

Final Evaluation of the
Office for Students
Learning Gain Pilot Projects
July 2019

Report to the OfS by King's College London
Author: Dr Camille B. Kandiko Howson
Learning Gain Pilot Projects Evaluator

The report reflects the views of the pilot project evaluator and not that of the Office for Students (OfS).

Table of Contents

List of abbreviations	3
Notes	4
1 Executive summary	5
Table 1. Summary of projects and strands of activity	13
2 Dimensions and development of measures of learning gain	17
Definition	17
Rationale and conceptual models	18
Dimensions of learning gain	19
Entry and outcome measures	21
3 Robustness and effectiveness	23
Project design and set-up	23
Student engagement	24
Non-linear learning	26
Subject-specific or generic measures	27
Standardised tests	29
Surveys using self-assessment	30
Using existing grades and attainment data	31
4 Suitability: Instrument uses	33
Use of instruments, methods and metrics	33
Enhancing teaching and learning	34
Employability and transition into work and further study	36
Quality, accountability and performance	37
5 Scalability: Pilot project analysis	41
Purpose and development	42
Usability and embeddedness	43
Analysis of scalability	43
Leadership and staff engagement	45
Data and ethics	46
Scalability of projects and strands of activity	48
Table 2. Summary of scalability of projects and strands of activity	49
Appendix 1: Background information	53
Appendix 2: Evaluation process	56
Appendix 3: Project methodology	61
Appendix 4: Evaluating validity and reliability	64
Appendix 5: Student engagement approaches	66
Appendix 6: Pilot project selection	68

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
CALOHEE	Measuring and Comparing Achievements of Learning Outcomes in Higher Education in Europe
CLA+	Collegiate Learning Assessment
DLHE	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
LEGACY	Learning and Employability Gain Assessment Community Project
NSS	National Student Survey
NSSE	National Survey of Student Engagement
OECD	Organisation for Economic Co-operation and Development
OfS	Office for Students
SJT	Situational Judgement Test
SEDA	Staff and Educational Development Association
SRHE	Society for Research into Higher Education
SSA	Student self-assessment
TEF	Teaching Excellence and Student Outcomes Framework
UKES	UK Engagement Survey
VALUE	Valid Assessment of Learning in Undergraduate Education
VLE	Virtual Learning Environment
VSA	Voluntary System of Accountability

Notes

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

Recommendations are directed at the funder of this work, the Office for Students, but are relevant for any agency, sector or individual designing learning gain measures. Conclusions summarise points relevant to the aims of the project and the evaluation.

This evaluation focuses on the remit of the pilot projects: *to test and evaluate the different methodologies*, and therefore does not report on findings of the projects, except where relevant to evaluating the methodology.

The pilot projects were not designed to explore whether learning gain metrics could be used in current regulatory and policy initiatives. However, this evaluation report offers suggestions for such purposes relevant at the time of writing.

Executive summary

- 1.1 This is the final report of the evaluation of the Office for Students' (OfS) learning gain pilot projects programme¹, which was inherited from its predecessor the Higher Education Funding Council for England (HEFCE). 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. This drew on a HEFCE-funded scoping study which provides an introduction to learning gain². Projects lasted between one and three years and summaries are available in Table 1 (page 13), with details available on the OfS website³.
- 1.2 The multifaceted nature and purpose of higher education leads to a breadth of conceptions of learning gain. Combining useful approaches from across the projects through this evaluation leads towards understanding learning gain as **the change in knowledge, skills, work-readiness and personal development, as well as enhancement of specific practices and outcomes in defined disciplinary and institutional contexts**. [Sections 2.2 – 2.3]
- 1.3 The aims of the pilot phase were to:
- test and evaluate the different methodologies currently used to measure and assess learning in higher education, and the suitability and potential scalability of these methods in England;
 - promote partnerships between institutions, departments and disciplines to facilitate the transfer of expertise and new ways of measuring learning gain more widely across the sector;
 - build knowledge and capacity for learning gain within the higher education sector in England; and
 - facilitate dialogue across the sector and with experts, and thereby to develop a shared understanding to underpin future developments (e.g. the development of national measures).
- 1.4 The pilot projects were evaluated on their selection criteria⁴ to explore questions about learning gain⁵ 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 Background information is available in Appendix 1

² Rand Europe report on Learning Gain:

<https://webarchive.nationalarchives.gov.uk/20180319114205/http://www.hefce.ac.uk/pubs/rereports/year/2015/learninggain/>

³ <https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/learning-gain-pilot-projects/>

⁴ Selection criteria is available in Appendix 6

⁵ Details on the evaluation are available in Appendix 2

Measuring learning gain

- 1.5 Measuring learning gain is complex and contested. Overall, what has emerged from the pilot projects and related international initiatives⁶ is that there is no simple ‘silver bullet’ metric that accurately and effectively measures student learning comparatively across subjects of study and institutional types, despite appetite from government ministries and the media⁷.
- 1.6 Across the projects, nearly 30 approaches were piloted; they identified and captured three dimensions of learning gain:
- measures of general **cognitive gain** – what students think and know;
 - measures of **soft skills development** – affective measures of attitudes and how students feel, and behavioural measures of students’ engagement; and
 - **employability and career readiness** – largely behavioural measures of activities students have undertaken in preparation for the world of work.
- [Sections 2.8 – 2.16]
- 1.7 Evaluations from the first two years of pilot projects⁸ identified major challenges to measuring learning gain including challenges of setting up and administering learning gain projects; student engagement in completing tests and surveys; the breadth of the definition of learning gain; and issues around data protection, data sharing and research ethics.
- 1.8 Evaluation of the projects identified different instruments and approaches that most effectively captured these three dimensions of learning gain, and their potential scalability. These dimensions capture the measures of learning gain explored in the projects and across related international initiatives. Recommendations are listed according to the approach suitable to the relevant dimension of learning. A summary of the scalability of the pilot projects and strands of activity is available in Table 2 (page 49).

Recommendation for developing measures of learning gain

Dimensions

Measures of learning gain identified by the projects are captured through cognitive measures – what students think and know; behavioural measures of what students do and how they engage; and affective measures of attitudes and how students feel. These dimensions should be specified when developing measures of learning gain. [Sections 2.8 – 2.16]

⁶ For summary see: <https://www.officeforstudents.org.uk/media/1386/evaluation-of-hefce-s-learning-gain-pilot-projects-year-2.pdf> p43 and for further detail: Hazelkorn E, Coates H, McCormick AC (Eds), 2018, ‘Research handbook on quality, performance and accountability in higher education’ Cheltenham: Edward Elgar.

⁷ Caspersen J, Smeby J-S, Aamodt PO, 2017, ‘Measuring learning outcomes’, European Journal of Education, 52, pp20–30.

⁸ <https://www.officeforstudents.org.uk/media/1386/evaluation-of-hefce-s-learning-gain-pilot-projects-year-2.pdf> and <https://webarchive.nationalarchives.gov.uk/20180322111250/http://www.hefce.ac.uk/pubs/rereports/year/2017/lgeval/>

Attainment data to measure cognitive gain

- 1.9 For measuring knowledge, or students' cognitive gain, existing attainment data from students' on-going assessment was identified by the projects as being the most efficient and effective approach to capture students' learning from their course and is the best starting point to drive enhancement⁹. Learning gain measures can be used to improve alignment of course-level student learning outcomes with existing quality frameworks and institutional assessment practices. They are also useful for identifying attainment gaps across student characteristics, modules and courses, as noted by the University of Reading.
- 1.10 Evidence from multiple projects, including The Open University, The Manchester College, the University of Reading and the University of East Anglia, found that attainment data provides measures of students' cognitive gain in relation to their subject. They also identified that large-scale trends in grading patterns and trajectories are linked with subject and institutional practices rather than students' performance, which limits the efficacy of using grades data as a scalable, national measure of learning gain.
- 1.11 To address concerns from the pilot projects about this approach, future efforts should focus on enhancing alignment and benchmarking across quality systems, at module, course, institution and subject levels. Future work should engage disciplinary academics focusing on quality regimes including benchmarks and frameworks at subject level. External models for this include the Valid Assessment of Learning in Undergraduate Education (VALUE) initiative in the US and the Measuring and Comparing Achievements of Learning Outcomes in Higher Education in Europe (CALOHEE) project.
- 1.12 Evaluation of the standardised tests used in the pilot projects by Birmingham City University, the University of Lincoln and the University of Reading shows that they do not offer a robust, feasible and value-for-money approach to satisfying desires for more information about the quality and content of student learning for a scalable, national measure across courses and institutions.

Recommendation for developing measures of learning gain

Cognitive gain

Cognitive gain is best captured through existing attainment data. Additional work in this area should focus on improving assessment, marking and alignment with quality frameworks. [Sections 2.8 – 2.11; 3.37 – 3.44; 5.9 – 5.10]

⁹ Jankowski NA, Timmer JD, Kinzie J, Kuh GD, 2018, 'Assessment that matters: Trending toward practices that document authentic student learning', Urbana, IL: National Institute for Learning Assessment.

Conclusion

Pilots of standardised tests carried out during the projects have not proven to be robust and effective measures of learning gain due to challenges of student engagement, differential scores across socio-demographic characteristics, subject differences and use of data. [Sections 3.29 – 3.33]

Surveys to measure soft skills development

- 1.13 Surveys most efficiently capture the soft skills and personal development students gain from higher education on a large scale. Pilot projects identified numerous instruments that can combine engagement data with affective measures that capture student self-reported gains. For scaling up this approach, a new instrument could draw on behavioural and affective elements from the tool developed by the University of Cambridge strand of the University of Warwick project, the UK Engagement Survey (UKES) (used by half of the projects), and additional newly piloted affective tools from the University of Portsmouth and the University of Manchester. Additional time is needed to understand the usefulness of these tools, as they were still being developed and tested when the projects concluded.
- 1.14 Evaluation of the projects suggests such survey data is most effective for enhancement purposes but could also be used to meet accreditation standards, as has been modelled with the National Survey of Student Engagement (NSSE) in the US. Data could be used for benchmarking comparisons across subject levels and student characteristics, as identified by the University of East London. Plymouth University noted such approaches also function as reflective tools for students when engaging with the instruments.
- 1.15 Development of a new instrument, or scaling up a newly developed or piloted instrument, needs to have a clear rationale for selecting items and a plan for dissemination and use of data. Any such instrument would need to be considered alongside existing surveys, such as the National Student Survey (NSS), for prioritisation, complementary approaches and use.

Recommendation for developing measures of learning gain

Soft skills development

Soft skills and personal development are most efficiently captured through surveys. Development or adoption of an instrument needs a clear rationale and needs to be embedded within the existing quality, accountability and performance frameworks. [Sections 2.12 – 2.13; 3.34 – 3.36; 5.11 – 5.13]

Measuring employability and career readiness through registration data

- 1.16 Work-readiness data has been usefully captured through a careers registration approach piloted by the Careers Group. It provides institutions with data both to support enhancement in teaching and learning and for internal strategic use, and offers opportunities for benchmarking and comparisons across subjects and student characteristics. Data from career readiness can be linked with emerging data from Graduate Outcomes to address quality, accountability and performance with regard to employability.

Recommendation for developing measures of learning gain

Employability and career readiness

Work-readiness data is usefully captured through careers registration and is already being widely scaled up across the sector. This provides useful data for institutions and opportunities for benchmarking across subjects and student characteristics.

[Sections 2.14 – 2.16; 3.34 – 3.36; 4.18 – 4.20; 5.14]

Principles for future work on measuring learning gain

- 1.17 Scalability of approaches to measuring learning gain are dependent upon the rationale. Indeed, 'The greatest challenge in developing learning indicators is getting consensus on what kind of learning should be measured and for what purpose a learning indicator is to be used.'¹⁰ Outcomes of the pilot projects, sector learning gain debates and events, and lessons from international pilots and initiatives to measure learning gain draw out several fundamental recommendations.

Purpose and rationale

- 1.18 Measures of learning gain need a clear rationale for development. Buy-in for scaling up any measure of learning gain depends on (a) its intended use: for improvement or accountability; and (b) the level of comparison (student, subject, institution). This drives the presentation and dissemination of data¹¹. For metrics to provide value they need to be able to produce data on what students know and can do for improvement and enhancement to ensure quality¹².
- 1.19 Recommendations for further activity on learning gain by institutions and sector bodies are dependent on the rationale. Given considerations of value for money and practicality, measures of learning gain need to benefit student learning. Students' engagement in completing tests and surveys depends on the usefulness of the outcomes to them and their learning. Developing measures and approaches with students ensures they are relevant, as showcased in the University of Cambridge strand of the University of Warwick project.

Recommendations for developing measures of learning gain

Rationale

Measures need a clear rationale for their development, use and audience. Measures should be developed in partnership with students, clearly support student learning and have a dissemination plan for clearly communicating findings with relevant audiences. [Sections 5.4 – 5.6]

¹⁰ Shavelson RJ, Zlatkin-Troitschanskaia O, Mariño JP, 2018, 'Performance indicators of learning in higher education institutions: An overview of the field', in Hazelkorn E, Coates H and McCormick AC (Eds), 'Research handbook on quality, performance and accountability in higher education', pp249-263. Cheltenham: Edward Elgar, p251.

¹¹ Lennen MC, Fukahori S, Edwards D. 2018, 'Global Assessments of Disciplinary Learning Outcomes: What We Learned from AHELO', in Hazelkorn E, Coates H and McCormick AC (Eds), 'Research handbook on quality, performance and accountability in higher education', pp264-274. Cheltenham: Edward Elgar.

¹² Kuh GD, Jankowski NA, 2018, 'Assuring High-Quality Learning for All Students: Lessons from the Field', in Hazelkorn E, Coates H and McCormick AC (Eds), 'Research handbook on quality, performance and accountability in higher education', pp305-320. Cheltenham: Edward Elgar.

Relevance for students

Students should be partners in developing learning gain measures. Measures of learning gain should benefit students' learning, self-reflection and development planning to support engagement and learning. To support engagement, measures of learning gain should be timetabled and embedded in the curriculum and institutional processes so that it is clear to students and staff that it forms a valued part of their education. [Sections 3.8 – 3.16]

Balancing rigour and practical delivery

- 1.20 Analysing the projects, there are two broad approaches to measuring learning gain. A key aspect of both approaches is balancing a sufficiently rigorous methodology that is practically deliverable, as is the case with all large-scale efforts to measure learning gain¹³. Projects identified major challenges around data and ethics to both approaches that need to be considered when developing measures of learning gain [Sections 5.20 – 5.28].
- 1.21 The first uses analysis of **existing data** from students' experiences at their institution, including prior educational attainment and learning analytics such as participation, engagement and experiences, progression and grades. A number of the projects have demonstrated this approach can be useful for uncovering patterns in progress and attainment across student groups or courses of study. Challenges include separating meaningful differences versus module or subject-specific idiosyncrasies. For projects analysing institution-level attainment data, there were large grading profiles and trajectories across modules, courses and institutions. This raises questions about the reliability of using grade trajectories as a comparative learning gain measure and as a validation tool for new learning gain measures.
- 1.22 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. A challenge of these is the appropriateness of using existing instruments and the robustness of new instruments. However, the major hurdle projects faced was student engagement: gathering sufficient meaningful data to be generalisable across student characteristics, subjects and institutional types. By far the most successful projects have embedded surveys, tests and questionnaires in the curriculum and standard institutional processes such as registration, as shown by the Careers Group.

