



EUROPE

Learning gain in higher education

Cecile Hoareau McGrath, Benoit Guerin, Emma Harte,
Michael Frearson, Catriona Manville

For more information on this publication, visit www.rand.org/t/rr996

Published by the RAND Corporation, Santa Monica, Calif., and Cambridge, UK

RAND is a registered trademark.

© Copyright HEFCE 2015

RAND Europe is a not-for-profit research institute whose mission is to help improve policy and decisionmaking through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Support RAND

Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org
www.randeurope.org

Table of Contents

Table of Contents	iv
Figures.....	vii
Tables.....	ix
Executive summary	xi
Acknowledgements	xix
1. Introduction	1
1.1.The rationale for this research	2
1.2.Key questions of the study.....	3
1.3.Our methodological approach	4
1.4.Outline of this report.....	5
2. Definitions of learning gain.....	7
2.1.A novel concept in England.....	7
2.2.What do students learn?	7
2.3.How do we define learning gain?	8
2.4.What can measures of learning gain be used for?.....	11
3. How can learning gain be measured?	13
3.1.What attributes contribute to a robust method to measure learning gain?.....	13
3.2.What are the key questions to bear in mind when assessing learning gain methods?.....	16
4. Overview of methods	21
4.1.Grades	21
4.2.Surveys	23
4.3.Standardised tests.....	28
4.4.Mixed methods	34
4.5.Other qualitative methods.....	36
4.6.Summary assessment of different methods	38

5. Suggestions for next steps.....	43
5.1.Awareness-raising activities and stakeholder consultations.....	43
5.2.Series of pilots around five methods	43
6. Conclusion	45
6.1.Definitions	45
6.2.Methodology	45
6.3.Summary of findings in light of research aims.....	45
6.4.4. Relevant considerations for a robust pilot	49
6.5.Next steps for the project	50
References	53
Appendix A: Summary of methods reviewed.....	63
Appendix B: Research method.....	71
Appendix B.1: Search protocol for the literature	79
Appendix B.2: Interview protocol for professional bodies.....	83
Appendix B.3: Interview protocol for higher education institutions.....	90
Appendix C: Evaluation framework.....	93
Appendix D: Glossary	101
Appendix E: List of acronyms.....	105

Figures

Figure 1: Representation of learning gain	9
Figure 2: Graphic definition of value-added	10

Tables

Table 1: Preliminary assessment of models considered within the study, based on characteristics of the methods themselves	17
Table 2: Advantages and disadvantages of different methods and tools of measuring learning gain	38
Table 3: Summary of methods	64
Table 4: Breakdown of reviewed papers.....	72
Table 5: Data extraction matrix used for the literature review.....	74
Table 6: Professional bodies selected for interview.....	77
Table 7: List of consulted experts	78
Table 8: Criteria for inclusion in the literature review.....	81
Table 9: Evaluation framework to assess the (potential) use of the method by the sector.....	96
Table 10: Evaluation framework to assess the potential of the method to be adopted by the sector	99

Executive summary

Context and rationale

There is growing interest around the world in measuring how much students in higher education learn, and the contribution that institutions, courses and particular educational activities make to that learning. Amongst other factors, such measures can be seen as important to the debates about the quality and impact of higher education, how we evidence the value of investment in it, and how we evidence students' skills acquisition for employers.

The question of how to measure the student learning gained in higher education has been much debated. The debate is healthy - any attempt to measure a process which is as multidimensional as learning can encourage a focus on what is effective - but it also risks distorting activity. There is an evident need to consider different approaches to the measurement of gains in learning and consider their impact on practice and on outcomes. Debate on the measurement of the learning students gain in higher education has been prominent internationally, but the concept has not been studied extensively in the English higher education context. This research explores the concept of learning gain, as well as current national and international practice, to investigate whether a measure of learning gain could be used in England.

Definitions

Learning gain is understood in a variety of ways by the higher education sector: as the difference in student performance between two stages of their studies, as a variant of the concept of 'value added' commonly used in school performance tables, or simply as 'learning'. Indeed it is important to note that whilst the assessment of learning gain in higher education is in its infancy in England, the country has acquired a position of international leadership in developing value added measures of school education, and has pioneered research and applications of value added models in secondary education.

For the purposes of this report, the concept of 'learning gain' is defined as the 'distance travelled', or the difference between the skills, competencies, content knowledge and personal development demonstrated by students at two points in time. This allows for a comparison of academic abilities and how participation in higher education has contributed to such intellectual development.

The concept of learning gain is often used interchangeably with the notion of value added. Learning gain is measured based on the difference between two measures of actual student performance, while value added is based on the comparison between performance predicted at the outset of studies and actual performance achieved.

Research aims

As the concept of learning gain has not been extensively explored in English higher education, the Higher Education Funding Council for England (HEFCE), working in partnership with the Department for Business, Innovation and Skills (BIS) and the Higher Education Academy (HEA), commissioned RAND Europe in 2014 to investigate the following key areas:

- a. In what ways and for what purposes are methods and tools for measuring learning gain already in use in English higher education?
- b. Analysis of the relative benefits of approaches to measuring generic skills independently of discipline-specific knowledge, and measuring generic skills in disciplinary contexts.
- c. Analysis of the applicability of methods and tools for measuring learning gain for different identified purposes such as to inform improvements to learning and teaching; to provide information to prospective students and their advisers; to investigate the impact of particular learning and teaching interventions and contextual factors; to assist in international comparison; or to form part of the quality assurance of learning and teaching.
- d. What are the relevant considerations for the use or adaptation of methods or tools for measuring learning gain in an English higher education context?
- e. What are the relevant considerations for the design and development of robust pilot activities for measuring learning gain that might be drawn from existing practice and literature?
- f. What relevant lessons can be drawn from research on value added undertaken in the UK schools sector?

Methodology

The report draws upon a review of the English and international literature relevant to the topic of learning gain; a call for information circulated by the HEA to the Network of Deputy Vice-Chancellors, Pro Vice-Chancellors and Vice-Principals; and interviews with personnel from selected professional bodies, academics and administrators across 13 higher education institutions (HEIs).

The research team used these different types of data to gain a deeper understanding of the sector's perception and understanding of the concept of learning gain, to report a sample of methods presented as measuring learning gain, as well as to assess the robustness of these methods.

Summary of findings

Purposes of learning gain

Measures of learning gain can be used for a number of different purposes. Learning gain measures can meet an enhancement purpose, supporting higher education professionals to increase their understanding

of how to facilitate students' learning. They can also be used to support accountability, promote transparency and enable comparability of the outcomes of higher education.

Sector response to learning gain

One hundred and thirty out of 147 respondents to the call for information launched via the HEA acknowledged that the concept of learning gain could be useful. Interviewees also demonstrated a willingness to learn more about learning gain.

Uncertainty regarding the purposes to which a measure of learning gain might be applied, particularly regarding purposes of transparency and accountability, gave rise to a certain level of cautiousness among respondents and interviewees.

Respondents, interviewees and experts also highlighted the need to be aware of the potential for unintended consequences arising from application of a measure of learning gain. An example of unintended consequence could include for example gaming of indicators, to the detriment of delivering their educational mission.

The need for sensitivity to institutional autonomy and institutional diversity was highlighted, as was the potential for resistance from the higher education sector if a measure were perceived as 'imposed' on the sector rather than sector-owned.

Critical overview of learning gain methods

Based on a review of the literature, the call for information and interviews, the research team has identified 14 methods, clustered in five groups, to measure learning gain. These methods include grades, student surveys, standardised tests, mixed methods, as well as other qualitative methods. The assessment of these groups of methods is summarised in Appendix A3.

The utilisation of methods in England is at a more experimental and smaller scale than in comparator countries around the world, especially the US. Many methods in use have different core purposes and/or only measure learning gain indirectly. In the US, methods have predominantly been used in the context of a liberal arts/general education; such approaches have now begun to establish themselves in Europe. This context is different to the more specialised nature of curricula in most HEIs in England.

There is a distinction to be drawn between direct measures of learning gain and proxy measures. Proxies, for example surveys, graduate employment rates or expected salaries, have been suggested by some in the English literature as proxies for one or more aspects of learning gain. Others are designed to act as substitutes for the concept of learning gain as a whole. These proxies could include some of the methods summarised below: engagement and experience for example. The use of proxies has been explored and used in the US, but has remained at the proposal stage in England as far as the research team is aware.

Using proxies to measure learning gain could prove efficient since they exploit existing metrics. However, this approach would need to address some experts' concerns regarding their validity. Some proxies, such as graduate employment, appear to depend on developments extraneous to higher education, such as the state of the labour market, and the behaviour of employers. All proxy measures, including surveys of

student engagement and experience, need to be validated by comparison with other direct assessments of learning if they were to be used for this purpose.

Grades

- One approach presented as measuring learning considers student achievement by taking the difference between students' grades at two points in time.
- This can be undertaken in different ways. These include: comparing the difference between actual grades at two points in time, using a standardised measure of these grades (such as the Grade Point Average (GPA)) as a comparison, or using a set of grades (standardised or not) to make predictions on future grades.
- The research team also found some evidence of the use of grades as part of mixed methods, for example, combined with surveys as well as standardised tests. There is evidence of each of these methods being used currently in the higher education sector in England.
- One of the advantages of using grades is that, from the onset, it provides a way directly to measure learning gain which relies on existing institutional capacity (all HEIs issuing grades).
- The key issue regarding the use of grades in the UK is comparability. Different practices in grading and assessment are used within and between disciplines, thus only allowing for broad comparability to exist between awards.
- Another issue is that a significant proportion of students tend to get results in a similar part of the honours degree classification system five point-grading scale (2:1 or 2:2). This may not allow enough variance to measure learning gain, for example if the intention were to use grades to compare learning gain across English HEIs. There is also debate on whether tariff scores/points are an adequate measure of starting point and the difficulties caused by ignoring some A level subjects and entrants with non-standard qualifications.

Surveys

- Another approach is to ask students to self-report the extent to which they consider themselves to have learned through a survey.
- Several surveys are in use in the higher education sector in England. Two of these surveys are run by our commissioning bodies, the National Student Survey (NSS) and the UK Engagement Survey (UKES). In addition, some surveys are administered by one or a handful of institutions, including the Careers Registration instrument from the University of Leeds (which contains questions on personal development related to employability) or the skills audit of the University of Durham (which measures students' progression in terms of acquired skills).
- These surveys are not primarily designed to measure learning gain. But, in the case of UKES and the NSS, they do contain some questions related to student development. The most recent UKES pilot included 12 items regarding students' perceived skills development and the NSS contains one question relating to personal development, and is piloting the inclusion of questions relating to student engagement. The limited number of targeted questions directly related to gains in UKES, and the NSS in particular, casts some doubt on the validity of using these surveys as currently constructed as a measure of learning gain.

- In general, surveys are considered to provide more indirect and less valid measures of learning gain than direct measures such as grades or tests as students are reporting their own assessment of their learning gain. There is debate in the US on the nature of the association between students' test scores and measures of engagement.
- Some surveys, including the NSS and UKES, are undertaken at only one point in time and would therefore need repeating within each cohort at the start and end of study to obtain longitudinal data suitable for measuring learning gain.
- Response rates to surveys can be highly variable and, were a survey with lower response rates to be used to measure learning gain, additional work would need to be undertaken to increase this (for example the Careers Registration tool at the University of Leeds ties in student responses to the annual registration process).
- Any survey would have to be validated as a proxy for learning gain in England as the Wabash Study has done for a suite of instruments in the US. This could be done through piloting to test the relationship between student self-reporting and progression.

Standardised tests

- Standardised tests provide a third established approach to measuring learning gain. Standardised tests measure the acquisition of certain skills, which may be generic or specialised. They are administered to students either as part of their formative or summative assessment for their degree or as an additional exercise alongside the course.
- There are two main types of standardised test:
 - Discipline-specific test, such as the Progress Testing administered in some schools of medicine in England.
 - Generic skills test, which are not discipline specific, such as the Collegiate Learning Assessment (CLA), which relies on open-ended questions and the Measure for Academic Proficiency and Progress (MAPP), which is a more traditional test relying on multiple choice questions. Both these examples originated and are applied in the US; as far as the research team is aware, these are not applied in England.
- Standardised tests are considered to be more objective as measures of learning gain than methods based on self-report such as surveys. Furthermore, several of the standardised tests mentioned in the report are designed specifically to measure 'distance travelled' with comparability in mind.
- Discipline-specific tests generally achieve a higher level of precision and hence validity than general tests. However, the outputs of a discipline-specific approach are difficult to compare across disciplines. Should broader comparability be desirable, work would need to be undertaken to explore the possibility of aggregating results up to a more generic level.
- Non-discipline specific tests are attractive to some given that they rely on complex psychometric techniques to compare the acquisition of generic skills, such as critical thinking. Others have raised concerns regarding their representativeness in application, as in the case of the CLA, which uses samples of students rather than the entire cohort.
- The CLA has also been criticised as measuring skills which are significantly too general to measure learning gain. Both CLA and MAPP have been criticised as not taking account of the

differences in programme emphasis across institutions. MAPP also assumes that complex skills can be disaggregated into multiple choice questions. Both can be used for institutional comparability, although again this is debatable.

- The purpose of non-discipline specific standardised tests – focusing on generic skills – would need to be clearly specified since there can be a somewhat negative perception of standardised tests across the sector when these are used to rank students and institutions.
- Developing standardised tests for England would require consideration of the level of resources needed to develop, initiate and fund and administer in steady state. For example, Progress Testing relies on an extensive pool of questions to be administered each year.

Mixed methods

- Mixed methods use a range of tools and indicators to track performance, for example through the combination of GPAs, student learning data and student surveys, or the combination of grades and a test to predict students' performance, which could be presented as akin to measuring learning gain.
- Methods can be combined to increase the robustness of the measurement; however, these methods are relatively sparse in England. Weaknesses in the individual methods can also transmit to the mixed methodology.
- Some approaches are likely to require broad institutional capacity to implement in terms of systems and expertise to undertake learning analytics.

Other qualitative methods

- In addition, the project also investigated a range of other methodologies of tracking student development currently in place in the sector, including qualitative methods.
- These methods include students writing a portfolio such as the Personal Development Portfolio (PDP), or initiatives to encourage students to think about their acquired skills and skills gaps such as 'skills awards'.
- These other methodologies tend to be particularly useful to encourage students to reflect on their learning, as well as to stimulate a formative exchange between students and their tutors. However, some of these methods may not be entirely representative since they tend to be taken up by a limited pool of students. They could also be difficult to compare across institutions, given that they do not currently have comparable questions and formats for replies.
- They are currently most suited to individual measurements of learning gain rather than institutional comparability.

Relevant considerations for a robust pilot

In light of the findings of the report, the research team encourages the commissioning bodies to take the following points into consideration when devising the next stages of the project:

- Clearly state the purpose of measuring learning gain for any particular application.
- Clearly define dimensions of analysis, be they generic skills or elements of personal development for example.

- Capture the complexity of measuring learning gain, especially if comparing institutions, including taking confidence intervals and the complex interactions between institutions (represented through interaction effects).
- Recognise the diversity of needs and goals of students, the state, or the mission and goals of an institution, or a higher education system.
- Adopt methods which have comparable inputs and outputs (and adjust for control variables).
- Remain practical and consider effectiveness of implementation as well as cost-effectiveness.
- Remain aware of unintended consequences, for example, to prevent ‘gaming’.
- Achieve buy-in from HEIs and other stakeholders.
- Acknowledge the limits of measuring learning gain, namely that HEIs are not the only drivers of learning gain in a student’s life.
- All of the methods identified include some aspects which could be considered to measure learning gain. They also experience common challenges, which would need to be addressed in a pilot, including the need appropriately to validate these methods, for example by controlling for different potential factors which could impact on the validity of the measure, such as socio-demographic or neighbourhood effects.

Conclusion: next steps for the project

On the basis of this overview, the research team suggests two next steps for this project.

Suggestion 1

The commissioning bodies may wish to engage with the sector to raise awareness of and promote debate about measures of learning gain and their purposes.

Options to consider are:

A series of symposia or conferences in order to inform the sector and broader public of the relevance of learning gain. These symposia and conferences would encourage further communication between different stakeholder groups.

A policy workshop to test the perception of the higher education sector towards measures of learning gain and validate the applicability of measures of learning gain.

Suggestion 2

The commissioning bodies may wish to conduct a series of pilots in order to explore further some of the considerations relating to the measurement of learning gain in the English context. These pilots could draw on the methods investigated in the report, exploring issues of viability and validity in their various aspects: practical, technical, methodological and financial. Consideration should be given to an impact assessment of the intended and unintended consequences of measuring learning gain.

These pilots could include:

- Methods which are currently applied in the English higher education sector, as well as their scalability. This approach could include piloting surveys for example or some discipline-specific tests, noting that the methods currently in place were designed to measure other aspects than learning gain, or to cover one dimension or discipline.
- Testing methods which have not yet been found to be applied in England, including generic standardised tests. Consideration would be needed of the implications of transferring such a test from one higher education system to another, and the sector’s potential reactions to using external measures.

- Designing an English measure of learning gain, possibly constructed as a test. This approach could yield sector-engagement if designed in cooperation with the sector and be tailored to the needs and structure of the English higher education sector; however, it would require a significant investment of time and resources. The examples of international comparisons, such as Assessment of Higher Education Learning Outcomes (AHELO), show that this approach comes with its own complexities and sector-based dynamics.

Acknowledgements

The authors would like to thank the technical and scientific advisory board, Hamish Coates, Ellen Hazelkorn and Camille Kandiko-Howson for their active and supportive involvement in this project.

This project would not have been possible without the work of additional members of the research team, including James Williams.

We are also grateful for the constructive guidance of the members of the HEFCE/BIS/HEA Steering Group, including Ellie Clewlow, Scott Court, Pamela Macpherson Barrett, Ross Hudson Rob Stroud, Charles Ritchie, Philippa Levy, Geoff Stoakes and Alex Buckley. Data collection was greatly improved by the support of the HEA, in particular through Philippa Levy, who helped to administer the call for information throughout the HEA network. We are also grateful to the Steering Group for their input to the report.

This report also greatly benefited from expertise provided by delegates who attended the workshop and from the written feedback of additional experts, including Harvey Goldstein, Bob Gilworth, Trey Miller, Mantz Yorke and Anna Vignoles. Finally, the report would not have been possible without the contribution of the interviewees and respondents who have provided information for this project. We are also grateful to Mark Langan and Chris Van Stolk for their reviews.

1. Introduction

This research project on learning gain in higher education in England has been commissioned by a steering group composed of the Higher Education Funding Council for England (HEFCE); the Department for Business, Innovation and Skills (BIS); and the Higher Education Academy (HEA). We report research, which constitutes the first phase of the Steering Group's project. Learning gain, by which we broadly mean the difference between the skills, competencies, content knowledge and personal development at two points in time, is a relatively new concept in England.¹ International debates on how to measure the quality of education have heightened interest in this originally US-based concept in England and elsewhere. International interest has grown because learning gain has the potential to say more about the extent to which higher education fulfils its key mission of contributing to students' knowledge, skills and understanding than traditional output indicators, such as the number of graduates.

This report aims to explore the concept of learning gain by investigating a number of objectives:

- **Objective 1:** Shedding light on the definition of learning gain.
- **Objective 2:** Describing the methods to measure learning gain in existence internationally as well as, if applicable, in England.
- **Objective 3:** Evaluating the suitability of these methods to the measurement of learning gain in England.

The report constitutes the first of three stages of the learning gain project. The second stage is likely to include a series of institutional pilots to test and assess a range of methods to measure learning gain. This will be followed by a comparative evaluation of the pilot activities.

This report seeks to both inform the Steering Group of the methods which are the most appropriate to take forward at the pilot testing phase, and to inform a broader audience of what learning gain is, and of its potential relevance in English higher education.

It is clear to us that the further exploration of learning gain will carry some sensitivities around whether the approaches selected for the institutional piloting phase are sophisticated enough to reflect the complexity of learning, and how to reflect the varied missions of autonomous higher education institutions (HEIs). The authors hope that this report will support the Steering Group in acquiring and

¹ Department for Business Innovation and Skills (BIS), 2011.

demonstrating an awareness of these potential concerns, in formulating a clearly stated purpose when measuring learning gain, and choosing methods that are fit-for-purpose.

