that the scores from the CLA+ test were only meaningful if they were aggregated at the course or institution level, but were not reliable measures to use for student-level decisions.

5.12   After testing of instruments and analysis of initial findings, many projects cut the number of questions, dropped inconsistent scales and reworded questions. Several projects amended questions to have more relevance across disciplines, and others created subject-specific elements of surveys. Some projects added in additional scales to explore.

5.13   Cognitive gain measures. Projects exploring cognitive gains have noted challenges relating to validity and reliability of measures, which is not surprising given the complexity of learning as an activity and the challenge of developing generic instruments in a subject-specific higher education system. As noted earlier, several projects have been exploring the validity of using grades as a measure of learning

gain. Grades are validated through internal and external quality assurance procedures at each institution, but projects have identified how marking and grading patterns are highly subject-specific. The University of East Anglia found different disciplinary patterns in measuring distance travelled due to subject-specific marking rubrics and marking profiles, subjectivity, variations in assessment design, and lack of cross-institution marking discussions.

5.14 Several projects utilised multi-level modelling statistical procedures for analysing grades data to take into account the clustered nature of the data. However, there are questions about the reliability of grading procedures, particularly across modules and institutions. Data analysis indicates there is substantial variation in module grades, which could be explained by inconsistent alignment of grade descriptors across a course and variations in marking within a module. Across several projects there was substantial variation across qualifications and modules that significantly influenced academic performance. Patterns emerged that show students receive lower marks as they progress through a qualification at one institution and noted that institutional factors influence grading outcomes over time, questioning the validity of longitudinal grade comparisons. Analysis from the Open University concludes that although grades may not be a useful proxy measure for learning gains, multi-level analyses can be extremely useful in identifying alignment and mismatches in the curriculum.

5.15 The standardised cognitive tests being piloted have raised the most concerns about validity and reliability. Focus groups with students about the CLA+ test have led to questions about the discipline-independence of the test. Students perceive a benefit from studying certain disciplines but there has been little evidence of this from the limited data collected so far. Birmingham City University also found high scores linked with high participation programmes, which raises questions about what the test is measuring, and how much depends on staff motivation for encouraging students to complete and put effort into the test. They also found lower scores for English as a Second Language students, raising questions about the wording and structure of the tasks. Findings from the University of Reading led them to question the reliability of the CLA+ to measure critical thinking and they found strong correlations with student socio-economic and demographic characteristics.

5.16 The SJT run by A&DC in the University of Lincoln project was dropped as an instrument because of difficulties with getting students sufficiently motivated to engage with the tasks. The data that had emerged was not suitable for use at the individual level but could be used to compare learning gain across schools, colleges or longitudinally. The project also noted the challenges of administering an SSA, because the instrument relies on self-reported data and there are inconsistencies in how students rate themselves in each area. They also found a limited spread of results which offers little opportunity to explore trends in the data compared to student profiles or to monitor an increase in gains over time. However, through qualitative testing, they found most students valued their university experience in preparing them for future employment citing increases in their confidence, maturity and employability.

5.17 <u>Soft skill development measures</u>. From the projects focusing on soft skill development, preliminary findings from the University of Manchester indicate associations between measures, group differences, and the impact of confidence

measures in models of learning gain. Similarly, in the self-efficacy strand of work in the University of East Anglia project, when students learn from each other in the classroom, their confidence at tackling similar problems in the future also increases (at student and class levels) and student grade performance also increases.

5.18   From cross-sectional analysis done by the University of Portsmouth, they found significant improvement on learner resilience, graduate capital, self-efficacy, surface and strategic approaches to learning, but they found no significance for implicit theories of intelligence, self-esteem or deep approaches to learning. The Open University project found the UKES survey to be valid and reliable, with areas of confidence, social interactions and personal development significant to students and graduates.

5.19   Employability and career readiness. From the projects focusing on employability, the University of Nottingham strand of the LEGACY project found the 'concern' factor, exploring a positive attitude towards and engagement with the future, significantly increased over eight months. They also noted initial gains acted as prompts for reflection and action in relation to career activities for students. The University of Birmingham strand of the same project found that costs were a major factor in non-engagement with international activities, such as study abroad.

5.20   In their preliminary analysis of their cross-sectional 'snapshot' data, the Careers Group showed that 46 per cent of all students were in the 'decide' phase of career thinking. For 'tracked' (partial-longitudinal) data, more work readiness learning gain was shown between Years 2 and 3 of a full-time undergraduate programme, and between Years 1 and 2 there were smaller changes in learning gain (within categories, rather than moving from e.g. 'decide' to 'plan' category responses). They also noted an increase of 18 per cent in 'compete' category career thinking responses between Years 1 and 3 of study. Interestingly, 59 per cent of students responding in Years 1 and 2 of their programme had no change in their career thinking. Of those that changed their career thinking, 23 per cent selected a higher ranked statement, and 18 per cent selected a lower ranked statement. In Years 2 and 3 61 per cent of students responding selected the same career thinking statement, a quarter of students selected a higher ranked statement and 13 per cent selected a lower ranked statement.

5.21   In their in-depth exploration of research methods learning, Plymouth University focused their reliability testing on developing questions that would be applicable across disciplines. They found self-efficacy and confidence scales had the most internal consistency, with moderate levels of consistency for research orientation feelings and learning motivations; low levels of consistency were reported for research orientations (perceptions of research methods). They concluded that the importance of discipline-specific aspects should not be underplayed in any use of research methods as a proxy method for measuring learning gain, and neither should the context of the programme of study or the institutional context. Similarly, findings from the use of concept inventories at the University of East Anglia indicate conceptual learning gain related to the module objectives; they found no gain in a different discipline where the instrument was not linked to learning objectives.

5.22   All three-year projects are continuing to analyse existing and new cohorts of longitudinal and cross-sectional data. Much of the work will be on analysis integrating various project strands of work and findings. There are also efforts to conduct more cross-disciplinary and cross-institutional comparisons as well as analyses across socio-demographic factors.

Key considerations: Findings and validity and reliability

Much of analysis and preliminary findings reported so far are part of testing the reliability and validity of measures and approaches. Examples of 'learning loss' may be due to the effectiveness of the instrument, growth in self-awareness and criticality, or accurate measures of students' learning.

Preliminary project findings are necessary for evaluating the validity and reliability of measures but should not be taken as indications of student or institutional performance in this pilot phase.

---

6 Suitability – instrument uses

6.1 Overall, the projects have largely focused on using metrics for local teaching and learning enhancement and are cautious about the comparability of data across broad subject areas. The projects have not, so far, recommended using measures for cross-institutional accountability purposes. Very few projects explore institutional comparisons, as it is contractually against some of the partnership agreements, methodologically challenging and politically sensitive.

6.2 The **unit level of analysis** is a concern for several projects, related to design and subsequent use of measures. For example, Birmingham City University is focusing on student-level outcomes of the CLA+ test, but the instrument was largely designed for institution-level use and the University of Reading did not find it valid for measuring individual student-level progress.

6.3 **Embeddedness** encompasses to what extent the projects, or the outcomes of the projects, have an impact on the activities of academic staff, professional staff and students, as well as leadership and management. The degree of embeddedness varies across strands of work within projects, throughout an institution (e.g. if the data is only used by the careers office) and across project partners.

6.4 Additional lenses of analysis include the **feasibility** of obtaining the measure, incorporating practicality and value for money, exploring whether the measure makes sense to students, staff and other stakeholders and whether the measure helps to support students and enhance teaching and learning. **Practicality** involves projects being able to successfully define and pilot a measure of learning gain, to get students to complete instruments and to be able to analyse and report on findings. These dimensions are explored through the analysis of use of the instruments and evaluation of scalability of individual projects. The notion of **value for money** will be explored in the Year 3 evaluation.

Use of instruments, methods and metrics

6.5 Due to the potential future relationship between learning gain metrics and the Teaching Excellence and Student Outcomes Framework (TEF), several projects found there was increasing institutional interest in their projects and many felt this would likely grow, based on how learning gain metrics fit into the evolving higher education regulatory environment.

6.6 Several projects noted that findings from their projects will be of more use within institutions at the end of the pilot phase, but there may be some on-going benefit. Some projects felt the need to wait to share the data for enhancement purposes until they felt they had confidence in the metrics and models. Challenges around student engagement meant many projects focused resources on incentivising students to participate rather than on dissemination. Some project teams are also more closely embedded with institutional decision-makers and operational areas of the university and have more opportunities to put data to use in practice. Being able to influence institutional policy also depends on whether there is a sense of senior management buy-in to the projects. However, there already are many examples of how measures

of learning gain are being used to help support students and improve teaching and learning.