Recommendations for developing measures of learning gain

Project development

Measuring learning gain within institutions is complex and takes extensive staff time to set up and run. It can take five years to run a longitudinal project, with six months to a year for set-up, three years to run, another six months until attainment data is available plus time for analysis. This time-frame and investment in resources needs to be built into any future projects. [Sections 3.4 – 3.7]

¹³ Coates H, 2016, 'Assessing student learning outcomes internationally: Insights and frontiers', *Assessment & Evaluation in Higher Education*, 41(5), pp662-676.

Measurement design

- 1.23 No single test score or quality-ranking system can accurately measure what students actually know and do. Best practice recommends multiple indicators of learning¹⁴. Projects noted the need for multiple measures of learning gain to account for the diversity of student learning and to prevent gaming of single metrics, which can move institutions away from meaningful teaching¹⁵. Using multiple measures also highlights the interconnectedness of measures, as evidenced by The Open University.
- 1.24 Subject-level differences, in terms of student understanding of measures, findings, relevance of measures and usefulness of comparisons, emerged across all the projects. For example, analysis at subject level identified differences in grading trajectories by The Open University, self-efficacy by the University of Manchester and employability by Ravensbourne. Students and academics reported the importance of integrating learning gain measures in students' disciplinary contexts.

Recommendations for developing measures of learning gain

Multiple measures

Learning gain has three main dimensions: affective, behavioural and cognitive. Multiple measures of learning gain are necessary to capture the diversity of student learning in higher education. [Sections 2.8 – 2.16]

Multiple points in time

Student learning occurs across many dimensions and varies over time and direction, and needs to be accounted for throughout a student's course experience at multiple points in time. [Sections 3.17 – 3.22]

Use of entry and outcome measures

Outcome measures need to clearly specify and communicate what is being explored through measuring learning gain. Entry and outcome measures need to be used contextually and carefully to support capturing learning gain measures and not as targets in themselves. [Sections 2.17 – 2.21]

Using learning gain data for enhancement

- 1.25 Measuring learning gain does not inherently provide benefit to students¹⁶; rather, as evidenced across projects, it depends on how the data is put into practice for enhancement purposes. These uses feed into scalability as they help build a rationale for development and support for student engagement. Three main uses of learning gain measures were identified: enhancing teaching and learning [see Sections 4.5 – 4.14 for examples]; employability and transition into work and further

¹⁴ Shavelson RJ, Zlatkin-Troitschanskaia O, Mariño JP, 2018, 'Performance indicators of learning in higher education institutions: An overview of the field', in Hazelkorn E, Coates H and McCormick AC (Eds), 'Research handbook on quality, performance and accountability in higher education', pp249-263. Cheltenham: Edward Elgar.

¹⁵ Dudas B, 2015, 'What lies ahead for students and learning gain in the UK?' National Union of Students. Presentation at learning gain in higher education – HEFCE/OfS Second National Conference, 10 November 2016, London, UK.

¹⁶ Kuh GD, Jankowski NA, 2018, 'Assuring High-Quality Learning for All Students: Lessons from the Field', in Hazelkorn E, Coates H, McCormick AC (Eds), 'Research handbook on quality, performance and accountability in higher education', pp305-320. Cheltenham: Edward Elgar.

study [Sections 4.15 – 4.19]; and quality, accountability and performance [Sections 4.20 – 4.32].

- 1.26 There is a trade-off between usability and scalability. The University of East Anglia project had two strands of activity that captured learning gain data relevant to specific courses which proved directly useful to the instructors to improve their teaching. However, these were not scalable measures across subjects or institutions. Conversely, several projects that developed new instruments for use across multiple subjects and institutions struggled to identify how the resulting data could be used to enhance teaching and learning.
- 1.27 Students and academic staff need support, advice and guidance to use learning gain data to enhance teaching and learning. Subject-level data on cognitive gains from attainment data, soft skills development data from surveys, and employability data can support students' reflective learning and development and aid staff to benchmark progress and enhance the curriculum to drive increased learning gain. Employability data can prompt students to reflect on their development and can be used by careers staff to support students.
- 1.28 Findings from the pilot projects highlighted areas for improvement in existing sector-wide mechanisms for quality, accountability and performance. Rather than develop additional frameworks for using learning gain measures, better value for money and impact would come from using learning gain metrics to improve and enhance existing mechanisms.

Recommendation for developing measures of learning gain

Embeddedness

Measures should be embedded into existing sectoral, institutional and discipline quality, accountability and performance frameworks. [Sections 5.7 – 5.8]

Conclusion

Contextual factors such as subject-level differences, institution type and student characteristics differences impact the transferability of measures of learning gain. These differences should be considered when designing and selecting learning gain measures; when analysing and presenting findings. Mediating effects need to be considered. [Section 3.23 – 3.28]

Table 1. Summary of projects and strands of activity

Project/strand	Approach and dimension of learning gain	Project context (supporting validity)	Methods for validity testing	Highlights	Challenges
The Careers Group	Four careers questions built into enrolment system; self-report data [Employability (behavioural)]	Operating in all subjects of study across a large proportion of higher education institutions	Face validity through focus groups with students; external and comparison validity	Developed standardised questions to explore career readiness; embedded into institutional systems	Limited use beyond career services staff; embedding into teaching and learning practices; data protection and data sharing
Birmingham City University	Administered Collegiate Learning Assessment (CLA+), an online critical thinking test (timed) [Cognitive]	Specific subject areas across similar institutions; concerns about student engagement and motivation	Validated in US and other contexts; exploring staff and student perspectives; comparing scores across years and genders	Explored use and interest of test of generic critical thinking outcomes in English higher education; learning gain identified across years of study	Low student engagement; concern that scores influenced by level of staff engagement with the project; not linked with institutional data
University of East Anglia: Grade modelling	Analysis of existing attainment data trends and trajectories [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	Developed institutional Grade Point Average (GPA) measure; identified areas of improvement in understanding and standardising grading practices	Differences in grade trajectories noted across years and subjects of study; strands of work across the project not integrated
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	Found increased learning gain through active- and peer-learning activities done in large lectures	Localised activity; not linked with other learning gain measures
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	Identified subject-specific cognitive learning gain in modules	Localised activity; not linked with other learning gain measures; dependent on relating to module learning outcomes
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	Developed measures that captured transition experiences and	High staff turnover; complex project administration; varying

Project/strand	Approach and dimension of learning gain	Project context (supporting validity)	Methods for validity testing	Highlights	Challenges
			validated; different items used each wave	trajectories of disadvantaged students	relevance of various measures and analysis across institutional partners
University of Lincoln	Piloted Situational Judgement Test (SJT), student self-assessment (SSA) and a local 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	Showed importance of senior management buy-in to embed learning gain measures for support, advice and guidance	Varying importance of employability measures for early-year students; relevance of measures across subjects; due to low student engagement not able to test conceptual model linking different strands of work
The Manchester College	Questionnaire on self-report engagement 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	Developed an integrated model to explore learning gain using data from surveys and existing institutional data across multiple institutions. Identified that learning gain increased over years of study	Difficult to capture relevant measures and standardise them across multiple institutions; models dependent on local data collection and reporting practices
University of Manchester	Subject-specific on-line questionnaires on affective measures; self-report data [Soft skills (affective)]	Across multiple subject areas in single institution	Psychometric testing of existing validated scales; linking with administrative data still underway	Developed a questionnaire on affective measures of learning gain, tailored to subject-specific contexts, and implemented it within institutional processes	Only collected two years of data so not able to link with attainment data; limited comparisons across subjects due to customisation of questions
Plymouth University	Self-report survey on research methods learning [Soft skills (affective)]	Piloted across multiple disciplines across multiple institutions, focusing on college-based higher education	Psychometric testing of method and approach	Identified multiple approaches for capturing students' self-reported understanding, skills, and confidence in research methods across levels of study	Found disciplinary differences in the way students perceive research and in the definition, focus and applications of research

Project/strand	Approach and dimension of learning gain	Project context (supporting validity)	Methods for validity testing	Highlights	Challenges
University of Portsmouth	Online questionnaire on non-cognitive gains; self-report data [Soft skills (affective)]	Across multiple institutions; concerns about student engagement	Continued psychometric testing of existing validated scales; still refining instrument	Found gains in non-cognitive learning, including affective measures on resilience, employability capital and identity. Complemented survey design with qualitative research from employers, students and parents	Delays to data collection limited institutional comparisons as well as linking with existing institutional data on grades and progression
Ravensbourne	Piloted multiple measures of work-based learning: Destinations of Leavers from Higher Education (DLHE) Triangulation ; DLHE Plus three years; Solent Capital Compass Model [Employability (behavioural)]	Across multiple subjects in multiple creative vocational specialist institutions	Validity testing of internal triangulation approaches; DLHE Plus largely qualitative	Used multiple measures to identify student employability gains from work-based learning and preparation activities; developed recommendations for sector bodies about wider measures of employability (resilience, career sustainability and career satisfaction) and the need for multiple survey points	Identified challenges of data collection across small, specialist institutions; difficulties of quantitative analysis across small programmes; various strands of work not brought together
University of Reading: CLA+ strand	Administered CLA+, an online critical thinking test (timed) [Cognitive]	Specific subject areas across similar institutions; concerns about student engagement and motivation	Validated in US and other contexts; exploring staff and student perspectives; comparing scores across years and genders	Critiqued approach of measuring generic learning gain in subject-specific setting; identified challenges of centrally administering a standardised test	Found correlations with student characteristics and prior attainment; due to low student engagement not able to test conceptual model linking strands of work
Grade modelling	Analysis of existing attainment data trends and trajectories [Cognitive]	Approaches across a variety of projects and institutions	Grades validated through internal and external quality assurance procedures, however institutional and disciplinary differences;	Used grade modelling to identify attainment gaps across modules, programmes and years of study	Differences in grade trajectories noted across years and subjects of study; very labour intensive to offer support for how data from grade

Project/strand	Approach and dimension of learning gain	Project context (supporting validity)	Methods for validity testing	Highlights	Challenges
University of Warwick: University of Cambridge strand	Online questionnaire capturing multiple measures of learning gain; primarily self-report data [Cognitive; soft skills (affective); employability (behavioural)]	Research-intensive institutions; piloted across subjects of business, chemistry, English and medicine	statistical testing of analytical approaches Concurrent; predictive; internal; external; and face validity; focus groups with staff and students; continued validation of student self-reporting	Developed an accessible instrument capturing multiple dimensions of learning gain, run across multiple institutions; instrument acts as pedagogical reflection tool for students	modelling can be used for enhancement Analysis with existing attainment data not complete; not clear how data would feed into quality enhancement, assurance and accountability processes
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	Provided multiple approaches to measuring employability and benchmarked data across institutions	The three strands of work were not brought together; of limited use beyond career services staff
The Open University: Grade modelling	Analysis of existing attainment data trends and trajectories [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	Developed models for grade modelling across modules, courses and institutions; conducted analysis across student demographics and characteristics; identified non-linear learning gain	Differences in grade trajectories noted across institutions, years and subjects of study
The Open University: Student engagement	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	Embedded measures of student engagement within institutional context; linked student engagement and attainment data	Usefulness of measures depends on how they are integrated with institutional priorities; low student engagement with surveys; difficult to triangulate data across different project strands

Final evaluation of the Office for Students learning gain pilot projects

Dimensions and development of measures of learning gain

- 2.1 Conceptual models or frameworks for measuring learning gain help contextualise definitions, offer a rationale for approaches and provide scope for inclusion of additional measures. Models 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 employment¹⁷. These conceptions are fundamental to developing indicators of learning gain as they drive the selection of what to measure¹⁸. The purposes, values and aims that underpin the projects can be broadly grouped by the three dimensions of learning gain detailed below¹⁹.

Definition

- 2.2 A challenge for the projects was the definition of learning gain used when the projects were commissioned, which was: **the ‘distance travelled’ or the difference between the skills, competencies, content knowledge and personal development demonstrated by students at two points in time**²⁰. Most projects went beyond the ‘two points’ to a wider focus on students’ time spent in higher education, which was adopted by the Higher Education Funding Council for England (HEFCE) in their working definition: **a change in knowledge, skills, work-readiness and personal development by students during their time in higher education**. The University of Cambridge strand of the University of Warwick project’s definition of learning gain was narrower: **students’ change in knowledge, skills, attitudes, and values that may occur during higher education across disciplines**. Projects also identified learning gain as: **the enhancement of specific practices and outcomes in defined disciplinary and institutional contexts**.
- 2.3 Differences in definitions shows the complex territory of learning gain and the debate over what ‘counts’ as learning gain and what is being measured²¹. However, the combination of broad and more focused definitions supports understanding learning gain as **the change in knowledge, skills, work-readiness and personal development, as well as enhancement of specific practices and outcomes in defined disciplinary and institutional contexts**. While there is general consensus across the sector in the definition, it remains very broad and open to wide interpretation; the challenge is in operationalising it²². This would be supported

¹⁷ Gossman P, Powell S, Neame C, 2018, ‘Pain, gain, mission’. Higher Education Pedagogies, 3(1), pp158-160.

¹⁸ Marshall S, 2017, ‘Rising to the challenges of tomorrow’. York: Higher Education Academy.

¹⁹ Further information on the project methodologies is available in Appendix 3.

²⁰ Rand Europe report on Learning Gain:

<https://webarchive.nationalarchives.gov.uk/20180319114205/http://www.hefce.ac.uk/pubs/rereports/year/2015/learninggain/> (pg xi).

²¹ Randles R, Cotgrave A, 2017, ‘Measuring student learning gain: A review of transatlantic measurements of assessments in higher education’. Innovations in Practice, 11(1), pp50–59.

²² Vermunt JD, Ilie S, Vignoles A, 2018, ‘Building the foundations for measuring learning gain in higher education: a conceptual framework and measurement instrument’, Higher Education Pedagogies, 3:1, pp266-301.

through developing a rationale, or conceptual model, to justify the selection and use of measures. This follows advice from similar evaluations of large-scale project piloting measures of learning gain²³.

Conclusion

Learning gain can be defined as **the change in knowledge, skills, work-readiness and personal development, as well as enhancement of specific practices and outcomes in defined disciplinary and institutional contexts.**

Rationale and conceptual models

- 2.4 There are different views on the purposes and uses of learning gain across the projects leading to different approaches and measures of learning gain. For example, through qualitative research, the University of Portsmouth revealed that employers, students and parents expect university to be a place where students can develop personally (e.g. resilience, confidence, independence) and in terms of their employability/career-readiness. Whereas the University of Cambridge strand of the University of Warwick project found students viewed critical thinking as one of the most important aspects of their learning across subjects, followed closely by time management and communication skills. Students also included an emphasis on adaptability, perseverance, resilience, and self-management as part of their learning gain from higher education.
- 2.5 The University of Cambridge strand of the University of Warwick project developed a comprehensive theoretical framework, incorporating cognitive, meta-cognitive, affective and socio-communicative components alongside dimensions of openness, moral reasoning and research. This developed from an intensive analysis of the literature and interviews with students and staff²⁴. This was a major output from the first year of the University of Cambridge strand of work and is comprehensive in that it covers a wide range of aspects of learning gain, beyond those specifically measured in their project. This inclusive model encompasses most of the aspects of learning gain covered across the projects and could function as a meta-model for future development.
- 2.6 The Open University's project was based on an 'ABC' model, incorporating affective, behavioural and cognitive elements of learning gain. This model broadly covers the different approaches used across the projects for measuring learning gain. The University of Reading created a research design model, particularly focused on their secondary data analysis. The Manchester College also developed a conceptual model for calculating the learning gain from grades and soft skills measures including skills development, interpersonal skills, contemporary skills, and employability skills, satisfaction and demographic data, which is particularly relevant for the further

²³ Lennon MC, Fukahori S, Edwards D, 2018, 'Global Assessments of Disciplinary Learning Outcomes: What We Learned from AHELO', in Hazelkorn E, Coates H, McCormick AC (Eds), 'Research handbook on quality, performance and accountability in higher education', pp264-274. Cheltenham: Edward Elgar.