The remainder of the introduction explains the rationale for this research, key questions to be raised and the methodological approach of the study.

1.1. The rationale for this research

The learning gain in England project is influenced by two principal drivers, as detailed below.

1.1.1. Higher education financing mechanisms

The UK Government's 2011 White Paper 'Putting students at the heart of higher education' introduced major changes in the higher education policy landscape within England. Changes to the financing mechanism in place from 2012-13 onwards have significantly shifted the balance towards increased student funding through tuition fees in conjunction with decreased government funding for higher education teaching through the Funding Councils. Starting in September 2012, universities and other HEIs in England have been able to charge students up to £9,000 per year in tuition fees for UK/EU students – fees which had previously been capped at £3,375 per year. In addition, recent years have seen a greater diversification of the higher education landscape, with more than 10,000 courses now on offer in over 160 higher education providers in the UK. The change in funding arrangements and diversification of provision has contributed to greater pressure to justify to students, employers and governments increased investment in higher education.

1.1.2. Quality and accountability in English higher education

Changes in financing of higher education have also served both to underline the importance of quality in higher education, and position student choice as a key concern for the sector. Students' expectations in terms of their course and experience are increasingly becoming a concern of universities and policy makers, and institutions have sought to provide more information to prospective students on the value of degrees. HEIs are now also required to provide more information to prospective students by making data available through 'key information sets'.²

Value is measured not only in terms of the qualification acquired, but also in terms of what is learned during higher education - including the growth in what students know/understand/can do throughout their degree programme. This has in part been addressed by measures to improve the recording of student achievement through the creation of the Higher Education Achievement Report (HEAR). The HEAR aims to provide data on student learning *per se*, rather than simply on the results of students' formal assessment.³ It does not, however, demonstrate progress made during the course of studies.

² See online, on such websites as UNISTATS, URL: <https://unistats.direct.gov.uk/>

³ Universities UK (UUK), 2007. See also Ewell, 2009.

1.1.3. International context

This interest in learning gain, which is relatively new in England,⁴ fits a broader international trend.

A prominent policy drive to develop learning outcomes (sometimes used as a synonym of learning gain, particularly in a US context) was begun with the Commission on the Future of Higher Education in the United States, also known as the Spellings Commission. The Commission's report questioned the learning outcomes from US higher education, and recommended that post-secondary institutions should measure and report 'meaningful student learning outcomes'.⁵ Further international attention was spurred with the publication of a number of books, including *Academically Adrift* by Arum and Roska (2011). This book claimed that 45 per cent of US students demonstrated no significant improvement in a range of skills – including critical thinking, complex reasoning and writing – during their first two years of college.⁶

A series of other international projects have also taken place relating to learning gain. The Organisation for Economic Co-operation and Development (OECD) aimed to find a way to compare generic and discipline-specific skills in economics and engineering across HEIs in a feasibility study called Assessment of Higher Education Learning Outcomes (AHELO). The Bologna Process, a European intergovernmental series of agreements, also covered the debate during discussions on learning outcomes.

Various international initiatives have also been launched in the past few decades to provide more evidence of students' learning. Major work on the topic is unfolding in large higher education systems, such as Canada, Germany, the People's Republic of China, Russia and the United States of America. Several other countries, such as Australia, Brazil and Colombia, have been involved in this field for a few decades.⁷

Clearly, interest in defining and measuring learning in higher education is not a new or small-scale phenomenon, nor is it shrinking in significance. This project was conducted to evaluate in what ways England might or should engage.

1.2. Key questions of the study

This report seeks to address a number of key questions under three overarching research objectives as detailed below. Box 1 details which research objectives and sections are covered by these questions.

- **Objective 1:** Shedding light on the definition of learning gain.

⁴ The issue was debated across the higher education sector in the mid-2000s (UUK, 2004).

⁵ This report led to the creation of the Voluntary System of Accountability (VSA), developed in 2007 by university leaders and higher education associations to provide transparent and comparable information to the public and policymakers on relevant aspects of higher education, including learning outcomes/learning gain. U.S. Department of Education, 2006, pp. 21–24.

⁶ The findings of Arum and Roska (2011) have led to an intense debate. For example, Astin (2011) underlined that Arum and Roska's (2011) findings were mostly about statistical significance, rather than absolute numbers. On the other hand, Pascarella et al. (2010) validated the results of this study using data from the Wabash National Study (WNS).

⁷ Coates (ed.), 2014.

- **Objective 2:** Describing the methods to measure learning gain in existence internationally as well as, if applicable, in England.
- **Objective 3:** Evaluating the suitability of these methods to the measurement of learning gain in England.

Box 1: Key questions of the report

Key questions of the report	Research objectives	Sections where these questions are addressed
a. How do we define learning gain?	Objective 1	Section 2
b. In what ways and for what purposes are methods and tools for measuring learning gain already in use in English higher education?	Objective 2	Section 4
c. What are the relative benefits of approaches to measuring generic skills independently of discipline-specific knowledge and to measuring generic skills in disciplinary contexts?	Objective 2	Section 3
d. What is the applicability of methods and tools for measuring learning gain for different identified purposes, such as to inform improvements to learning and teaching; provide information to prospective students and their advisers; investigate the impact of particular learning and teaching interventions and contextual factors; assist in international comparison; or form part of the quality assurance of learning and teaching?	Objective 3	Appendix A and Appendix C
e. What are the relevant considerations for the use or adaptation of methods or tools for measuring learning gain in an English higher education context?	Objective 3	Section 4, Appendix A and Appendix C
f. What are the relevant considerations for the design and development of robust pilot activities for measuring learning gain that might be drawn from existing methodology and literature?	Objective 3	Section 4 and Appendix C
g. What relevant lessons can be drawn from research on value added undertaken in the UK school sector?	Objective 3	Section 3

1.3. Our methodological approach

The methodology used for this research sought to best inform the authors of the substantial conceptual and methodological debates relating to learning gain, as well as practices currently in place in English HEIs to measure it. A full description of the approach undertaken is described in Appendix B of this report. In summary, the approach included:

- A review of the international and British literature relating to learning gain, including close to 200 publications, in order to understand the main debates and practices in operation. The search protocol for this review is described in Appendix B.1.
- A call for information to evaluate perceptions of learning gain and methods currently in place in England. This call resulted in 147 responses from individuals within the Network of Deputy

Vice-Chancellors, Pro Vice-Chancellors and Vice-Principals and individual academics featuring on the HEA's disciplines lists.

- Interviews with six professional bodies, in order to get an overview of perceptions and methods by discipline. The interview protocol is described in Appendix B.2.
- Interviews with staff of 13 HEIs, using the protocol described in Appendix B.3.

1.4. Outline of this report

The report begins with a summary of different conceptions of learning gain, followed by an analysis of current mechanisms in place and their potential robustness in the English higher education context. The report then seeks to provide suggestions regarding the next steps of the project; the main findings are summarised in the conclusion. The report also showcases examples of methods put forward by interviewees across HEIs in England as of potential relevance to measuring learning gain. The appendices provide more detail regarding the methodology used for this project as well as a potential evaluation framework for practitioners to assess the use of each of the presented methods to measure learning gain.

2. Definitions of learning gain

2.1. A novel concept in England

The concept of learning gain, and its definition, is in its infancy in higher education (HE) in England. Interpretations of the term remain vague and related to a number of different issues and factors. Although a total of 86 per cent of respondents to our call for information provided a definition of learning gain, the diversity of responses showed that conceptions of learning gain varied between stakeholders across HE in England. The following chapter sheds light on the definition of learning gain, and its relationship with other variants and concepts including value-added, and seeks to give an overview of how learning gain can be used.

2.2. What do students learn?

The first step to understanding learning gain is to understand what it seeks to measure: the difference in student performance between two stages of their studies. This performance – i.e. what students learn – can be thought of in a number of ways.

2.2.1. *Content knowledge*

Content knowledge represents a significant part of what students learn in higher education.⁸ This can include curriculum-based knowledge, or more generally a body of information related to a particular field. Students traditionally acquire this knowledge through their classes and other study at university.

2.2.2. *Skills and competencies*

Students also improve their skills and competencies in higher education. Skills are generally defined as the abilities or proficiencies to apply cognitive or practical knowledge and use know-how to complete tasks and solve problems in certain areas.⁹ A competence, in addition to including a skill, covers abilities, behaviours and knowledge that are fundamental to the use of such skill. A competence is a proven ability to use knowledge, skills and personal, social and/or methodological abilities, in work or study situations and in professional and personal development.¹⁰

⁸ Content knowledge is defined in Appendix D.

⁹ Evers et al., 1996: 44.

¹⁰ European Parliament, 2008. Skills and competences are further defined in Appendix D.

The development of skills and competencies during higher education study has been widely debated within the literature.¹¹ For example, the OECD defines these skills as ‘critical thinking, analytical reasoning, problem-solving, and written communication’.¹² HEFCE also considers the following additional skills as part of higher education: ‘imaginative understanding’, ‘seeing the broader perspective’, ‘seeing relationships within what students have learned’, ‘an enquiring, independent approach’, ‘critical self-awareness’, ‘learning how to learn’, and ‘personal fulfilment’.¹³ Others also seek to underline employability as a relevant skill. For example, about half of interviewees from professional bodies agreed that employability skills were a priority for them and for their stakeholders. Skills and competencies are either discipline specific or non-discipline specific.

2.2.3. *Personal development*

As well as the enhancement of skills and content knowledge, the learning gain experienced by students undertaking higher education could also be seen in terms of broader personal and psychological development. This development could be seen in terms of citizenship or confidence, or touch upon values such as respect for diversity.

2.3. How do we define learning gain?

2.3.1. *Learning gain defined as ‘distance travelled’*

Perhaps one of the simplest ways to understand learning gain is as ‘distance travelled’ or learning acquired by students at two points of their academic career. This definition relates to definitions proposed by the OECD (2012) and Rodgers (2007).

The definition is summarised by Figure 1, where learning gain (LG) is represented by the distance between points A’ and B.¹⁴ B represents the student’s performance (TS2) at a second point in time (Y_{x+1}). A’ represents the student’s performance at Y_{x+1} assuming that it is equal to his performance at the first point in time.

$$LG = B - A'$$

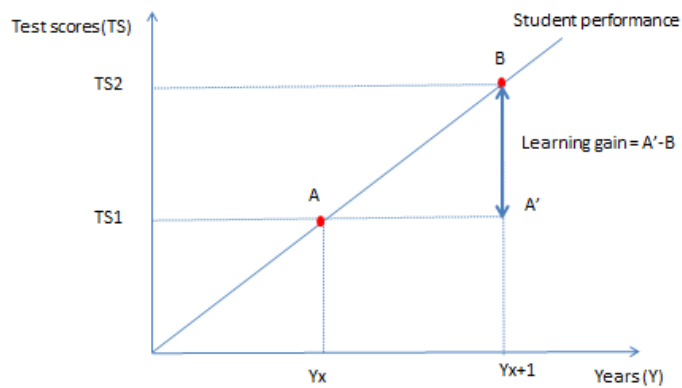
¹¹ Evers, Frederick and O’Hara, 1996.

¹² OECD, 2012:18.

¹³ HEFCE, 2014a: 28

¹⁴ The assumption here is that gain is measured across two years, using test scores.

Figure 1: Representation of learning gain



Source: Author's own

2.3.2. How is learning gain different from other similar terms?

Learning gain is often conflated with other parallel terms.

What is the difference between learning gain and learning outcomes?

The concept of learning outcomes often appears in the literature relevant to learning gain, particularly in the US. Learning outcomes are understood as expected outputs, that is, 'written statements of what a learner is expected to know, understand or be able to do at the end of a period of learning'.¹⁵ Therefore, while learning gain compares performance at two points in time, learning outcome concentrates on the output level.

What is the difference between learning gain and value added?

In a number of ways value added is a very similar concept to learning gain.¹⁶ The concept of 'value added' has a long history of development in economics and a more recent history of use in education.¹⁷ However there is a seemingly minor but very important difference, in that it uses a different statistical premise. Instead of relying on a difference between scores at two points in time, as is the case for learning gain, value added is typically measured by the difference between a student's actual test score at a second point in time and his or her *predicted* test score at a first point in time (which could be based on median - or middle - output point score achieved by others with the same or similar starting point).¹⁸

Figure 2 provides an illustration of value added measurements. Figure 2 represents a student's performance across three points. Point A is the student's expected performance at Year x. Point B represents the student's expected performance at Year Y_{x+1} based on her performance at Year x. Point B' reflects the student's actual performance at Year Y_{x+1} .

¹⁵ OECD, 2011. Learning outcomes are further defined in Appendix D.

¹⁶ UUK, 2004: 13; Harvey, 2004: 12.

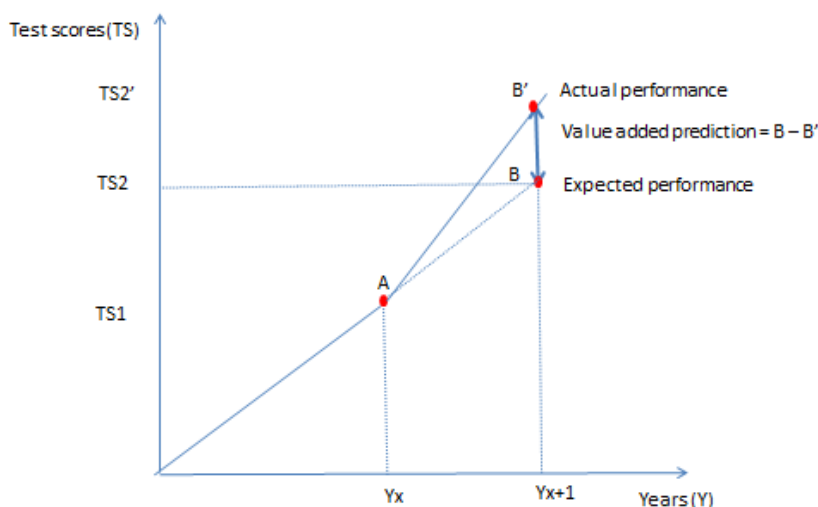
¹⁷ OECD, 2008.

¹⁸ Department for Education, 2014a.

Hence, value added (VA) could be defined as:

$$VA = B' - B$$

Figure 2: Graphic definition of value-added



Source: author's own

Measuring value added is accomplished through statistical models of various levels of complexity.¹⁹

Within England, value added modelling has gained prominence as an indicator of school performance. Indeed England has acquired a position of international leadership in developing value added measures of school education²⁰ and has pioneered research and applications of value added models in secondary education.

Box 2: Value added in secondary education in England

Value added models are used in performance tables and systems for school improvement. These models are used to inform school inspection, to help schools with particular initiatives and to inform policy makers of the effectiveness of particular schools in England. The Department for Education uses value added measures to compare schools' and colleges' ability to help students improve progression.

An example of a value added model includes the 'Level 3 Value'.²¹ This model measures value added by comparing a student's actual outcome at Level 3 qualification to the student's estimated outcome (the outcome achieved by students of similar ability nation-wide). The difference between the student's actual and estimated performance gives the student a value added score in the Level 3 qualification.

¹⁹ For more information, see: Doran & Izumi, 2004; Doran & Lockwood, 2006; Tekwe et al. 2004; Wainer, 2004; Wiley, 2006; Fulcher & Willse, 2007; Steedle, 2012; Hoon-Ho & Lalancette, 2013; Kim & Lancette, 2013.

²⁰ Tymms, 1999; Tymms & Dean, 2004.

²¹ Level 3 value added models are further defined in Appendix D.

Students' value added are then averaged across schools and colleges at Level 3 qualification to help identify the schools and colleges that are helping their students make more or less progress than average.²²

Original value added models did not control for factors such as children's circumstances and prior performance. Hence, the Office for Standards in Education, Children's Services and Skills (Ofsted), the UK school inspector and regulator, suggested a method called Contextual Value Added (CVA) scores.²³ The aim of this method was to show how much progress students had made from the start of a 'key stage' to the end, as a measure of educational quality.

But CVA models have been criticised because they exclude some potentially relevant information, such as the number of privately educated pupils, contextual information on background factors or direct measures of social class, or information linking results to teachers or subject departments. In addition, CVA scores estimate how effective schools were for the cohort of children who have already completed their GCSE examinations.²⁴ This can lead to issues because the CVA cannot predict how effective the school will be for parents selecting schools for their children, who are moving through the education system on average seven years after the cohort of children measured in the CVA.²⁵

Additional criticisms of value added in secondary education include the fact that the statistical model used in secondary education to create school performance tables has no absolute interpretation: schools are compared with an average performance, but there is little indication of how to interpret differences in terms of points. And educational institutions were prone to 'gaming', i.e. favouring student groups that provide them with the most significant value added. The intensity of gaming is assumed to have been lower when value added was used in a formative way in the experimental case of the Hampshire value added project of 1996. During this project, value added scores remained confidential to the school and the local education authority.²⁶

It is important to note that unintended consequences observed in the context of value added measurements in secondary education, such as 'gaming', could also apply to the context of higher education if the objective is to use the measure for accountability purposes. HEIs could concentrate on students who increase their learning gain scores, rather than concentrate on achieving the provision of a quality student experience for the broadest student population. It would therefore be necessary for researchers and policy makers to be sensitive in developing and applying these measures, particularly with regard to comparison between institutions.

2.4. What can measures of learning gain be used for?

Measuring learning gain could fit one of a number of broad purposes.

2.4.1. *Enhancing teaching and learning and informing students*

A first purpose is formative. Using learning gain could help to facilitate the enhancement of teaching and learning. This includes using data and predictions on learning gain to help HEIs exchange best practices,

²² Department for Education, nd.

²³ Contextual Value Added models are further defined in Appendix D.

²⁴ Ray, 2006.

²⁵ Leckie & Goldstein, 2011.

²⁶ Goldstein et al., 2014.

support students in their learning, or institutions in their decisions on applicants. This corresponds to a growing trend for learning analytics in the sector.²⁷

Indicators of learning gain could also help to inform student choice and increase transparency in terms of the outputs of the higher education sector. A majority of respondents to the call for information emphasised the usefulness of using learning gain in this manner. Respondents detailed that learning gain measures could be used to help verify that students have learned or achieved what their degree was designed for, and could be utilised to explore which approaches to learning and teaching are more effective (in order to amend curricula or teaching methods and enhance accountability).

2.4.2. Increasing the accountability of higher education

Indicators and measures of learning gain could also be used to help increase accountability in higher education (as they have in secondary education, for example) and could help governments justify value for money of investments in higher education to both students and other stakeholders.

The evidence gathered in the call for information exercise showed that this specific use of learning gain measurement is a source of concern for 18 per cent of respondents. Respondents were concerned by the potential unintended consequences of learning gain, including potentially discouraging effects on learning itself. Assessing institutions that serve different missions according to a similar standard was also a source for concern. For example, students from different institutions, with differing missions will have very different learning experiences. For example a student enrolled in a research-intensive institution on a highly selective courses, will have a very different experience to a student enrolled at teaching-intensive institution. As such it is difficult to compare learning in this context.²⁸

For similar reasons, learning gain should not be considered as a sole determinant of the quality of higher education. Quality is relative to the many missions and goals that higher education has - there is both a broader conception of what is meant by quality of learning than solely learning gain, and there are more aspects to what is understood as quality of higher education in institutions, than just quality learning.²⁹ In addition, while learning gain is important, it constitutes only one aspect of several key missions of higher education, including, for example, research productivity, innovation and improving the capacity and diversity of the workforce for the labour market.

²⁷ The term 'learning analytics' is defined in Appendix D.

²⁸ This diversity of missions implies that policy makers should ideally work towards an alignment between national objectives and those of the institutions, according to Miller (forthcoming).

²⁹ Gibbs, 2010.

3. How can learning gain be measured?

A number of methods could be used to measure learning gain. In order to analyse which methods would be the most appropriate and robust, the research team first identified a number of basic properties that learning gain methods should fulfil. The team then devised a series of questions to evaluate these methods according to these properties.