6.7   Student-level use. Several projects took an approach of providing students with data to be used for personal development. However, Birmingham City University found very few students chose to collect or use their results. They have been working to incorporate the test scores as a diagnostic tool. Personalised reports for participating students were adopted as part of the student recruitment strategy at the University of Portsmouth and the University of Lincoln. Through the employability strands in the University of Warwick project, they provide feedback to students and careers support follow-up. However, as the University of Reading found, without the measures being embedded in a student's course, few took up opportunities to attend workshops or discuss their results. Data from Plymouth University could be used to enhance student engagement in their own development supported by a tutorial element which ensures that individual students can gain future value from self-reflective activity.

6.8   Classroom-level use. The project at the University of East Anglia trialled and tested some new pedagogical approaches in specific classroom settings. These strands provided feedback to students and offered outcome data of innovative practices to share within and beyond the institution, such as 'flipped classrooms', involving instructional content delivered on-line and traditional 'homework' activities done in groups during timetabled sessions. In one part of the project, they trialled extensive flipped classrooms (all sessions run in a flipped format) but found it was 'too much' for some students. Through qualitative and quantitative evaluation, some students reported that their learning needs were not being addressed by such an active lecture theatre experience. In response the project scaled back the changes and offered a balance of flipped classrooms and traditional sessions.

6.9   The aim of the Plymouth University project was to develop a prototype toolkit that can be implemented within institutions through appropriate learning technologies. An implementation strategy has been developed to provide guidance on the most effective use of the toolkit resources.

6.10   Course-level use. Data from the projects has the potential to provide tailored information to inform specific programme design. However, as Plymouth University found, the project needs a sufficient level of responses at programme level to have valid data to offer to support teaching staff. They did find that programme leads and teaching staff valued the opportunity to discuss and reflect upon the role of research methods within their programme. The University of Manchester found that some schools have used the data to evaluate their teaching and learning methodology and have embedded the learning gain data collection processes into their teaching provision. Through engagement with the director of Teaching and Learning, participating schools received bespoke results and workshops presenting data. They noted that one challenge for embedding learning gain measures is competition from other national surveys.

6.11   Although not continued at an institutional level, the University of Lincoln's School of Pharmacy has used data from the SJT as a proxy for the kind of psychometric testing to be used in future as a gateway to pre-registration professional training. Similarly, at Birmingham City University one particular programme has made the CLA+ a

compulsory part of a first-year professional development module, with the results acting as a piece of evidence of development in a professional portfolio. The University of East London is making survey data available to academic staff online through an online app. They are currently working with directors of Teaching and Learning and programme leaders to identify how this will inform teaching and learning. Data from Plymouth University could be used as a curriculum review and design tool as it enables programmes to map research methods and associated knowledge and skills development activity and learning outcomes.

6.12 <u>Institution-level (service use).</u> Several projects provide data that helps support services like careers offices and skills and training units to tailor services and target specific students. The employability strands of the LEGACY project offer feedback to participants, careers support and developmental on-line questionnaires. The University of Warwick has embedded staff development workshops on using the tools and data. The University of Nottingham suggests their tool could be used for careers service delivery to specific groups, with skills awards or curricular employability input for students to assess their adaptability and make plans to strengthen their capabilities. It could also be used to help students develop career preparedness self-awareness. The University of Birmingham approach could be used for staff development and resources for students' international experiences and to develop models for on-campus internationalisation.

6.13 The University of Lincoln and the Careers Group both plan to use their data to bridge employability initiatives between careers services units and academic departments and suggest the data could be used by careers staff for strategic planning and tailoring and targeting support for students. The Careers Group have found that institutions have been using the data both strategically and operationally to:

- engage with academic departments;
- inform institutional policy-making decisions;
- persuade new and existing employers of student interest in their sectors;
- promote relevant events and support to students from widening participation backgrounds; and
- target event marketing at students based on interests expressed in the survey.

6.14 A strand of the project at Ravensbourne involved following up with alumni three years after graduation. They had positive feedback from alumni contacted, to such an extent that the work was 'mainstreamed' as part of institutional activities. This helps with alumni fundraising, mentoring, and careers and placement support.

6.15 <u>Institution-level (strategic use).</u> It is too early in the projects for the data to inform senior management decision making, but the analyses have been able to help shape targeted areas within institutions. The Open University found their data visualisations and modelling sparked internal quality enhancement discussions. Similarly, the modelling of student progression done at the University of Reading allows staff and senior managers to compare attainment across programmes and between student groups within programmes. This has already highlighted where some programmes need to reconsider their assessments and marking criteria. Data from The Manchester College has offered the partner colleges working within the project more

insight into the composition of their student body and their students' soft skills development. Some of the initial outcomes have prompted managerial discussions at The Manchester College on future changes based around such findings.

6.16   Several projects are planning for learning gain data to inform strategic decision making. The Cambridge strand of the LEGACY project will be sending benchmarked institutional reports, but it is not necessarily clear how these would be used. The University of Portsmouth plans to use some of the tools as a method of demonstrating improvement in the recently revised Hallmarks of a Portsmouth Graduate project. Both the University of Portsmouth and Birmingham City University may link learning gain work with related HEFCE-funded Catalyst bid projects around retention and Black and minority ethnic (BME) attainment[5]. Birmingham City University also suggested that the CLA+ results could potentially also be used as an enhancement tool informing curricular design in the university if programmes are showing little to no learning gain. Data from Plymouth University could provide a standard measure of learning gain at a programme level as a measure of student engagement with research methods and associated learning.

6.17   The Open University has active engagement from senior management to use data to better understand grading practices and student journey modelling. As part of the Open University project, Oxford Brookes University has embedded running the UKES survey, and the project has led to strategic decisions about data warehousing, data anonymisation, and data sharing and linking. Across the institution, the Oxford Brookes Enhancing the Student Experience Strategy embeds graduate attributes (integrative across disciplines and years) and links with learning gain data. Also through the Open University project, the University of Surrey is conducting a university-wide review of a strategy for enhanced learning environment, from pre-entry through to employment. Learning gain research has been complemented by projects on employability, placement research and widening participation.

---

[5] http://www.hefce.ac.uk/funding/catalyst/projects/

Level of use	Purpose of use	Challenges
Prospective students	Informed choice on value for money	Learning gain is not necessarily what students want information about
Students	Provide data for reflection, awareness raising	Without support, advice and guidance, students do not use the data
Classroom	Pedagogical enhancement, data for staff to tailor information to students	Requires local staff awareness and engagement
Course	Management, pedagogical enhancement	Challenging to get buy-in from staff
Institutional (service delivery)	Enhance and tailor services (e.g. careers services)	May not be embedded in core institutional teaching and learning delivery
Institutional (strategic)	Programme review, inform strategy, enhancement, evaluate programmes, staff reward and recognition	Priority given to existing metrics and accountability frameworks
Cross-Institutional	Benchmarking, comparisons	Measures may not be suitable for rankings
Employers	Recruiting graduates, diversifying workforce	Data needs to target employer needs and be easily understood
Government	Accountability, regulation, quality assessment, market indicators	Data is complex and needs to be contextualised

Table 5. Analysis of levels of use of learning gain metrics

Analysis of use of instruments, methods and metrics

6.18 Across the projects there is a general lack of senior management engagement, except in cases where they are part of running a project. Overall, senior managers are generally reactive to metrics and accountability agendas, and there is a lack of proactive interest in measuring learning gain or evaluating effective educational practices. The institutional focus is on existing accountability measures, such as current TEF and league table metrics, and institutions are adopting a 'wait and see' approach to which metrics will be used for future regulatory and accountability purposes.

6.19 For academic staff, the pattern found at Birmingham City University is generally found across most projects. There was generally low engagement from staff participating in the project among a multitude of different programmes. However, there were high levels investment from engaged programme staff where they managed to successfully embed measures. Once learning gain measures are embedded, staff generally find them useful. However, staff are wary of potential

accountability measures and do not see the immediate benefit of learning gain metrics, and like senior managers, prioritise existing accountability metrics, such as the National Student Survey (NSS).

Key considerations: Data use

Learning gain data has multiple uses, but data used for local enhancements may not be scalable for comparative use across the sector. For example, some data may be useful to identify patterns of student attainment across subjects, for curriculum redesign or to design interventions for at-risk students.

---

7   Scalability: Pilot project analysis

7.1   Exploring scalability draws on analysis of data, conclusions from each of the stages of evaluation, incorporating which purposes the measures are being used for, for which audiences (level of use) and in what contexts regarding students, subjects and institutions. Three major themes impacting scalability emerged as the projects developed: challenges of student engagement; data ethics; and staff time, financial and opportunity costs.

7.2   <u>Student engagement</u>. Student engagement has been the greatest challenge for the projects so far, impacting on the ability of some projects to pilot, test and evaluate some measures. For all measures and approaches involving additional data collection, the major factor in scalability will be incentivising student participation or making it compulsory.