²⁴ Vermunt JD, Vignoles A, Ilie S, 2016, 'Defining learning gain in higher education – exploring the student perspective'. Paper presented at the Society for Research into Higher Education, in Symposium: 'A cross-institutional perspective on merits and challenges of learning gain for Teaching Excellence Framework'. Paper 0230. Retrieved from https://www.srhe.ac.uk/conference2016/downloads/SRHE_ARC_2016_Programme.pdf

education sector. However, the project only collected two years of data so was not able to test the model with final attainment data.

- 2.7 Conceptual models and theoretical frameworks are very helpful for communicating and explaining approaches to measuring learning gain and help to address a missing argument about 'why' learning gain is being measured and the links to 'how' it is being measured²⁵. There was more external interest in projects that had a clear conceptual model because diverse stakeholders could understand the rationale for the selection of measures. The models were designed to be transferrable, although they would need adapting as they were designed to meet the needs of certain institutional types. A model provides justification for what dimensions of learning gain to measure, which were categorised through the evaluation and detailed below.

Recommendation for developing measures of learning gain

Conceptual models

Conceptual models should be included in the development of learning gain measures as they provide a clear, transparent and justified rationale for the selection of learning gain measures and the purpose of using them.

Dimensions of learning gain

- 2.8 Cognitive gain measures. Measures of general cognitive gain explore the variation in non-subject-specific learning through academic years. 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, learning²⁶.
- 2.9 Measures of cognitive abilities explored by the projects 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.
- 2.10 Developing rigorous new tests and surveys of such cognitive measures is both very time consuming and challenging to ensure high quality of psychometric properties, validity and reliability. Most projects used existing instruments, and several created bespoke surveys using selected scales and items from existing surveys.

²⁵ Peseta T, Barrie S, McLean J, 2017, 'Academic life in the measured university: Pleasures, paradoxes and politics'. Higher Education Research and Development, 36(3), pp453–457.

²⁶ Hazelkorn E, Coates H, McCormick AC, 2018, (Eds), 'Research handbook on quality, performance and accountability in higher education'. Cheltenham: Edward Elgar.

- 2.11 Some projects explored 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 (discussed in detail in Sections 3.37 – 3.44 *Using existing grades and attainment data*).
- 2.12 Soft skills development. Measurements of soft skills draw out data that is useful for individual students as well as courses, institutions and employers²⁷. Soft skills development 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. However, analysis of soft skills can be sensitive to groups of students, local contexts and specific types of higher education provision. Projects found that within institutions, comparable measurements of learning gain in diverse disciplines cannot be a valid measurement without an element of contextualisation. To this end, several projects attempted to amalgamate this perspective with a contextualised appreciation of institutional missions, subject areas and pedagogical approaches that underpin students' learning experiences.
- 2.13 Measures of soft skills, or non-cognitive development, used across the projects 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 mathematics; behavioural confidence; disposition to complete chosen course, disposition towards critical thinking; attitudinal/affective measures; openness, confidence and attitude towards research; need for cognition; motivation; and competence in applying critical reasoning skills. 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.
- 2.14 Employability and career readiness. Employability measures can support the development of individual students, institutional strategies and support careers services and cross-institutional benchmarking. Combining employability measures with progression and attainment data allows for better understanding of 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.
- 2.15 Employability measures overlap with several soft skills measures including self-awareness, self-efficacy, resilience, motivation, concern, control, curiosity and confidence. Specific competencies explored in the projects include: global citizenship, agility, commercial awareness, influencing, leadership, and emotional intelligence. In the US literature these are referred to as meaningful learning outcomes, focused on application in the context of work²⁸. Career readiness is linked with its effectiveness as a predictor of employment outcomes for graduates; the

²⁷ Heckman JJ, Kautz T, 2012, 'Hard evidence on soft skills'. *Labour Economics*, 19(4), pp451-464.

²⁸ Bates R, Holton EF III, Hatala JP, 2012, 'A revised learning transfer system inventory: Factorial replication and validation'. *Human Resource Development International*, 15(5), pp549–569.

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.

- 2.16 The three broad dimensions of learning gain, and the multiple approaches, signal that learning gain is multifaceted. Multiple indicators of learning gain are strongly recommended in the literature²⁹ and this is supported by the findings from the pilot projects.

Recommendation for developing measures of learning gain

Multiple measures

Learning gain has three main dimensions: cognitive measures, affective measures of soft skills development and behavioural measures of employability and career readiness. Multiple measures of learning gain are necessary to capture the diversity of student learning in higher education.

Entry and outcome measures

- 2.17 Most projects linked new survey or test data with existing student records data, providing information on student background characteristics and entry qualifications. Student background characteristics have been shown to have an effect on learning gains³⁰, across cognitive³¹ and affective measures³². This is particularly the case for students from disadvantaged backgrounds³³.
- 2.18 Half of the projects used tariff points to account for student differences on intake. However, there is concern about using indicators based on entry qualifications too crudely in high-stakes learning gain measures as institutions can easily lower them³⁴. Several projects noted specific challenges of using standardised baseline 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. This lack of standardisation emanates from the diversity of the sector but also hinders wide-scale comparison.

²⁹ Shavelson RJ, Zlatkin-Troitschanskaia O, Mariño JP, 2018, 'Performance indicators of learning in higher education institutions: An overview of the field', in Hazelkorn E, Coates H, McCormick AC, (Eds), 'Research handbook on quality, performance and accountability in higher education', pp249-263. Cheltenham: Edward Elgar.

³⁰ Crawford I, Wang Z, 2015, 'The impact of individual factors on the academic attainment of Chinese and UK students in higher education'. *Studies in Higher Education* 40 (5), pp902–920.

³¹ Rogaten J, Rienties B 2018, 'Which first-year students are making most learning gains in STEM subjects?', *Higher Education Pedagogies* 3:1, pp161-172.

³² Pampaka M, Swain D, Jones S, Williams J, Edwards M, Wo L, 2018, 'Validating constructs of learners' academic self-efficacy for measuring learning gain', *Higher Education Pedagogies* 3:1, pp118-144.

³³ Ashwin P, 2017, 'Making sense of the Teaching Excellence Framework (TEF) results' Centre for Global Higher Education policy briefings, Lancaster. Retrieved from <http://www.researchcghe.org/publications/making-sense-of-the-teaching-excellence-framework-tef-results/>

³⁴ Yorke M, 2007, 'Grading student achievement in higher education: Signals and shortcomings', *Key Issues in Higher Education*. New York: Routledge.

- 2.19 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, DLHE; employment data) for outcomes measures. Like concerns about institutions’ ability to lower entry qualifications, institutions can also raise final attainment scores,³⁵ so caution is needed in how they are used in high-stakes learning gain measures to prevent grade inflation.
- 2.20 Although often used as an umbrella term for a wide variety of outcomes including cognitive gain, critical thinking and problem-solving skills, ‘employability’ is also sometimes used very narrowly in terms of skills required to get a job, such as CV-writing or job interview skills. The predominance of ‘employability’ and career-oriented terms being used as an umbrella term for learning gain outcomes runs the danger of masking the breadth of outcomes from higher education. The employability focus of the projects evaluated here was noted in contrast to other large-scale learning gain initiatives³⁶.
- 2.21 There is also conflation in the sector of ‘employability’ and ‘employment’ leading to a view that the learning gain agenda is focused solely on jobs³⁷. More exploration, nuance and communication of what is covered by the different outcome measures (and what is being explored under the term ‘employability’) would help broaden the understanding of what is being explored through learning gain. Learning gain data can be linked with grades and entry and exit measures to conduct longitudinal and cross-sectional value-added models of learning gain. However, there are considerable variations in how value-added is measured³⁸.

Recommendation for developing measures of learning gain

Use of entry and outcome measures

Outcome measures need to clearly specify and communicate what is being explored through measuring learning gain. Entry and outcome measures need to be used contextually and carefully to support capturing learning gain measures and not as targets themselves.

³⁵ Brown R, 2018, ‘Changing Patterns of Accountability in the UK: from QA to TEF’, in Hazelkorn E, Coates H, McCormick AC (Eds), ‘Research handbook on quality, performance and accountability in higher education’, Cheltenham: Edward Elgar.

³⁶ Measuring Employability Gain – approaches from different countries. A Joint SRHE/LEGACY Seminar, 21 November 2017, Society for Research into Higher Education, London.

³⁷ Behle H, 2017, LEGACY ‘Exploring Learning Gain in Higher Education’ Seminar. 8 February ??WHAT YEAR??. University of Nottingham.

³⁸ Kim H, Lalancette D, 2013, ‘Literature review on the value-added measurement in higher education’. OECD. Retrieved from <http://www.oecd.org/edu/skills-beyond-school/Litterature%20Review%20VAM.pdf>

Robustness and effectiveness

- 3.1 This chapter summarises the main themes that emerged across the projects regarding robustness and effectiveness. Validity and reliability of instruments and measures are essential to ensure the effectiveness and appropriateness of measures of learning gain. Each project explored the technical aspects of validity and reliability in relation to their study design and instruments used (a summary analysis is available in Table 1, page 13). Further detail from several projects are available in articles in a special issue of *Higher Education Pedagogies*³⁹. Detail on evaluating validity and reliability is available in Appendix 4.
- 3.2 Most projects used existing instruments which have been previously tested and trialled. As part of testing and developing instruments, following 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.
- 3.3 Detail on the robustness and effectiveness of measures aims to support future research into learning gain⁴⁰. Additionally, contextualisation and analysis of findings, taking into consideration the purpose and level of use, is a fundamental part of establishing validity and reliability. This evaluation focuses on the remit of the pilot projects: *to test and evaluate the different methodologies*, and therefore does not report on findings.
- Project design and set-up*
- 3.4 The timing of contracts being confirmed and the funding transferred meant most projects got off to a delayed start on their project plan. This resulted in lower than anticipated Year 1 data collection, including several projects moving a planned second Year 1 data collection point to Year 2, or adding additional cohorts to complement longitudinal studies.
- 3.5 Collecting baseline data, before students commence their studies, allowed measures to accurately capture the effect of education on students⁴¹. Only one project, The Careers Group, was able to collect new baseline data on incoming students, through asking questions at the time of enrolment before students had begun classes. This was possible because some institutions had been piloting the initiative before the learning gain project started, but this was only done at half of their partner institutions.
- 3.6 Due to challenges in setting up projects, four of the longitudinal projects were only able to collect two years of data (out of a planned three): The Manchester College, University of East London, University of Portsmouth and University of Manchester. These projects were not able to deliver a final validated instrument at the time of

³⁹ <https://www.tandfonline.com/toc/rhep20/3/V1>

⁴⁰ Dawson P, Dawson SL, 2016, 'Sharing successes and hiding failures: "Reporting bias" in learning and teaching research'. *Studies in Higher Education*, 1–12.

⁴¹ Fraenkel JR, Wallen NE, Hyun HH, 2011, 'How to design and evaluate research in education', New York: McGraw-Hill Humanities/Social Sciences/Languages.

evaluation or link findings between new data from tests and surveys with existing data on final degree attainment. Therefore, it is not possible to fully evaluate the validity and robustness of their methodology. The University of Portsmouth plans to continue to run its project internally to full completion.

- 3.7 Several projects are continuing to analyse existing data and new cohorts of longitudinal and cross-sectional data to test the validity of their instruments. There are also efforts to conduct more cross-disciplinary and cross-institutional comparisons as well as analyses across socio-demographic factors. However, across projects, only very small numbers of students were tracked longitudinally, due to challenges of student engagement and retention detailed below.

Recommendation for developing measures of learning gain

Project development

Measuring learning gain within institutions is complex and takes extensive staff time to set up and run. It can take five years to run a longitudinal project, with six months to a year for set-up, three years to run, another six months until attainment data is available plus time for analysis. This time-frame and investment in resources needs to be built into any future projects.

Student engagement

- 3.8 Student participation. Collecting new data for measuring learning gain, through tests or surveys, is premised on the ability to get students to meaningfully engage with the instruments. The feasibility of getting students engaged was the greatest concern that emerged from the projects. There were significant challenges for all projects collecting new data (see Appendix 3).
- 3.9 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 impacted on testing whole instruments, and across most projects there was not sufficient engagement for validating disciplinary breadth, across student characteristics or longitudinal analysis. The projects from University of Reading and University of Lincoln were not able to get sufficient student engagement to have enough data to link new and existing data and test the conceptual models proposed in their projects. However, they were able to test subsets of approaches and instruments.
- 3.10 Additionally, projects faced challenges in retaining students to test instruments longitudinally. Projects that tried to capture students at multiple points in time struggled with this, such as the University of East Anglia with its concept inventories. It was also challenging to get individual students to complete several different activities, as with the University of Lincoln. Projects with lengthy surveys or long tests also found that students did not fully complete them, such as the University of Portsmouth and Plymouth University.
- 3.11 Given the relatively low response rates across most projects, some as low as 5 per cent, and small sample size the longitudinal samples are not representative in terms of student characteristics or in levels of motivation and engagement. For example, several projects indicated that mainly highly engaged students have responded and that response rates of 10-15 per cent are not representative of student populations,

which can limit the reliability of any comparative analysis⁴². Similarly, within institutions, often schools or subjects with highly engaged staff were selected for pilot target areas, which may not be representative of the institution. Several projects are continuing analysis of their sample population to be able to fully judge the representativeness, including the University of Cambridge strand and the University of Portsmouth.

- 3.12 A key factor that emerged in engaging students is their perception in relation to those delivering the test; when delivered by a team located centrally within the university the ability to recruit is limited. Projects had more success when they went directly into classrooms and engaged with teaching staff, such as at the Manchester College. Participation in a research project is considerably less attractive to students than being offered the opportunity of a developmental learning resource. Engagement was higher when the aims of the measurement were made explicit and relevant to students. A summary of approaches to the challenge of student engagement with learning gain are detailed in Appendix 5; further analysis on the challenge of student engagement was undertaken by the University of Lincoln⁴³.
- 3.13 Methodologically, there is a need to understand better how to engage with students. Stakeholders suggested ways to overcome the lack of student engagement by utilising existing data such as those gained in assessments or by relating learning gain to supporting students in their course of study, leading to the conclusion that embedding measures in curriculum design is the most effective approach for collecting data for measuring learning gain.
- 3.14 Student involvement and understanding. Projects found it very useful to work with students in conceptual and explanatory workshops and focus groups, which allowed for testing and understanding student responses. Students were involved in developing metrics, approaches and making sense of the data in various ways across all of the projects⁴⁴. The Careers Group worked with students to test their methodology and refine their categories of career readiness. The Open University worked with students who gave their perception of their attainment data trajectories.
- 3.15 Focus groups highlighted that students began to appreciate the value of wider curricular activities only when they considered them within a self-evaluation environment, such as that provided in focus groups or tutorial sessions⁴⁵. The University of Manchester included as part of its instrument the degree to which learning experiences are perceived to be helpful for academic learning; employment;

⁴² Bennett R, Kane S, 2014, 'Students' interpretations of the meanings of questionnaire items in the National Student Survey', *Quality in Higher Education*, 20(2), pp129–164.