3.1. What attributes contribute to a robust method to measure learning gain?

Robust methods to measure learning gain would need to fulfil four properties described below.

3.1.1. *Longitudinal or cross-sectional design*

A basic requirement for measuring learning gain (within one institution or group of students) is that the methods employed produce comparable data. Perhaps the most robust method to achieve this is through longitudinal data, i.e. data on the same group of students over at least two points in time. These assessment points might for example be at entry into higher education and graduation. Longitudinal designs could cover several types of data, such as student grades, scores in other assessments or self-reports.

An alternative to this approach, which is possibly more practical in certain circumstances, would be to compare a cross-section of student cohorts over a number of years. However this method may not be as reliable as longitudinal data, because it would be more difficult to control for differences since the two groups of students would have different characteristics, and would be performing under different conditions.³⁰

3.1.2. *Validity*

An important property of any statistical method is its validity. That is to say, does the method employed actually measure what it intends to measure?³¹ Validity is of particular concern in analysis of the concept of learning gain. Students learn a broad range of valuable things during their time in higher education, including a considerable amount outside of their day-to-day attendance of their course or degree programme. Indeed, learning occurring outside of studies may be difficult to capture through formal

³⁰ Cunha & Miller, 2012.

³¹ Validity is further defined in Appendix D.

assessments or tests, but remains an important aspect of the value provided by higher education (according to UUK, 2004, many aspects of, for example, employability fall into this category).

Validity could also be affected by a number of other ‘omitted’ variables. If not properly taken into account, these variables will distort the interpretation of the results of a particular assessment method. These omitted variables include student characteristics,³² motivation (some students may lack the motivation to do well on certain tests or to fill in surveys accurately), student mobility, dropout, course changes and other factors, each of which can make it difficult to keep track of students. Finally, changes in learning gain scores could be due to changes in the characteristics of institutions, which could include change to the assessment instruments, changes in each institution’s administrative or financial situation and/or changes in student body composition.³³

3.1.3. Representativeness

One of the most common issues related to representativeness is missing data. Missing data make certain methods unrepresentative of the student population, or distort estimates in one direction.

Missing data is a problem particularly if respondents are asked to take part in the measurement on an opt-in or voluntary basis and if the outcome assessment is low-stake.

Some statistical manipulations exist to correct for missing data, including imputation, exclusion and weighting. Another way to solve the problem of missing data is to concentrate on a subset of the student population, i.e. a sample. However, models can overestimate or underestimate programme or sampling effects, and sampling errors can occur. Therefore a third solution could include an administrative integration of learning gain measurements to university-wide student data systems. Section 4 provides further details regarding these approaches.

3.1.4. Comparability

Comparison between student performance, disciplines, institutions and countries are all potentially valid uses of a measure of learning gain.

Comparability across disciplines

A key dimension to this debate is concerned with whether learning gain can be compared across disciplines, given the different requirements and professional practices of each discipline.

In general, it is easier to measure knowledge-specific skill in one discipline rather than several. The measurement of learning gain is especially facilitated by disciplines which are vocational, applied or rely on clearly defined professional standards, given that these disciplines allow for a clear definition of the knowledge and skills students need to acquire.

Several discipline-specific tests exist. These tests include, for example, Progress Testing (a longitudinal test used in medical sciences which will be further covered in Section 4) to assess a mix of generic and knowledge-specific skills. Within the literature there has been an intense debate as to whether generic

³² Pascarella et al, 2010: 21; Cunha & Miller, 2012.

skills and competencies are comparable across disciplines. The literature also debates the definition of generic skills – are these skills ‘general skills’ and hence the same across disciplines? Are they less amenable to formal educational intervention in order to be improved? For example, one could ask, should communication skills be assessed according to the same standard between a medical doctor and a barrister? In addition, the significant differences in student performance across subject areas also reinforce the case against a cross-disciplinary measurement.³⁴

Despite these differences, there is some between the Subject Benchmark Statements published by the Quality Assurance Agency (critical thinking, or using data, for instance). In addition, nearly all discipline-specific tests agree on key content areas: English, maths, science, reading, writing, critical thinking, analytical reasoning, etc.³⁵ Interviewees from different professional bodies appeared to agree on their definition of generic (or employability) skills, which tended to include a mix of numeracy, literacy, problem-solving, critical thinking, or project management skills.

Non-skills-related tests also exist to measure performance across disciplines. For example, several employers have increasingly used psychometric measurements and personality tests.³⁶ (Further examples of non-discipline-specific tests are available in Section 4).

Comparability across institutions

Learning gain could provide more information on the progress made by students (and hence presumably educational effectiveness) than output measures, which reach limits when comparing institutions. For example, an output measure based on student achievement would assess an institution where students achieve top grades highly. However, these students could have been recruited at an already high level, which does not allow researchers to infer much regarding the effectiveness of the learning undertaken, for example.

In order to address this, it would be necessary to take into account contextual and demographic characteristics, though - as demonstrated in Box 2 - the models currently being used in England suggest that it may be difficult and controversial to determine the exact characteristics to take into account.

In addition, this comparison remains challenging in higher education. There is more diversity of paths in higher education than in secondary education. For example, students can pass through numerous subjects, with dozens of teachers and many students change directions after commencing their study, making it even more difficult to record learning-relevant settings and disciplinary areas. For this reason, some methods only measure learning gain at an aggregate level across departments or faculties.

Comparisons between institutions can also be difficult because of the different disciplinary mix between institutions. In addition, contrary to the situation regarding other levels of education in England, there is no national curriculum for higher education. Any measure of learning gain would need to acknowledge this distinction.

³⁴ Banta & Pike, 2012; Steedle & Bradley, 2012.

³⁵ Millett et al., 2007

³⁶ These tests include, for example, the Occupational Personality Questionnaire (OPQ) or the Myers-Briggs Type Indicator (MBTI).

Comparability across countries

Previous attempts have also been made to compare what students learn across countries, for example the OECD's AHELO study. Finding a method that is suited to international comparisons and interpreting the results of this method can be difficult, given the differences in the educational context in different countries. For example, the United States has followed a liberal arts education path which is used only in a minority of institutions in Europe. In contrast, higher education in England, follows a similar format to that in many other European countries over the past decades (i.e. through the Bologna process and European Higher Education Area), which favours early specialisation. Higher education in the United States starts with a general education period of about two years, with students specialising later than in typical undergraduate degrees in England. The general skills measured by several US-based skills tests are hence designed to suit the general education model, and there is some uncertainty regarding their applicability to England.

Concerns about the transferability of methods also relate to the potential misuse of results. The higher education sector is particularly wary of comparative exercises, such as international university rankings.³⁷ Section 5 provides further details regarding how these methods have been applied in the English context.

3.2. What are the key questions to bear in mind when assessing learning gain methods?

The research team, in conjunction with the steering group and the delegates of the expert workshop, have developed a mapping framework, presented in Table 1 below. The objectives of this framework are to provide a clear set of questions, based on the four properties identified in Section 3.1, to evaluate the 14 methods to measure learning gain identified in this project. The methods have been evaluated by the research team using the empirical material gathered in the project. Section 4 provides more details regarding the evaluation of each method.

The first properties test the validity of the measure as well as its comparability. To test these properties, the framework includes such questions as whether the test is objective (based on standardised grades or tests) or subjective (i.e. based on self-reported surveys, internal grades or portfolios) and whether these methods are internally or externally validated. The assessment shows that objective and externally validated methods (such as standardised tests and shared assessments) have higher validity than subjective and internally valid methods (such as student portfolios and surveys) and that their comparative power across units of analysis may also be higher.

An additional property that the framework seeks to assess is whether the method gathers longitudinal data. As such, the framework includes questions about whether the method aims to measure learning gain or value added, or whether it simply acts as a measure of output or input. Methods which measure a difference longitudinally tend to be less common in England.

Finally, the framework tests how representative a particular method is. This set of questions shows that several methods have been criticised on the basis of low sample size.

³⁷ Olds & Robertson, 2011.

Table 1: Preliminary assessment of models considered within the study, based on characteristics of the methods themselves³⁸

				Grades			Mix		Surveys				Standardised tests			Qualitative		
				Comparing scores	Predicted pass rate	Output ³⁹	APTT	ALIS	NSS	UKES	Careers Registration	Other (skills audit)	CLA	MAPP	Progress Testing	PDP	Skills award	
				Measurement level														
Validity and comparability	Objective and externally validated	External tests	Test that is standardised across institutions and implemented as a separated or stand-alone activity			✓								✓	✓	✓		
		Shared assessment	Externally validated assessment embedded into everyday assessment practices													✓		
	Objective and internally validated	Internal grades	Internal grades captured via everyday assessment practice using a wide range of assessment resources and practices	✓	✓	✓	✓	✓										
	Subjective and internally	Portfolios	Rich qualitative artefacts that document a student's output during time spent in the course				✓						✓				✓	

³⁸ Please note that this table is an assessment of how the models considered within the study are currently implemented. In some cases changes could be made. Practices highlighted in yellow are international and were not reported in the list of sample institutions. A checkmark (✓) indicates that an answer to the question is included in the framework. A cross (✗) indicates that the practice also answers the question, but that this answer may be a cause for concern.

³⁹ Output indicators include, for example, earning differentials, expected degree class and employment rates.

				Grades			Mix		Surveys				Standardised tests			Qualitative		
				Comparing scores	Predicted pass rate	Output ³⁹	APTT	ALIS	NSS	UKES	Careers Registration	Other (skills audit)	CLA	MAPP	Progress Testing	PDP	Skills award	
Measurement level																		
	validated	Student surveys	Student self-reports on learning or development gain, captured using surveys						✓	✓	✓						✓	
Statistical analysis																		
Longitudinal data	Analysis of difference	Learning gain	Analysis of difference between performance over time without modelling expectations	✓		✓					✓			✓	✓			
		Value added	Analysis of expected performance based on individual characteristics, such as prior achievement		✓		✓					✓						
		Dimension of analysis	Difference in content knowledge		✓	✓	✓		✓						✓			
			Difference regarding skills (generic or specific)					✓	✓		✓			✓	✓	✓	✓	✓
			Difference regarding personal development					✓		✓	✓	✓			✓	✓		
	Analysis of point in time	Analysis of outcomes	Analysis of measures of learning outcome only (not including inference made from qualification-level accreditation)	✓		✓	✓		✓					✓				
		Analysis of inputs	Analysis of data used for admissions and other priors available to the institution		✓							✓						
Response rate		Has the student response rate been recognised as 'appropriate'?	✓	✓		✓	✓				✓			✓				

				Grades			Mix		Surveys				Standardised tests			Qualitative	
				Comparing scores	Predicted pass rate	Output ³⁹	APTT	ALIS	NSS	UKES	Careers Registration	Other (skills audit)	CLA	MAPP	Progress Testing	PDP	Skills award
Measurement level																	
			Has the literature flagged any 'bias' in patterns of respondents?			x			x	x		x	x		x	x	

4. Overview of methods

This section summarises 14 methods which could be used to measure learning gain that the research team identified, based on the call for information and interviews. These methods are summarised in five groups: grades, surveys, standardised tests, qualitative methods and other mixed methods. The focus of current projects in England is mostly on qualitative approaches, although examples of most of these methods have been found in England. This summary is illustrative of the main types of methods rather than exhaustive.

4.1. Grades

Using grades constitutes an intuitive method to measure learning gain. This method relies on students' grades being comparable at two points in time and across units of analysis, for example, in the context of degree exams.

4.1.1. Assessment of the use of grades

Grades provide a way to directly measure learning gain that relies on existing institutional capacity since all HEIs issue grades. Grades theoretically operate within the framework of subject benchmarks which set out expectations about standards of degrees across a range of subject areas.

The main difficulty with the use of grades relates to comparability across time and between disciplines.

Is it possible to compare grades across two points in time?

Comparison of grade results at two points in time can be difficult, because standards are assessed using different metrics, for example at entry and final exam levels.

At the level of entry, grades are contingent on different requirements by institutions, including, for example, A levels, AS levels, Scottish Highers, and BTEC National diplomas. The UCAS 'tariff scores' specify how many points are required for entry to study for specific degrees or in certain institutions. But tariff scores are not universally available entry measures.

At the exit level, final exams are typically graded across a five-level scale. University exit grades or exit percentages (first, 2:1, 2:2, third and fail) remain very general categories, since most of the grade distribution is concentrated along two of the five levels on the scale (the 2:1 and 2:2 axes). The fact that these categories are broad means that it would be difficult to find much variance between entry and final grades.

Some institutions, particularly those in England, use a credit system, based on the completion of a module or course. However, the award of credit does not provide detailed information regarding the quality of the student's performance.⁴⁰

Given that graduation metrics may be difficult to compare, some have suggested using output indicators other than grades. These proxies could include earning differentials or employment rates. However, the use of these indicators has been criticised in England. For example, entry profiles of individual HEIs may be too different for comparisons to be valid when comparing institutions with reference to earning differentials.⁴¹ In addition, subject variation, institutional status and reputation have an influence on recruitment and employment patterns, thereby making what is being measured unclear, and assessing graduate employment would not be able to account for graduates who are under-employed.^{42 43}

4.1.2. Examples of use of grades in England

Three of the English HEIs interviewed mentioned relevant methods with regard to grades and similar proxies as ways to assess a student's progress and, by extension, learning gain. These institutions incorporate grades in a variety of ways. These can include, for example, comparing a standardised grade called the Grade Point Average (GPA; which has been piloted at a number of institutions through a national project managed by the Higher Education Academy, and is currently in use at institutions including Oxford Brookes University) or using grades to issue predictions on the student's likelihood of passing a module (such as at the Open University, called the Predicted pass rate model). Predicted pass rate models have also been reported at the University of Birmingham and at Queen Mary University. Vignette 1 provides an example of how a particular institution (the Open University) uses grades to predict the pass rate of students.

Vignette 1: Predicted pass rate

⁴⁰ UUK, 2004: 34.

⁴¹ Allen, 2005.

⁴² Hazelkorn (forthcoming): Chapter 2.

⁴³ Smith et al., 2000; Dill & Soo, 2005: 509.

In the early 2000s the Open University (OU) set up the Predicted pass rate methodology – an approach which relies on grades to predict the likelihood of students passing a module. Although this methodology was not designed to measure learning gain, it could be seen as measuring various elements relevant to learning gain, since it incorporates an element of distance travelled by predicting future performance. This measure operates by producing coefficients of various features that would affect a student's chance of passing a given module, based on personal demographic characteristics, characteristics of the module, and the student's performance on assessment on previous OU modules (if any). This information is used by OU to produce a predicted pass rate for a module, with the data produced separated out by students' experience and level (e.g. undergraduate or postgraduate). These predictions are used by the institutions as part of the review process. This review includes a monitoring of modules which require attention (using an amber light) as well as a graphical representation of the most important factors that are shown to predict students' pass rates as part of a broad-scale review. The predicted pass rates can also be used to compare faculties. The results are shared only within the university and not intended for diffusion to a broader audience.

4.2. Surveys

A second way to measure learning gain could include student surveys.

4.2.1. *Assessment of the use of surveys*

Student surveys are a commonly used method to obtain data from students, including asking them about their perceived gains during their higher education programme. As such, using data extracted from particular questions of existing surveys in operation in England could provide a useful way to measure learning gain.

There are, however, a number of considerations to bear in mind. Surveys are subjective, and rely on students' self-reports, which could be influenced by a variety of factors, and students could be subject to incentives to misrepresent their learning, in order to portray their institutions in a positive light. Some of the literature questions the correlation between students' self-reports and objective measures of learning,⁴⁴ and details that survey data would require further tests to be validated. This validation process could include comparing the results of self-reports against achievement data or measures of students' performance. In addition, existing surveys would need to be undertaken at more than one point in time, and be sufficiently representative of the pool of students within an institution, to be considered as valid. Finally, the profile of answers (which could be Likert scale) means that variance may appear as relatively limited.⁴⁵

4.2.2. *Examples of use in England*

This section analyses a number of the main student surveys currently in use in England, including student engagement and student experience surveys,⁴⁶ as well as smaller scale examples, including career readiness and skills audit surveys.

⁴⁴ Bowman, 2010; OECD, 2012.

⁴⁵ Term defined in Appendix D.

⁴⁶ Terms defined in Appendix D.

Student experience and student satisfaction surveys

A number of surveys ask students about their experience and satisfaction with the provision of higher education. These surveys include the nation-wide National Student Survey (NSS) – a large scale survey implemented in the UK, which has had a response rate of above 50 per cent since 2008.⁴⁷

These surveys are not primarily concerned with the measurement of learning gain, but do include a small number of relevant questions. For example, the current version of the NSS has three questions/items related to personal development, which could be used as a proxy measure of learning gain. These questions are included in Box 3 below.

Box 3: Example of relevant questions in the NSS

(For each question, show the extent of your agreement or disagreement by putting a cross in the one box which best reflects your current view of the course as whole. Categories of answers are 'definitely agree', 'mostly agree', 'neither agree nor disagree', 'mostly disagree', 'definitely disagree', and 'not applicable')

Personal development

Q19. The course has helped me to present myself with confidence

Q20. My communication skills have improved

Q21. As a result of the course, I feel confident in tackling unfamiliar problems.

Although relevant, these questions provide a relatively narrow view of learning gain and do not measure skills, for example. In addition, the NSS is designed to be administered with students only once, at the end of a course of studies. To measure learning gain more effectively, the survey would have to be modified to be administered at several points in the course of a student's studies.

HEFCE's (2014) review of the NSS made recommendation that we are currently considering and testing, including the survey measuring learning gain more directly, through the inclusion of questions addressing student engagement⁴⁸ The NSS is currently under review once more, and a new questionnaire is due to be issued in 2017.

Student engagement surveys

Student engagement surveys constitute another type of survey which could be considered to measure learning gain. The main student engagement survey in England is the UK Engagement Survey (UKES), which is based on the US National Survey of Student Engagement (NSSE)⁴⁹. UKES primarily aims to

⁴⁷ HEFCE, 2008. Other experience surveys used internationally include the Student Experience Survey in the Research University, coordinated by the University of California, Berkeley, for a network of research universities; the University Experience Survey (UES). See Douglass et al. (2012), Graduate Careers Australia (2010) and Radloff et al. (2012). A review of the role of the HEA on the NSS is provided by Ramsden (2008) and Buckley (2012).

⁴⁸ Callender et al., 2014: 27.

⁴⁹ The NSSE has an international reach. Australia runs the Australian Survey of Student Engagement (Radloff et al. 2012). Student engagement surveys have expanded to other countries, including Ireland, where the Irish Survey of Student engagement (ISSE) was launched in 2013. In China, tools used to measure student engagement include the

provide institutions with data about their students' levels of engagement and inform efforts to enhance learning and teaching. As such it is not primarily designed as a measure of learning gain.

UKES is smaller in scale than the NSS. The survey engaged 32 institutions in 2014⁵⁰, and is currently entering its third administration (after the two pilot administrations, in 2013 and 2014). UKES includes one question (of 12) which covers skills development, as detailed in Box 4 below.

Box 4: Questions relevant to the measurement of knowledge, skills and personal development in UKES

<p>Skills Development</p> <p>How much has your experience at this institution contributed to your knowledge, skills and personal development in the following areas? (Response categories: very much/quite a bit/some/very little)</p> <ol style="list-style-type: none">1. Writing clearly and effectively2. Speaking clearly and effectively3. Thinking critically and analytically4. Becoming an independent learner5. Being innovative and creative6. Working effectively with others7. Developing or clarifying personal values or ethics8. Understanding people of other backgrounds9. Being an informed and active citizen10. Analysing numerical and statistical information11. Acquiring job- or work-related knowledge and skills12. Solving complex real-world problems.

The broader UKES questionnaire also includes questions related to conceptions of skills, notions of citizenship, ethical behaviour and diversity, and therefore seeks to highlight the broad relevance of higher education to an individual's development.

In order to contend as a valid measure of learning gain, UKES would need to address issues of comparability, representativeness and validity. Issues of longitudinal comparability would also need to be

National Survey of Student Engagement China (NSSE-C) and the High School Survey of Student Engagement (HSSE-C). NSSE-C was, for example, used by Tsinghua University to make changes and improvements, notably to the student–faculty relationship. For an international overview, see Coates & McCormick (2014).