7.3   A second point on student engagement is that given the relatively low response rates across most projects, some as low as 5 per cent, there is a concern that the samples are not representative, in terms of student characteristics or in levels of motivation and engagement. For example, several projects have indicated that mainly highly engaged students have responded. A few projects are conducting analyses in Year 3 to determine if response rates of 10-15 per cent are representative of student populations. Similarly, within institutions, often schools or subjects with highly engaged staff were selected for pilot target areas, which may not be representative.

7.4   <u>Data ethics</u>. The second major area of concern for the projects is about the use of data, drawing together issues around data protection, data sharing and research ethics. Several projects highlighted the importance of keeping 'students at the centre' in such discussions, particularly around the use of individual-level data, asking who would be able to access it and for what purposes. These concerns impacted upon student engagement, with some students worried about participating in projects because of what would be done with and who would see their results. Students were particularly concerned about their teaching staff knowing their results and this impacting upon their perceptions of students or on their course marks.

7.5   Additional concerns about the use of data focused on the role of data in wider accountability agendas. Several projects highlighted the importance of institutional autonomy, and institutions having the ability to choose what and how to teach and that this should not be dictated by externally imposed learning gain metrics (with fears of 'teaching to the test' as seen in other areas of the educational system). Relatedly, several projects expressed concern about future use of learning gain data for national accountability purposes, institutional 'surveillance', institutional performance management and monitoring.

7.6   While student engagement was a major concern for projects collecting new data, for projects doing secondary analysis, using 'big data' or learning analytics approaches, they found it is important to have a rationale for what questions are being asked from the data. Although institutional systems contain a huge amount of data, the projects found most measures to be highly dependent on context, including student characteristics, subject of study and institution. This led to challenges in trying to compare findings of secondary data analysis, such as grade trajectories, because of

the need to consider subject and institutional factors in contextualising 'big data' approaches.

7.7   Lastly, across all projects there is the challenge of compliance with new data protection laws coming into effect. Gaining consent at multiple stages of use of data has implications for the ability of projects to continue to validate their findings (such as linking different datasets) and for more widely scaling up the use of data that was not expressly given for the purpose of measuring learning gain beyond the context in which the data was provided.

7.8   Costs. Evaluation of the staff time, financial and opportunity costs associated with setting up the projects highlights how measuring learning gain is not currently embedded within institutions. This includes strategic leadership, academic buy-in, and technological and support services. It took many projects a full year to begin data collection. The time intensity of managing projects, collecting and analysing data and disseminating findings are relevant considerations for scaling up any measure, with some requiring specific research, methodological or analytical expertise as well. There are also significant opportunity costs in staff and student time spent on developing or taking part in learning gain measures.

Projects and strands of activity

7.9   Given that 10 of the 13 pilot projects have another year remaining, the analysis of scalability is only tentative at this stage. Projects with separate strands of activity that have not been brought together are analysed separately. Initial indications for scalability are divided into four categories (see Paragraphs 7.10-7.13 and Tables 6-9):

7.9..1.1   Potential for scalability at subject-level across institutions. A group of approaches have moderate levels of potential scalability. Continued analysis in Year 3 of the projects will help determine the degree of subject-level difference and the ability to conduct meaningful cross-institutional comparisons. Projects would also need to be further piloted across a wider range of institutions to ensure validity across the sector.

7.9..1.2   Potential for amalgamation of indicators and wider scalability across institutions. A group of projects have piloted instruments with similar approaches. There could be benefit in analysis of the most useful elements across the projects and possible development of a single instrument focusing on soft skills development, drawing on affective and behavioural dimensions of learning gain. All approaches need further analysis in Year 3 of subject-level differences and potential for meaningful cross-institutional benchmarking and comparisons. This area of work could also be usefully analysed alongside the existing UKES, which features across half of the projects and has been previously validated in the English higher education context.

7.9..1.3   Low scalability; approach requires local embedding; potential for data to be used for cross-institutional benchmarking. Several strands of work focused on approaches relevant at the classroom or course level. These offer potential for

useful benchmarked data of similar modules and programmes across institutions, but may be of more use as evaluation tools for individual teachers and programme leads. Development of toolkits and guides for implementation would aid in wider take-up of such approaches.

7.9..1.4   <u>Low scalability; useful data for enhancement embedded within institutions</u>. A few projects and strands of work have identified approaches to measuring learning gain that provide useful data for analysis and enhancement within institutions but which may not necessarily be comparable across subjects or institutions. Statistical modelling approaches to analysing existing data highlight some of the challenges of what specifically attainment and progression data are representing and the lack of standardisation across the sector, both theoretically (What are the intended trajectory patterns?) and practically (How are subject-level differences managed institutionally?). Other strands of work explore areas relevant to specific service or support units within institutions, for which analysis through Year 3 may identify useful benchmarking opportunities.

Project/strand	Approach (methods)	Characterisation of learning gain	Context (validity)	Methods for validity testing	Potential scalability	Benefits	Limitations	Additional information
University of Cambridge strand of LEGACY project	On-line questionnaire; primarily self-report data	Cognitive; soft skills (affective); employability (behavioural)	Research-intensive institutions; piloted across subjects of business, chemistry, English and medicine	Concurrent; predictive; internal; external; and face validity; focus groups with staff and students; continued validation of student self-reporting	Moderate; dependent on findings from complete longitudinal study	Robust underpinning conceptual framework; low cost; non-proprietary; testing time average 22 minutes	Piloted in specific institutional type (research-intensive); multi-year analysis not completed yet; small monetary incentive to participate (£5); external validity work still underway	Variance at subject and institutional levels; questions of scalability across institutional types; longitudinal usability; unknown relationship with other outcome measures
Birmingham City University/ University of Reading CLA+ strand	On-line test (timed)	Cognitive	Specific subject areas across similar institutions; concerns about student engagement and motivation	Validated in US and other contexts; exploring discipline independence in UK context	Moderate; dependent on findings from longitudinal study	Designed to be discipline-independent (but awaiting longitudinal data to explore in English context)	Proprietary and expensive; high costs (staff training for assessment and test administration); correlation of scores with high response rates; testing time average 45 minutes; not designed to capture outcomes of	Variance at subject and institutional levels; longitudinal usability; unknown relationship with other outcome measures

							English degrees	
The Careers Group	Four questions built into enrolment system; self-report data	Employability (behavioural)	Operating in all subjects of study across a large proportion of the higher education sector	Face validity through focus groups with students; external and comparison validity still underway	Highly scalable; some inconsistency in question wording across institutions	Low burden; clear and easy to use data outputs	Limited and utilitarian view of learning gain; competition with other institutional priorities for enrolment question focus; desire for institutional customisation	Variance at subject and institutional levels; unknown relationship with other outcome measures; question usefulness of measure for enhancement and accountability

Table 6. Potential for scalability at subject-level across institutions

Project/strand	Approach (methods)	Characterisation of learning gain	Context (validity)	Methods for validity testing	Potential scalability	Benefits	Limitations	Additional information
University of Portsmouth	On-line questionnaire; self-report data	Soft skills (affective)	Across multiple institutions; concerns about student engagement	Continued psychometric testing of existing validated scales; still refining instrument	Moderate; very dependent on full findings from longitudinal study	Non-cognitive focus so scalable across disciplines	Limited data collection and cross-institutional analysis so far; long instrument	Variance at subject and institutional levels; questions of longitudinal usability; unknown relationship with other outcome measures
University of Manchester	Subject-specific on-line	Soft skills (affective)	Across multiple	Continued psychometric	Possible; not tested across	Use as a formative tool	Limited coverage of	Variance at subject level;

	questionnaires; self-report data		subject areas in single institution	testing of existing validated scales; linking with administrative data still underway	other institutions; comparability across subjects being explored	for students; 10-15 minutes completion time	learning gain concepts; single institution use; outputs not user friendly (yet)	questions of longitudinal usability; unknown relationship with other outcome measures
The Manchester College	On-line and hard copy questionnaire; self-report data linked with administrative data	Soft skills (affective); employability (behavioural)	Across multiple subjects at multiple further education institutions	Continued psychometric testing of existing validated scales; construct validity and focus groups with students	Moderate (highly scalable across higher education in further education contexts)	Questionnaire and mathematical model for integrating administrative and grades data	Piloted in higher education in further education institutions; have not tested institutional benchmarking and comparisons; similar questions to UKES survey	Variance at subject and institutional levels; unknown relationship with other outcome measures
UK Engagement Survey	Self-report survey on student engagement	Soft skills (affective); employability (behavioural); some cognitive areas	Previously validated in UK context; piloted as part of many projects	Psychometric testing of method and approach; noted disciplinary differences	Moderate; already widely used across the sector	Robust conceptual underpinning; useful as formative tool; provides data for enhancement; extensive international use	Disciplinary differences; proprietary	Relationship with other outcome measures; variance at subject and institutional levels

Table 7. Potential for amalgamation of indicators and wider scalability across institutions