⁴³ Speight L, Crawford K Haddelsey S, 2018, 'Towards measures of longitudinal learning gain in UK higher education: the challenge of meaningful engagement', *Higher Education Pedagogies*, 3:1, pp196-218.

⁴⁴ Rosenfeld M, Rosenfeld S, 2011, 'Illustrating a complementary paradigm for styles research: From a third-person to a second-person perspective'. in S Rayner, Cools E, (Eds.), 'Style differences in cognition, learning, and management: Theory, research and practice', pp143–159, New York: Routledge.

⁴⁵ Kandiko Howson CB, 2018, 'Closing attainment gaps through personal tutoring: Putting learning gain data to use' (0360) Society for Research into Higher Education Annual Conference. 5-7 December ??WHAT YEAR??. Celtic Manor, Newport, Wales.

critical thinking; and life. This trend serves to underpin the contention that students begin to appreciate their self-development when encouraged to self-evaluate⁴⁶.

- 3.16 Projects concluded that for measures to be meaningful to students they should have opportunities to be centrally involved in research design, data collection and analysis, and be debriefed on the outcomes of any data gathering exercise; these activities should also not be solely the preserve of final-year students. Students acting as partners⁴⁷ in designing measures helps to gain student buy-in, makes the measures meaningful for students and validates approaches. Meaningful student engagement is needed to judge the effectiveness of measures of learning gain.

Recommendation for developing measures of learning gain

Relevance for students

Students should be partners in developing learning gain measures. Measures of learning gain should benefit students' learning, self-reflection and development planning in order to support engagement. Measures of learning gain should be timetabled and embedded in the curriculum so that it is clear to students and staff that it forms a valued part of their education.

Non-linear learning

- 3.17 Research across the projects reinforced that students do make progress in their knowledge and understanding. However, they can also have set-backs and critical reflective moments of 'learning how much you do not know'. As noted from analysis across the University of Warwick project, accepting the pedagogic staple that learning is non-linear is essential. Otherwise measurements at particular points in time will not accurately reflect a student's final level of development. For example, they identified that a student's 'career development' may change over time and drop as a student develops greater self-awareness.
- 3.18 Rather than thinking of a linear path of learning and 'distance travelled' attainment, a more accurate approach to measuring learning gain should focus on the complex ways in which students develop while at university. The University of Manchester project suggested that learning gain be reconceptualised and described as a multi-dimensional vector. Similarly, the University of East London noted the need for multiple measures, and pathways, to account for the reality of students' learning (and lives).
- 3.19 Data from several measures, including grades, self-assessments and the R2 Strengths employability tool, initially found indications of 'learning loss'. Projects explored if this is due to challenges of the reliability of the measure or whether this is a true measure of students' experiences. Across analysis of existing data, multiple tests, surveys and qualitative approaches, projects found that students indeed have

⁴⁶ Hattie J, 2014, 'Self-concept'. Psychology Press.

⁴⁷ Healey M, Flint A, Harrington K, 2014, 'Engagement through partnership: Students as partners in learning and teaching in higher education'. York: Higher Education Academy. Retrieved from https://www.heacademy.ac.uk/system/files/resources/engagement_through_partnership.pdf

non-linear learning patterns, confirming existing research^{48,49,50}. Analysis of grades data also noted a non-linear trend across cohorts of students.

- 3.20 This is particularly common in the first year as students adjust to new learning environments. The University of Warwick strand, through trying to use simple five-point scales to capture gain, found that learning is neither straightforward nor linear but more complex, requiring reflexivity. The qualitative interviews produced far richer evidence of non-linear learning gain, though fewer students were captured.
- 3.21 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. From the projects focusing on employability, the University of Nottingham strand of the University of Warwick project found the 'concern' factor, 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. Furthermore, The Open University and the University of East London found learning to be multifaceted, finding that 'loss' in one measure, for example confidence, may be off-set by gain in another, such as cognitive gain.
- 3.22 This suggests that a re-conceptualisation of the term 'gain' is necessary to consider the widest definition of learning as relevant to all students in all contexts. This is particularly pertinent to measurement over time, considering the frequency and timing of measurement and whether measurement is able to capture any unintended learning gain which may occur.

Recommendation for developing measures of learning gain

Multiple points in time

Student learning occurs across many dimensions and varies both over time and in direction and needs to be accounted for throughout a student's course experience at multiple points in time.

Subject-specific or generic measures

- 3.23 A key factor in selecting measures of learning gain in higher education is whether they are subject-specific, focusing on the outcomes of a particular field of study, or generic, such as problem-solving⁵¹. Subject-specific metrics are more closely aligned to students' programmes of study and better map to the structure of English higher education⁵², but can be costly and burdensome to develop. Generic measures are

⁴⁸ LeCun Y, Bengio Y, Hinton G, 2015, 'Deep learning', *Nature*, 521(7553), p436.

⁴⁹ Haggis T, 2004, 'Meaning, identity and "motivation": expanding what matters in understanding learning in higher education', *Studies in Higher Education*, 29(3), pp335-352.

⁵⁰ Fraser SW, Greenhalgh T, 2001, 'Complexity science: Coping with complexity: Educating for capability'. *BMJ: British Medical Journal*, 323(7316), p799.

⁵¹ Shavelson RJ, Zlatkin-Troitschanskaia O, Mariño JP 2018, 'Performance indicators of learning in higher education institutions: An overview of the field', in Hazelkorn E, Coates H, McCormick AC (Eds), 'Research handbook on quality, performance and accountability in higher education', pp 249-263, Cheltenham: Edward Elgar.

⁵² Pollard E, Williams M, Williams J, Bertram C, Buzzeo J, Drever E, Coutinho S, 2013, 'How should we measure higher education? A fundamental review of the performance indicators', Synthesis Report, London: HEFCE. Retrieved from

easier to develop and use across the sector, although subject-level differences are still evident; the University of Cambridge strand of the University of Warwick project which developed a non-subject specific instrument, recommended using it at the subject-level. Similarly, in a large-scale learning gain assessment in Brazil, much greater gains were found on subject-specific tests than general knowledge tests⁵³.

- 3.24 Significant subject-level differences were noted across projects both in terms of how students responded and what is valued or considered important (which affects the degree to which students engage). The Birmingham City University project found that the Collegiate Learning Assessment (CLA+) instrument and its value as a standardised measurement of critical thinking is more attractive to certain disciplines, or perhaps to key academics running programmes within those disciplines. Focus groups with students led to questions about the discipline-independence of the test. Students perceive a benefit from studying certain disciplines but there was little evidence of this from the data collected. This perception matters because it influences how students engage with the instrument and how they interpret the results.
- 3.25 Across projects, there are more differences in outcomes between subjects than institutions. As the University of Cambridge strand of the University of Warwick project noted, learning gain appeared most relevant at the subject level. Context was paramount and included the importance of: measurement which was integrated into the discipline of study; measurement which utilised prior relationships at local levels; and detailed knowledge and practice within service provision to ensure applicability and impact. These factors highlight the bounds of scalability and the limitations of a one-size-fits-all approach which should be balanced against considerations of comparability and consistency.
- 3.26 In their in-depth exploration of learning research methods, 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.
- 3.27 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 when testing in a different discipline where the instrument was not linked to learning objectives. This raises questions about the effectiveness of cognitive tests outside of the curriculum.
- 3.28 Summarising data from focus groups across the projects, students viewed content knowledge and subject-adjacent skills, abilities and competencies as inextricably

<https://webarchive.nationalarchives.gov.uk/20180322111239/http://www.hefce.ac.uk/pubs/rereports/year/2013/ukpireview/>

⁵³ Dalmon DL, Fonseca I, Avena CP et al. High Educ?? 2019. <https://doi.org/10.1007/s10734-019-00376-6>

linked. They viewed much of their development of these latter skills, abilities and competencies as implicit (i.e. learning gain that is not subject-specific occurs alongside the development of content knowledge), and recognised that any explicit focus on aspects of learning gain (e.g. through courses, sessions, workshops, etc. that seek to develop a certain skill) would be supported by a clear link to their respective academic subject.

Conclusion

Contextual factors such as subject-level differences, institutional type and student characteristics differences impact the transferability of measures of learning gain. These differences should be considered when designing and selecting learning gain measures, and when analysing and presenting findings, and mediating effects need to be considered.

Standardised tests

- 3.29 Drawing on the point above, national standardised tests, such as the CLA+, are generic measures of learning gain. They do not capture the outcomes from disciplinary specialisation found in England as they do not cover subject-specific knowledge⁵⁴; this was raised as a concern by academics from Birmingham City University and further noted by the University of Reading. Pilots of the CLA+ found it to be bureaucratic and expensive in terms of licensing costs and the training and paying of markers for the tests, and projects struggled with buy-in across institutions. 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, across all year groups and with a large effect size, 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 other measures of academic achievement; some of these are more readily available, cheaper to obtain and do not involve the administrative burden of student recruitment and testing. They also found correlations with student socio-economic and demographic characteristics.
- 3.30 The location of variance is another issue affecting the validity and reliability of standardised 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 individual student-level decisions⁵⁵. This raises questions about using instruments that may provide useful comparative information across courses or institutions but have less benefit for individual students.

⁵⁴ Coates H, Richardson S, 2012, 'An international assessment of bachelor degree graduates' learning outcomes', *Higher Education Management and Policy*, 23(3), pp1-19.

⁵⁵ Aloisi C Callaghan A, 2018, 'Threats to the validity of the Collegiate Learning Assessment (CLA+) as a measure of critical thinking skills and implications for Learning Gain', *Higher Education Pedagogies*, 3:1, pp57-82.

- 3.31 The University of Reading concluded that on an individual level, the longitudinal reliability is too low to detect changes in student performance consistently. One could administer the CLA+ several times and look at the overall trend, but this would be costly. The same trade-off would apply to any measure that is valid for course or institutional comparison but not at the individual student level.
- 3.32 However, there have been updates to the CLA+ test and scoring since the commencement of the learning gain projects which impact on the recommendations and feasibility of administering the test. Developments are underway to automate assessment and de-couple the different elements of the test. While this makes it quicker to administer, it loses personalisation and expert judgement. The CLA+ does have potential as a diagnostic tool⁵⁶, as suggested by Birmingham City University.
- 3.33 The Situational Judgement Test (SJT) run by Assessment and Development Consultants (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 diagnostically.

Conclusion

Pilots of standardised tests from the projects have not proven to be robust and effective measures of learning gain due to challenges of student engagement, differential scores across socio-demographic characteristics, subject differences and use of data.

Surveys using self-assessment

- 3.34 Several projects were based on students self-assessing the development of their knowledge, skills and attributes through surveys. The UK Engagement Survey (UKES) provides benchmarked data and has a well-developed evidence-base for enhancing teaching and promoting student learning⁵⁷ and is already used as a proxy measure for learning gain⁵⁸. It was used fully, partially or in an amended form by six projects. Regarding non-linearity in learning, the University of Lincoln noted the challenges of administering a student self-assessment (SSA), because the instrument relies on self-reported data and there were inconsistencies in how students rated themselves in each area, which echoes findings from wider research⁵⁹. The University of Lincoln concluded that self-assessment is a skill that students develop, and first-year students may not yet be able to reliably self-assess their learning. This is an important consideration on the reliability of measures based on tracking self-assessment data longitudinally.

⁵⁶ Dill D, 2018, 'Can Public Policy Promote Academic Quality? An Assessment of Policy Instruments for Instruction and Research', in Hazelkorn E, Coates H, McCormick AC (Eds), Research handbook on quality, performance and accountability in higher education, pp42-55, Cheltenham: Edward Elgar.

⁵⁷ 'Engaging with Engagement Data: Education, Enhancement and Excellence'. Student Experience Network Seminar. 28 June 2016, Society for Research into Higher Education, London.

⁵⁸ Neves J, Stoakes G, 2018, 'UKES, learning gain and how students spent their time'. Higher Education Pedagogies, 3(1), pp219-221.

⁵⁹ Porter S, 2013, 'Self-reported learning gains: A theory and test of college student survey response'. Research in Higher Education, 54, pp201-226.

- 3.35 Further, they found that it was hard to capture meaningful career data for first-year students. Similarly, the Careers Group found greater change between Year 2 to Year 3 students than between Years 1 and 2. Additionally, first-year students have limited and sporadic engagement with career service units⁶⁰. This, alongside findings of non-linear trajectories in the employability projects, raises questions about how accurately students are able to self-assess their career readiness when they start higher education. However, the projects did not identify better ways to capture students' affective and behavioural measures of learning gain.
- 3.36 The University of Lincoln found it was difficult to rate and usefully quantify engagement in co-curricular activities and challenging to quantify broad types of engagement data and make that data usable. 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.

Conclusion

Self-assessment

Measures based on longitudinal surveys using self-assessment need to be used with caution to account for students' reflective development.

Using existing grades and attainment data

- 3.37 The evaluation of attainment patterns focuses on the use of grades trajectories (change over time), not the quality of grades as status indicators (e.g. final mark), although using measures based on existing attainment data are only as valid as the assessments they are based on⁶¹. Projects analysing institution-level attainment data found large grading profiles and varying trajectories across modules, courses and institutions. Several projects utilised multi-level modelling statistical procedures for analysing grades data to take into account the clustered nature of the data. Projects identified the need to explore grading procedures, particularly across modules, subjects and institutions⁶².
- 3.38 Individual and module-level performance. There are validity challenges to using grades data, as change depends on individual performance. Low-performing first-year students will have more opportunity to show change than high-performing first-year students; this is also relevant to cohorts of students. Related 'ceiling effects' were noted in the University of Reading pilot of the CLA+, where very high-performing first-year students had no opportunity to improve their scores over time.

⁶⁰ AGCAS 2015, 'First-year student careers and employability survey report'. Retrieved from http://www.agcas.org.uk/agcas_resources/823-First-Year-Student-Careers-and-Employability-Survey-Report

⁶¹ Shavelson RJ, Zlatkin-Troitschanskaia O, Mariño JP, 2018, 'Performance indicators of learning in higher education institutions: An overview of the field', in Hazelkorn E, Coates H, McCormick AC (Eds), Research handbook on quality, performance and accountability in higher education, pp249-263, Cheltenham: Edward Elgar.

⁶² Pokorny H, 2016, 'Assessment for learning' in Pokorny H, Warren D, (Eds.), 'Enhancing teaching practice in Higher Education' pp69–90, London: Sage.

Further, it also makes such measures easily 'gameable' as students can be encouraged to perform poorly at the start to allow for more for more apparent improvement.