⁵⁰ Buckley, 2014.

addressed, since UKES is currently administered once across cohorts of students, and at a different point in time in different institutions (the survey window runs from the start of February to mid-June).

Comparability across groups of students would also need to be controlled for. UKES collects uncoordinated responses from students of all years (which could result in variations across types of students by year across institutions).

In addition, the response rate for UKES remains relatively low, averaging 13 per cent in the second pilot⁵¹. Taking UKES forward would therefore require institutions to increase representativeness, including, for example, linking survey response to registration.

In terms of validity, there is some evidence of a relationship between engagement and learning gain. In the US, researchers have found a relationship between NSSE benchmark scales and self-reports on engagement, and some dimensions of problem solving and personal development.⁵² In England, the cognitive testing report for the first pilot of UKES, held in 2013, showed that students who were most likely to have a positive learning experience were strongly engaged with their university activities.⁵³ Because this evidence tends to compare self-reports, it would, as discussed in earlier sections, need to be further tested through a correlation with objective measures of student attainment in order to be considered as valid in the English context. (The relationship has been found to be weak in the US context: Ewell (2002) showed that there was only a modest statistical association between ACT CAAP test scores⁵⁴ and student engagement, when using data from six South Dakota public colleges and universities, for example).

Vignette 2 below presents the example of the use of UKES in a HEI, namely, Oxford Brookes University.

Vignette 1: UK Engagement Survey

UKES was piloted in Oxford Brookes in the latter half of the 2013-14 academic year on first- and second-year students (not final-year students, in order not to be confused with the NSS). The implementation of UKES ran alongside the implementation of other initiatives, including the Academic Progress Tracking Tool (APTT) and the GPA. UKES has allowed for a comparison of cohorts across disciplines. Oxford Brookes uses UKES on a voluntary basis as part of the national HEA initiative, although the HEA does not define how to use NSSE.

The HEA then released a comparison with other institutions taking part in the pilot. According to an interviewee, UKES, although it does not measure learning *per se*, helps one to understand the factors which facilitate learning and engagement. Oxford Brookes University has approved the use of UKES for implementation in specific programmes to improve programmes and facilitate student learning.

⁵¹ Buckley, 2014: 16.

⁵² Pascarella et al., 2010; Callender et al., 2014; Tam, 2004; and publications from the Wabash National Study of Liberal Arts Education.

⁵³ See also correlation tables in the data annex of Buckley, 2014.

⁵⁴ ACT CAAP is a standardised test that measures academic achievement in general education, generally at the end of the second year of study at university in the United States (American College Test, 2014).

Surveys of workplace skills and career readiness

Another type of survey concentrates on measuring the acquisition of skills for employability. Comparing workplace skills could provide a relevant dimension to measure learning gain. One example from England is Careers Registration: a careers readiness survey in operation at the University of Leeds. The applicability of developing a similar version of this survey is currently being investigated in a number of other institutions, including the University of Nottingham and King's College London.

The Careers Registration survey has an advantage in that it is repeated across all students at the same point in time every year. It is also designed to limit the burden on students and provide them with an incentive to reply (in the form of built-in careers advice).

In order to address learning gain, the Careers Registration survey would need to be further validated, for example through a correlation between this survey and attainment data, or against student employment rates after graduation. In addition, the survey reaches some limits in terms of comparability across disciplines, because although it is designed to be a cross-disciplinary measurement, a number of disciplines require a tailored approach to the survey (e.g. in medicine, because students have already selected a general career path and issues of choice are more about specialisms). The small sample size of institutions using this method means that inter-institutional comparability is limited. Finally, career readiness is not linear and its results need to be interpreted carefully. Vignette 3 below summarises the Careers Registration survey.

Vignette 2: Careers Registration survey

The University of Leeds launched the Careers Registration survey in 2011–12 as part of its employability strategy. This survey obtained two awards from the Association of Graduate Careers Advisory Services.

The Careers Registration survey assesses the degree of students' preparedness for their careers along three dimensions: whether students have made a decision, whether they are prepared for their chosen career, and to what extent they can compete – the assumption being that students will reach the 'compete' phase as they reach the completion of their studies. The questionnaire includes 10 common sense questions about career planning (put to students as statements which resemble a Likert scale) and asks students about their work experience.

The assumption of this questionnaire is that career motivation and career awareness were the main drivers for career success, above the accumulation of skills.

Formally, this is undertaken via a survey repeated year-on-year between August and September, which makes it possible to gain an idea of the distance travelled. By focusing on employability and career readiness, the survey touches on such issues as the type of generic skills students have and enables comparison across disciplines.

The survey is linked to students' registration, and as such includes data for all students. The survey is used for a formative purpose. Managed by the careers centre, its results are discussed with each faculty as part of their employability working groups. Students' personal tutors are also able to access the data and students receive online recommendations based on their replies to the survey.

Skills audit

Another self-assessment survey instrument is the skills audit method, which aims to get students to both self-assess their skills and to identify gaps.

Some advantages of this approach include the fact that skills audits provide a useful source of data regarding how to design degree programmes to encourage skills awareness and skills acquisition and they encourage students to reflect on their skills acquisition in a comprehensive manner. Students may also provide informative material for teachers to adapt their courses to students' needs and to identify learning gaps.

However, the data produced in the skills audit serves mainly as a working document between the supervisor and the student, rather than as a robust measure to compare gain year-on-year, and it may be difficult therefore to achieve comparability or validity. Vignette 4 summarises this method in place at Durham University.

Vignette 3: Skills audit

At Durham University undergraduate students undergo an annual skills audit process, which seeks to assess and provide students with information on how their skills have progressed over the course of their academic cycle.

This approach begins by sending students a booklet before they arrive at the university, which contains information on Durham University 'graduate qualities'. These include discipline-specific skills, study skills, interpersonal skills, entrepreneurial behaviours, personal effectiveness and understanding the world.

Students are then given the opportunity to take a 'skills audit' in their own time. Through the university's virtual learning environment, students rate their confidence in the given skills area, prompted by a series of questions. This generates a personalised skills list that identifies areas that require further development, as well as highlighting available resources to support their personal development.

Once students are assigned an academic adviser, they have an opportunity to discuss potential gaps in their skills at an early stage in their academic career, and thereby start thinking about their skills. These conversations enable students to prioritise their training and development. Although there appears to be an uneven spread of engagement across departments, students from every department take part in the skills audit. It is also deemed fairly 'light touch', while covering a wide range of skills.

Postgraduate students are also offered a similar skills analysis process which involves a discussion with the supervisory team about four areas: discipline knowledge, advanced research skills, general/transferable skills and career plans.

Skills needs analysis was the object of a large scale nation-wide research project five years ago (sponsored by the UK Research Councils). According to colleagues interviewed in the research for this scoping study, the amount of funds invested in research in this area stimulated a 'feeling of change' in the sector and contributed to the adoption of the initiative at Durham University.

4.3. Standardised tests

Standardised tests can be used to measure the acquisition of certain skills or other elements of personal development or content knowledge, all of which may be seen in generic or specialised terms. Standardised tests are often used as a way to obtain data which are comparable across years as well as institutions.

Using standardised tests is a much more widespread practice in the US than in England. The company Educational Testing Service (1993) provided a review of 237 tests used in the US. Several tests exist to measure 'distance travelled', although many of the registered 237 tests occur at one point in time and may not entirely be suited to measuring learning gains. The introduction of standardised tests has previously

been investigated in England, albeit as an entry test rather than a measure of learning gain. For example, in the past few years, BIS looked into the feasibility of the creation of a standardised test for entry into higher education. The BIS study found that a standardised aptitude test such as the Scholastic Aptitude Test (SAT) had some predictive power, but did not add any additional information over and above GCSEs and A-levels.⁵⁵

4.3.1. Assessment of standardised tests

Standardised tests are objective in the sense that they assess the acquisition of certain attributes rather than ask students to report how they think they have acquired these attributes. These tests also have the advantage of being tailored to measure the acquisition of defined attributes and therefore may provide a broader and more accurate measure of learning gain than grades, which focus on content knowledge. In addition, these tests fulfil different purposes, since they can be administered to students either at the beginning or end of their degree programme.

Standardised tests can either be discipline specific or non-discipline specific. Discipline-specific tests are the only type that the research team found in use in higher education in England.

Assessment of discipline-specific standardised tests

Discipline-specific standardised tests tend to achieve a higher level of precision and hence validity than more generalised tests, but their outputs are difficult to compare between disciplines. Should broader comparability be desirable, work would need to be undertaken to explore the possibility of aggregating results up to a more generic level.

A number of countries are investigating the development of discipline-specific tests to measure learning gain. For example in Germany, the German Federal Ministry for Education and Research has developed a project entitled *Kompetenzen im Hochschulsektor*. This project provides funding for 23 academic alliances, and aims to model and measure competencies across a set of disciplines. These include teacher training in the STEM subjects (science, technology, engineering and mathematics), as well as educational sciences, economic sciences and social sciences. The OECD project AHELO also looked into the feasibility of developing a comparable measure of learning outcomes, notably in civil engineering and economics. In addition, there are some examples of nations implementing generic skills and discipline-specific tests, including for example Colombia, which administered a test as part of the national end-of-tertiary-education exam, called SABER PRO.⁵⁶

Example of use in England

The section below presents one example of a discipline-specific standardised test used in England in medical sciences, known as Progress Testing. Progress Testing, which originated in the Netherlands and USA relies on the repetition of a similar test to students four times a year, every year. The test works on the principle that it is expected that first-year students will fare less well in the exam, but that these

⁵⁵ BIS, 2010: 4; Kirkup et al., 2010.

⁵⁶ OECD 2012: 166.

students will progress throughout their degree and achieve higher marks by the end of the degree programme.

In the UK Progress Testing is in use in a number of institutions, including the Universities of Cardiff, Swansea and Manchester, as well as the National Board of Medicine Examiners (NBME) for St. Bartholomew's Hospital (Barts), St. George's Hospital Medical School (St. George's London), Leeds University and Queens University Belfast. It is also in the process of being implemented at the University of Liverpool, but each of these institutions has their own rather than a common test.

Progress Testing has the advantage of measuring real, rather than rote, learning and allows for students to understand their 'distance travelled' from their first year. It has been detailed as well that Progress Testing may also help students to take on more responsibility in their learning. However, the further application of Progress Testing methods would need further investigation, including to better understand if teachers 'taught to the test' rather than focusing on developing the students' broad knowledge acquisition.

One debate around this test includes whether Progress Testing could be extended to other disciplines. According to some interviewees, parts of the standardised tests used in medical sciences, such as the situational judgement test, could be used in other disciplines. However, further consideration would need to be given to the ways in which Progress Testing would need to be adapted to fit a modular programme with greater optional choice.

To be an effective measurement of learning gain, Progress Testing may necessitate that the test is altered to better fit around a modular based programme.

The implementation of a national progress test would need a wide bank of questions to be set for students, which would be outside the capacity of single institutions. In addition, given that each of the 33 medical schools in the UK has its own exam, Progress Testing may not be suited for adoption across a wide group of medical schools (if this was the intent of the measure).

An example of the use of a discipline-specific test at Peninsula College of Medicine and Dentistry in Plymouth is given in Vignette 5 below.

Vignette 4: Progress Testing

This vignette describes the use of Progress Testing at Peninsula College of Medicine and Dentistry. This test runs alongside other medical tests, including the Objective Structured Clinical Examination, where students are assessed based on their ability to perform clinical tasks in 20–25 work stations.

At Peninsula School of Medicine, all students studying medical sciences (from first to fifth year) take the same test at the same times each year. The test uses 125-item tests and is administered four times a year (which requires a different 125 items each time).

Questions are drawn from a question bank of between 3,500 and 5,000 items. To create a database of questions, Peninsula went into a partnership with the Medical Schools Council and formed an assessment alliance with a number of medical schools who started to write items and share them. They also bought approximately 5,000 items from a consortium based in the medical school in Hong Kong – although the items had a wide variation in standards – and they employed doctors to create questions in order to build up a bank that they could draw from year-on-year. All questions are designed as multiple choice questions; Peninsula College of Medicine and Dentistry uses items that are written as ‘single best answer’, considered best practice in multiple-choice questions, as opposed to yes/no answers.

The test is set at a higher standard (namely, the standards for a junior doctor) than the standard at which the student is expected to perform. Students are expected to improve over the years: at Peninsula Medical School, first-year students would not obtain a good score (with an approximate average of eight-15 per cent), while second-year students would show improvement (scoring approximately 15-20 per cent) and third-year students would reach between 20 and 35 per cent, and so on.

The Progress test is not directly related to the syllabus but instead seeks to make students reflect upon medicine more globally. This method enables school administrators or faculty to develop a progress chart/growth chart for both class and individual progression over five years and to single out a given student’s progress against the cohort mean.

Progress Testing was one of the results of a changed focus towards more integrated learning at Peninsula College. In order to integrate such a programme, the college hired new staff to implement the new curriculum and initiated changes to the quality assurance procedures.

4.3.2. General skills standardised tests

Some standardised tests aim to be non-discipline specific. These tests tend to measure general skills, such as critical thinking.

Assessment of general skills standardised tests

Non-discipline-specific tests are attractive to some because they rely on complex psychometric techniques to compare the acquisition of generic skills, such as critical thinking. These tests are also attractive because they allow for comparison across groups of students and institutions.

The purpose of non-discipline-specific standardised tests – focusing on generic skills – would however need to be clearly specified, since there can be a somewhat negative perception of standardised tests across the sector, especially when these are used to rank students and institutions.⁵⁷

Others have raised concerns about whether these tests were not too ‘general’ to provide a significant measure of learning gain. There was also wariness of the representativeness of tests which used samples of

⁵⁷ Foley & Goldstein, 2012.

students rather than the entire cohort. Therefore general skills tests would need to be carefully validated before being applied to the English context.

Examples of general skills standardised tests

The two tests outlined below constitute examples used in US institutions - the Collegiate Learning Assessment (CLA) and the Measure of Academic Proficiency and Progress (MAPP). The sample of HEIs included in this research did not include any examples of non-discipline-specific standardised tests in England. The purpose of the CLA is to provide an assessment of the value added by the university/college's instructional and other programmes with respect to certain defined learning outcomes. As such this measure aims to help institutions understand how much their students are improving and whether that improvement is in line with the gains acquired by comparable students at other institutions.

The Collegiate Learning Assessment

The CLA has several advantages over other methodologies. The CLA is one of the first value added measures to enable direct comparisons of an institution's contribution to student learning.⁵⁸ Its burden on students is relatively limited, given that, although it is administered twice, the CLA administers separate components of the test to different groups of students following a 'sampling matrix' assessment. The CLA aims to achieve a 'holistic' approach to measuring student learning, through open-ended questions. Endorsed by several national higher education commissions, including the Association of American Colleges and Universities⁵⁹ and the American Association of State Colleges and Universities⁶⁰ in the United States, and is gaining traction in a number of European countries. The CLA has also been used in high profile studies, such as the publication of Arum and Roska (2011) suggesting that a large proportion of US undergraduates showed no significant learning gain during their undergraduate career.⁶¹

The visibility of the CLA has also contributed to a wide debate, including whether it is too US-centric to be easily applicable outside the United States. The relationship between CLA scores and actual test scores is also debated. On the one hand, Klein et al. argue that the 'average improvement in scores between entering freshmen and graduating seniors on the CLA is more than one standard deviation'.⁶² On the other hand, other researchers have shown that there was a strong relationship between entry level test scores from the Scholastic Aptitude Test (SAT) and the GPA. This relationship could therefore be seen to indicate that students may not achieve a significant learning gain in higher education.⁶³ It could also point to the fact that the CLA may measure skills which are too general to be used as an appropriate measure of learning gain.⁶⁴ In addition, in many institutions, particular disciplines place different emphasis on the importance of general skills.⁶⁵

⁵⁸ Glenn, 2010.

⁵⁹ AACU, 2005.

⁶⁰ AASCU, 2006.

⁶¹ See introduction for an additional discussion of these debated findings.

⁶² Klein et al., 2007; 423

⁶³ Arum & Roska, 2011.

⁶⁴ Shermis, 2008.

⁶⁵ Banta & Pike, 2012.

Finally, critics have argued that the CLA may not facilitate comparison between institutions because much of the variation in test scores reflects individual differences rather than the quality of education.⁶⁶

Vignette 6 provides further detail regarding the content of the CLA.

Vignette 6: Collegiate Learning Assessment

The CLA, currently run by the Council for Aid in Education, has been used by around 700 institutions in the United States and internationally.⁶⁷ It is administered to second- and fourth-year undergraduate students, but focuses on the institution as the level of analysis (rather than the student).

Two components are included: the first consists of tasks measuring the ability to read tables, weight and analyse documents, and develop a cogent response to a question, and the second consists of two analytical writing measures on developing a good argument. The test uses a set of tasks, which students are randomly assigned to, that aim to measure generic skills, including writing, higher order thinking, problem solving and quantitative reasoning.⁶⁸ These tasks include two performance tasks of 60 minutes each to analyse and evaluate the information presented in a library of documents and craft a written response to a scenario-based problem. In addition, test takers have to reply to 25 multiple choice questions to be taken within a time period of 30 minutes.

The CLA focuses on 'value added' by (i) examining how well a student performs compared with 'similarly situated' students and (ii) using a pre-/post-test model to examine progress made during tertiary studies.

The Measure of Academic Proficiency and Progress (MAPP)

- The US company 'Educational Testing Service' runs MAPP, which was introduced in 2006. Previous studies (based on MAPP and its predecessor, the Academic Profile Test) have shown that MAPP was able to differentiate the performance of freshmen and seniors, meaning that MAPP could potentially act as a measure of learning gain.⁶⁹ In addition, MAPP provides differentiated scores in the humanities, social sciences and natural sciences, and therefore provides a measure of what experts agree an academic programme should teach.
- However, there is some disagreement on whether the content covered in these assessments is appropriate given differences in programme emphasis across institutions. For example, some maths degrees are more applied than others. In addition, the design of MAPP has been criticised for being too simplistic, in that it assumes that complex skills can be disaggregated in the form of items related to multiple choice questions and then re-aggregated into a composite score.⁷⁰ Finally, although MAPP can be repeated over time, it was not originally designed to be.

Vignette 7 below provides more details regarding the content of MAPP).

⁶⁶ For instance, see Pike, 1992. For an additional review of the advantages and disadvantages of the CLA, see Klein et al., 2007.

⁶⁷ Council for Aid in Education. The Council for Aid in Education has recently launched the CLA+, to be administered around graduation time and adopted by 200 HEIs in the United States.

⁶⁸ Hardison & Vilamovska, 2009; Miller, 2014.

⁶⁹ Marr (1995) found that the higher scores of juniors and seniors could be explained by their completion of more of the core curriculum. Completion of advanced courses beyond the core curriculum had relatively little impact on Academic Profile scores. For another review of MAPP, see Liu (2011).

⁷⁰ Klein et al., 2007.

Vignette 7: Measure of Academic Proficiency and Progress (MAPP)

- MAPP asks students to answer a number of multiple choice questions, and then disaggregates different components of analysis by a number of dimensions (critical thinking, reading, writing and mathematics). MAPP has two forms: the standard form has 108 items, with 27 items in each of the four skill areas measured, and takes two hours to complete. The short form has 36 items and takes about 40 minutes to complete.
- MAPP results are then aggregated to form a score. More precisely, MAPP leads to two 'either/or' scores. In the standard form, one score includes a total score, scaled sub-scores and proficiency classifications at the group and individual student level. In the short form, the score includes group data, as well as a single total score for individual students.
- This test is designed for two- and four-year colleges and universities to improve the quality of instruction.

4.4. Mixed methods

Another approach to measuring learning gain involves using a range of different methods in conjunction. This involves combining several tools and indicators to track performance - for example, through the combination of GPAs, student learning data and student surveys. Such an approach could be presented as akin to measuring learning gain.