Project/strand	Approach (methods)	Characterisation of learning gain	Context (validity)	Methods for validity testing	Potential scalability	Benefits	Limitations	Additional information
University of East Anglia: self-efficacy strand	Build self-efficacy measures into classroom peer learning	Soft skills (affective)	Piloted in individual modules in specific institutional setting	Psychometric testing of method and approach; discussions with students	Low; specific pedagogical approach	Supports active learning; provides actionable data on innovative pedagogical practices	Needs individual teacher buy-in; needs to be embedded within the curriculum; not as effective without champion; requires multiple assessment points	Examples of use in additional settings; usefulness of comparative findings
University of East Anglia: concept inventory strand	Embed disciplinary-based concept inventories measures into curriculum	Cognitive	Piloted in individual modules in specific institutional setting	Extensively validated by disciplinary experts	Low; individual inventories are time consuming and expensive to develop	Provides standardised, objective measures of students' disciplinary learning;	Limited disciplinary coverage; challenging instrument, risk of low student motivation and engagement; proprietary	Complete list of available concept inventories
Plymouth University	Self-report survey on research methods learning	Soft skills (affective)	Piloted across multiple disciplines across multiple institutions, focusing on college-based	Psychometric testing of method and approach	Moderate; would require embedding in local curriculum practices	Focus on research methods applicable across subjects; pedagogical enhancement benefits	Student engagement is dependent on tool being a supportive aspect of learning, rather than a summative measurement;	Examples of use in additional settings; usefulness of comparative findings

| | | | | higher education | | | time intensive to embed | |

Table 8. Low scalability; approach requires local embedding; potential for data to be used for cross-institutional benchmarking

Project/strand	Approach (methods)	Characterisation of learning gain	Context (validity)	Methods for validity testing	Potential scalability	Benefits	Limitations	Additional information
Grade modelling: The Open University; University of East Anglia; University of Reading	Analysis of existing attainment data	Cognitive	Approaches across a variety of projects and institutions	Grades validated through internal and external quality assurance procedures, however institutional and disciplinary differences; statistical testing of analytical approaches	Scalable approach but not clear what is being measured if comparing across subjects and institutions	Data already collected for all students	Not clear that cross-institutional comparisons would be differentiating learning gain from institutional marking patterns; disciplinary differences; complicated modelling techniques	Further exploration of how comparative grades data informs learning gain
LEGACY employability strands	Questionnaires and qualitative data supporting careers services	Employability (behavioural)	Across multiple subjects at a range of research-intensive institutions	Projects used existing, validated instruments and approaches	Scalable approaches across careers service units	Provide benchmarked data about employability initiatives	Need to be embedded with accompanying support, advice and guidance	What approaches are useful for enhancement and/or comparison and benchmarking
Ravensbourne	DHLE Triangulation; DHLE Plus 3 years; Solent	Employability (behavioural)	Across multiple subjects in multiple	Validity testing of internal triangulation	Scalable approaches (but may be less relevance	Engagement with alumni; rounded view of	Approach relevant to specific institutional	Additional use beyond new Graduate

	Capital Compass Model		creative vocational specialist institutions	approaches; DHLE Plus largely qualitative	to other institutions); some points embedded into new Graduate Outcomes	employability outcomes	type; not clear how multiple approaches fit together	Outcomes approach
University of Lincoln	SJT, SSA, Transition questionnaire	Employability (behavioural)	Tested across multiple disciplines at one institution, small pilot at partner institution	Validity, reliability and use of SJT and SSA not found to be robust	Low, but showcases integrated approach to using data to inform, support and evaluate institutional strategy	Approaches are embedded into institutional strategy	SJT proprietary	Piloting conditions for success in other institutions
University of East London	On-line self-report questionnaire	Soft skills (behavioural)	Tested across multiple disciplines across several institutions	Using parts of existing instruments but whole questionnaire not validated; different items used each wave	Low, but highlights need to account for student characteristics	Focusing on transitions and trajectories of disadvantaged student groups	Significant differential BME responses to questions about academic behaviour, confidence and need for cognition	Analysis by student characteristics and trajectories across other approaches

Table 9. Low scalability; useful data for enhancement embedded within institutions

8   Scalability: Policy analysis

8.1   This is a broad analysis of how measuring learning gain fits into the broader policy context, which impacts on evaluating effective measures of learning gain and for making recommendations about potential uses of learning gain data. Drawing on best practice recommendations, guiding principles should ensure that measures of learning gain:

- are locally relevant;
- externally generalisable;
- advance transparency of validity and reliability; and
- make efficient use of time and money.

International context: measuring student outcomes and learning gain

8.2   The drive for transparency, accountability and value for money in higher education is occurring internationally. Initiatives exploring measuring learning gain provide context for evaluating the pilot projects and inform recommendations for scalability of different approaches.

8.3   Several years ago, the Organisation for Economic Co-operation and Development (OECD) undertook a feasibility study of the Assessment of Learning Outcomes in Higher Education (AHELO) across multiple countries and subjects of study. They faced challenges around questions of what to measure, with international, cultural and subject-level differences emerging. Due to concerns about data quality and use, the project was not continued.

8.4   In the US, the Association of American Colleges and Universities (AAC&U) Valid Assessment of Learning in Undergraduate Education (VALUE) project offers rubrics to externally assess students' in-course assignments against nationally standardised learning outcomes. This has extensive institutional buy-in but is resource and time-intensive. In another approach in the US, the Council for Aid to Education (CAE) developed the Collegiate Learning Assessment Plus (CLA+) test. The test measures generic critical thinking skills that are outcomes of the general education approach in the US, which differs from subject-specific degrees in England. This is being piloted through the projects and discussed below.

8.5   The US has also developed a holistic accountability framework, the Voluntary System of Accountability (VSA), which was created to provide greater accountability through accessible, transparent, and comparable information. The VSA was introduced by the Association of Public and Land-grant Universities (APLU) and AASCU based on the premise of offering straightforward, flexible, comparable information on the undergraduate experience, including student progress and learning outcomes[6]. This provides a model for how multiple measures of student learning can be incorporated into a customisable portal.

---

[6] http://www.voluntarysystem.org/

8.6   The European Commission is supporting the Measuring and Comparing Achievements of Learning Outcomes in Higher Education in Europe (CALOHEE) project as part of the Tuning framework[7]. This work is underway but focuses on aligning frameworks for course design rather than student outcomes. There are also national research projects in Germany[8], Brazil[9], Italy[10] and Columbia[11] on student learning outcomes which have raised concerns about student engagement, breadth of focus across sectors of higher education and practical challenges.

Policy context: Drivers for measuring learning gain

8.7   Measuring learning gain is complex, involving philosophical questions of what to measure and scientific questions on how can it be measured, often involving high-level statistical analysis. This is in a broader political context of asking 'why' measure learning gain related to the relationship between higher education and funding, and the state, the economy and individuals.

8.8   Purpose. Learning gain is broadly considered to relate to the improvement in knowledge, skills, work-readiness and personal development made by students during their time spent in higher education. What aspects are focused on depends on views on the purpose of higher education, which vary across the sector. For example, different measures would be used if one thinks higher education is for:

- developing flexible knowledge workers for the 21st century economy;
- creating an engine for social mobility;
- cultivating a passion for a subject and skills for life-long learning;
- training professionals to participate in a functioning society;
- fostering democratic citizenship and contributing to a broader community; or
- investment in skills for accrual of individual capital.

8.9   There are multiple stakeholders with differing opinions on the purpose of higher education, including students, academic staff, senior managers, the government and the wider public. And as with all social science policy research, there are inevitable trade-offs between what is conceptually possible, methodologically rigorous and practically feasible.

8.10   Competition. There is a policy drive for market competition and informed student choice to raise quality in the sector. Changes to tuition fee policies and a greater burden of cost put directly on students has raised questions about value for money for individual students and for the sector. These policies influence the analysis of the approach, validity, suitability and scalability of different measures. However, the definition of value for money, particularly notions of value and quality, are highly contested.

---

[7] https://www.calohee.eu/
[8] Modelling and Measuring Competencies in Higher Education. http://www.kompetenzen-im-hochschulsektor.de/index_ENG.php
[9] https://link.springer.com/article/10.1007/s10734-015-9963-x
[10] https://link.springer.com/article/10.1007/s40888-017-0075-1
[11] http://www.tandfonline.com/doi/abs/10.1080/02602938.2016.1168772?journalCode=caeh20

8.11   Quality. In higher education, quality has a strong foundation in institutional reputation, drawing heavily on institutional age, research performance, student selection and graduate salaries. Reputation may be only partially, or conversely, 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 scale (first, upper second, lower second and third) and through rises in grade inflation there is little differentiation amongst graduates. Questions of how quality is judged have led to the desire for better metrics to account for what students have gained from their time in higher education, and what added value institutions provide.

8.12   Social mobility. Widening participation and social mobility agendas feature in debates on measuring learning gain. 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 has led to a desire for more information to differentiate the quality of graduates beyond measures highly correlated with prior socio-economic status.