- 3.39 Data analysis indicates there is substantial variation in module grades, which could be explained by variations in marking within a module. The variability of assessment in higher education makes comparisons between students and institutions unreliable measures and raises questions about the use of grades as the sole outcome measure of higher education⁶³ and suggests the need for multiple measures.
- 3.40 Subject-level differences. Grades are validated through internal and external quality assurance procedures at each institution, but projects have identified how 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.
- 3.41 The University of East Anglia analysed the use of grades in their project, concluding that the nature of subjects presented 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.
- 3.42 Non-linearity. Projects found patterns that show grades may decrease over time – which was not seen to be negative learning gain, or students unlearning, but rather students receiving lower marks as they progress. This was more common in some subjects, particularly hard sciences. The University of Reading identified three main trajectories of grades: positive, negative and flat. However, across multiple projects there was a lack of consensus of 'ideal' grade trajectories. This adds to findings about non-linearity of student learning and the limitations of only using two time points.
- 3.43 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, calling into question the validity of longitudinal grade comparisons.
- 3.44 Student perceptions. The Open University concluded measuring learning gains using secondary attainment data is currently not reliable and that results from the secondary data analysis did not always overlap with the lived experiences from the students. They found that module satisfaction as a proxy for affective learning gain was not an effective measure and that online engagement, drawing on the Virtual

⁶³ Bloxham S, Boyd P, 2012, 'Developing effective assessment in higher education: A practical guide', Maidenhead, Berkshire: Open University Press.

Learning Environment (VLE), was not a useful proxy for behavioural learning gain measures as its usage led to inconsistent findings. They found it of more use for cognitive gains but it is dependent on course teaching and marking, and they were disciplinarily and institutionally influenced and determined. The Oxford Brookes University strand of that project found that students did not view their gains in terms of assessment performance, which highlights the need to go beyond grades in assessing student learning.

Suitability: Instrument uses

- 4.1 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 recommended using measures for cross-institutional accountability purposes, but measures may be more suitable for cross-institutional comparisons at the subject-level. Very few projects explore institutional comparisons, as it was contractually against some of the partnership agreements, methodologically challenging and institutionally sensitive due to reputational effects.

Use of instruments, methods and metrics

- 4.2 There was limited interest in using learning gain data from senior managers or academics, outside of those engaged in the projects, across most institutions. This has severely limited the impact of the instruments and methodologies on student outcomes. Additionally, projects that focused on developing and testing instruments and approaches have not had time to validate their methodology, collect data, put that into use and assess the impact on students. Many projects noted that interest and use could grow, depending on how learning gain metrics fit into the evolving higher education regulatory environment.
- 4.3 Accountability, teaching enhancement and student learning have been identified as key drivers for learning gain⁶⁴. The projects concluded that the data flow should be bottom up, exploring what is useful and relevant to students, rather than top-down with a focus on something that can be measured for comparison's sake. However, there is debate across the sector about what comes first: do you measure what is valued or do you value what is measured?⁶⁵ Evidence suggests that accountability largely drives efforts to enhance student learning⁶⁶. This highlights the interconnectedness of uses of learning gain, but also sounds a note of caution of how uses for one purpose may lead to consequences for other areas of the sector.
- 4.4 Some projects focused on developing and evaluating tools and methodologies to measure learning gain and felt the need to wait to share their data to be used for enhancement purposes until they felt they had confidence in the metrics and models, such as the University of Portsmouth. Some project teams were also more closely embedded with institutional decision-makers and operational areas of the university

⁶⁴ Varsavsky C, Matthews KE, Hodgson Y, 2014, 'Perceptions of science graduating students on their learning gain', *International Journal of Science Education*, 36 (6), pp929–951.

⁶⁵ LEGACY National Learning Gain Conference, 28 September 2017, Warwick Business School. The Shard, London.

⁶⁶ Stensaker B, Sweetman R, 2014, 'Impact of assessment initiatives on quality assurance', *Higher Education Learning Outcomes Assessment*, pp237-259.

and had more opportunities to put data to use in practice, as seen at Plymouth University. Being able to influence institutional policy also depends on whether there is a sense of senior management buy-in to the projects, as shown at the University of East Anglia. However, there is a small but growing body of research on how measures of learning gain are being used to help support students and improve teaching and learning,⁶⁷ drawn into three categories.

Enhancing teaching and learning

- 4.5 Student-level use. Several projects took an approach of providing students with data to be used for personal development. However, some, including Birmingham City University, found very few students chose to collect or use their results. They found it useful to incorporate the test scores as a diagnostic tool, providing baseline data to help understand student starting points. This type of data may be particularly useful at institutions with diverse student in-takes with a variety of prior qualifications. At Birmingham City University the main interest in the CLA+ for staff and students was students being able to analyse their position in relation to the skills being tested, thus informing their ongoing development activities.
- 4.6 Personalised reports for participating students were adopted as part of the student recruitment strategy for the projects at the University of Portsmouth and the University of Lincoln, although with little success. Through the employability strands in the University of Warwick project, feedback was provided to students and for 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. Plymouth University suggested their data 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. The University of East London found the weekly time allocation and personal development subscales of UKES worked well as synoptic measures for end-of-year evaluation and adaptive provision.
- 4.7 Several projects initially planned to feed learning gain measures to personal tutors to help make use of the data, as students need support, advice and guidance to understand what the learning gain data means for them and how to use it to improve their learning^{68,69}. However, they found personal tutors were usually not successful in doing this as they did not understand the data or what they were supposed to do with it. This was particularly the case for centrally run projects that were not embedded in courses; projects suggested linking the learning gain measures into the curriculum could help support student and staff engagement.
- 4.8 There were also notes of caution about the ethics of simply providing students with data. Some measures are deliberately meant to be low for incoming students (to

⁶⁷ See Higher Education Pedagogies special issue for more examples:
<https://www.tandfonline.com/toc/rhep20/3/V1>

⁶⁸ Kandiko Howson CB, 2016, 'Excellence in Tutoring: What do we know, what do we need to know?' Third National Tutoring Seminar. 26 October 2016, University of Central Lancashire.

⁶⁹ Kandiko Howson CB, 2018, 'Learning gain metrics and personal tutoring: Opportunities and ethics', Society for Research into Higher Education (SRHE) seminar 'Getting Personal: Conceptualising and Understanding Personal Tutoring'. 21 June 2018. SRHE, London, UK.

provide opportunities to show gain) but this may be discouraging for first-year students, even if it is normal.

- 4.9 Classroom-level use. The project at the University of East Anglia trialled and tested 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 online 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.
- 4.10 Learning gain measures can support a research-informed approach to pedagogy⁷⁰, but it can be challenging to raise awareness and build support across frontline teaching staff. It is also essential that the measures relate to module content or spurious results can be found⁷¹. There are also notes of caution to be sounded regarding students’ rights and the development of learning gain measures, and their ability to choose how they want to engage in their higher education experience⁷².
- 4.11 Course and subject-level use. Data from the projects has the potential to provide tailored information to inform specific programme design. 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 was developed to provide guidance on the most effective use of the toolkit resources. However, as Plymouth University found, the project needs a sufficient level of response 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.
- 4.12 The University of Manchester found that some schools have used 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. The priority given to the National Student Survey (NSS) within institutions had a significant impact on projects getting buy-in from their institutions, particularly to run additional surveys such as UKES.
- 4.13 The University of Cambridge strand of the University of Warwick project concluded that it is at the within-university subject level that the measurement instrument is best

⁷⁰ Evans C, Waring M, Christoudolou A, 2017, ‘Building teachers’ research literacy: Integrating practice and research’, *Research Papers in Education*, 32(4), pp303–423.

⁷¹ Asikainen H, Gijbels D, 2017, ‘Do students develop towards more deep approaches to learning during studies? A systematic review on the development of students’ deep and surface approaches to learning in higher education’, *Educational Psychology Review*, 29, pp205–234.

⁷² Macfarlane B, 2016, ‘The performative turn in the assessment of student learning: A rights perspective’. *Teaching in Higher Education*, 21 (7), pp839–853.

used. Uses of the measurement instrument therefore appear to be related to its function as a diagnostic tool at the subject-stage interface within each institution, with potential for teaching, structure and other such aspects of a course to be addressed in response to repeated measures taken with the instrument. However, the Birmingham City University project noted that at course level, learning gain measures can seem redundant: there are already other skills analyses and personal development planning aspects built into most courses and so the CLA+ was seen as an addition (by tutors too) to things that were already in place. Also, most students said they would not want their scores to be shared with their course team of teachers, which echoed concerns students had across a number of projects about how their scores could influence teachers' perceptions of them. However, students also stated that they would want to know how their scores compared with those of other students who had undertaken the test for peer-comparison.

- 4.14 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 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. 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. The University of East London made survey data available to academic staff online through an online app. They worked with directors of teaching and learning and programme leaders to identify how this could inform teaching and learning. This highlights that when the purpose of a measure is not clear users can struggle with what to do with the resulting data.

Conclusion

Students and academic staff need support, advice and guidance to use learning gain data to enhance teaching and learning. Subject-level data on cognitive gains from attainment data, soft skills development from surveys and employability data can support students' reflective learning and development and aid staff to benchmark progress and enhance the curriculum to drive increased learning gain.

Employability and transition into work and further study

- 4.15 Institution-level (service use). Several projects provide data that helps support careers offices and skills and training units to tailor services and target specific students. The employability strands of the University of Warwick project offer feedback to participants, careers support and developmental online questionnaires. The University of Warwick strand has embedded staff development workshops on using the tools and data. The University of Nottingham strand 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.

- 4.16 The University of Lincoln and the Careers Group propose using 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. Learning gain data can help professional services staff ‘get to know’ students they may not regularly engage with but need to support. At the University of Lincoln, the Get Set pre-arrival survey, added as an additional tool to their project, has highlighted the ability to provide tutors and professional services with additional student-generated data in order to enable them to better support individual students and to help inform the provision of appropriate skills-development opportunities. Professional service staff and senior tutors who received the data from Get Set reported that it was “really useful”.
- 4.17 Although there were noted challenges in using the data to impact on institutional practices, the Careers Group have found examples that some 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.
- 4.18 The Careers Registration methodology is being used at over 60 UK institutions, but the work of the employability-focused projects was mostly isolated to use within careers service units and not part of teaching and learning activities. Some extended analysis (e.g. with Destinations of Leavers from Higher Education (DLHE) data) with potential institutional impact was affected by a switch to the Graduate Outcomes record⁷³. However, the high voluntary take-up of this approach signals institutional commitment and the usefulness of the approach.
- 4.19 Alumni engagement. 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.

Conclusion

Employability data can prompt students to reflect on their development and can be used by careers staff to support students. It can enhance teaching and learning if shared with teaching staff and embedded into the curriculum. When linked with institutional data it can provide information on differential patterns across students and linkages with employability.

Quality, accountability and performance

- 4.20 Regulatory use. Although separate strands of activity, the learning gain pilot projects coincided with the development of a new quality mechanism, the Teaching

⁷³ <https://www.graduateoutcomes.ac.uk/>

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’⁷⁴. There have been debates that 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^{75,76,77}. However, no single metric emerged that could be used as an institution-level core metric in the TEF.

- 4.21 Widening Participation and social mobility agendas feature in debates on measuring learning gain⁷⁸. In the US, learning gain data is widely used to promote equity, through disaggregating outcomes by student characteristics⁷⁹. Learning gain data could also support policy aims in England through Access and Participation Plans⁸⁰. Projects used learning gain data to explore differences between students at module level (The Open University); across Widening Participation characteristics (University of East London); across years of study (University of Reading); and institutionally (University of Manchester). Projects found they could identify differences in access, success and progression of underrepresented groups of students, but the data did not offer a solution. However, they could identify where gaps start or grow, and also identify where there are no gaps. Learning gain data could be part of a suite of activity to eliminate differential attainment.
- 4.22 From the projects focusing on soft skill development, findings from the University of Manchester indicate associations between measures, group differences, and the impact of confidence measures in models of learning gain. Their research has shown that students with different background characteristics bring with them different academic dispositions. These dispositions can sometimes be a key predictor of learning gain and must therefore be taken into account alongside attainment indicators. 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.
- 4.23 Institution-level (strategic) use. Being able to influence institutional policy in part is related to how embedded the projects are and whether there is a sense of senior

⁷⁴ <https://www.officeforstudents.org.uk/advice-and-guidance/teaching/what-is-the-tef/>

⁷⁵ ‘Teaching excellence – what works, and how can we make it happen?’ Wonkfest 2017 conference. 7 November 2017. Ravensbourne, London, UK.

⁷⁶ LEGACY National Learning Gain Conference, 28 September 2017, Warwick Business School, The Shard, London.

⁷⁷ Inaugural meeting of the Spinnaker Group, 13 July 2017, University of Portsmouth.

⁷⁸ Mountford Zimdars A, Sabri D, Moore J, Sanders J, Jones S, Higham L, 2015, ‘Causes of differences in student outcomes’. London, UK: HEFCE. Report to HEFCE by King’s College London, ARC Network and the University of Manchester. Retrieved from <https://webarchive.nationalarchives.gov.uk/20180322111249/http://www.hefce.ac.uk/pubs/rereports/year/2015/diffout/>

⁷⁹ Jankowski NS, Timmer JD, Kinzie J, Kuh GD, 2018, January, ‘Assessment that matters: Trending toward practices that document authentic student learning’. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. Retrieved from <http://www.learningoutcomesassessment.org/documents/NILOA2018SurveyReport.pdf>

⁸⁰ <https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/access-and-participation-plans/>

management buy-in to the data⁸¹. 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 allowed staff and senior managers to compare attainment across programmes and between student groups within programmes. This 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. Outcomes have prompted managerial discussions at The Manchester College on future changes based around such findings. As part of researching learning gain, the University of East Anglia rolled out Grade Point Average (GPA) for all students.

- 4.24 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 embedded running UKES, 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 conducted a university-wide review of a strategy for enhanced learning environments, from pre-entry through to employment. Learning gain research has been complemented by projects on employability, placement research and widening participation.
- 4.25 Several projects have plans for learning gain data to inform strategic decision making. The University of Cambridge strand of the University of Warwick project developed benchmarked institutional reports, but it is not yet clear how these would be used in practice. The University of Portsmouth used some of their tools as a method of demonstrating improvement in the recently revised Hallmarks of a Portsmouth Graduate project.
- 4.26 Both the University of Portsmouth and Birmingham City University linked learning gain work with related HEFCE/OfS-funded Catalyst bid projects around retention and Black and minority ethnic (BME) attainment⁸². 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 or no learning gain. Data from Plymouth University project could provide a standard measure of learning gain at a programme level as a measure of student engagement with research methods and associated learning.
- 4.27 Staff involved in projects have been able to share their experiences of trialling measures of learning gain and how this may support other institutional activities. The University of East London found the expertise created by the project has informed

⁸¹ Seifert TA, Gillig B, Hanson JM, Pascarella ET, Blaich CF, 2014, 'The conditional nature of high impact/good practices on student learning outcomes', *The Journal of Higher Education*, 85(4), pp531–564.

⁸²

<https://webarchive.nationalarchives.gov.uk/20180319120251/http://www.hefce.ac.uk/funding/catalyst/projects/>

their TEF Readiness and Learner Analytics working group and evaluation processes within schools.