4.4.1. Assessment of mixed methods

In theory taking a mixed methods approach increases the robustness of the measurement. These methods, given their mixed character, include an element of objectivity. However, weaknesses in individual methods could transmit to the mixed method approach, and some approaches are likely to require broad institutional capacity to implement in terms of systems and expertise to undertake learning analytics. Finally, mixed methods tend to have a limited scope for comparability across the sector, since they are currently only applied on a small scale.

4.4.2. Examples of use in England

Mixed methods approaches are not in frequent use in institutions in England. Nevertheless, this section gives an overview of two that are in operation - the Academic Performance Tracking Tool (APTT) and the Advanced Level Information System (ALIS).

Academic Performance Tracking Tool

This method started as an institution-wide pilot at Oxford Brookes University in 2012 and is now fully embedded across the university.

The APTT covers a number of dimensions applicable to learning gain in that it includes measures such as a comparison of grades through a standardised GPA and personal development elements through the NSS survey. In addition, the APTT is able to track students longitudinally through year-on-year measurement, and can therefore help to maximise the learning environment of students.

It is, however, important to note that the APTT was only initiated in 2012, and that further testing regarding its sustainability would be necessary to further assess its viability. It is also important to underline that APTT does not focus specifically on 'generic skills', but instead monitors

module/programme outcomes. Vignette 8 below explains how the how the APTT is used at Oxford Brookes University.

Vignette 5: Academic Performance Tracking Tool (APTT)

The APTT was developed in 2012 at Oxford Brookes University. The APTT encompasses a live update of module and programme outcomes, in addition to other metrics reported annually. These metrics include grades, measured as a GPA, which provides a progression metric aligned to the module outcomes achieved from the first year of students' undergraduate programmes. The APTT provides an overview of GPA outcomes, survey results (through the NSS), qualitative methods through portfolio review metrics and academic adviser meetings, as well as other data, such as enrolment and retention.

The APTT is fed from student records, and students can access the APTT via their student portal. The APTT is used both as a summative and formative tool, and the data informs the student's records, allowing students to see their progression via their personal portals. The APTT and GPA are part of both the university's student experience strategy and quality assurance process, and in this respect has therefore benefited from support from the institution, staff and students.

Advanced Level Information System

Another example of mixed methods approach is the value added model of the ALIS, which is used at Hull College, with further education students, as summarised in Vignette 9 below.

ALIS relies on a combination of different methods, including a comparison of grades, survey results and portfolios, and links this data to other indicators from students' records.⁷¹ This model enables staff to measure distance travelled through regular data generation, and has the benefit of focusing on skills, such as students' approaches to work, confidence building, and other transferable skills.

However, a potential disadvantage is that the research team has only found evidence of its use with further education students in England. This may be a source of concern, since students who enter the higher education sector do so at different stages in their lives and at different levels of maturity, and will therefore have followed different types of education routes. In addition, the tool itself takes time to produce results: a whole iteration is needed to obtain an accurate measure per individual, which, as an interviewee noted, could take about three years.

⁷¹ Institution-wide practices have also been reported in other institutions, such as Nottingham Trent University, where institution-wide tools are set up to analyse student progress and learning gain which are shared with schools within the university.

Vignette 6: Advanced Level Information System (ALIS)

The ALIS is used at Hull College to attribute a score to students reflecting their likely achievement on their programme.

The ALIS score, originally run by the Centre for Evaluation and Monitoring of Durham University, is based on two types of scores. It is computed based on GCSE results (using the average GCSE score as the baseline). For students who have not taken GCSEs, including students from overseas or those attending international schools, the ALIS score is based on a standardised test called the computer-adaptive baseline test (CABT). The CABT includes vocabulary, math and non-verbal tests. In addition the ALIS also includes a survey, including a subject-specific teaching and learning survey and a general welfare survey.

More generally, this method fits into the growing awareness that Hull College has developed on learning gain as part of a benchmarking exercise by a group of institutions in the same region.

4.5. Other qualitative methods

Most of the reports on learning gain methods in England refer to outputs from qualitative methods that have been developed to try and understand how much students learn.

4.5.1. *Assessment of qualitative methods*

Qualitative measures seek to provide an opportunity for students to reflect on their learning, and are used as a formative tool both to encourage students to reflect on their learning, and to stimulate a formative exchange between students and their tutors. Some educators use these methods to encourage the development of generic skills through team work or presentation skills as core to their modules. Others also use these methods to offer regular feedback to students and thereby report improvements in the uptake and refinement of skills, although this process is not rigorously or consistently recorded in any formal manner.

However, qualitative methods can be difficult to compare across institutions because they typically rely on rich open text answers to different questions. In addition, these methods may not be entirely representative, since they tend to be taken up by a limited group of students, suffer from biases due to self-reporting, and rely more on the initiative of the lecturer than forming part of an institution-wide year-on-year comparison.

Additional difficulties in comparability come from the fact that institutions would potentially not have comparable questions or formats of replies. Qualitative methods are hence currently more suited to measuring individual learning gain than for institutional comparability.

4.5.2. *Examples of use in England*

Examples of the use of qualitative methods in England include initiatives such as the Personal Development Portfolio (PDP), or others that encourage students to think about their acquired skills and gain awareness of their skills gaps, such as skills awards.

Personal Development Portfolio (PDP)

One of these qualitative methods is the PDP. The use of PDPs, which are used in order to inform teaching and learning, appears well developed in the higher education sector. In some circumstances PDPs have been integrated into certain programme modules, and there had previously been an expectation that UK universities establish and develop PDPs as part of their quality review regime. PDPs enable students to assess their skills and better understand where their strengths and weaknesses lie, and can include a learning log, where students reflect on their learning, and/or a collection of their assignments.

While they enable students and supervisors to have an overview of progress over their period of study, PDPs tend to rely on self-reporting and may therefore lead to perverse incentives, since they are not always graded. In addition, PDPs may be difficult to compare across units of analysis and do not, in their current form, appear to lend themselves to a rigorous longitudinal measurement.

PDPs are used in various institutions, including the University of Southampton, Oxford Brookes University and Durham University. Vignette 10 below summarises the use of PDPs at Durham University and the University of Bristol.

Vignette 7: Personal development portfolios (PDPs)

PDPs are in use in some Durham University courses. They are particularly in use for courses with a more vocational angle (e.g. pharmacy or education), where they are tied in with accreditation for the student.

A variant of the PDP approach is used at the University of Bristol in the medical sciences, where each student has a logbook with sections covering different types of skills, which are observed repeatedly, year on year, in a variety of settings. Students are required to record teachers' and professors' comments on their application of these skills in their personal logbook, which they review together with their academic supervisor after the five years of their degree.

Skills awards

Other qualitative measures include incentivising students to gain awareness of their skills developments through awards. Awards encourage students to engage with learning processes by stepping out of their disciplinary knowledge and to understand how the skills they are acquiring through their degree can be applied to a different context.

One issue with this approach is that no baseline assessment is undertaken, since the application form does not measure baseline skills but motivation to complete. An additional concern is self-selection bias, in that students who are in need of additional support, or have jobs or family commitment, or who are less engaged or motivated, do not engage in this process.

Institutions that have implemented skills awards programmes include Durham University and King's College London. The skills awards are called the Durham Award and the King's Leadership and Professional Skills Award, respectively. Vignette 11 below summarises the example of the Leadership and Professional Skills Award in practice at King's College London.

Vignette 11: Leadership and Professional Skills Award (LPSA)

King’s College has set up a Leadership and Professional Skills Award (LPSA) to help students focus on the development of their employability skills. The award includes a co-curricular programme of 10 sessions, of which students can select to attend five or more in order to develop their portfolio.

The programme covers a range of skill areas that are considered important to employability. These include: teamwork, leadership, analytical thinking, problem solving, etc. The skill areas were identified in consultation with employers and students, and each session is designed to help students to ‘unpack’ the skill area in order to engage with it critically to deepen understanding, identify how employers assess these skills in recruitment and selection processes, and reflect on their own development in each area based on personal experiences. The sessions are co-delivered with employer representatives to provide additional insights (e.g. TeachFirst, PwC). Receipt of the award is based on the submission of a reflective portfolio which is assessed using standardised marking criteria.

This award was originally set up by the careers service. According to an interviewee, the reception of the award by participants has been extremely positive.

4.6. Summary assessment of different methods

The following table summarises the assessment of different methods. For each group of methods, this summary provides a general assessment as well as an additional assessment, which provides further advantages and disadvantages specific to each example covered in Section 4.

Table 2: Advantages and disadvantages of different methods and tools of measuring learning gain

Method group 1: grades		
General assessment	Advantages	Disadvantages
	<ul style="list-style-type: none"> Relies on existing institutional capacity 	<ul style="list-style-type: none"> Comparability may be problematic
Assessment of specific examples	Advantages	Disadvantages
Predicted pass rate (Open University)	<ul style="list-style-type: none"> Predicts students’ likelihood to pass a course Provides useful data on student performance in specific disciplines Enables longitudinal analysis Appears to achieve suitable representativeness (given that it applies to all students) 	<ul style="list-style-type: none"> Was not originally designed as a measure of learning gain Uncertainty as to how this data is used institution-wide No data collected on student performance before university
GPA (Oxford Brookes University), grades (Hull College)	<ul style="list-style-type: none"> See assessment of ALIS under mixed method 	<ul style="list-style-type: none"> See assessment of ALIS under mixed method

Method group 2: survey		
General assessment	Advantages	Disadvantages
	<ul style="list-style-type: none"> • Some surveys contain questions of relevance to measure learning gain 	<ul style="list-style-type: none"> • Not primarily designed to measure learning gains • More subjective, indirect and less valid measures than grades or tests • Applicability to measure learning gain would have to be validated in an English context • No survey measures all aspects of learning gain
Assessment of specific examples	Advantages	Disadvantages
UKES (30 institutions across the UK)	<ul style="list-style-type: none"> • Has a formative element through the exchange of results among its participatory institutions • Collects additional data, for example, on students' levels of engagement, that could assist in interpretation of gains data 	<ul style="list-style-type: none"> • Comparability over time and across institutions may be difficult • Representativeness limited by relatively low response rate to date • Little incentive for students to complete
NSS (nation-wide)	<ul style="list-style-type: none"> • Is already being used by most HEIs in the UK with relatively high response rate • Collects data on students' perceptions of their course, which could help interpretation of gains data 	<ul style="list-style-type: none"> • Comparability over time difficult • Relatively limited items which could be applied to measure learning gain
Careers Registration (University of Leeds, University of Lancaster, Liverpool John Moores University; University of Ulster due to go live in 2015; under consideration by numerous other HEIs)	<ul style="list-style-type: none"> • Includes personal development/employability questions which could help to measure learning gain • Comparability appears straightforward: repeated each year across disciplines for the entire student population • Appears to be simple and efficient 	<ul style="list-style-type: none"> • Potential cultural barriers to replicability • Extensive validation would be necessary before wider implementation • Only measures one aspect of learning gain
Skills audit (Durham University)	<ul style="list-style-type: none"> • Measures a relevant aspect of learning gain (skills) • Repeated at two points in time to facilitate comparability 	<ul style="list-style-type: none"> • Uncertainty regarding response rate • Comparability across institutions may be limited

Method group 3: standardised tests		
General assessment	Advantages	Disadvantages
	<ul style="list-style-type: none"> • Considered as an objective measure of learning gain • Facilitates comparability • Can be used for different purposes • Some tests specifically designed to measure learning gain 	<ul style="list-style-type: none"> • Trade-off between comparability across disciplines and level of precision regarding measurement of learning gain • Use would need to be validated in England, especially for tests which are not currently in application in England • Likely to require a considerable level of resources to initiate and develop (particularly for some tests) • Purpose would need to be clarified to avoid a negative perception from the sector
Assessment of specific examples	Advantages	Disadvantages
Progress Testing, discipline-specific test (University of Cardiff, University of Exeter, University of Swansea, University of Plymouth, Peninsula College Medical School, University of Manchester, considered at the University of Bristol)	<ul style="list-style-type: none"> • More precise than discipline-independent tests, especially for content knowledge • Allows for comparison across institutions (in medical sciences) and years • High response rates 	<ul style="list-style-type: none"> • Only applicable to one discipline • Requires a very large bank of questions, and it may be difficult for one institution to develop those questions by itself • UK medical schools tend to have school-specific exams
CLA (discipline-independent); no example found in England	<ul style="list-style-type: none"> • Measures generic skills in a 'holistic' manner • Allows for comparison across institutions and samples of students across years • One of the earliest tests to measure value added/learning gain 	<ul style="list-style-type: none"> • Would require incentives to motivate students to take the CLA • The test aims to measure generic skills originally designed for a US higher education environment, which heightens need for validation • Response rates can be low (CLA uses samples of students rather than the entire cohort)
MAPP (discipline-independent); no example found in England	<ul style="list-style-type: none"> • Measures generic skills with a breakdown by general group of disciplines • Allows for comparison across institutions 	<ul style="list-style-type: none"> • Multiple choice questions could be seen as an 'oversimplification' • Would require incentive to motivate students to take the test • Would need to be further validated in the context of English higher education, especially given that the test was originally designed for a US higher education

		environment
--	--	-------------

Method group 4: combination of methods		
General assessment	Advantages	Disadvantages
	<ul style="list-style-type: none"> • In theory achieves higher validity by combining different methods • Allows for further data analysis as part of a formative process 	<ul style="list-style-type: none"> • Weaknesses of each method may remain • Some approaches require broad institutional capacity to implement
Assessment of specific examples	Advantages	Disadvantages
APTT (Oxford Brookes University)	<ul style="list-style-type: none"> • Allows for longitudinal comparisons and learning analytics • Tied to student records, so allows for high representativeness • Increasingly accepted: developed in 2012 and currently routinely used across the university 	<ul style="list-style-type: none"> • May require broad institutional capacity to implement • Self-reporting data part of APTT may be subjective • Would need further validation to ensure control for socio-economic background
ALIS (Hull College)	<ul style="list-style-type: none"> • Allows for longitudinal comparisons • Representativeness should be high if ALIS is presented as a requirement 	<ul style="list-style-type: none"> • For further education; not piloted in higher education • Provides score of likely achievement on the programme (rather than actual learning gain)

Method group 5: other qualitative methods		
General assessment	Advantages	Disadvantages
	<ul style="list-style-type: none"> • Encourages students' self-reflection and formative exchange between students and tutors • Appears to be applied in many institutions in England 	<ul style="list-style-type: none"> • May not entirely be representative due to low response rate and self-selection bias • May be difficult to compare across institutions
Assessment of specific examples	Advantages	Disadvantages
PDPs(various institutions)	<ul style="list-style-type: none"> • Helps students and staff by identifying the learning outcomes students find most difficult 	<ul style="list-style-type: none"> • Largely tailored to individual and text-based. Hence may be difficult to derive rigorous longitudinal and cross-institutional

RAND Europe

		measurement.
Skills awards (Durham University and King's College London)	<ul style="list-style-type: none">• Same as above	<ul style="list-style-type: none">• Same as above

5. Suggestions for next steps

This project scoped various conceptualisations of learning gain and its potential purposes, and has given an overview of a number of methods in use in England and internationally to measure and quantify it. In order to test whether a given method of learning gain would be suitable in England, the research team recommends that the commissioning bodies HEFCE, BIS and HEA engage in a series of consultations and awareness-raising activities to further assess the scalability of methods used to measure learning gain. This section provides an overview of a number of recommended key next steps for the project, including raising awareness among major stakeholders, and setting up small scale pilots for specific methods.

5.1. Awareness-raising activities and stakeholder consultations

Given that the concept of learning gain is relatively novel and unexplored in England the research team recommends that the commissioning bodies organise a series of symposia or conferences to inform the sector and broader public of the relevance of learning gain. These conferences would seek to encourage further communication among different stakeholder groups which include HEIs, official bodies and major representatives.

In a follow-up stage of this project, a comparative evaluation could include a broad stakeholder consultation to further understand the purposes of measuring learning gain and to assess the scalability of potential methods. This comparative evaluation could make use of the implementation framework drafted by the research team and included in Appendix C.

5.2. Series of pilots around five methods

The research team also suggests that the commissioning bodies conduct a series of pilots in a second stage of this project, in order to test and achieve a more refined understanding of the methodologies currently used to assess learning gain and related proxies, as outlined in this report.

A piloting phase would be useful for two reasons. First, as the assessment in this report details, further validation of methodologies would be necessary to understand how scalable learning gain measurement methods are in England. Second, these pilots would allow the commissioning bodies to better understand the applicability of these methods, and the consequences of using them.

To investigate the implementation of a measure of learning gain, the next stages of the project, including the piloting process could follow one (or more) of number of courses.

A) Use of methods already in place to measure learning gain, and related proxies.

In the first instance, the Steering Group could for example explore the use of student surveys such as UKES or Careers Registration (depending on the purpose of the method) or methods such as APTT and more fully analyse their applicability as measures of learning gain. It is, however, worth bearing in mind that none of these methods was originally designed to measure learning gain.

Such an approach could start by examining the validity of the measures in England, through a series of standard statistical analyses. This could, for example, include correlating and analysing the relationship between the student survey data with a measure of student achievement across a number of years. Each measure would also be the object of a sensitivity analysis, which would seek to control for student socio-demographic background, for example, in order to understand how sensitive a particular measure is to other parameters.

B) Piloting a test that has been used internationally but not applied to England

This could include CLA or MAPP. The report has shown that directly testing students on their knowledge, skills, competencies and personal developments (through in-house exams and assessments) is considered more objective and valid than individual self-reports. The research team also found that using grades and other output metrics could constitute a way to measure learning gain. Using data on grades and other proxies could, however, also prove problematic given that grades correspond to very general categories in England, are not issued in a standardised way, and that proxies may be influenced by other factors. It is important to also note that in England the research team could only find evidence of discipline-specific tests (Progress Testing) or of practices used in further education (ALIS).

C) Develop a new assessment of learning gain, based on academic collaboration around key disciplines.

Such an approach was used recently in other countries, such as Germany, and has been attempted by the OECD (AHELO). Rather than impose external assessment routines, such an initiative could involve strengthening assessment practices via a range of institutional collaborations, in order to respect the principle of institutional autonomy. Collaboration is already woven into the research function of higher education as well as core education processes, such as curriculum development, but it has not frequently been rolled out to generalised assessment. Importantly, such collaborations yield more transparent information on learning, academic processes, and contexts. Collaborations of this kind could be used to strengthen institutional governance and leadership of assessments; help to develop, validate and quality-assure a number of assessment tasks; and produce cross-institutional benchmarking reports that include comparative analysis. In addition such collaborations could help to produce learning gain assessment communities that could in turn produce new methodologies and instruments.

6. Conclusion

This report investigated the methods for measuring student learning gain in higher education, and the contribution that institutions, courses and particular educational activities make to learning gain, both nationally and internationally.

This study reported on the state of this debate – which is more prominent internationally than it is in England – and explored the concept of learning gain, as well as existing and potential methodologies and practices currently used to measure it and its related proxies.

6.1. Definitions

The report showed that learning gain, defined as ‘distance travelled’, or the difference in student performance at two points in time, was understood in a variety of ways by the higher education sector. It is conceived by some as the difference in student performance between two stages of their studies, by others as a variant of the concept of value added, or simply as ‘learning’.

6.2. Methodology

The report drew upon a review of the English and international literature relevant to the topic of learning gain, as well as results from a call for information circulated by the HEA to the PVC network and subject leads; and interviews with personnel from selected professional bodies, and academics and administrators across 13 HEAs.

The research team used these different approaches to try to gain a deeper understanding of the sector’s perception and understanding of the concept of learning gain, to introduce a selection of methods presented as measures of learning gain, and to assess the robustness of these methods.

6.3. Summary of findings in light of research aims

As the concept of learning gain has not yet been extensively explored in English higher education, the report investigated the following issues:

6.3.1. Purposes of measuring learning gain

The report sought to provide answers to the terms of reference for the work, including the ways in which methods for measuring learning gain are currently used in English higher education and the applicability

of these methods for different purposes. It showed that measures of learning gain in English higher education are currently mostly used for quality improvement; such measures can serve other goals, including facilitating students' learning, or fostering accountability, transparency and comparability of higher education outcomes.