8.13   In response many employers now design in-house recruitment mechanisms. These are often methodologically flawed and burdensome tests, which creates high inefficiencies for employers and graduates. This situation has led to a desire for metrics which allow for better criteria for targeting students, courses and institutions that demonstrate the knowledge, skills and attributes that employers are looking for and the economy needs.

8.14   TEF. Policy concerns around competition, quality and social mobility have been drawn together through a new quality mechanism, the 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'[12]. The TEF has been piloted with noted limitations of existing metrics as proxy measures of the assessment criteria. Some of the currently used measures are satisfaction, retention and completion, employment and salary data, all of which are benchmarked to take account of the backgrounds of students studying at each provider. These are not necessarily measures of student learning, or may even be linked with lower levels of student learning and academic challenge. Several are also highly dependent on institutional and student characteristics. Measures of learning gain could, in theory and if developed to a suitable level of robustness, provide more directly relevant metrics to assess areas of teaching and learning quality.

8.15   The potential inclusion of learning gain measures in a national accountability system has raised interest and concerns about the pilot projects. There are arguments that it is too difficult to measure the complexity of student learning. However, although it may be challenging, the learning gain pilot projects focus on exploring measures that better capture student learning outcomes than existing measures. Measures of learning gain have the potential to contribute to a virtuous cycle, through which holding institutions accountable while activities undertaken to raise outcomes would lead to improvements in teaching and learning and the student experience.

---

[12] http://www.hefce.ac.uk/lt/tef/whatistef/

8.16   Overall, what has emerged from pilot projects and international initiatives is that there is no simple, 'silver bullet' metric that accurately and effectively measures student learning comparatively across subjects of study and institutional types. There is no single metric that could be appropriately scaled up to the institutional level and slotted into the existing TEF metrics. Particular challenges emerged around pilots of standardised tests and grades.

8.17   <u>Standardised tests</u>. National standardised tests, such as the CLA+, do not capture the outcomes from disciplinary specialisation found in England. Pilots of the CLA+ found it to be bureaucratic and expensive in terms of licensing costs, training and paying of markers for the tests, and pilots struggled with buy-in across institutions. The design of the test for institutional, rather than student-level use, led to challenges of student engagement and motivation.

8.18   <u>Grades</u>. Several projects explore to what extent grades provide an effective comparative measure of learning gain. The conclusions thus far show that grades are subjective measures, with variations across individual markers, modules, courses, disciplines and institutions. Different trends on grading patterns are linked with local practices rather than students' performance. Furthermore, there is a lack of confidence in the degree classification system due to external pressures to increase and inflate grades, through quality assurance processes, league table rankings, retention and completion agendas, and student satisfaction.

8.19   Given the multiple purposes of higher education, the breadth of institutional types and missions and the subject-specific nature of degrees, a selection of tools and methodological approaches may offer a more viable approach to measuring learning. The pilot projects are developing several tools that have the potential to offer valid and robust measures of learning gain, at least within specific institutional, subject and pedagogical circumstances, and that are contextualised for use at the appropriate level.

8.20   Allowing for multiple metrics supports institutional diversification through evidence-based competition by selecting what metrics are most relevant for the mission, subject-mix and portfolio of provision. Such data allows prospective students to make informed choices about what and where to study and provides current students additional information relevant to their experience to support their learning and develop their employability. It also offers employers greater opportunities to select diverse graduates and to target specific courses and institutions relevant to their needs.

8.21   Drawing on other national approaches such as the VSA in the US, one possible approach would be to establish an expert panel to approve tools and methods, oversee modifications and new approaches.

8.22   Institutions could select from the approaches and be required to embed measures in institutional practices and the curriculum. This allows for institutional autonomy to select the metrics that are most appropriate for the institutional mission and subject mix and places responsibility for delivery on institutions. Measures embedded within

institutions better capture institutional contribution to learning gain beyond socio-economic and selection characteristics. This approach also embeds institutional commitment and buy-in.

8.23 To ensure measures were being collected and analysed appropriately, the Office for Students, or another designated body, could reserve the right to audit institutional processes related to collection and analysis of data. Alternatively, this service could be offered through a third-party provider.

8.24 Measurement of learning gain could, if developed to a suitable standard, feed into the TEF in multiple ways. Specific metrics could be part of additional data provided on each institution to help inform decisions about rating. Evidence from learning gain metrics could also be required, or encouraged, as evidence incorporated into the provider submission, allowing institutions to contextualise their data relative to mission and subject mix. Alternatively, following proposed developments of the VSA approach, learning gain metrics could be incorporated into future provision of information for students, offering a customisable, searchable portal of data on student outcomes.

8.25 Subject-level TEF provides additional opportunities, though is also subject to many of the same limitations. Some approaches to measuring learning gain may be selected to be used across certain subjects, and could be used as benchmarked measures relevant to the appropriate assessment criteria.

Key consideration: Multiple approaches to measuring learning gain

A toolkit of multiple metrics approach to learning gain is most viable given the breadth of purposes and different institutional and subject contexts, though implementation would be dependent upon overcoming a number of significant and complex challenges. There is no 'one' measure of learning gain, but rather multiple measures that are relevant for different contexts.

Appendix 1: Evaluation of individual projects

Birmingham City University

Overview	
Summary	Testing and tracking individual students across different subjects at different institution in their three years of study using the CLA+ test and the UKES.
Partners	Coventry University, Liverpool John Moores University, Staffordshire University

Methodology	
Project type	Three-year longitudinal project; added cross-sectional element
Process measures	General cognitive gain; and student engagement
Outcome measures	General cognitive gain
Methods	Test (CLA+) and survey (UKES)

Data collection	
Numbers of students	Year 1: 845 students in selected faculties across four institutions (Nursing is one of the faculties)
	Year 2: 693 students from 2016 cohort; 201 students for cross-sectional element; interviews and focus groups with students

Project development	
Partners	High engagement through existing partnerships; semi-regular site visits, project meetings and on-line scorer training sessions
Progress	Two data collection points; added cross-sectional collection. Analysis of the data is on-going.
Institutional embeddedness	Partner institutions working as separate data collection sites
Plans beyond project	Possible continued engagement from Business School at a partner institution

Challenges	
	Staff time for assessing exams (useful if blocked out in advance)
	Staff concerns about how data will be used (the project is not using data for institutional comparisons)

	Student engagement with completing tests
	Project start date and initial data collection
	Partnership agreements
	External contractor agreements
	Delayed receipt of results from external contractor

Critical analysis

| | The project had to liaise with CAE to adapt US-based survey questions for UK context (mainly demographic questions). Overall a very ambitious project that had a challenging start but managed to make progress in the first year. It will be useful to track continued engagement of students, disciplinary differences and any linking with other surveys or datasets. |
| | The project noted the importance of buy-in from senior academics and programme-level staff with whom students identify strongly. This worked much better than messaging from staff perceived as external to the students' programme of study. |

The Careers Group

Overview

| **Summary** | Each institution asking two to four career-focused questions at compulsory student registration across the timespan of the project, exploring work readiness – a student's preparedness for the process of choosing and planning their career, for obtaining work beyond graduation and for successfully managing their career in the long term |
| **Partners** | Aberystwyth University (added Year 2)
City, University of London (added Year 2)
Goldsmiths, University of London
King's College London (King's)
Lancaster University
Liverpool John Moores University (LJMU)
Queen Mary, University of London (QMUL)
Royal Veterinary College (RVC)
School of Oriental and African Studies (SOAS)
St George's University of London (SGUL)
St Mary's University
University of Bristol
University of Exeter
University College London (UCL) |

| | Ulster University
University of Edinburgh |
Methodology	
Project type	Three-year longitudinal project; longitudinal and cross-sectional
Process measures	Employability-enhancing experiences
Outcome measures	Employability: Career readiness
Methods	Link questions from a brief compulsory survey incorporated into students' registration process with secondary data on students and progression from institutions. The questions differ slightly at each institution but are broadly comparable
Data collection	
Approach **Numbers of students**	Whole cohort of students (extending to postgraduate taught and postgraduate research) Year 1: 171,297 student responses across all subjects at eight institutions Year 1-2: 308,000 student responses were collect across the first two years of the project
Project development	
Partners **Progress**	Different institutions joining each year of the project. Regular updates provided from partner institutions. Partners sharing in best practice dissemination. Sub-groups of partners engaged with sub-projects focusing on different project elements (related to success criteria). Collected data from eight of 14 institutions in Year 1. Preliminary analytics were completed on a sample of our data set comprised of 118,378 responses from four partner institutions in Year 2. Continuing data collection and meta-analysis and benchmarking activities. Developing benchmarks for careers registration data for national use Conducted pilot focus groups to review the wording of statements, the order of statements and the categories defining progression with their career planning, along with the activities they think should be included as options for the employability-enhancing experience question
Institutional embeddedness	The survey questions were compulsory at most institutions so the engagement from students was very high. Varying

Plans beyond project	levels of use of the data within partner institutions, particularly beyond the Careers Service. Suggest setting up a national centre coordinating the on-going career-readiness work across the sector, currently 32 institutions using the data, possibly up to 58, and 18 plan to do so
Challenges	
	Infrastructure, sensitive data-sharing and protection issues Slow start due to delays in getting project staff hired Implementation failure at some institutions Limited number of questions asked
Critical analysis	
	A large-scale and comprehensive project covering all disciplines and student characteristics. Overall a very straightforward project, although focusing on a single narrow aspect (career readiness). Given the nature of the project it may elicit more useful data for targeting employability and enhancement work within institutions than as a broad measure of learning gain. It will be interesting to see the results of meta-analysis across institutions (currently underway) as well as analysis through linking with other datasets.