- 4.28 Sector use. The potential inclusion of learning gain measures in a national accountability system has raised interest and concerns about the pilot projects, echoed across multiple national higher education systems⁸³. There are arguments that it is too difficult to measure the complexity of student learning, but these are increasingly dismissed by those inside and outside the higher education sector by noting the need for multiple and diverse measures⁸⁴. However, although it may be challenging, the learning gain pilot projects focus on exploring measures that better capture student learning outcomes than existing measures.
- 4.29 Measures of learning gain have the potential to contribute to a virtuous cycle, through holding institutions accountable while activities undertaken to raise outcomes would lead to improvements in teaching and learning and the student experience⁸⁵. This was a key outcome⁸⁶ from the Wabash National Study (2006-2012) led by the Center of Inquiry (2016)⁸⁷, which fed into the design of the OfS learning gain programme. Projects noted this is in contrast to the approach of chasing some current metrics and rankings which can lead to practices which detract from students' learning experiences⁸⁸.
- 4.30 Questions of how quality is judged, and the lack of differentiation of graduates due to grade inflation⁸⁹, 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. However, caution is needed when developing high-stakes metrics as they can lead to perverse actions and unintended consequences^{90,91}. Broad and generic metrics can lead to privileging certain types of knowledge and knowing⁹².
- 4.31 Prospective student-level use. The US has 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

⁸³ Douglass JA, Thomson G, Zhao C-M, 2012, 'The learning outcomes race: The value of self-reported gains in large research universities', *Higher Education*, 64, pp317–335.

⁸⁴ Baume D, 2018, 'Towards a measure of learning gain. A journey. With obstacles', *Higher Education Pedagogies*, 3:1, pp51-53

⁸⁵ LEGACY Final Conference, 'Learning gain: An agenda for change'. 27 September 2018, Woburn House, London.

⁸⁶ Pascarella ET, Blaich C, 2013, 'Lessons from the Wabash national study of liberal arts education'. *Change: The Magazine of Higher Learning*, 45(2), pp6–15.

⁸⁷ Center of Inquiry, 2016, 'Wabash national study 2006–2012'. Retrieved from <http://www.liberalarts.wabash.edu/study-research/>

⁸⁸ Hazelkorn E, 2015, 'Rankings and the reshaping of higher education: The battle for world-class excellence'. Springer.

⁸⁹ Higher Education Academy, 2015, 'Higher education achievement report'. York: HEA. Retrieved from <http://www.hear.ac.uk/about/history-development>

⁹⁰ Edwards MA, Roy S, 2017, 'Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition', *Environmental Engineering Science*, 34(1), pp51–61.

⁹¹ Gray D, Micheli P., Pavlov A, 2015, 'Measurement madness, recognising and avoiding the pitfalls of performance measurement', West Sussex, UK: Wiley.

⁹² Brooks P, 2012, 'Outcomes, testing, learning: What's at stake?' *Social Research*, 79, pp601–611.

Universities (APLU) and the American Association of State Colleges and Universities (AASCU) based on the premise of offering straightforward, flexible, comparable information on the undergraduate experience, including student progress and learning outcomes⁹³. The College Portrait initiative of the VSA offered information for students, families, policy-makers, academic faculty and staff, the general public and other higher education stakeholders. This provides a model for how multiple measures of student learning can be incorporated into an external-facing customisable portal.

- 4.32 Following the VSA College Portrait approach, learning gain metrics could be incorporated into future iterations of the Key Information Set, offering a customisable, searchable portal of data on student outcomes. However, evidence suggests prospective students do not know how to make good use of student learning data when making decisions about higher education^{94,95,96}. Additionally, as Arum and Roska (2011) noted in their high-profile study on learning gain, there “is no reason to expect that students and parents as consumers will prioritize undergraduate learning as an outcome”⁹⁷.

Conclusion

Quality, accountability and performance across the sector can be enhanced through using learning gain data at appropriate and relevant levels for benchmarking, comparison and differentiation.

Scalability: Pilot project analysis

- 5.1 Drawing on best practice recommendations⁹⁸, guiding principles should ensure that measures of learning gain:
- are locally relevant;
 - are externally generalisable;
 - advance transparency of validity and reliability; and
 - make efficient use of time and money.

⁹³ <http://www.voluntarysystem.org/>

⁹⁴ Bowes L, Evans J, Hughes D, Olewe-Richards S, Roberts J, Birkin G, 2015, ‘UK review of information about higher education: Report on the review of the key information set and Unistats’, (Annex B: Research on information use by students and their advisers). A report to the UK higher education funding bodies by CFE Research. Retrieved from <https://webarchive.nationalarchives.gov.uk/20180322111514/http://www.hefce.ac.uk/pubs/year/2015/201527/>

⁹⁵ Diamond A, Vorley T, Roberts J, Jones S, 2012, ‘Behavioural approaches to understanding student choice’. York: Higher Education Academy. Retrieved from https://www.heacademy.ac.uk/system/files/resources/student_choice.pdf

⁹⁶ Nuffield Foundation, 2017, ‘Improving social well-being through education, research and innovation’. Retrieved from <http://www.nuffieldfoundation.org/news/young-people%E2%80%99s-subject-choices-influences-and-impact>

⁹⁷ Arum R, Roska J, 2011, ‘Academically adrift’, Chicago, USA: University of Chicago Press, p 127.

⁹⁸ Coates H, 2016, ‘Assessing student learning outcomes internationally: Insights and frontiers’, *Assessment & Evaluation in Higher Education*, 41(5), pp662-676.

- 5.2 As noted by the University of Warwick project, learning gain measurement requires careful consideration of a) the nature of learning being measured and the associated impact; b) the importance of context and the relationship to the subject/discipline; c) difficulties in student participation which can be addressed by a sense of ownership; and d) the methodological rigour of exploratory measurement which needs to be made relevant to policy and practice.
- 5.3 Three major concerns impacting scalability emerged as the projects developed: challenges of student engagement; data ethics; and staff time, financial and opportunity costs. The pilot projects developed 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. The following points frame the evaluation of scalability.

Purpose and development

- 5.4 Through evaluation of the projects, given considerations of value for money and practicality, it was noted that because something *can* be measured does not mean it *should* be measured, particularly at a large scale. Indeed, “The greatest challenge in developing learning indicators is getting consensus on what kind of learning should be measured and for what purpose a learning indicator is to be used”⁹⁹. Broad recommendations for future development and scaling up of measures of learning gain are drawn from the outcomes of the pilot projects; sector learning gain debates and events; and lessons from international pilots and initiatives to measure learning gain and student outcomes.
- 5.5 Measures of learning gain need a clear rationale. Buy-in for scaling up any measure of learning gain depends on its intended use: for improvement or accountability; this drives the presentation and dissemination of data¹⁰⁰. Projects found student engagement in completing tests and surveys depends on the usefulness of the measure to them and their learning. Developing measures and approaches with students ensures they are relevant and result in data that is useable.
- 5.6 Projects recommended developing measures of learning gain to benefit student learning; for metrics to provide value they need to be able to produce data on what students know and can do for improvement and enhancement to ensure quality¹⁰¹.

⁹⁹ Shavelson RJ, Zlatkin-Troitschanskaia O, Mariño, JP, 2018, ‘Performance indicators of learning in higher education institutions: An overview of the field’, in Hazelkorn E, Coates H, McCormick AC (Eds), ‘Research handbook on quality, performance and accountability in higher education’, pp249-263. Cheltenham: Edward Elgar, p251.

¹⁰⁰ Lennon MC, Fukahori S, Edwards D, 2018, ‘Global Assessments of Disciplinary Learning Outcomes: What We Learned from AHELO’, in Hazelkorn E, Coates H, McCormick AC (Eds), ‘Research handbook on quality, performance and accountability in higher education’, pp264-274. Cheltenham: Edward Elgar.

¹⁰¹ Kuh GD, Jankowski NA, 2018, ‘Assuring High-Quality Learning for All Students: Lessons from the Field’, in Hazelkorn E, Coates H, McCormick AC (Eds), ‘Research handbook on quality, performance and accountability in higher education’, pp305-320. Cheltenham: Edward Elgar.

Recommendation for developing measures of learning gain

Rationale

Measures need a clear rationale for their development, use and audience. Measures should be developed in partnership with students, clearly support student learning and have a dissemination plan for clearly communicating findings with relevant audiences.

Usability and embeddedness

- 5.7 Findings from the pilot projects highlighted areas for improvement in existing sector-wide mechanisms for quality, accountability and performance. Rather than develop additional frameworks for using learning gain measures, better value for money and impact would come from using learning gain metrics to improve and enhance existing mechanisms.
- 5.8 Several projects that developed complex models to measure learning gain integrated local data, including that from learning management information systems, so specific models would not be easily transferrable across institutions. However, the general principles of the approach could be transferable to other contexts. All projects found cleaning the data and conducting necessary statistical manipulation was very time consuming and required high-skill expertise. Although relevant to higher education in general, projects found they needed to work with statistical experts who also understood the variations across courses, such as non-traditional entry, placement years and module selection options. The models developed in the projects would need to find ways to be aligned with standardised data sets for scalability including:
- the University of Reading attainment model
 - the Open University cognitive gain model
 - the Manchester College learning gain model
 - the University of East Anglia attainment strand

Recommendation for developing measures of learning gain

Embeddedness

Measures should be embedded into existing sectoral, institutional and discipline quality, accountability and performance frameworks.

Analysis of scalability

- 5.9 Drawing on the points made above, recommendations for scalability of the pilot projects is based on the definition of learning gain and how to capture relevant information on students' knowledge, skills, work-readiness and personal development.
- 5.10 For measuring knowledge, or students' cognitive gain, existing attainment data from students' on-going assessment is best placed to capture students' learning and best

placed to drive enhancement¹⁰². Learning gain measures can be used to improve alignment of course-level student learning outcomes within existing quality frameworks and institutional assessment practices. To address concerns from the pilot projects about limitations of this approach, efforts should focus on enhancing alignment and benchmarking across systems, at module, course, institution and subject levels. Future work should engage disciplinary academics who are focusing on quality regimes including benchmarks and frameworks at subject-level. External models for this include the Valid Assessment of Learning in Undergraduate Education (VALUE) initiative and the Measuring and Comparing Achievements of Learning Outcomes in Higher Education in Europe (CALOHEE) project. Standardised tests do not offer a robust, feasible and value for money approach to desires for more information about the quality and content of student learning; projects concluded this is better captured through additional analyses and improvements to existing attainment data.

Recommendation for developing measures of learning gain

Cognitive gain

Cognitive gain is best captured through existing attainment data. Additional work in this area should focus on improving assessment, marking and alignment with quality frameworks.

- 5.11 Projects identified that surveys most efficiently capture the skills and personal development students gain from higher education. Pilot projects identified numerous instruments that can combine engagement data with affective measures that can capture gains students make in these areas. Additional time is needed to understand the usefulness of some newly developed tools, some of which are being developed and tested. A new instrument could draw on behavioural and affective elements from the University of Cambridge strand instrument, UKES and additional newly piloted affective tools. This should be used primarily for enhancement purposes, but used to meet accreditation standards, as has been modelled with the National Survey of Student Engagement (NSSE) in the US. There could be useful comparisons across subject levels and student characteristics.
- 5.12 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. However, across multiple projects (University of Reading, Birmingham City University, University of Portsmouth, University of Lincoln and the Open University) there was a lack of ability to determine if UKES is a sufficient proxy measure for learning gain in England due to extremely low response rates and inability to link sufficiently with robust data sets, although the survey has wide use in Scotland, Wales and Northern Ireland¹⁰³.
- 5.13 Development of a new instrument, or scaling up a newly developed or piloted instrument, needs to have a clear rationale for selecting items and a plan for dissemination and use of data. Any such instrument would need to be considered

¹⁰² Jankowski NA, Timmer JD, Kinzie J, Kuh GD, 2018, 'Assessment that matters: Trending toward practices that document authentic student learning'. Urbana, IL: National Institute for Learning Assessment.

¹⁰³ Coates H, McCormick A (Eds.) 2014, 'Engaging university students: International insights from system-wide studies', London: Springer.

alongside existing surveys, such as the NSS, for prioritisation, complementary approaches and use.

Recommendation for developing measures of learning gain

Soft skills development

Soft skills and personal development are best captured through surveys. Development or adoption of an instrument needs a clear rationale and needs to be embedded within the existing quality, accountability and performance frameworks.

- 5.14 Work-readiness data has been usefully captured through careers registration. It provides institutions with data for internal use and provides opportunities for benchmarking and useful comparisons across subject levels and student characteristics. The data could be widely used to support enhancement in teaching and learning as evidenced by development of modules based on categories of readiness and through online learning activities by the Careers Group¹⁰⁴. Data from career readiness can be linked with emerging data from Graduate Outcomes to address quality, accountability and performance of employability.

Recommendation for developing measures of learning gain

Employability and career readiness

Work-readiness data has been usefully captured through careers registration and is already being widely scaled up across the sector. This provides useful data for institutions and opportunities for benchmarking across subjects and student characteristics.

- 5.15 Several strands of work focused on approaches relevant to enhancing teaching and learning 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 encourage wider take-up of such approaches; references to these are available in Table 2 (page 49).

Leadership and staff engagement

- 5.16 To facilitate scaling up learning gain data, two further areas need consideration. This analysis draws from sector-wide events, conferences, debates and institutional engagement and investment in learning gain. Across the projects there is a general lack of senior management engagement, except in cases where they were part of running a project. Projects found their institutions' focus to be on existing accountability measures, such as current TEF and league table metrics, and institutions are adopting a 'wait and see' approach to which metrics may be used for future regulatory and accountability purposes. Learning gain is not a continuing priority area for most project institutions and partners, shown through non-continuation of staff beyond projects and very limited continued institutional funding.

¹⁰⁴ For examples see <https://www.thecareersgroup.co.uk/research/research-projects/careers-registration-learning-gain-project/>

- 5.17 In contrast, there has been extensive interest and engagement from the academic and educational development communities in the projects, the findings and in uses of learning gain data. This included learning gain being the theme of a Staff and Educational Development (SEDA) conference¹⁰⁵ and in an Annual National Teaching Fellowship Conference¹⁰⁶, and featuring at several events of the Higher Education Academy's Network of Deputy Vice-Chancellors, Pro-Vice Chancellors and Vice-Principals. Learning gain also featured in the annual teaching and learning conferences that many institutions host.
- 5.18 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 of 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 NSS.
- 5.19 There was greatest support for learning gain metrics where data supports the existing curriculum or feeds into on-going institutional enhancement initiatives, such as at The Manchester College, The Open University and the University of Reading. However, changes in senior management, learning gain project staff and key institutional staff hindered embeddedness across many projects.

Data and ethics

- 5.20 A major area of concern for the projects was 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.
- 5.21 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 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 across institutions of secondary data analysis, such as grade trajectories, because of the need to consider subject and institutional factors in contextualising 'big data' approaches.
- 5.22 Across all projects there is the challenge of compliance with data protection legislation. Gaining consent at multiple stages of use of data had implications for the ability of projects to validate their findings (such as linking different data sets) and for

¹⁰⁵ Kandiko Howson CB, 2017, 'Measuring learning gain', SEDA Spring Teaching and Learning Conference 11-12 May 2017. Manchester, UK.

¹⁰⁶ Kandiko Howson CB, 2017, 'Measuring learning gain', 12th Annual National Teaching Fellowship Conference, 11 April 2017. Aston University.

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.