A total of 130 out of 147 respondents to the call for information (mainly subject leads and members of the PVC network) acknowledged that a more fully developed analysis of the concept of learning gain could be useful. However, some uncertainty remains with regard to the application of measures of learning gain, especially when it comes to transparency and accountability. This gave rise to a certain level of cautiousness among respondents and interviewees.

Respondents, interviewees and experts also highlighted the need to be aware of the potential for unintended consequences arising from the application of measures of learning gain, such as the temptation for stakeholders or institutions to 'game' the system. This partly responds to Question 7 of the terms of reference regarding the relevant lessons drawn from research on value added in the UK schools sector.

The need for sensitivity to institutional autonomy and institutional diversity was highlighted, as was the potential for resistance from the higher education sector, if a measure were perceived as 'imposed' on the sector rather than sector-owned.

6.3.2. Critical overview of learning gain methods

Based on a review of the literature, the call for information and interviews, the research team has identified 14 methods used to measure learning gain, clustered into five groups. These methods include grades, student surveys, standardised tests, mixed methods, as well as other qualitative methods. The assessment of these groups of methods is summarised in the form of a table in Appendix A.

The utilisation of methods in England is at a more experimental and indeed smaller scale than in other countries, especially the United States. A number of methods in use have different core purposes and/or only measure learning gain indirectly. In the United States, methods have predominantly been used in the context of a liberal arts/general education. This US context is different to the more specialised nature of curricula in most English HEIs.

There is a distinction to be drawn between direct measures of learning gain and proxy measures. Some English scholars have suggested other concepts measured in surveys, such as graduate employment rates or expected salaries, as proxies for one or more aspects of learning gain. These proxies could include some of the methods summarised below, including engagement and experience, for instance, and whilst the use of these proxies has been explored and used in the United States, it has remained only at the proposal stage in England (as far as the research team is aware).

Using proxies to measure learning gain could prove efficient since it involves relying on existing metrics, but this approach would need to allay concerns regarding their validity. For example, some proxies, such as graduate employment, appear to depend on developments extraneous to higher education, such as the state of the labour market. Furthermore, all proxy measures, including surveys of student engagement and experience, would need to be compared against other direct measures of assessment of learning gain.

Grades

Assessing academic grades over time is an approach presented by some as a measure of learning gain. It considers student achievement by taking the difference between students' grades at two points in time.

- One approach presented as measuring learning considers student achievement by taking the difference between students' grades at two points in time.
- This can be undertaken in different ways. These include: comparing the difference between actual grades at two points in time, using a standardised measure of these grades (such as the Grade Point Average (GPA)) as a comparison, or using a set of grades (standardised or not) to make predictions on future grades.
- The research team also found some evidence of the use of grades as part of mixed methods, for example, combined with surveys as well as standardised tests. There is evidence of each of these methods being used currently in the higher education sector in England.
- One of the advantages of using grades is that, from the onset, it provides a way directly to measure learning gain which relies on existing institutional capacity (all HEIs issuing grades).
- The key issue regarding the use of grades in the UK is comparability. Different practices in grading and assessment are used within and between disciplines, thus only allowing for broad comparability to exist between awards.
- Another issue is that a significant proportion of students tend to get results in a similar part of the honours degree classification system five point-grading scale (2:1 or 2:2). This may not allow enough variance to measure learning gain, for example if the intention were to use grades to compare learning gain across English HEIs. There is also debate on whether tariff scores/points are an adequate measure of starting point and the difficulties caused by ignoring some A level subjects and entrants with non-standard qualifications.

Surveys

Another approach is to ask students to assess the extent to which they consider themselves to have learned through self-reports (such as through a survey).

- Several surveys are in use in the higher education sector in England. These include the NSS and the UKES, both of which are administered by the commissioning bodies for this report. In addition, some surveys are administered by one or a handful of institutions, including the Careers Registration instrument from the University of Leeds (which contains questions on personal development related to employability) or the 'skills audit' employed at Durham University (which measures students' progression in terms of skills acquired).
- These surveys are not primarily designed to measure learning gain. However in the case of UKES and the NSS, they do contain some questions related to student development. The most recent UKES pilot included 12 items (clustered in one question) regarding students' perceived skills development, and the NSS contains one question relating to personal development. The limited number of targeted questions directly related to gains in these surveys casts some doubt on the validity of using these surveys as currently constructed as a measure of learning gain.
- In general, surveys are considered to provide more indirect and less valid measures of learning gain than other methods, since students report their own assessment of their learning gain. There

is some debate in the United States on the nature of the association between students' test scores and measures of engagement.

- In addition, some surveys, including the NSS and UKES, are conducted at a single point in time and would therefore need repeating within each cohort at the start and end of the course of study to obtain longitudinal data suitable for measuring learning gain.
- Response rates to surveys can be highly variable. Were a survey with lower response rates to be used to measure learning gain, additional work would need to be undertaken to increase the response rate. For example the Careers Registration tool at the University of Leeds ties in student responses to the annual registration process.
- Any survey would have to be validated as a proxy for learning gain in England, as the Wabash Study has done for a suite of instruments in the United States. This could be done through piloting to test the relationship between student self-reporting and actual data on student progression.

Standardised tests

Standardised tests provide a third established approach to measuring learning gain. They measure the acquisition of certain skills, which may be generic or specialised, and are administered to students either as part of their formative or summative assessment for their degree, or as an additional exercise alongside the programme.

- There are two main types of standardised test:
 - *Discipline specific* tests, such as the Progress Testing administered in some schools of medicine in England.
 - *Generic skills* tests, which are not discipline specific, such as the CLA, which relies on open-ended questions, and the MAPP, which is a more traditional test relying on multiple choice questions. Both these examples originated and are applied in the United States; as far as the research team is aware, these are not applied in England.
- Standardised tests are considered to be more objective as measures of learning gain than methods based on self-reporting, such as surveys. Furthermore, several of the standardised tests mentioned in the report are designed specifically to measure 'distance travelled', with comparability in mind.
- Discipline-specific tests generally achieve a higher level of precision and hence validity than general tests. However, the outputs of a discipline-specific approach are difficult to compare across disciplines. Should broader comparability be desirable, it would be necessary to explore the possibility of aggregating results up to a more generic level.
- Non-discipline-specific tests are attractive to some given that they rely on complex psychometric techniques to compare the acquisition of generic skills, such as critical thinking. Others have raised concerns regarding their representativeness when applied, as in the case of the CLA, which uses samples of students rather than the entire cohort.
- The CLA has also been criticised as measuring skills which are too general to significantly measure learning gain. Both CLA and MAPP have been criticised as not taking enough account of the differences in programme emphasis across institutions. MAPP also assumes that complex

skills can be disaggregated into multiple choice questions. Both can be used for institutional comparability, although again this is debatable.

- The purpose of non-discipline-specific standardised tests – focusing on generic skills – would need to be clearly specified since there can be a somewhat negative perception of standardised tests across the sector when these are used to rank students and institutions.
- Developing standardised tests for England would require consideration of the level of resources needed to develop, initiate, fund and administer the test in a steady state. For example, Progress Testing relies on an extensive pool of questions to be administered each year.

Mixed methods

- Mixed methods use a range of tools and indicators to track performance, for example, through the combination of GPAs, student learning data and student surveys, or the combination of grades and a test to predict students' performance, which could be presented as akin to measuring learning gain.
- Methods can be combined to increase the robustness of the measurement. However, these methods are used on a relatively small scale in England. Weaknesses in the individual methods can also affect the mixed methodology.
- Some approaches are likely to require broad institutional capacity to implement in terms of systems and expertise to undertake learning analytics.

Other qualitative methods

- In addition, the project also investigated a range of other methodologies of tracking student development currently in place in the sector, including qualitative methods.
- These methods include students creating a portfolio, such as the PDP, or initiatives to encourage students to think about their acquired skills and skills gaps, such as skills awards.
- These other methodologies tend to be particularly useful to encourage students to reflect on their learning, as well as to stimulate a formative exchange between students and their tutors. However, some of these methods may not be entirely representative, since they tend to be taken up by a limited pool of students. They could also be difficult to compare across institutions given that they may not have comparable questions or formats of replies.
- They are currently most suited to individual measurements of learning gain rather than institutional comparability.

6.4.4. Relevant considerations for a robust pilot

In light of the findings of the report, the research team encourages the commissioning bodies to take the following points into consideration when devising the next stages of the project:

- Clearly state the purpose of measuring learning gain – for any particular application.
- Clearly define dimensions of analysis, be it generic skills or elements of personal development, for example.

- Capture the complexity of measuring learning gain, especially if comparing institutions, including taking confidence intervals and representing the complex interactions among institutions through interaction effects.
- Recognise the diversity of needs and goals of students, the state, or the mission and goals of an institution or a higher education system.
- Adopt methods that have comparable inputs and outputs (and adjust for control variables).
- Remain practical and consider effectiveness of implementation as well as cost-effectiveness.
- Remain aware of unintended consequences to, for example, prevent ‘gaming’.
- Secure buy-in from HEIs and other stakeholders.
- Acknowledge the limits of measuring learning gain, namely, that HEIs are not the only drivers of learning gain in a student’s life.
- Recognize that all of these methods include some aspects which could be considered to measure learning gain. They also suffer from common issues which would need to be addressed in a pilot, including the need to appropriately validate these methods, for example, by controlling for different potential factors that could impact on the validity of the measure, such as socio-demographic or neighbourhood effects.

6.5. Next steps for the project

On the basis of this overview, the research team suggests two next steps for this project.

6.5.1. Suggestion 1

The commissioning bodies may wish to engage with the sector to raise awareness of and promote debate about measures of learning gain and their purposes.

Options to consider are:

- A series of symposia or conferences in order to inform the sector and broader public of the relevance of looking at learning gain. These would encourage further communication among different stakeholder groups.
- A policy workshop to test the perception of the higher education sector to a measure of learning gain and to validate the applicability of a measure of learning gain.

6.5.2. Suggestion 2

The commissioning bodies may wish to conduct a series of pilots in order to explore further some of the considerations relating to the measurement of learning gain in the English context. These pilots could draw on the methods investigated in the report, exploring issues of viability and validity in their various aspects – practical, technical, methodological and financial. Consideration should be given to an impact assessment of the intended and unintended consequences of measuring learning gain.

These pilots could include:

- Assessment of methods that are currently applied in the English higher education sector, as well as their scalability. This approach could include piloting surveys, for example, or some discipline-specific tests, all the while taking into account that the methods currently in place were designed to measure aspects other than learning gain or to cover just one dimension or discipline.
- Testing methods that have not yet been found to be applied in England, including generic standardised tests. Consideration would be needed of the implications of transferring such a test from one higher education system to another and of the sector's potential reactions to using external measures.
- Designing an English measure of learning gain, possibly articulated as a test. This approach should seek to ensure sector engagement and could be more tailored to the needs and structure of the English higher education sector. However, it would require a significant investment of time and resources. The examples of international comparisons, such as the AHELO, show that this approach comes with its own complexities and sector-based dynamics and would constitute a fairly ambitious way forward.

References

- Adamson, I., M. Berbero, P. Cullen, L. Gonzalez-Vega, J.J. Sobrino, & N. Ryan. 2010. 'Quality Assurance and Learning Outcomes.' Brussels: European Association for Quality Assurance in Higher Education, 2010, Helsinki.
- Allen, J. 2005. 'Grades as Valid Measures of Academic Achievement of Classroom Learning', *The Clearing House* 78 (5): 218–23.
- American College Test (ACT). 2014. 'ACT-CAAP.' As of 14 December 2014: <http://www.act.org/caap/>
- Association of American Colleges and Universities (AACU). 2005. 'Liberal Education Outcomes.' Washington: Association of American Colleges and Universities.
- American Association of State Colleges and Universities (AASCU). 2006. 'Value added Assessment.' In *Perspectives*. Washington: American Association of State Colleges and Universities.
- Arum, R., & J. Roska. 2011. *Academically Adrift: Limited Learning on College Campuses*. Chicago: University of Chicago.
- Association of American Colleges and Universities (AACU), American Association of State Colleges and Universities (AASCU), and Association for Public and Land-Grant Universities (APLU). 2014. 'Rising to the Challenge: Meaningful Assessment of Student Learning.'
- Association of American Colleges and Universities (AACU). 2009. 'Learning and Assessment: Trends in Undergraduate Education – A Survey among Members of the Association of American Colleges and Universities.' Washington: Association of American Colleges and Universities.
- Astin, A.W. 2011. 'Data Don't Back Up Sweeping Claim.' *The Chronicle of Higher Education*, 14 February. As of 14 January 2015: <http://chronicle.com/article/Academically-Adrift-a/126371/>
- Australian Council for Educational Research (ACER). 2001. *Graduates Skills Assessment: Summary Report*. Occasional Paper 01/E. Canberra: ACER. As of 14 December 2014: http://www.acer.edu.au/files/GSA_SummaryReport.pdf
- Babbie, E. 2001. *The Practice of Social Research*. London: Thomson Learning.
- Banta, T., & G. Pike. 2012. 'Making the Case Against – One More Time.' In *The Seven Red Herrings about Standardized Assessments in Higher Education*, 24–30. Occasional Paper 15. Champaign, IL: National Institute for Learning Outcomes Assessment, University of Illinois. As of 20 November 2014: <http://www.learningoutcomesassessment.org/documents/HerringPaperFINAL.pdf>

- Bearer-Friend, J. 2009. *Measuring Student Achievement at Postsecondary Institutions*. Washington: National Governors Association Center for Best Practices. As of 15 December 2014: <http://www.nga.org/files/live/sites/NGA/files/pdf/0911MEASURINGACHIEVEMENT.PDF>
- Bowman, N. 2010. 'Can 1st Year College Students Accurately Report Their Learning and Development?' *American Educational Research Journal* 47 (2): 466–96.
- Buckley, A. 2012 "Making it count: Reflecting on the National Student Survey in the process of enhancement" As of 15 January 2014: <https://www.heacademy.ac.uk/making-it-count-reflections-national-student-survey-nss-process-enhancement>
- Buckley, A. 2014. *UK Engagement Survey: Second Pilot Year*. London: The Higher Education Academy.
- Callender, C., P. Ramsden, & J. Griggs. 2014. 'Review of the National Student Survey.' London: HEFCE.
- Carini, R. M., G. D. Kuh, & S. P. Klein. 2006. 'Student Engagement and Student learning: Testing the Linkages.' *Research in Higher Education* 47 (1): 1–32.
- Coates, H., & S. Richardson. 2012. 'An International Assessment of Bachelor Degree Graduates' Learning Outcomes.' *Higher Education Management and Policy* 23 (3): 51–69.
- Coates, H., ed. 2014. *Higher Education Learning Outcomes Assessment: International perspectives*. Frankfurt: Peter Lang Publishers.
- Coates, H., & M. Mahat. 2013. Assessing Student Engagement and Outcomes: Modelling Insights from Australia and Around the World. *International Journal of Chinese Education* 2 (2): 241–64.
- Coates, H., & A. McCormick, eds. 2014. *Engaging University Students: International Insights from System-Wide Studies*. Dordrecht: Springer.
- Coates, H. 2015. Assessment of Learning Outcomes. In *Teaching, Learning and Student Engagement*, edited by M. Klemencic & P. Ashwin. Dordrecht: Springer.
- Cole, J.S., M. Kennedy, & M. Ben-Avie. 2009. 'The Role of Precollege Data in Assessing and Understanding Student Engagement in College.' *New Directions for Institutional Research* 2009 (141) : 55-69.
- Cunha, J., & T. Miller. 2012. 'Measuring Value added in Higher Education: Possibilities and Limitations in the Use of Administrative Data.' *Economics of Education Review* 42 (2014): 64–77.
- Department for Business, Innovation and Skills (BIS). 2010. 'Use of an aptitude test in university entrance: a validity study', BIS research paper no 26 As of 15 January 2015: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32423/10-1321-use-of-aptitude-test-university-entrance-validity-study.pdf
- Department for Business, Innovation and Skills (BIS). 2011. 'Putting Students at the Heart of Higher Education.' Government announcement. June 28, 2011.

- Department for Business, Innovation and Skills (BIS). 2014. *Methodological Issues in Estimating the Value added for Further Education, Higher education and Skills: A Review of Relevant Literature*. London: BIS.
- Department for Education. n.d. 'A Guide to Level 3 Value added in 2013 Schools and Colleges Performance Table.' London: Department for Education.
- Department for Education. 2014a. 'School performance tables: value added technical information', As of 15 January 2015:
http://www.education.gov.uk/schools/performance/archive/schools_04/sec3b.shtml
- Department for Education. 2014b. 'Key Stage 2 to Key Stage 4 (KS2-KS4) Contextual Value added Measure (CVA) including English and Maths.' As of 13 January 2015:
http://www.education.gov.uk/schools/performance/archive/schools_10/s3.shtml
- Dill, D., & M. Soo. 2005. 'Academic Quality, League Tables and Public Policy: A Cross-National Analysis of University Ranking Systems.' *Higher Education* 49 (4): 495–537.
- Doran, H.C., & L.T. Izumi. 2004. 'Putting Education to the Test: A Value added Model for California.' San Francisco: Pacific Research Institute.
- Doran, H., & J. Lockwood. 2006. 'Fitting Value added Models in R.' *Journal of Educational and Behavioral Statistics* 31 (2): 205–30.
- Douglass, J.A., G. Thomson, & C. Zhao. 2012. 'The Learning Outcomes Race: The Value of Self-Reported Gains in Large Research Universities. *Higher Education: The International Journal of Higher Education and Educational Planning* 64 (3): 317–35.
- Downes, D., & O. Vindurampulle. 2007. 'Value added Measures for School Improvement.' Melbourne: Department of Education and Early Childhood Development.
- Educational Testing Service. 1993. *Annotated bibliography of tests: achievement tests, college level*, Princeton: Educational Testing Service.
- European Parliament. 2008. 'Recommendation of the European Parliament and of the Council of 23 April 2008 on the Establishment of the European Qualifications Framework for Lifelong Learning.' *Official Journal of the European Communities* [111, 6.5.2008]. As of 15 December 2014:
http://europa.eu/legislation_summaries/internal_market/living_and_working_in_the_internal_market/c11104_en.htm
- Evers, Frederick T., & S. O'Hara. 1996. 'Educational Outcome Measures of Knowledge, Skills, and Values by Canadian Colleges and Universities.' *Educational Quarterly Review* 3 (1): 43–56.
- Ewell, P.T. 2002. *An Analysis of Relationships between NSSE and Selected Student Learning Outcomes Measures for Seniors Attending Public Institutions in South Dakota*. Boulder, CO:
- National Center for Higher Education Management Systems.
- Ewell, P.T. 2009. *Assessment, Accountability, and Improvement: Revisiting the Tension*.

- Occasional Paper 1. Champaign, IL: National Institute for Learning Outcomes Assessment, University of Illinois.
- Wales. Victoria, British Columbia: Centre for Curriculum, Transfer and Technology.
- Foley, B., & H. Goldstein. 2012. *Measuring Success: League Tables in the Public Sector*. London: British Academy Policy Centre.
- Fulcher, K.H., & J.T. Willse. 2007. 'Value added: Back to Basics in Measuring Change.' *Assessment Update* 19 (5): 10–12.
- Gibbs, G. 2010. *Dimensions of Quality*. York: Higher Education Academy.
- Gibbs, G. 2012. 'Implications of Dimensions of Quality in a Market Environment.' York: Higher Education Academy.
- Glenn, D. 2010. 'A Measure of Education Is Put to the Test.' *Chronicle of Higher Education* 57 (5): A1–A9.
- Goe, L., C. Bell, & O. Little. 2008. *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis*. Washington D.C.: National Comprehensive Center for Teacher Quality.
- Goldstein, H., P. Huiqi, T. Rath, & N. Hill. 2014. *The Use of Value added Information in Judging School Performance*. London: Institute of Education, University of London. As of 15 December 2014: <http://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/using-value-added-information.pdf>
- Gorard, S. 2008. 'The Value added of Primary Schools: What Is It Really Measuring?' *Educational Review* 60 (2): 179–85.
- Graduate Careers Australia. 2010. *Postgraduate Research Experience 2009: The Report of the Postgraduate Research Experience Questionnaire*. Melbourne: Graduate Careers Australia.
- Green, M.F. 2011. Lost in Translation: Degree Definition and Quality in a Globalized World. *Change* 43: 18–27.
- Halpern, D.F. 2013. 'A Is for Assessment: The Other Scarlet Letter.' *Teaching of Psychology* 40 (4): 358–62.
- Hardison, C.M., & A.-M. Vilamovska. 2009. *The Collegiate Learning Assessment: Setting Standards for Performance at a College or University*. Santa Monica, Calif.: RAND Corporation. TR663. AS of 14 January 2015: http://www.rand.org/pubs/technical_reports/TR663.html
- Harvey, L. 2004–14. 'Analytic Quality Glossary.' Quality Research International. As of 3 July 2012: <http://www.qualityresearchinternational.com/glossary>
- Harvey, L., & D. Green. 1993. 'Defining Quality.' *Assessment & Evaluation in Higher Education* 18 (1): 9–34.