University of Lincoln

Overview	
Summary	Longitudinal, mixed method study following students across selected disciplinary areas. Students are asked to complete SJTs and SSAs, which are linked with academic performance data, student engagement and WP data to create individual student profiles.
Partners	University of Huddersfield

Methodology	
Project type	Three-year longitudinal, mixed method study (two cohorts)
Process measures	Situational judgement; Co-curricular activities and engagement; Student self-assessment (SSA)
Outcome measures	Employability
Methods	SJT with data on student participation in training, democracy, work experience, and extra-curricular activities and secondary data on academic performance and widening participation
Instruments	Graduate Dilemmas SJT by A&DC

Data collection	
Numbers of students	Year 1: 669 students across selected faculties at two institutions
	By Year 2: Level 1 cohort 1: 272 students completed SJT, SSA or both; Level 1 cohort 2: 397 completed some elements (669 Level 1 students in total); survey on student expectations completed by 1,865 cohort 1 and 2,852 cohort 2 students

Project development	
Partners	The University of Huddersfield is a secondary partner, functioning as a data-collection site. Plans to repeat use of instruments with selected groups of students, allowing for institutional comparisons. So far facing similar engagement issues; routine liaison
Progress	Added a second cohort to the longitudinal design; continuing data collection and analysis
Institutional embeddedness	High embeddedness at lead institution and trailing engagement from partner institution. Plan to run 'Get Set' student expectations survey beyond the project

Challenges	
	Student engagement
	Selecting test provider
	Initial partner institution dropped out
	Student motivation (Offered students workshops and personal tutor meetings which have not been taken up; attempted to liaise with course staff but failed to retain students, will try to track students into Year 3)

Critical analysis	
	This is a well-organised, comprehensive project. Given how much the project is linked in with the institutional agenda it will be interesting to see how the project progresses at the partner institution. Also with the different data points being linked to individual student profiles it will be useful to track if this is a scalable exercise.

The Manchester College

Overview	
Summary	Linking data from an amended version of UKES with secondary institutional data, developed conceptual and mathematical models of learning gain in the further education context
Partners	The Blackpool and Fylde College, Bradford College, City College Norwich, Doncaster College, Grimsby Institute of Further and Higher Education, Hull College, Leeds City College, New College Durham, Rotherham College, Tameside College, South Devon College, Wakefield College, York College (dropped Gloucestershire and West Herts College in Year 2 due to lack of engagement)

Methodology	
Project type	Three-year longitudinal study
Process measures	Student engagement (adapted UKES)
Outcome measures	Level progression between further and higher education; grades
Methods	Link engagement survey data with secondary data analysis

Data collection	
Numbers of students	Year 1: 185 students across subjects completed the pilot survey
	Year 2: collected 988 surveys (with over a 50% response rate) from 13 institutions and institutional data from three partner colleges. Conducted validation interviews with students.

Project development	
Partners	Mixed; high engagement through existing regional networks; two dropped out
Progress	Amended Year 1 to a pilot testing and developing instruments. Currently collecting and analysing data and refining models. Learning gain measures related to soft skills have been calculated. Learning gain measures related to grading outcomes have been calculated. The overall learning gain model measure has been calculated
Institutional embeddedness	Medium; within selected areas of institutions
Plans beyond project	The Manchester College intends to continue the work of the learning gain project to the end of the academic year 2018-19. This final year will focus on continuing to research the

cohort who started in 2016-17 so that a full three-year study can be concluded, the research will also look at the data cross-sectionally to examine learning gain from as many angles as possible. For the final year of the project only data from the Manchester College will be collected and used.

Challenges

Timing of start of project

Partner engagement (one dropped out so far)

Hiring project staff

Scheduling was difficult, changes to class times and dates and finding convenient slots for every class were the main concerns. This was addressed by closely liaising with department heads and tutors to create an in-depth timetable for the surveys.

Certain colleges struggled to collect survey data. To help with this we created an online version of the survey and also had a meeting in which we gave them pointers as to how we had gone about collecting the data and planning the collection. The online survey did not garner a significant amount of further results, but partner colleges did find the pointers we gave at the meeting helpful in collecting their survey data

Critical analysis

This project has focused on developing a mathematical model for estimating learning gain (grades) and satisfaction using academic challenge, engagement data and data on soft skills development. The outcomes from this project will be interesting to see if there is change over time, what emerges as being significant in the model and the robustness of the model. It will be useful to evaluate if this model is specific to the higher education in further education sector or replicable across wider higher education context. It is helpful to note that measuring learning gain is not new to the further education sector, as much of their regulation follows that of schools.

The Open University

Overview	
Summary	Develop an Affective-Behaviour-Cognition model of learning, to broaden the concept of learning gain, to develop, test, implement and evaluate a range of measurements for learning gains at each of the ABC levels
Strands of work	1) Conduct a literature review of learning gains research and evaluation; 2) Quantitative analysis of student satisfaction data; Virtual Learning Environment (VLE) engagement and grades; and 3) Triangulation with qualitative data
Partners	Oxford Brooks University, University of Surrey

Methodology	
Project type	Three-year longitudinal, and cohort mixed method study
Process measures	Satisfaction, VLE engagement
Outcome measures	Grades (GPA)
Methods	Learner analytics; qualitative diaries and interviews

Data collection	
Numbers of students	Over 200,000 students across all Open University subjects. Conducted student focus groups as part of project design and validity-testing

Project development	
Partners	Secondary partners, bulk of work at lead institution. All partners contributed quantitative and qualitative data
Progress	Conducted analyses on lead institution secondary data, developed mathematical models; comprehensive literature review. Grade-modelling work is complete. Undertaking triangulation of data and in-depth interviews and student log-books
Institutional embeddedness	High embeddedness at the lead institution and trailing engagement from partner institutions
Plans beyond project	Continue quality-enhancement work related to grading practices. A continued partnership project exploring learning gains of international students at University of South Africa

Challenges	
	Infrastructure data sharing

Data-sharing agreements

Research ethics for different elements of the project

Logistical challenges of working across different VLE platforms at partner institutions

Importance of programme, discipline and institutional effects on data means it cannot be easily simplified into simple metrics or proxy measures

Critical analysis	
	Preliminary findings noted: From a literature review, the Open University found that students using self-reported cognitive retrospective measures of learning gains report higher learning gains than pre-post-test studies. Preliminary data analysis suggests non-linear development of learning gain over time.
	This is a very comprehensive, large-scale project. The use of learner analytics is relatively new to the sector and it will be useful to consider how the findings can be used for enhancement and accountability purposes, and raises the question of whether this data is useful as an 'end' or to identify areas for investigation. It will be useful to compare the findings from the analyses done at the partner institutions with different delivery models.

Plymouth University

Overview	
Summary	Student progression in their self-reported understanding, skills, experience and confidence in research methods
Partners	Bridgwater College, City College Plymouth, Cornwall College, Petroc College of Further and Higher Education, South Devon College, Strode College
Methodology	
Project type	Two-year longitudinal mixed methods study
Process measures	Confidence, skills self-assessment
Outcome measures	Research methods skills and confidence; Level progression
Methods	A survey to capture students' understanding of and experience with research methods; staff semi-structured interviews and programme documentation and student reflective logs; links with module performance and additional secondary data analysis
Data collection	
Numbers of students	Year 1: 205 survey completions and 39 student reflective logs across selected subjects (history, business, law, public services, education, marine biology) at seven institutions
Project development	
Partners	High engagement through existing regional partnerships; engagement has been challenging as project is not key part of some institutions' missions
Progress	Adding a cross-sectional element to the survey design; conducted three rounds of data collection (beginning of Level 5 reflecting on Level 4; end Level 5/beginning Level 6, reflect on Level 5; end Level 6 reflecting on Level 6), staff interviews and student reflections, instrument shortened (or different versions for different subject groupings), developed into toolkit
Institutional embeddedness	Medium; embedded in narrow aspects of the institution
Challenges	
	Student engagement with reflective logs
	Project start timing
	Student retention, particularly for final data collection point when students have switched institutions
Critical analysis	

Preliminary findings noted: there were disciplinary differences in the way students perceive research.