- 5.23 Data access and data sharing were significant and often unanticipated challenges. Several projects were extensively delayed while trying to set up data sharing agreements across projects. Some agreements were signed on the principle that data would be anonymised before it was shared. Some projects agreed not to share institutional data with each other, but only see their own compared to a benchmark of other projects. Several institutions also encountered challenges accessing and linking internal data sets. Several institutions dropped out of projects because of concerns about data access and data sharing. This has consequences for future data sharing and benchmarking across institutions.
- 5.24 There were on-going practical challenges of IT systems and processes. There are infrastructure issues related to two key areas. One is the amount of data being shared – institutions have struggled to identify safe and secure mechanisms for sharing large volumes of data between projects. The Open University has struggled to link VLE data across three institutions with different systems. There were also challenges with linking institutional data and partner data given the scale of the project. The second infrastructure challenge is linking data sets once they are shared. Ravensbourne ran into issues about making sure the correct fields were shared across all partners. The Careers Group ran into challenges of sharing large data sets with sensitive information.
- 5.25 Projects noted the importance of key contacts within institutions to facilitate linking data sets, and a few faced major setbacks with staff turnover. These challenges show how institutions are at the very early stages of using existing data, and far from developing new metrics around student learning. Data teams and expert analysis functions within institutions are often disconnected from learning and students, and can be a barrier for integrating data.
- 5.26 There is a need to build in large scale analysis from the beginning of designing IT systems. In addition to challenges linking data across systems internally and across institutions, basic student data is collected differently across the sector – different systems, formats and granularity. For example, The Open University found some institutions collected module-level grades and others collated this at the end of the year. On data sharing there were differences in data governance, ethics, legal and data security across partners in projects.
- 5.27 Projects sought research ethics approval as part of their new data collection. Most secondary data analysis was covered under existing institutional data use agreements. Due to the scale and complexity of some of the projects, this took an extensive amount of time for some projects. Some institutions did research ethics approval at the lead institution, which covered collection across partners. Other institutions did separate ethical approval for each partner. The latter approach held up some projects, as with the University of Portsmouth where partners had not acquired approval through most of the first year. Securing approval from the University of Manchester's internal ethics board was a lengthy and challenging process. A key component of the research design required the linking of survey data to the wealth of administrative and academic data held by the university.

- 5.28 Several projects had high ambitions for linking multiple approaches to measuring learning gain but did not get sufficient numbers of responses or were not able to obtain Year 1 data to track students longitudinally including projects that attempted to integrate data from UKES. Integrated plans for combining different learning gain metrics could still be explored through further research, including the University of Lincoln's attempt to combine outputs from a standardised psychometric test and reflective student self-assessments with data on academic achievement, attendance, engagement in extra-curricular activities, and work experience. By combining this data into individual 'student profiles', it was hoped not only to track student development in terms of competence and self-perception, but also to identify potential correlations between extra-curricular activity and growth in confidence and ability.

Scalability of projects and strands of activity

- 5.29 Each project has reported concerns about scaling up their approach. Most concerns focus on the need for appropriate contextualisation to remain valid and retain their educational purpose and robustness as measures of learning gain. These concerns draw on associated problems with consequences of using metrics without contextualisation in rankings¹⁰⁷. Scalability of projects and strands of activity are summarised in Table 2 below. Projects with separate strands of activity that have not been brought together are analysed separately; similar approaches across projects are grouped together.

¹⁰⁷ Hazelkorn E, 2015, 'Rankings and the reshaping of higher education: The battle for world-class excellence', Springer.

Table 2. Summary of scalability of projects and strands of activity

Project/strand	Challenges to scalability	Benefits	Limitations	Recommendation
The Careers Group	Some inconsistency in question wording across institutions; variance at subject and institutional levels; unknown relationship with other outcome measures; question usefulness of measure for enhancement and accountability	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; challenges of data sharing and data protection	Sector-wide coordination of institutional use of career-readiness questions to facilitate benchmarking and subject-level comparisons; comparable across student characteristics; methodology of questions embedded into registration process could be adopted for other areas of inquiry
Birmingham City University/ University of Reading: CLA+ strand	Variance at subject and institutional levels; longitudinal usability; unknown relationship with other outcome measures; challenges of student engagement	Designed to be discipline-independent and comparable across institutional types	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 discipline-specific outcomes of English degrees	Institutional use as a diagnostic tool; may be of interest to specific subjects
University of East Anglia: Self-efficacy strand	Specific pedagogical approach; would need examples of use in additional settings	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	Effective approach to enhancing teaching and learning. Would be useful to track findings from similar efforts
University of East Anglia: Concept inventory strand	Individual inventories are time consuming and expensive to develop; limited to specific fields and uses in classroom settings	Provides standardised, objective measures of students' disciplinary learning;	Limited disciplinary coverage; challenging instrument, risk of low student motivation and engagement; proprietary	Effective approach to enhancing teaching and learning in specific disciplinary contexts

Project/strand	Challenges to scalability	Benefits	Limitations	Recommendation
University of East London	Narrow focus, 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; unclear rationale for inclusion of various measures; challenges of student engagement; not clear how data leads to enhancement	Approach to focus on experiences of disadvantaged students useful across future learning gain work.
University of Lincoln	Low scalability, but showcases integrated approach to using data to inform, support and evaluate institutional strategy	Approaches are embedded into institutional strategy	SJT proprietary; surveys institutionally-specific; challenges of student engagement at multiple points in time; challenges of meaningful engagement with challenging instruments	Transition questionnaire may be useful across other institutions for local enhancement
The Manchester College	Possibly scalable across higher education in further education contexts; variance at subject and institutional levels; unknown relationship with other outcome measures	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; not clear how data leads to enhancement	Modelling approach may be useful for enhancement activities across other institutions but would need adapting to local context
University of Manchester	Not tested across other institutions; comparability across subjects being explored; variance at subject level; questionable longitudinal usability; unknown relationship with other outcome measures	Use as a formative tool for students; 10-15 minutes completion time; tailored for subject-specific use	Limited coverage of learning gain concepts; single institution use; outputs not user friendly; tailoring of survey to different subjects limits data comparability and scalability; challenges of buy-in across subjects	Items could be considered for development of a national survey covering behavioural and affective domains
Plymouth University	Requires 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; time intensive to embed in the curriculum	Effective approach to enhancing teaching and learning at course level. Would be useful to track findings from similar efforts

Project/strand	Challenges to scalability	Benefits	Limitations	Recommendation
University of Portsmouth	Dependent on full findings from longitudinal study; variance at subject and institutional levels; questionable longitudinal usability; unknown relationship with other outcome measures	Non-cognitive focus so scalable across disciplines	Limited cross-institutional analysis so far; few findings reported so not possible to fully evaluate the efficacy of the approach; long instrument; challenges of institutional buy-in and student engagement	Items could be considered for development of a national surveys covering behavioural and affective domains
Ravensbourne	Scalable approaches (but may be less relevance to other institutions); some points embedded into new Graduate Outcomes	Engagement with alumni; rounded view of employability outcomes	Approach relevant to specific institutional type; not clear how multiple approaches fit together; focus on small specialist institutions	Was useful for informing Graduate Outcomes
University of Warwick: University of Cambridge strand	Variance at subject and institutional levels; questionable scalability across institutional types; longitudinal usability; unknown relationship with other outcome measures	Robust underpinning of 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	Items could be considered for development of a national survey covering behavioural and affective domains
University of Warwick: <i>Employability strands</i>	Scalable approaches across careers service units	Provide benchmarked data about employability initiatives	Need to be embedded with accompanying support, advice and guidance; challenges of student engagement	Approaches would be of interest to other careers services units
University of Lincoln; The Manchester College; The Open University; University of Portsmouth; University of Reading; University of East London: UK Engagement Survey	Already widely used across the sector; relationship with other outcome measures; variance at subject and institutional levels	Robust conceptual underpinning; useful as formative tool; provides data for enhancement; extensive international use	Disciplinary differences; proprietary; challenges of student engagement and prioritisation with national surveys such as the NSS and local module evaluation surveys	Items could be considered for development of a national survey covering behavioural and affective domains

Project/strand	Challenges to scalability	Benefits	Limitations	Recommendation
The Open University; University of East Anglia; University of Reading: Grade modelling	Scalable approach but not a direct measure of learning gain	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	Useful for identifying subject and institutional trends to support aligning the curriculum and identifying differential outcomes

Appendix 1: Background information

- 1.1 This is the final report of the evaluation of the OfS learning gain pilot projects programme^{108,109}. 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. OfS took over managing the projects in 2018. The projects ran between one and three years, although some projects will be extending their work using internal funding.
- 1.2 In addition to the pilot projects, there were separate strands of the learning gain programme. These included:
- 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;
 - the Higher Education Learning Gain Analysis (HELGA) project, an assessment of the potential application of national data sets to learning gain issue; and
 - capacity building and networking events.
- 1.3 Information on learning gain was gathered in an independent scoping study carried out by RAND Europe¹¹⁰ in 2015. Drawing on the RAND report and evaluation of the pilot projects, learning gain is broadly conceptualised as **change in knowledge, skills, work-readiness and personal development, as well as enhancement of specific practices and outcomes in defined disciplinary and institutional contexts** [see Sections 2.2 – 2.3 *Definition* for further discussion].
- 1.4 On the basis of the scoping study, five broad approaches to measuring learning gain were identified, and were tested and analysed through the pilot projects:
- Grades – measuring the progress in students' achievement by comparing the difference between grades at two points in time. This could include using a standardised measure (such as the GPA), or using a set of grades (standardised or not) to predict future grades.
 - Self-reporting surveys – asking students to report the extent to which they consider themselves to have gained knowledge and developed skills, through

¹⁰⁸ Evaluation of HEFCE's learning gain pilot projects: Year 1 report
<https://webarchive.nationalarchives.gov.uk/20180322111250/http://www.hefce.ac.uk/pubs/rereports/year/2017/lgeval/>

¹⁰⁹ Evaluation of HEFCE's learning gain pilot projects: Year 2 report
<https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/learning-gain-pilot-projects/>

¹¹⁰ Rand Europe report on Learning Gain:
<https://webarchive.nationalarchives.gov.uk/20180319114205/http://www.hefce.ac.uk/pubs/rereports/year/2015/learninggain/>

a survey administered at a number of points throughout their degree programme.

- Standardised tests – measuring the acquisition of certain generic or specialised skills, through a test that could be administered to students either as part of their formative or summative assessment for their degree, or as an additional exercise alongside the course.
- Other qualitative methods – including encouraging students to reflect on their learning, acquired skills and remaining skills gaps, and to facilitate a formative exchange between students and their tutors.
- Mixed methods – using a combination of methods and indicators to track improvement in performance, for example through a combination of grades, student learning data and student surveys.

- 1.5 Coinciding with the projects being transferred from HEFCE to OfS, the learning gain programme has evolved to also support the objectives of OfS, namely **experience** and **outcomes**¹¹¹.

International learning gain activities

- 1.6 Course-level measures of learning gain are the focus of several international initiatives. The European Commission is supporting the CALOHEE project as part of the Tuning framework¹¹². This work is underway but focuses on aligning frameworks for course design rather than student outcomes. There are also national research projects in Germany¹¹³, Brazil¹¹⁴, Italy¹¹⁵ and Columbia¹¹⁶ on student learning outcomes which have raised concerns about student engagement, breadth of focus across sectors of higher education, and practical challenges.
- 1.7 In the US, the AAC&U) VALUE project offers rubrics to externally assess students' in-course assignments against nationally standardised learning outcomes¹¹⁷. This is an academic staff-led initiative, has extensive institutional buy-in but is resource and time-intensive. Investment in large-scale subject-based learning outcome initiatives indicates this is a level of use with higher potential, including possibilities for international comparison.
- 1.8 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

¹¹¹ <https://www.officeforstudents.org.uk/about/our-strategy/>

¹¹² <https://www.calohee.eu/>

¹¹³ Modelling and Measuring Competencies in Higher Education. http://www.kompetenzen-im-hochschulsektor.de/index_ENG.php

¹¹⁴ <https://link.springer.com/article/10.1007/s10734-015-9963-x>

¹¹⁵ <https://link.springer.com/article/10.1007/s40888-017-0075-1>

¹¹⁶ <http://www.tandfonline.com/doi/abs/10.1080/02602938.2016.1168772?journalCode=caeh20>

¹¹⁷ Drezek McConnell K, Rhodes TL, 2017, 'On solid ground', VALUE report 2017. Washington, DC: Association of American Colleges and Universities.

and subject-level differences emerging. Due to concerns about data quality and use, the project was not continued¹¹⁸.

¹¹⁸ OECD, 2013a, 'Assessment of higher education learning outcomes. Feasibility study report. Volume 2: Data analysis and national experiences'. OECD. <http://www.oecd.org/education/skills-beyond-school/AHELOFSReportVolume2.pdf>

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

Appendix 2: Evaluation process

Evaluation approach

- 1.1 As part of the pilot projects, this independent evaluation of the pilot projects has been funded through the learning gain programme with a view to identifying best practice and supporting evidence. This includes determining the extent to which the different methods piloted would represent useful measures of learning gain that could be applied across the English higher education sector, and contains advice on the strength and weaknesses of the different methodologies and approaches used across the range of projects.
- 1.2 The aims of the evaluation were 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;
 - 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 advice from OfS to Government on future learning gain policy.
- 1.3 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. Further information on the individual projects can be found on the OfS website¹¹⁹.

Evaluation framework

- 1.4 Overview of evaluation framework:
- Development of a measure/proxy of learning gain
 - What approach was used?
 - How was learning gain measured?
 - Robustness and effectiveness
 - Validity and reliability
 - How many students were involved?
 - How did the project develop over time?
 - How was the measure of learning gain judged and assessed?

¹¹⁹ <https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/>

- Suitability
 - Was the measure feasible to obtain (practicality and value for money)?
 - Does the measure make sense to students and academics and other stakeholders?
 - Does the measure help support students and improve teaching and learning?
- Scalability
 - Was data and information shared across institutions?
 - Was/is the measure replicable across disciplines, student groups and at other institutions?

- 1.5 Development of measures of learning gain. 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.
- 1.6 The next step 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 projects¹²⁰.
- 1.7 Robustness and effectiveness. 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.
- 1.8 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 survey items and scales.
- 1.9 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? A further consideration is 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 be valid for

¹²⁰ Evaluation of HEFCE's learning gain pilot projects: Year 1 report
<https://webarchive.nationalarchives.gov.uk/20180322111250/http://www.hefce.ac.uk/pubs/rereports/year/2017/lgeval/>

comparison across institutions. Robustness and effectiveness were major areas of focus for the Year 2 evaluation of the pilot projects¹²¹.

- 1.10 **Suitability**. The third stage of evaluation, suitability, explores contextualised validity through feasibility and usability. Suitability builds on the robustness and effectiveness of measures, draws on lessons from the projects and wider practice and research. Analysis considers potential benefits from using measures as well as possible consequences. Suitability is also relative to context rather than absolute. The **unit level of analysis** is a concern for several projects, related to design and subsequent use of measures. For example, Birmingham City University focused 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.
- 1.11 **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.
- 1.12 Additional lenses of analysis include the **feasibility** of obtaining the measure, 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. Some projects tested instruments with hundreds of items, which may result in robust data but very few students would complete it.
- 1.13 **Usability** explores how measures are 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 were major areas of focus for the Years 2 and 3 of the evaluation.
- 1.14 There was little engagement with the notion of **value for money**. Qualitative feedback has shown that students recognise and value both the opportunities for self-development and the skills they have learned. However, additional work would need to be undertaken before it would become possible to translate these findings into an assessment of value for money.
- 1.15 **Scalability**. 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. Drawing on the first stage of evaluation (the approach to measuring learning gain), scalability explores what could be done with the metrics,

¹²¹ <https://www.officeforstudents.org.uk/media/1386/evaluation-of-hefce-s-learning-gain-pilot-projects-year-2.pdf>

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.