- Hazelkorn, E. 2013. 'What We Don't Know about Quality.' Discussion paper for the IA-HERO meeting, Washington.
- Hazelkorn, E. Forthcoming. *Rankings and the Reshaping of Higher Education: The Battle for World Class Excellence*. 2nd ed. London: Palgrave Macmillan.
- Hidden Curriculum. 2014. In *The glossary of education reform*, edited by S. Abbott. As of 15 January 2015: <http://edglossary.org/hidden-curriculum>
- Higher Education Funding Council for England (HEFCE). 2008. *Best ever response rate for National Student Survey*. wired-gov.net, 11 September. As of 15 January 2015: <http://www.wired-gov.net/wg/wg-news-1.nsf/lfi/DNWA-7JDCSF>
- Higher Education Funding Council for England (HEFCE). 2014a. *Review of the National Student Survey*. London: HEFCE. As of 15 December 2014: <http://www.hefce.ac.uk/pubs/rereports/year/2014/nssreview/#d.en.87657>
- Higher Education Funding Council for England (HEFCE). 2014b. 'Glossary.' As of 15 December 2014: <http://www.hefce.ac.uk/glossary/>
- Higher Education Statistics Agency (HESA). 2014. 'UK Performance Indicators in Higher Education'. hesa.ac.uk, n.d. As of 12 January 2015: <http://www.hesa.ac.uk/pis/>
- Hoon-Ho and Lalancette, K.. 2013. *Literature Review on the Value-added Measurement in Higher Education*. Paris: OECD.
- Hussain, I., S. McNally, & S. Telhaj. 2009. *University Quality and Graduate Wages in the UK*. Centre for the Economics of Education, London School of Economics. As of 15 December 2014: <http://cee.lse.ac.uk/cee%20dps/ceedp99.pdf>
- Kandiko, C., & F. Matos. 2013. *Engagement for Enhancement: Full Qualitative Testing Report for a UK Pilot*. London: Higher Education Academy.
- Kapteyn, A., J. Smith, A. Van Soest, & H. Vonkova. 2011. *Anchoring Vignettes and Response Consistency*, Santa Monica, Calif.: RAND Corporation. WR-840. As of 15 January 2015: http://www.rand.org/pubs/working_papers/WR840.html
- Kim, H., & D. Lancette. 2013. *Literature Review on the Value added Measurement of Higher Education*. Paris: OECD.
- Kirkup, C., R. Weather, J. Morrison, B. Durbin, & M. Pomati. 2010. 'Use of an Aptitude Test in University Entrance: A Validity Study.' BIS Research Paper 26. London: Department for Business, Innovation and Skills.
- Klein, S., R. Benjamin, R. Shavelson, & R. Bolus. 2007. 'The Collegiate Learning Assessment: Facts and Fantasies.' *Evaluation Review* 31: 415-440.

- Langan, A.M., et al. 2013. 'Applying Models to National Surveys of Undergraduate Science Students: What Affects Ratings of Satisfaction?' *Education Sciences* 3 (2): 193–207.
- Law School Survey of Student Engagement. 2006. *Engaging Legal Education: Moving Beyond the Status Quo*. Bloomington, IN: Law School Survey of Student Engagement, Indiana University School of Law. As of 15 December 2014:
http://lssse.iub.edu/2006_Annual_Report/pdf/LSSSE_2006_Annual_Report.pdf
- Learning Analytics and Knowledge. 2011. 'First Conference on Learning Analytics and Knowledge 2011, LAK 2011.' As of 14 December 2014: <http://tekri.athabascau.ca>
- Leckie, G., & H. Goldstein. 2011. 'Understanding Uncertainty in School League Tables.' *Fiscal Studies* 32 (2): 207–24.
- Lenkeit, J. 2013. 'Effectiveness Measures for Cross-sectional Studies: A comparison of Value added Models and Contextualised Attainment Models.' *School Effectiveness and School Improvement* 24 (1): 39–63.
- Liu, O.L., & Educational Testing Service. 2009. *Measuring Learning Outcomes in Higher Education*. ETS R&D Connections 10. Princeton, New Jersey: Educational Testing Service.
- Liu, O.L. 2011. 'Measuring Value added in Education: Conditions and Caveats.' *Assessment & Evaluation in Higher Education* 36 (1): 81–94.
- Liu, O.L., Bridgeman, B., Adler, R., and Educational Testing Service. 2012. 'Measuring Learning Outcomes in Higher Education Motivation Matters.' *Educational Researcher* 41 (9): 352–62.
- MacLeod, S. 2008. 'Likert scale', URL: <http://www.simplypsychology.org/likert-scale.html> As of 15 January 2015.
- Mallier, T., & T. Rodgers. 1995. 'Measuring Value added in Higher Education: A Proposal.' *Education Economics* 3 (2): 119-132.
- Marr, D. 1995. *Validity of the Academic Profile: An Investigation of the Construct Validity of the Long Form of the Academic Profile*. Princeton, NJ: Educational Testing Service.
- Martin, R., L. Delaney, & H. Colm. 2009. 'Enhancing the Comparability of Self-Rated Skills-matching Using Anchoring Vignettes.' *Irish Economic Association*, 2009-01-16.
- Matthews, D. 2013. 'Degrees of Comparability.' *Times Higher Education* (2094): 24–25.
- Miller, M.A. 2014. 'Editorial: Transferrable Thoughtfulness.' *Change* 46 (2): 4–5.
- Miller, T. Forthcoming. *Ensuring Quality in the Context of Outcomes based Funding*. Washington: Lumina Foundation.
- Millett, C.M., et al. 2007. *A Culture of Evidence: Critical Features of Assessments for Postsecondary Student Learning*. Princeton, NJ: Educational Testing Service.
- Millett, C.M., et al. 2008. *A Culture of Evidence: An Evidence-centered Approach to Accountability for Student Learning Outcomes*. Princeton, NJ: Educational Testing Service.

- National Survey of Student Engagement. 2002. *From Promise to Progress: How Colleges and Universities Are Using Student Engagement Results to Improve Collegiate Quality*. Bloomington, IN: Indiana University Center for Postsecondary Research.
- National Survey of Student Engagement. 2007. *Experiences that Matter: Enhancing Student Learning and Success*. Bloomington, IN: Indiana University Center for Postsecondary Research.
- OECD. 2008. *Measuring Improvements in Learning Outcomes: Best Methods to Assess the Value added of Schools*. Paris: OECD.
- OECD. 2011. *A Tuning-AHELO Conceptual Framework of Expected Desired/Learning Outcomes in Engineering*. OECD Education Working Papers 60. Paris: OECD Publishing.
- OECD. 2012. *AHELO Feasibility Study Report: Volume 1: Design and Implementation*. Paris: OECD.
- OECD. 2014. *Skills beyond School: Testing Student and University Performance Globally: OECD's AHELO*. Paris: OECD. As of 23 November 2014: <http://www.oecd.org/edu/skills-beyond-school/testingstudentanduniversityperformancegloballyoecdsahelo.htm>
- Olds, K., & S. Robertson. 2011. 'On Being Seduced by the World University Rankings (2011–12).' *insidehighered.com*, n.d. As of 5 October 2011: <https://www.insidehighered.com/blog/280>
- Oxford dictionaries. Oxford University Press. As of 15 January 2015 <http://www.oxforddictionaries.com/>
- Pace, C.R. 1982. *Achievement and the Quality of Student Effort*. Washington: National Commission on Excellence in Education. As of 15 December 2014: <http://eric.ed.gov/?id=ED227101>
- Pascarella, E.T. 2001. 'Identifying Excellence in Undergraduate Education: Are We Even Close?' *Change* 33 (3): 18–23.
- Pascarella, E.T., & P.T. Terenzini. 2005. *How College Affects Students: A Third Decade of Research*. San Francisco: Jossey-Bass.
- Pascarella, E., T. Seifert, & C. Blaich. 2010. 'How Effective Are the NSSE Benchmarks in Predicting Important Educational Outcomes?' *Change Magazine* (December). As of 14 December 2014: <http://www.changemag.org/Archives/Back%20Issues/January-February%202010/full-how-effective.html>
- Pascarella, E.T., et al. 2011. 'How Robust Are the Findings of Academically Adrift?' *Change* 43 (3): 20–24.
- Pike, G.R. 1992. 'The Components of Construct Validity: A Comparison of Two Measures of General Education Outcomes.' *Journal of General Education* 41: 130–59.
- Pollard, E., M. Williams, J. Williams, C. Bertram, J. Buzzeo, E. Drever, J. Griggs, & S. Coutinho. 2013. *How Should We Measure Higher Education? A Fundamental Review of the Performance Indicators*. Synthesis Report. London: HEFCE. As of 15 December 2014: http://www.hefce.ac.uk/media/hefce/content/pubs/indirreports/2013/fundamentalreviewoftheukpis/2013_ukpireview1.pdf

- Pollard, E., M. Williams, J. Williams, C. Bertram, J. Buzzeo, E. Drever, J. Griggs, & S. Coutinho. 2013b. *How Should We Measure Higher Education? A Fundamental Review of the Performance Indicators: Part Two: The Evidence Report*. London: HEFCE. As of 15 December 2014: <http://www.employment-studies.co.uk/pubs/report.php?id=hefce1213a>
- Provezis, S.J. 2010. *Regional Accreditation and Learning Outcomes Assessment: Mapping the Territory*. Occasional Paper 6. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Quality Assurance Agency (QAA). 2002. *Medicine*. Gloucester: The Quality Assurance Agency for Higher Education. As of 15 December 2014: <http://www.qaa.ac.uk/en/Publications/Documents/Subject-benchmark-statement-Medicine.pdf>
- Quality Assurance Agency (QAA). 2007a. *History*. Gloucester: The Quality Assurance Agency for Higher Education. As of 15 December 2014: <http://www.qaa.ac.uk/en/Publications/Documents/Subject-benchmark-statement-History.pdf>
- Quality Assurance Agency (QAA). 2007b. *Law*. Gloucester: The Quality Assurance Agency for Higher Education. As of 15 December 2014: <http://www.qaa.ac.uk/en/Publications/Documents/Subject-benchmark-statement-law.pdf>
- Quality Assurance Agency (QAA). 2007c. *Chemistry*. Gloucester: The Quality Assurance Agency for Higher Education. As of 15 December 2014: <http://www.qaa.ac.uk/en/Publications/Documents/Subject-benchmark-statement-Chemistry-2007.pdf>
- Quality Assurance Agency (QAA). 2007d. *General Business and Management*. Gloucester: The Quality Assurance Agency for Higher Education. As of 15 December 2014: <http://www.qaa.ac.uk/en/Publications/Documents/Subject-benchmark-statement-General-business-and-management.pdf>
- Quality Assurance Agency (QAA). 2010. *Subject Benchmarking Statement: Engineering*. The Quality Assurance Agency for Higher Education. As of 15 December 2014: <http://www.qaa.ac.uk/en/Publications/Documents/Subject-benchmark-statement-Engineering-.pdf>
- Quality Assurance Agency (QAA). 2012. *UK Quality Code for Higher Education*. Gloucester: The Quality Assurance Agency for Higher Education.
- Radloff, A., H. Coates, R. James, & K. Krause. 2012. *Development of the University Experience Survey (UES)*. Canberra: Department of Education, Employment and Workplace Relations.
- Ramsden, P., & C. Callender. 2014. *Appendix A: Literature Review*. Prepared for The UK Higher Education Funding Bodies.
- Ramsden, P. 2008. *The Future of Higher Education: Teaching and the Student Experience*. York: The Higher Education Academy. As of 23 November 2014: <http://www2.le.ac.uk/offices/ssds/projects/student-retention-project/dissemination/resources/paulramsden-teaching-and-student-experience.docx/view>

- Ray, A. 2006. *School Value added Measures in England*. London: Department for Education and Skills.
- Richardson, S., & H. Coates. 2014. 'Essential Foundations for Establishing Equivalence in Cross-National Higher Education Assessment.' *Higher Education* 68 (6):825–36.
- Rodgers, T. 2005. 'Measuring Value added in Higher Education: Do Any of the Recent Experiences in Secondary Education in the United Kingdom Suggest a Way Forward?' *Quality Assurance in Education: An International Perspective* 13 (2): 95–106.
- Rodgers, T. 2007. 'Measuring Value added in Higher Education: A Proposed Method for Developing a Performance Indicator Based on the Economic Value added to Graduates.' *Education Economics* 15 (1): 55–74.
- Seifert, T., K. Goodman, S. Edvalson, J. Laskowski, E. Pascarella, & C. Blauch. 2007. *Why Longitudinal Research Is Critical in Studying College Impact: An Example of Between-College Effects*. Paper presented at the annual meeting of the Association for the Study of Higher Education, Louisville, KY.
- Shermis, M. 2008. 'The Collegiate Learning Assessment: A Critical Perspective.' *Assessment Update* 20 (2): 10–12.
- Smith, J., A. McKnight, & R. Naylor. 2000. 'Graduate Employability: Policy and Performance in Higher Education in the UK.' *The Economic Journal* 110 (464): 382–411.
- Soundarajan, N. 2004. 'Program Assessment and Program Improvement: Closing the Loop.' *Assessment & Evaluation in Higher Education* 29 (5): 597–610.
- Steedle, T. 2012. 'Selecting Value added Models for Postsecondary Institutional Assessment.' *Assessment & Evaluation in Higher Education* 37 (6): 637–52.
- Steedle, J., & M. Bradley. 2012. Selecting Value added Models for Postsecondary Institutional Assessment. *Assessment & Evaluation in Higher Education* 37 (6): 637–52.
- Steedle, T. 2014. 'Motivation Filtering on a Multi-Institution Assessment of General College Outcomes.' *Applied Measurement in Education* 27 (1): 58–76.
- Tam, M. 2004. 'Using Students' Self-reported Gains as a Measure of Value added.' *Quality in Higher Education* 10 (3): 253–60.
- Tekwe, C., et al. 2004. 'An Empirical Comparison of Statistical Models for Value added Assessment of School Performance.' *Journal of Educational and Behavioural statistics* 29 (1): 11–36.
- Trowler, P., & V. Trowler. 2010. *Student Engagement Evidence Summary*. York: The Higher Education Academy.
- Tymms, P., C. Merrell, & P.I. Jones. 2004. 'Using Baseline Assessment Data to Make International Comparisons.' *British Educational Research Journal* 30 (5): 673–89.
- Tymms, P., & C. Dean. 2004. *Value added in the Primary School League Tables: A Report for the National Association of Head Teachers*. Durham: CEM.

- Tymms, P. 1999. 'Baseline Assessment, Value-added and the Prediction of Reading.' *Journal of Research in Reading* 22 (1): 27–36.
- UNISTATS. (homepage). 2014. As of 15 December 2014: <https://unistats.direct.gov.uk/>
- Universities and Colleges Admissions Service (UCAS). n.d. 'HE Subject Group'. UCAS.com, n.d. As of 15 December 2014: <http://www.ucas.com/data-analysis/data-resources/data-tables/he-subject>
- UniversitiesUK. 2004. *Measuring and Recording Student Achievement*, London: Universities UK.
- UniversitiesUK. 2007. *Beyond the Honours Degree Classification: The Burgess Group Final Report*. London: Universities UK.
- U.S. Department of Education. 2014. 'College Ratings and Paying for Performance.' As of 15 December 2014: <http://www.ed.gov/college-affordability/college-ratings-and-paying-performance>
- U.S. Department of Education, 2006. 'A Test of Leadership: charting the Future of US higher education, Final report on the future of higher education', As of 15 January 2015: <http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf>
- Van der Vleuten C., G. Verwijnen, & W. Wijnen. 1996. 'Fifteen Years of Experience with Progress Testing in a Problem-based Learning Curriculum.' *Medical Teacher* 18 (2): 103–10.
- Van der Vleuten, C., & L. Schuwirth. 2005. 'Assessing Professional Competence: From Method to Programme.' *Medical Education* 39 (3): 309–17.
- Verhine R., Vinhaes Dantas, L. and Soares, J. 2006. 'Do Provão ao ENADE: uma análise comparativa dos exames nacionais utilizados no Ensino Superior Brasileiro' (The Provão to ENADE: an analysis comparative national tests used in Higher Education Brazilian, Ensaio: aval. pol. públ. Educ., Rio de Janeiro, v.14, n.52, p. 291-310, jul./set. 2006, As of 15 January 2015: <http://www.scielo.br/pdf/ensaio/v14n52/a02v1452.pdf>
- Verhoeven, B.H., et al. 2000. 'The Effect on Reliability of Adding a Separate Written Assessment Component to an Objective Structured Clinical Examination.' *Medical Education* 34 (7): 525–29.
- Wabash National Study of Liberal Arts Education. n.d. *How Do Students Change over Four Years of College?* Crawfordsville, IN: Wabash College. As of 15 December 2014: <http://www.liberalarts.wabash.edu/storage/4-year-change-summary-website.pdf>
- Wainer, H. 2004. 'Introduction to a Special Issue of the Journal of Educational and Behavioural Statistics on Value added Assessment.' *Journal of Educational and Behavioural Statistics* 29 (1): 1–3.
- Wass, V., et al. 2001. 'Assessment of Clinical Competence.' *The Lancet* 357 (9260): 945–49.
- Watson, J.E. 2010. 'A Facelift for the GRE.' *Diverse: Issues in Higher Education* 27 (16): 7–8.
- Wiley, E.W. 2006. *A Practitioner's Guide to Value added Assessment*. Educational Policy Studies Laboratory Research Monograph. Tempe, AZ: Arizona State University.
- Young, J. 2007. *Validity of the Measure of Academic Proficiency and Progress (MAPP) Test*. Princeton, NJ: Center for Validity Research Educational Testing Service.

Appendix A: Summary of methods reviewed

The Table below briefly introduces all methods reviewed during this project, as well as their relative benefits and drawbacks.