This is an interesting but narrow project, exploring students' self-reported understanding, skills, experience and confidence in research methods. The focus on research methods is interesting, both in the college-based higher education sector, and more broadly across the higher education sector as a differentiator from schools-based education. Definitions, focus and applications of research varies across disciplines and it will be useful to track how these are interpreted by staff and students and whether this is a scalable approach.

University of Portsmouth

Overview	
Summary	Explore the development of students' non-cognitive skills, including resilience, employability capital and identity, and self-theories and implicit theories (implicit theories of intelligence, self-esteem and self-efficacy)
Partners	University of Southampton; Royal Holloway University; University of the Arts, London (additional); University of Roehampton (dropped due to lack of engagement)

Methodology	
Project type	Three-year longitudinal, mixed methods study.
Process measures	Resilience, employability capital and identity, implicit theories of intelligence, self-esteem and self-efficacy, student engagement
Outcome measures	Employability capital; grades
Methods	Self-reported questionnaire data, using some pre-existing questionnaires (UKES, ASSIST, Dweck's Implicit Theories of Intelligence), as well as developing new psychometric tools

Data collection	
Numbers of students	In Year 1: 199 Year 1 students; cross-sectional follow-up with 94 Year 2 and 125 Year 3 students across subjects at one institution

Project development	
Partners	Added an additional partner to help with data collection; low engagement from some partners; have held project meetings, devolved data collection (leading to different survey windows)
Progress	Due to lack of Year 1 engagement across partner institutions, running data collection at different times across projects. Qualitative data targets nearly complete, analysis underway. First phase of quantitative data collected.
Institutional embeddedness	Low embeddedness; not planned for ways the project could enhance reaching and learning
Plans beyond project	Portsmouth (alone so far) plans to collect follow-up data from participants in their final year

Challenges	
	Partner contracts
	Timing of start of project and hiring staff

	Partner engagement
	Student engagement (increased incentives and running additional cohorts)
	Parent engagement
	Partner changes and data sharing; student engagement and retention

Critical analysis

This is an interesting project focusing on one strand of learning gain: non-cognitive skills (affective measures). It will be useful to see how this relates to other data (such as grades). The project seems to have significant senior management buy-in but has been largely run as an independent research project and has struggled to get students and partners on board. So far there has been little investment in making use of the data for enhancement purposes, but this may change. This is similar to the University of Manchester project and it will be interesting to compare the outcomes from the two projects.

Ravensbourne

Overview	
Summary	Measure gain from work-based learning and work preparation activities learning in the creative industries and the effects of these activities on employability
Strands of work	1) Triangulating DHLE-data across demographic groups and work-based learning activities; 2) Exploring graduates' outcomes three years out via a phone survey with alumni; and 3) Surveying students before and after a work-placement via an Employability Self-Evaluation Test
Partners	Trinity Laban Conservatoire of Music and Dance, Southampton Solent University, Rose Bruford College, Falmouth University, Norwich University of the Arts, Arts University, Bournemouth

Methodology	
Project type	One-year project with multiple strands on work-based learning
Process measures	Work-placements
Outcome measures	Career sustainability; Employability
Methods	Analysis of existing data (including DHLE); interviews with alumni; student employability survey before and after a work placement

Data collection	
Numbers of students	2883 responses for the DHLE Triangulation strand; 211 respondents to qualitative DHLE follow-up survey; and 261 respondents to ESE (Employability Self Evaluation Test) survey across selected subjects at seven institutions

Project development	
Partners	High engagement through existing partnerships
Variations across strands	Varying engagement across the different work streams
Progress	Narrowed the project scope to be able to deliver on three strands of work; carrying on drawing three strands of work together in analyses
Institutional embeddedness	Medium, although at such small institutions engagement is likely to percolate

Challenges	
	Maintaining focus and scope of project
	Delivering project in time

Staff time and scalability

Navigating institutional structures

Data sharing

Findings and recommendations	
	A number of recommendations emerged from the one-year project at Ravensbourne: 1. In taking forward measures of learning gain, it is recommended that policy makers recognise that 'qualifications on entry' are not always an appropriate starting point for measurement of distance travelled. 2. It is recommended that the Higher Education Statistics Agency (HESA) Student Record be amended to enable vocationally focused institutions or programmes of study to measure learning gain from work preparation by including optional fields on work preparation activity, enabling comparison with eventual DLHE outcomes. 3. It is recommended that further work is undertaken on the development of measures that capture information from graduates on resilience, career sustainability and career satisfaction in our belief that these will reflect more accurately the current reality of graduates working lives post-higher education, both in the creative disciplines and beyond. 4. Ravensbourne found that during the three years of a BA degree student's course their employability improves significantly. Ravensbourne also found from the outcomes from the Longitudinal DLHE and the Solent Capital Compass Model work package that both illustrate that 'longitudinal' measurements should mean multiple survey points. They further suggest that learning gain is best measured incrementally rather than in big leaps (qualifications on entry matched to salary, for example).
Critical analysis	
	This has been a successful one-year project. They narrowed the scope of the project to three stands of work, each of which produced interesting findings. However, they did not draw the stands of work together and provide a synthesis of the approaches and findings.

This project highlights the complexity of entry-level measures, and the challenge of trying to standardise them across the sector. In regard to process measures, this project highlights the importance of capturing the diversity of students' learning experiences (such as work placements). On outcome measures, it is important to consider the diversity of students' post-higher education plans, aims and goals. This is particularly relevant for the review of the DHLE survey. There is a key methodological finding in the lack of a linear development of gains, with the recommendation to include multiple measurement points over time (rather than limit to entry and exit measures).

Most of the findings of this project, based in the creative subjects, would be relevant across the higher education sector, particularly for all vocationally oriented and professional courses. The findings will continue to be incorporated into subsequent project evaluation reports.

University of Reading

Overview	
Summary	To test and evaluate a number of methods (UKES, Wellbeing Survey, progression and degree result) alongside a commercial critical thinking test (CLA+)

Methodology	
Project type	Single institution, three-year longitudinal and cross sectional
Process measures	Well-being, engagement, general cognitive gain
Outcome measures	Grades, general cognitive gain
Methods	Analysing existing student characteristics, progression and academic performance data across all students and linking with new data collection (UKES, Wellbeing Survey and CLA+)

Data collection	
Numbers of students	Conducted statistical modelling on 20,000 students from years 2008-2015. This was complemented by conducting additional surveys but only 39 students completed the CLA+ and there was a 5% survey response rate (for UKES and Wellbeing)

Project development	
Progress	Moved away from randomised sampling to open calls to try to get higher numbers of students participating; added a cross-sectional element
Institutional embeddedness	Low to start, run more as an independent research project in Year 1. Shared grade modelling data with programmes in Year 2.

Challenges	
	Student engagement
	Lack of partners and no engagement with potential partners
	Timing of start of project
	External survey contracts

Preliminary findings	
	Preliminary findings noted: programmes of study vary in terms of student performance at the end of their first year; and, that they mostly manage to increase student scores by the end of the third year; and that the trend is non-linear.
	At the University of Reading, a ceiling effect for the CLA+ was found, whereby more than half of participants have

already little or no room for improvement. They also noted that there certainly is a gender gap in the sample's performance, and it was argued that it could be explained by the weight given by the CLA+ to reading and writing. They found there seems to be little or no relationship between prior achievement and CLA+ scores, but a strong one with being fluent in English. This suggests that final achievement could be for a sizeable part influenced by skills unrelated to critical thinking. Further, there does not seem to be any relationship between results in the two parts of the assessments; therefore, it is not clear why two scores are averaged or what underlying ability the overall mastery level is trying to capture, other than the assessment criterion.

Critical analysis

As with most of the other multi-strand projects, it is not clear how the different elements of this study will fit together. The analyses of existing data strand are similar to part of the Open University project, and raise similar questions of whether is it useful data for external evaluation or more useful to indicate areas to follow up. It will be useful to track the progress of the CLA+ strand of work across Year 3 to see if enough data for analysis is collected.

University of East Anglia

Overview	
Summary	Experiment with, and evaluate, three different approaches to identifying and measuring learning gain using data from cohorts of students across different discipline areas across two institutions
Strands of work	1) Concept inventories, 2) Self-efficacy assessments and 3) Student marks and grades
Partners	City College Norwich

Methodology	
Project type	Two-year mixed method project
Process measures	Self-efficacy; disciplinary cognitive gain
Outcome measures	Level progression; grades; GPA
Methods	One strand focuses on discipline-based concept inventories which function like a standardised test taken at multiple points, trialled in two science disciplines. This data is analysed alongside secondary data on student marks and grades. The other strand is based on self-efficacy assessments, also analysed alongside secondary data on student marks and grades.

Data collection	
Numbers of students	Year 1: The self-efficacy strand had 950 respondents across four subjects at two institutions. The concept inventories strand had 390 students across three subjects at one institution.