- 1.16 The importance of contextualisation and transferability are 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.

Evaluation analysis

- 1.17 Challenges to evaluating learning gain are reflected throughout the report. Firstly, there was no clear purpose for measuring learning gain established at the outset. Projects were encouraged to consider relevant purposes and were brought together to share and discuss. These evolved over the duration of the pilot projects, with projects responding to policy developments. The evaluation of the projects suggests relevance of different metrics and approaches given different purposes proposed by the projects and sector.
- 1.18 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. For some instruments, this indicates the approach is not recommended, as the University of Lincoln found with using the SJT as a longitudinal measure. For others, additional research would be needed to judge the efficacy of the approach. This is particularly the case with some of the conceptual models that projects delivered which aimed to link multiple sources together, such as that from the University of Reading and The Manchester College.
- 1.19 Regarding scalability, all the projects are conducted in specific institutional and disciplinary contexts so wider testing would be needed to fully evaluate wider scalability. And finally, the philosophical orientation of those developing measures may vary from that of the end-user, such as senior management teams or prospective students, and particularly for potentially high-stakes accountability purposes. The purpose of what is being explored and the audience drive which aspects of learning gain data are used and what would be most appropriate, which are subjective, value-laden and part of continued debate and discussion across higher education around the world¹²².

Evaluation role

- 1.20 My role with the learning gain work has evolved over the project. Given the diversity of the projects funded (in terms of size, scope, scale, organisation, design) and the varying format, degree of specification and structure of the bids, it took substantial time to understand what the different projects were seeking to measure and how this would be carried out. Links were developed and established with all lead institutions, other appropriate stakeholders, as well as most of the partner institutions involved in

¹²² Hazelkorn E, Coates H, McCormick AC, (Eds.) 2018, 'Research handbook on quality, performance and accountability in higher education'. Cheltenham: Edward Elgar.

the projects. Projects were formally visited annually, and coordination with project leads occurred throughout the year, and concerns were gathered to be addressed at annual pilot project conferences. Further liaison occurred with them at national conferences and events. Seminars and conferences were organised which brought in project leads to present and disseminate findings of their projects. Project leads were also suggested as speakers at numerous events. Liaison with HEFCE and OfS staff and the project Steering Group occurred regularly, and outcomes reported out through annual external reports and interim internal reports. I have endeavoured to champion learning gain, provide clarity to a complex area of research and support both the aims of HEFCE, OfS and the pilot projects.

Appendix 3: Project methodology

- 1.1 Detail on project methodologies is available from the Year 2 evaluation report¹²³ and project case studies on the OfS website. Out of the 13 projects, 10 were funded for three years to be able to develop measures and track students over time. Plymouth University, a two-year project, was also longitudinal in design. Ravensbourne, the only one-year project, adopted a cross-sectional design and made extensive use of existing data, going back several years. The University of East Anglia's two-year project was also cross-sectional in design but involved some testing of students at multiple points in time. Two of the three-year projects also planned a cross-sectional element: the University of Warwick and the Careers Group.
- 1.2 In the spirit of the design of the pilot projects, a range of methods was used. All of the projects combined newly collected data with secondary data analysis of existing institutional data, accounting for entry data, student demographics and characteristics, and student progress, continuation and attainment data (grades).
- 1.3 Due to challenges of project start times and student engagement several projects had to amend their methodology. This included removing data collection points in Year 1 or switching to additional data collection points in Year 2. Some longitudinal-based projects, tracking individual students over time, also added a cross-sectional element, tracking additional incoming cohorts of students, to make up for low student engagement in Year 1. This allowed projects to collect enough data to test instruments but greatly increased the complexity of the analysis and findings and made drawing firm conclusions of usefulness and scalability difficult.
- 1.4 Using existing data. Two projects focus on analysis of all students across cohorts as well as change over time: The Open University and the Careers Group. Strands of work in two other projects involve analysing whole data sets. One strand of the work at the University of Reading involves analysis of all students using existing data, going back several years. Similarly, Ravensbourne did an analysis of existing data related to DLHE data.
- 1.5 Learner analytics. Learner analytics involves the collection, analysis and reporting of data about learners and their educational environments. This largely involves secondary data analysis (analysing existing data). Examples include analysing entry scores and eventual student grades; exploring the impact of students engaging in a module VLE and outcomes on assessments; and categorising attainment gaps across student characteristics. In institutions, learner analytics are generally useful for identifying trends and patterns (such as a BME or gender attainment gap; or the impact on grades of students submitting dissertations), that can then be followed up through qualitative methods.
- 1.6 The main strand of work in the project by The Open University collects data from satisfaction surveys (affective measures), data from VLE (behavioural measures), and on academic performance (cognitive measures). The behavioural elements include attendance in class, discussion forums, chats, essays and student

¹²³ <https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/learning-gain-pilot-projects/>

evaluations. This data is matched with central records to account for student characteristics and entry measures. Similar analyses were done in strands of work in the University of Reading and University of East Anglia projects, analysing existing student characteristics, progression and academic performance data.

- 1.7 Projects analysing of large amounts of existing data face the challenge of uncovering meaningful patterns, trends and areas for further investigation. However, a further challenge is that when interesting findings are found, the data often indicates a relationship but does not explain why, thus requiring further qualitative analysis. This may be useful for institutional improvement but may be less effective for the accountability use of learning gain.
- 1.8 *Collecting new data.* The rest of the projects focus on capturing new data and analysing it by specific groups of students and in some cases tracking them over time. Some projects in the future may analyse data sets of whole cohorts to get benchmark and comparative data, particularly on representativeness of entry scores and modelling outcome data, but none who planned this were able to complete this analysis in the time-frame of the projects. Unlike the projects using existing data, these have a specified area of inquiry. However, the challenge of these projects is gathering sufficient data to be generalisable across student characteristics, subjects and institutional type.
- 1.9 Surveys. Four projects broadly consist of combining survey data with secondary institutional data. The Careers Group project linked questions from a brief 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. Similarly, The Manchester College project links UKES data with secondary data from institutions on progression and attainment, with the analysis focused on the profiles and pathways of students in further education settings.
- 1.10 The University of East London project combined survey data including scales on Need for Cognition, Academic Behaviour Confidence Scales and Predict Your Grade, along with a partial administration of UKES and linked this with secondary institutional data on socio-demographics, Widening Participation, non-continuation and attainment. The University of Portsmouth's project focused on non-cognitive skill development using self-reported questionnaire data through 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.
- 1.11 Another strand of the University of East Anglia project is based on self-efficacy assessments in economics, also analysed alongside secondary data on student marks and grades (measured as GPA).
- 1.12 Tests and surveys. Half of the projects combine survey data with a standardised test. Two projects are using the CLA+. Birmingham City University administered the CLA+ test and attempted to link it with data from UKES but did not get sufficient response rates. The University of Reading project focused on analysis of existing student data (demographics and average marks), complemented with primary data from CLA+, UKES, and internal wellbeing and careers surveys. As the CLA+ is externally

administered, the projects were somewhat constrained in their data collection due to the availability of testing windows and the cost of administration.

- 1.13 The University of Cambridge strand of the University of Warwick project involved the measurement of learning gain in higher education and across disciplines through the development of a survey instrument. One strand of the University of East Anglia project focused on discipline-based concept inventories which function like a standardised test taken at multiple points in time and are gaining interest in physics and other disciplines¹²⁴. These were trialled in chemistry and biology in the first year and were expanded into pharmacy.
- 1.14 The University of Manchester's project created a 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; and confidence on generic learning outcomes, all linked with existing entry and attainment data. The University of Lincoln's project aimed to combine data from an SJT, with data on student participation in training, democratic involvement in the students' union, work experience, and extra-curricular activities and secondary data on academic performance and Widening Participation.
- 1.15 Multiple measures of a specific theme. Two projects focus on a specific theme and incorporate multiple approaches to measuring it. Plymouth University's project focuses on student progression in their self-reported understanding, skills, experience and confidence in research methods. Data collection includes a survey to capture students' understanding of and experience with research methods; staff semi-structured interviews; programme documentation; student reflective logs; links with module performance; and additional secondary data analysis. The project outcome measures include attainment and progression through further education and into university-level study.
- 1.16 With a focus on work-based learning in the creative vocational sector, the Ravensbourne project aimed to understand whether the learning gain achieved via work-based learning methods can be evaluated robustly and, if so, what (additional) data and methods is needed in order to achieve this at scale. This was explored through three work packages: triangulating DLHE data across demographic groups and work-based learning activities; exploring graduates' outcomes three years out via a phone survey with alumni; and surveying students before and after a work-placement via an Employability Self-Evaluation Test (ESE).
- 1.17 Three strands of the University of Warwick project focused on employability. The Realise 2 Strengths strand, led by the University of Warwick, involves a survey and one-to-one coaching 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.
- 1.18 All of the projects incorporated data from student and staff interviews and incorporate secondary data analysis including grades.

¹²⁴ Sands D, Parker M, Hedgeland H, Jordan S, Galloway R, 2018, 'Using concept inventories to measure understanding', *Higher Education Pedagogies*, 3:1, pp173-182

Appendix 4: Evaluating validity and reliability

- 1.1 All the projects explored the validity of different measures of learning gain in their individual success criteria. There are multiple approaches to judging the validity and reliability (see references for further information^{125,126,127}) of measures of learning gain, through statistical techniques, interviews with stakeholders such as staff and students, and analysis of findings.
- 1.2 Content and face validity. 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.
- 1.3 Construct validity. Projects developing new questions used multiple approaches to determine the validity of the approach for the concepts explored. This included interviewing groups of higher education students, as well as panels of experts and managers in higher education learning and experienced tutors from different courses and disciplines across institutions.
- 1.4 Concurrent and predictive validity. Most projects used 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 then 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.
- 1.5 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.
- 1.6 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

¹²⁵ Cozby PC, 2001, 'Measurement Concepts'. Methods in Behavioral Research (7th ed.). California: Mayfield Publishing Company.

¹²⁶ Cronbach LJ, 1971, 'Test validation', in Thorndike RL (Ed.), Educational Measurement (2nd ed.). Washington, D. C.: American Council on Education.

¹²⁷ Litwin M, 1995, 'How to Measure Survey Reliability and Validity'. Thousand Oaks, CA: Sage Publications.

this analysis is still underway to measure the effectiveness of measures for longitudinal analysis.

- 1.7 Consequential validity. There could be significant impacts across the sector depending on how learning gain metrics are used. There are lessons from the schools sector about 'teaching to the test' and the impact of high-stakes outcome measures on how students are taught. What might be the consequences of unleashing these measurements on the ecology of assessment in higher education?
- 1.8 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. All the projects conducted focus groups and interviews with students and other stakeholders, such as staff, parents and employers to explore the validity and reliability of instruments.

Appendix 5: Student engagement approaches

- 1.1 Compulsion. The project with the most success in engaging students was the Careers Group, where filling out the survey questions was mandatory as part of students' annual course registration. This meant minimal questions could be asked but very high numbers across all subjects and demographics were reached. Amongst the partners for the Birmingham City University project, the institution that chose to imply to students that the test was compulsory, and gave students the least notice, secured the best attendance.
- 1.2 Engaging student-facing staff. Projects were much more effective in engaging students when they worked with student-facing staff. Plymouth University found buy-in from the programme leads and lecturers to be an important determinant of student engagement across partner institutions. Similarly, Birmingham City University found the key relationship determining the extent of student recruitment is most probably that between the students and a key academic heavily involved with their programme cohort. Several projects noted there was high engagement in areas with committed frontline staff, however, success in one area was not necessarily replicable across institution, as the University of East Anglia found.
- 1.3 Workshops. Another successful tactic was to schedule testing and surveying sessions in booked rooms. The University of Lincoln found completion rates were the highest when the tests were undertaken in workshop conditions, rather than sending emails and hoping students would fill out tests in their own time. The University of Reading also found more engagement when they booked testing rooms for students.
- 1.4 Incentives. Overall, incentives were not deemed that helpful in getting students to participate. Some projects used them to some success but many students who completed tests and surveys did so because of personal interest (although they appreciated the incentive). Birmingham City University found the value of incentives for engagement debateable. Small incentives for all students to participate, like printer credit, were more successful than prize draws.
- 1.5 Projects tried numerous different approaches to incentives. Plymouth University used Amazon vouchers, prize draws, and encouraging students to include time spent on the project towards developing their application for the 'Plymouth Award'. At the University of Lincoln, after recruitment struggles, an incentive of £10-worth of printer credits was introduced for all students who completed all three elements. This was in addition to a prize draw for £100. All of the institutions in the University of Portsmouth project faced low take-up. They offered incentives (raffles) and personalised reports. They also tried offering printer credits after struggling to recruit for focus groups. Several institutions also tried to highlight the link between the project and institutional engagement or employability awards, as well as offering personalised reports to students.
- 1.6 Students' union engagement. Several projects engaged with their students' union. The University of Reading had a students' union representative on their steering group. The University of Manchester is liaising with its students' union when developing the project and instruments, and the University of Lincoln has worked in partnership with their students' union to collect a wide range of data on students'

activities and engagement. While this has been helpful, engaging with the students' union did not seem to help boost student engagement in the projects or raise response rates.

Appendix 6: Pilot project selection

Application criteria

- The extent to which the proposal has the potential to demonstrate that it will deliver a valid measure or measures of learning gain
- The extent to which the proposal has demonstrated innovation, and the potential viability, suitability and scalability of the method or methods in England
- Value for money, including likely sustainability
- Whether the pilot project will promote partnerships between institutions, departments and disciplines, and where relevant with employers, in order to facilitate the transfer of expertise.

Criteria	Description
A	The proposal has the potential to demonstrate that it will deliver a valid measure or measures of learning gain
B	The proposal has demonstrated innovation and the potential viability, suitability and scalability of the method or methods in England
C	The proposal demonstrates value for money, including likely sustainability
D	The proposal will have/has the potential to promote partnerships between institutions, departments and disciplines, and where relevant with employers, in order to facilitate the transfer of expertise

- a. The need to ensure programme-wide evaluation and concurrently to explore national data and trends in order independently to validate individual project findings.
- b. The prevalence of mixed-method and longitudinal projects among those highly graded and therefore the need to encourage projects individually and collectively to consider cross-sectional approaches.
- c. The value of at least some projects yielding negative results, given the purpose of the call to inform future approaches and policy.
- d. An imperative for individual projects to test or prove that their project's approach may be applicable across the sector, including through the regular capacity building and evaluation meetings.
- e. That the capacity building network should be open to institutions beyond those funded through the pilot phase, with a view to building up a research-informed consensus around shared typologies of, and approaches to, learning gain.
- f. That a number of institutions had signalled their appetite to work with others, so the funded projects should be alerted to those institutions who had expressed interest or been unsuccessful and that had similar interests.