Project development	
Partners	High engagement through existing partnerships; secondary partner with limited data collection
Progress	Made good progress on Year 1 targets. Continuing to roll out inventories and surveys more widely. Project success criteria met; local engagement with instruments will continue and analysis of data from the project may continue. Did not bring multiple strands of work together.
Institutional embeddedness	High, linked with institutional strategies, plans and staffing

Challenges	
	Slow start to project
	Student engagement with follow-up concept inventories
	Difficult to get staff to buy into new pedagogical processes

Preliminary findings

Preliminary findings from self-efficacy strand noted: a new pedagogy introduced allows students to develop good self-assessment skills. They also found that positive learning gain is associated with confidence gain. When students learn from each other in the classroom, their confidence at tackling similar problems in the future also increases. This strand of work provides empirical evidence for linking specific educational practices with positive outcomes and gains for students.

Critical analysis

This is a well-organised, multi-strand project. It is not clear how the different strands of work fit together. The concept inventories strand is interesting. There is a question of scalability as the concept inventories are more suited to some disciplines than others and are not universally developed (and they are not 'fixed' and need to be updated over time). The self-efficacy strand focuses on affective measures and may have more use for enhancement within institutions. The GPA piloting is helpful as this allows for greater granularity in the analysis.

Multiple strands of work not brought together.

University of East London

Overview	
Summary	Develop survey of academic and employability skills that is illuminating and responsive to subject, widening participation and institutional diversity
Partners	Brunel University London, University of Roehampton

Methodology	
Project type	Three-year longitudinal study (with two years of data)
Process measures	Confidence, engagement
Outcome measures	Level progression; grades
Methods	Combines survey data including scales on Need for Cognition, Academic Behaviour Confidence Scales and Predict Your Grade, along with a partial administration of UKES and links this with secondary institutional data on socio-demographics, widening participation, non-continuation and attainment

Data collection	
Numbers of students	Year 1: As part of piloting the survey instrument 150 students from two subjects at one institution completed survey

Project development	
Partners	Trailing partners, low engagement in strands of work
Progress	Progress has been about developing and testing an instrument. Conducted Year 2 data collection, preliminary analyses and dissemination milestones have been met. Nearing the end of third phase of data collection and have a programme of internal and external dissemination events for the next few months. New shorter instrument has proven easier to administer to large groups of mixed ability students. Five to seven learning gain items will be included in next year's enrolment task at UEL, their exact formulation will depend on this year's results
Institutional embeddedness	Low; engaging with partner institutions as independent elements and data collection opportunities
Plans beyond project	At UEL the project will be embedded in the recently established Centre for Student Success, and their implementation of Learning Gain through Employability and Learner Support Services. Currently working with

	colleagues to ensure the successful migration and embedding of the project
Challenges	
	Hiring staff
	Staffing changes
	Project start time
	Student churn between measurement points causes over-estimates of learning gain measures and reduced representativeness at the cohort level. This problem runs in parallel to the attendance problem that all higher education institutions face.
	Compulsory completion was pursued but proved unsuccessful this year for a project of this priority level. However, inclusion of learning gain questions in the 2018-19 UEL enrolment task shows a level of institutionalisation
Critical analysis	
	This project got off to a very slow start, not making much progress in the first year. They have ambitious targets for Years 2 and 3 data collection, which will be useful to track and monitor. It will be useful to see how the different elements of the survey are pulled together for analysis when the data is available. There does not seem to be much engagement with partner institutions so far. This is one of the few projects with a specific focus on widening participation students.

University of Manchester

Overview	
Summary	Mixed method study measuring and modelling learning gain including 1) Quantitative analysis of secondary data; 2) Questionnaires on transition, disposition, support and confidence; and 3) Critical reasoning skills standardised test

Methodology	
Project type	Three-year mixed method longitudinal study
Process measures	Transition experience; disposition to learning; confidence; student engagement; critical reasoning skills
Outcome measures	Grades
Methods	Competence scale to be developed for use as a critical reasoning skills (standardised) test and is administered alongside questionnaires on: disposition; transition; perceptions of support and pedagogic practices; confidence on generic learning outcomes, all linked with existing entry and exit data

Data collection	
Numbers of students	Year 1: 70 students engaged as part of instrument development and testing.

Project development	
Progress	Conducted interviews and focus groups with students; development of surveys that capture student's perceptions of higher education, dispositions and critical thinking skills across both Science, Technology, Engineering and Mathematics (STEM) and non-STEM disciplines; engagement with stakeholders; instruments developed, two waves of data collection, merged datasets and developed preliminary models. At third data-point collection, longitudinal data follow-up, and standardising entry/exit measures and conducting interviews
Institutional embeddedness	High embeddedness planned (linked in with institutional strategies, plans and staffing) but due to a lack of data collection in Year 1 this is not possible to judge yet

Challenges	
	Hiring staff
	Research ethics approval
	Gaining school support across the institution

Survey fatigue and low response rates for on-line surveys. Face-to-face collection gets responses but is staff-time intensive

Extensive time negotiating access to schools and students

Critical analysis

This project got off to a slow start, taking a year to recruit a PhD student. The first year was spent developing instruments and building relationships within the institution so there is potential for the project to have benefits for enhancement within the institution. This project focuses mostly on non-cognitive affective measures, but also includes critical reasoning. It will be useful to track how the different elements of the survey are brought together in analysis, once data has been collected.

University of Warwick

Overview	
Summary	Four strands of activity: 'Measurement of Learning Gain in HE and across Disciplines' (15 partners), 'Realise 2 Strengths' (six partners), 'Career Adaptabilities' (seven partners) and 'International Experience and Mobility' (seven partners)
Strands of work	Four work packages, one in curriculum areas and three related to employability
Partners	University College London, University of Birmingham, University of Bristol, University of Cambridge, University of Durham, University of Exeter, University of Leeds, University of Liverpool, University of Manchester, University of Nottingham, University of Southampton, University of Sheffield, University of York, Imperial College London, King's College London, Newcastle University, Queen Mary University of London

Methodology	
Project type	Three-year longitudinal, mixed method
Process measures	Cambridge strand developing a bespoke measurement instrument with: cognitive component; meta-cognitive component; affective component; socio-communicative component and cross-cutting dimensions of: openness; research and moral reasoning.
Outcome measures	General cognitive gain; Career adaptabilities; Employability
Methods	The 'Cambridge strand' involves the measurement of learning gain in higher education and across disciplines through the development of a survey instrument. Realise 2 Strengths strand (CAPP-R2 Strengths profile), led by the University of Warwick, involves a survey and one-to-one sessions; the Career Adaptabilities strand led by the University of Nottingham involves a survey; the International Experience and Mobility strand led by the University of Birmingham involves a survey and focus groups

Data collection	
Numbers of students	Year 1: Interviews with 33 students (University of Warwick-Cambridge strand); 57 interviews completed with Work Abroad and Study Abroad groups (University of Warwick-Birmingham strand).
	Year 2: Warwick: 524 'before' questionnaires and 400 'after' completed; interviews transcribed; training for 24 staff

	across six institutions. Nottingham: 431 responses (October 2016), 155 responses (April 2017), 14 interviews 6-7/2017. Birmingham: interviews 56 (each of three groups)
Project development	
Partners	Medium engagement with partners. Different institutions participating across the different strands; one institution dropped out altogether, some dropping out of strands. Multi-layered management structure with key strand leads
Progress	Conceptual framework developed, instruments piloted. Cambridge: two waves of data collection, on target; final data collection September 2018. Warwick: complete. Nottingham: final data collection and interviews spring 2018; running parallel cohorts in final year at three institutions. Birmingham: final data collection and being analysed. Report due March 2018
Institutional embeddedness	Low embeddedness of the Cambridge strand of the University of Warwick project. Medium embeddedness of other strands – largely within narrow elements of the institution.
Challenges	
	Timing of start of project
	Hiring staff
	Partner drop out/overlap with other projects
	Data-sharing agreements
	Student engagement and retention (increased incentives and running additional cohorts)
Critical analysis	
	The conceptual framework developed in the Cambridge strand is very helpful; it could be used to draw the work of many of the projects together. They usefully engaged students in the design of the framework, taking their views on board. Multiple strands of work not brought together.

Appendix 2: Evaluation Framework

1. Development of a measure/proxy of learning gain
   a. What approach was used?
   b. How was learning gain measured?
2. Robustness and effectiveness
   a. Validity and reliability
   b. How many students were involved?
   c. How did the project develop over time?
   d. How was the measure of learning gain judged and assessed?
3. Suitability
   a. Was the measure feasible to obtain (practicality and value for money)?
   b. Does the measure make sense to students and academics and other stakeholders?
   c. Does the measure help support students and improve teaching and learning?
4. Scalability
   a. Was data and information shared across institutions?
   b. Was/is the measure replicable across disciplines, student groups and at other institutions?