Table 3: Summary of methods

	Examples	Where it is used	How it works	Advantages	Disadvantages
Grades	Predicted pass rate	Open University	Predicts individual student's pass rate to a given module	<ul style="list-style-type: none"> • Predicts likelihood of student passing a course • Provides useful data on student performance in specific disciplines • Enables longitudinal analysis • Appears to achieve suitable representativeness (given that it applies to all students) 	<ul style="list-style-type: none"> • Was not originally designed as a measure of learning gain • Uncertainty about how this data is used institution-wide • Grades problematic to use for comparative purposes, because they are not similar for a given individual at entry and graduation rate and that there are no common standards across institutions • In terms of the predicted pass rate, in particular, no data are collected on a student's performance before university, and the model focuses on comparing modules rather than individual performance

	Grade point average (GPA)	Oxford Brookes University	Represents an annual average of students' grades	<ul style="list-style-type: none"> • See assessment of Advanced Level Information System (ALIS) under mixed method 	<ul style="list-style-type: none"> • See assessment of Advanced Level Information System (ALIS) under mixed method
	Output indicators	Proposal	Uses output indicators, such as graduate employment rates or salaries, instead of grades upon graduation in England	<ul style="list-style-type: none"> • Compensates for the absence of comparability in grades at the level of entry and exit 	<ul style="list-style-type: none"> • Prone to criticisms regarding their validity: output indicators need to control for various factors, including institutional prestige or the socio-economic environment rather than learning gain, and would hence need to be further validated

Surveys	UK Engagement Survey (UKES)	30 UK HEIs	A survey containing 28 items focusing on student engagement and containing two questions on skills development	<ul style="list-style-type: none"> • Includes skills development questions • Has a formative exchange of practices element through the exchange of results among its participatory institutions • Collects additional data, for example, on students' levels of engagement, that could assist in interpretation of gains data • Is used by 30 institutions across the UK • Self-reports on engagement have been reported as correlated with various measures of student achievement in the United States 	<ul style="list-style-type: none"> • Not originally designed as a measure of learning gain • US research has questioned the association between students' test scores and measures of engagement • Uses cohort data rather than longitudinal data • Response rates have been relatively low • Only administered at one point (which could vary across institutions) during the course of a degree • Further validation required
	National Student Survey (NSS)	England-wide	A survey of English HEIs, focusing on the student experience, employment prospects, among other aspects, containing one question on personal development	<ul style="list-style-type: none"> • Includes one personal development question related to the measurement of learning gain • Administered UK-wide 	<ul style="list-style-type: none"> • Not a measure of learning gain • Only taken upon graduation • Surveys take time for students to fill in

	Careers Registration	Currently at University of Leeds, University of Lancaster, Liverpool John Moores University; due to go live in 2015 at University of Ulster; under consideration by numerous other HEIs	Asks students to assess their readiness in terms of career	<ul style="list-style-type: none"> • Includes personal development/employability questions • Repeated every year for the entire student population • Targets employability • Appears to be simple and efficient • Cuts across disciplines 	<ul style="list-style-type: none"> • Potential cultural barriers to replicability • Extensive validation would be necessary before wider implementation
	Skills audit	Durham University	Students take a questionnaire in order to be able to understand their skills and skills gap	<ul style="list-style-type: none"> • Measures one aspect of learning gain (skills) • Repeated at two points in time 	<ul style="list-style-type: none"> • Measures only one aspect of learning gain • Uncertainty regarding response rate
Standardised tests	Progress Testing (discipline specific)	University of Cardiff, University of Exeter, University of Swansea, University of Plymouth (originates from the Netherlands and the United States)	Students take the same exam of 125 items multiple times over the course of their studies so that progress can be observed over time; applied in medical schools	<ul style="list-style-type: none"> • More precise than discipline-independent tests, especially for content knowledge, and hence considered as valid • Allows for comparison across medical sciences • Enables longitudinal comparison across years for specific students • Use of national selection tests for medical schools could provide entry data • Response rate is representative 	<ul style="list-style-type: none"> • Fragmented • Requires a large bank of questions, which may be difficult for one institution to develop by itself • UK medical schools tend to already have school-specific exams

	Collegiate Learning Assessment (CLA) (independent from discipline)	US	The CLA is a standardised test including two performance task questions and 25 multiple choice questions to help assess general skills such as critical thinking and problem solving at two points in time using open type of questions	<ul style="list-style-type: none"> • Measures generic skills in ‘holistic’ manner • Allows for comparison across institutions • Allows for longitudinal measurement 	<ul style="list-style-type: none"> • Not piloted in England; hence would need to be further validated • Response rates can be low (CLA uses samples of students rather than the entire cohort)
	Measure of Academic Proficiency for Progress (MAPP) (independent from discipline)	US	MAPP is a standardised test including between 36 and 108 questions measuring general skills per general group of discipline	<ul style="list-style-type: none"> • Measures generic and discipline-specific skills 	<ul style="list-style-type: none"> • Not piloted in England • Multiple choice questions an ‘oversimplification’
Mixed methods	Academic Progress Tracking Tool (APTT)	Oxford Brookes University	This performance tracking tool covers a wide range of indicators, including the GPA and student surveys; it enables staff to have an overview of outcomes for a specific programme and to engage in learning analytics	<ul style="list-style-type: none"> • Combines various metrics (GPA, surveys) and students’ learning data, hence should be more robust than using a unique method • Allows for longitudinal comparisons • Tied to student records, so allows for high representativeness • Provides outcomes aligned with key performance indicators for honours students and features data on discipline-specific skills • Used institution-wide (after piloting stage in 2012) 	<ul style="list-style-type: none"> • May require broad institutional capacity to implement • Based on self-reporting data, and as is often the case with metrics, student characteristics such as gender or ethnic group correlate with responses • Shortcomings of each individual method may still apply
	Advanced Level Information system (ALIS)	Hull College	ALIS combines grades and a test to predict students’ performance	<ul style="list-style-type: none"> • Tool to measure value added based on grades and tests, hence should be more 	<ul style="list-style-type: none"> • For further education and has not been piloted in

				<p>robust than using a unique method</p> <ul style="list-style-type: none"> • Allows for longitudinal comparisons • Representativeness should be high if ALIS is presented as a requirement 	<p>higher education</p> <ul style="list-style-type: none"> • Provides score of likely achievement on the programme (rather than actual learning gain) • Shortcomings of each individual method may still apply
Qualitative methods	Personal development portfolio (PDP)	Various institutions	A log constituted by students over their studies; can be assessed or simply used for supervision purposes	<ul style="list-style-type: none"> • Students get an opportunity to reflect on their learning • Helps students and staff by identifying the learning outcomes students find most difficult 	<ul style="list-style-type: none"> • Difficult to compare across units of analysis • Difficult to derive rigorous longitudinal measurement from PDPs

Appendix B: Research method

This section consists of the method deployed to conduct this work, in accordance with the initial proposal submitted by RAND Europe.

This method comprises an overview of the literature, which allowed the research team to identify the different ways to measure learning gain internationally and in England. This rapid evidence review was complemented by a call for information issued to relevant networks by the HEA. This call for information provided the research team with valuable information regarding the perception and understanding of learning gain across the sector, as well as the methods used to measure learning gain. In order to obtain more details regarding these methods, the team conducted a series of interviews. A first round of interviews was conducted with members of professional bodies and of relevant associations (Association of Graduate Careers Advisory Services), who provided discipline-oriented perspectives as well as the perspective from employers. A second round of interviews was conducted as part of institutional case studies, in order to provide more information regarding the use of the methods in institutions and the dynamics through which these methods have emerged.

This method provided a way to achieve a great scope of research, combining international overview with sector-wide information. It is, however, important to bear in mind that the method relies on a sample of disciplines and HEIs. The HEIs covered by this report are those who replied to the call for information and agreed to an interview. Additional ways to measure learning gain may exist in institutions that were not covered by this research. In addition, there may be a self-selection bias in the institutions that responded to the call for information and agreed to the interviews: those who replied may be more positive and active regarding the concept of learning gain.

Literature review

A large part of the method for this project relies on an overview of the international and UK-based literature. The RAND team undertook a comprehensive and systematic literature review, called a rapid evidence review, of the major documents on learning gain, value added measurement and student engagement, both in the UK and internationally. The rapid evidence review also touched on the methods available in the secondary education sector and their potential for transferability to higher education. In

terms of process, the rapid evidence review relied on existing literature summaries and major literature reviews (such as the recent work conducted on the NSS⁷²) and expanded the search to other major authors and topics. The review also served to identify relevant experts, concepts and proxies in this field. A brief breakdown of the literature review process can be found below.

In the first stage of the review, the team coordinated with the specialist education librarian in RAND’s headquarters to tailor the search strategy and search terms. The team worked iteratively with the librarian to develop the most suitable search terms for this topic and its sub-themes. The breakdown of papers reviewed is included in Table 9 below.

Table 4: Breakdown of reviewed papers

Category	Content	Additional information	Number of articles
Background information	Paper title		
	Papers meeting inclusion criteria	Partly	48
		Fully	138
	Country/countries	UK	38
		Australia	11
		United States	88
		International	18
		N/A	54
	Study type	Grey literature	90
		Peer review	75
News		21	
Learning gain measurement	Methods to measure learning gain		139
	Use in English higher education		27
	Benefits and approaches to measurement of generic skills		131
Applicability	For improvement of learning		90
	For prospective students and their advisers		37
	For the impact of particular learning and teaching interventions		44
	To assist in international comparisons		42
	To form part of the quality assurance of learning and teaching		59
	For other use		2
Transferability and pilot design/implementation	Transferability to UK context		85
	Remarks on pilot design and implementation		16

⁷² Callender et al., 2014.

Category	Content	Additional information	Number of articles
Other	Value added in UK schools		23
	Experts		22
	Other notes		182
	Articles from experts		22

During the second stage, which focused on data extraction, the team endeavoured to analyse each text against a data extraction matrix developed in line with the original study questions presented in the invitation to tender. Each title and abstract in the ‘core set’ of 178 texts included was reviewed a final time against (i) whether it answered any of the study questions and (ii) what the potential quality of the data in the article was (i.e. was the data source a news article, a peer reviewed study, a case study, etc.). Owing to time constraints, the team then prioritised data extraction for a subset of articles. This implies that only the most relevant articles are covered in this literature review. The articles shortlisted for full review within the core set were selected based on how seminal they were for each sub-theme within learning gain (i.e. student engagement, value added measurement, standardised testing, and the pros and cons of specific methods for measuring learning gain, among others), and the level of expertise of the author (based on publications and participation in specialist workshops or in projects). For each of the articles selected for full review (about 65 within the core set), data were extracted according to the various dimensions of the matrix, which can be found in Table 10 below. The review also included non-peer reviewed papers at this initial stage, in order to get a broad overview of the field. The final stage followed from the data extraction, and comprised writing up the literature review based on the data from the articles. The template for the data extraction matrix is included in Table 10 below.

Table 5: Data extraction matrix used for the literature review

1. Background information					2. Learning gain measurement			3. Applicability				
Paper title (briefly)	Inclusion	Country	Study type	Study aims/data analysis	Method to measure learning gain	Use in English higher education	Benefits of approaches to 'generic' skills	Applicability of tools (1)	Applicability of tools (2)	Applicability of tools (3)	Applicability of tools (4)	Applicability of tools (5)
	Does the paper meet our inclusion criteria? 1. Relevance to the study questions 2. Data quality (opinion vs. peer reviewed)			Very <i>succinct</i> description; use " " to quote the article if need be, and square brackets [] to input your own interpretation	A. What are the available discipline-independent and discipline-specific methods for measuring learning gain?	B. In what ways and for what purposes are methods for measuring learning gain already in use in English higher education?	C. Analysis of the relative benefits of methods to measuring generic skills independently of discipline-specific knowledge, and measuring generic skills in disciplinary contexts	D.1. Analysis of the applicability of methods for measuring learning gain for different identified purposes, such as to inform improvements to learning and teaching	D.2. Analysis of the applicability of methods for measuring learning gain to provide information to prospective students and their advisers	D.3. Analysis of the applicability of methods for measuring learning gain to investigate the impact of particular learning and teaching interventions and contextual factors	D.4. Analysis of the applicability of methods for measuring learning gain to assist in international comparison	D.5. Analysis of the applicability of methods for measuring learning gain to form part of the quality assurance of learning and teaching

4. Transferability and pilot design/implementation	5. Other			
Transferability to English context	Remarks on pilot design and implementation	Value added in schools in England	Experts	Other notes
E. What are the relevant considerations for the use or adaptation of methods or tools for measuring learning gain in an English higher education context?	F. What are the relevant considerations for the design and development of robust pilot activities for measuring learning gain that might be drawn from existing methods and the literature?	G. What relevant lessons can be drawn from research on value added undertaken in the UK schools sector?	H. Who are the experts and practitioners on learning gain in England and internationally, and what type of expertise or advice might they be in a position to offer in steering the direction of subsequent stages of the project?	Is there anything else the article is helpful for? Any other information we should keep? Please provide a page number if possible.

Call for information to stakeholders

The second main method employed by the team to generate evidence as part of this project was to consult stakeholders in the field of higher education in England. Questions for the call for information were developed in agreement with HEFCE, BIS and the HEA.

The call for information aimed to retrieve data on any methods in place in England to measure learning gain, but also to consult stakeholders on their thinking in the area. The call for information focused on three major themes, (i) the definition of learning gain, (ii) methods to measure learning gain that were in place in the respondent's institution or field, and (iii) the pros and cons of learning gain measurement.

The call for information was launched on 20 August 2014 by the HEA. It was distributed to more than 33,000 individual academics, 202 pro-vice-chancellors or equivalents, and a number of academic registrars. The final number of responses received was 150; after removing duplicates and blank responses, 147 valid responses were analysed. The research team has some uncertainty regarding the position of the respondents. However, the research team assumes that a large number of respondents were individual academics.

Respondents appeared to include faculty staff, doctoral students, and administrators. To lessen the burden on respondents, they were not asked to specify their job title or function. By the end of the call for information, out of 147 responses analysed, about 13 came from further education colleges and 19 from institutions located outside of England (in the UK or beyond).⁷³ About 10 per cent of all respondents were affiliated with organisations which were not HEIs (i.e. sixth form colleges, foundations, NHS trusts). The international respondents were included in the answers provided below, except in the section on methods used for measuring learning gain, where the point of reference is HEIs in England only.

Interviews

To elicit expert knowledge on the topic of learning gain and its measurement in England, the team planned to conduct two rounds of interviews, with professional associations and with HEIs.

The first set of interviews consisted of discussions with representatives from a range of professional bodies, partly to identify which HEIs have distinguished themselves by focusing on the measurement of learning gain. The list of professional bodies is given in Table 6 below. After discussions with the commissioning bodies, the team identified a number of subject areas, whose representatives were selected for interview. Four subjects or stakeholder groups were selected for interviews (conducted face-to-face where possible): law, medicine, engineering and cross-professional bodies.

These disciplines were selected on the basis of the existing amount of work and research done on either learning outcomes, value added or learning gain; the degree of definition of the destination of graduates and expectations of employers; and the level of precision of standards in the profession.⁷⁴

In addition, four subject areas or stakeholder groups were selected for input via email: chemistry, accountancy, history, and the arts. These disciplines were selected given their representativeness and diversity of practices and in order to achieve a broader overview of disciplines.

⁷³ These international contributions ranged from Sweden to Nigeria to Australia.

⁷⁴ See, for example, OECD (2014: 78).

Interviews with representatives from professional bodies were conducted in August and September 2014, and provided useful insights into several methods and institutions the team sought to further engage with over the next phases of the research. Although the interviewees spoke on behalf of diverse organisations, several core themes were identified through the interviews.

Table 6: Professional bodies selected for interview

Type of input	Theme	Institution	Interview date
Face-to-face interviews	Cross-professional	Professions for Good	N.A.
	Cross-professional	Professional Associations Research Network	21/08/2014
	Cross-professional	Chartered Institute of Personnel & Development	N.A.
	Law	The Law Society	21/08/2014
	Law	The Bar Council	N.A.
	Medicine	Royal Society of Medicine	N.A.
	Medicine	General Medical Council (GMC)	21/08/2014
	Engineering	Royal Academy of Engineering	22/08/2014
Email input	Accountancy	Consultative Committee of Accountancy Bodies	28/08/2014
	Accountancy	Chartered Accountants	15/08/2014
	History	Historical Association	N.A.
	Arts	Arts Council	N.A.
	Chemistry	The Royal Society of Chemistry	01/09/2014

The interview schedule developed for professional bodies revolved around a selection of core strands, focusing on (i) familiarity with learning gain, (ii) tools for measuring learning gain in the field and awareness of data used for this purpose, (iii) any learning gain measures used by the employers represented by the professional body interviewed, and (iv) any further recommendations for the study. A copy of the interview protocol for professional bodies can be found in Appendix B.2 of this report.

In addition, on the commissioning team’s advice, the team contacted the Association of Graduate Careers Advisory Services (AGCAS) and arranged an additional interview with an expert from the University of London. This contact led to further contacts at the University of Leeds.

The interview schedule used in interviews with HEIs had a specific focus on methods and the higher education context. It comprised four main themes: (i) an overview of the method in place as described in the call for information; (ii) an institutional assessment of the extent to which this method is widespread, which were the enablers or barriers to its implementation, etc.; (iii) a series of questions on the potential transferability of the method across the sector and within English HEIs; and, finally, (iv) any further recommendations for the study. A copy of the interview protocol for HEIs can be found in Appendix B.3 of this report.

The research team also conducted 20 interviews with HEIs. Interviews with HEIs were arranged with a sample of 13 institutions. The majority of these institutions were implementing relevant institution-wide

methods. Four of the institutions interviewed used faculty-wide methods, and one was selected because of a relevant classroom method.

Expert consultation

Finally, the research team also benefited from the input of an expert workshop (and written consultation). The workshop was held on 10 November 2014 in London.

This workshop was the occasion for experts to present their specialist views on the topic, as well as for the research team to obtain some constructive feedback on the draft evaluation framework.

The list of consulted experts is summarised in Table 7 below.

Table 7: List of consulted experts

Attendees to the workshop	
Name	Institution
Bob Gilworth	University of London
Harvey Goldstein	University of Bristol
Ellen Hazelkorn	Dublin Institute of Technology
Camille Kandiko-Howson	King’s College London
Mark Langan	Manchester Metropolitan University
Individuals who provided written feedback	
Trey Miller	RAND Corporation
Anna Vignoles	University of Cambridge
Mantz Yorke	Lancaster University

The drafting process

This report has been drafted by RAND Europe, and reviewed for quality assurance in accordance with ISO (International Organisation for Standardisation) requirements. An advisory board including three leading academics provided expert advice.

The drafting process took place between mid-November 2014 and mid-January 2015. The commissioning bodies, HEFCE, BIS and HEA issued some feedback regarding the draft final report on two occasions and suggested some changes in the formulation of the text in the executive summary. These suggested changes to the executive summary were largely editorial.

Experts referred to in Table 7 commented on a draft of the report and were consulted. Interviewees, having shared their institutional practices, also had the opportunity to provide comments on the report. These comments have been logged and are available upon request.

Appendix B.1: Search protocol for the literature

This appendix summarises the search protocol used by RAND Europe to conduct the rapid evidence review.⁷⁵

It also summarises the population, intervention(s), comparison(s), outcome(s), and study types of interest used as part of the review.

Participants

The review included studies which were conducted on students in higher education; value added models based on pupils in secondary education; and studies conducted on graduates of HEIs in the UK and internationally. The studies used concentrated on particular methods used as proxies for the measurement of learning gain.

International comparisons

The review also includes publications with reference to international comparisons. International comparisons of particular relevance include countries which have relevant examples of learning gain measures. These countries include, in particular, the United States, Australia, Ireland, as well as Brazil, Columbia, China and the Netherlands.

An analysis of the grey literature among international organisations, such as the OECD (for the feasibility study on the measurement of learning outcomes) and the European Commission, was also relevant.

Study designs

The review included publications following the following designs:

- Review pieces related to the general UK context, including quality debates.
- Papers which present a proposed or applied method to measure learning gain.
- Papers which assess existing measures of learning gain and value added.
- Papers which discuss value added models in secondary education in the UK.

⁷⁵ A rapid evidence review is a comprehensive, systematic and critical assessment of the scope and quality of available evidence. Rapid evidence reviews follow the same structure and are as replicable and transparent as systematic literature reviews, yet have the advantage of being less resource intensive. This is achieved by formally constraining the types of research to be sourced on the basis of where the research was published, in which language it was conducted, and during which time period it took place.

Timeline

The review concentrated on papers published between 1990 and 2014. The start date of 1990 was decided in collaboration with the advisory board, given that debates related to value added in secondary education took place in the 1990s in the UK.

Search strategy

Conducting the searches

Search terms that were used to cover the topic included the following:

- **Primarily:** learning gain, value added measurement, outcomes, entry criteria, graduate test, employment/career testing, student assessment, skill testing (generic skill testing), standardised tests
- **Secondarily:** student engagement surveys, student satisfaction surveys, student experience surveys, collegiate learning assessment, personality profiling, standardised tests

The searches were conducted in three rounds, each resulting in a set of articles. More information about each set, such as the search terms used and number of results retrieved, can be found below.

First set

The citations in the set are search results from the databases ERIC and Education Abstracts. The focus was primarily four-year colleges and student assessment (as opposed to programme or college assessments).

- Search terms: “value added”, “learning gain”, “outcomes”, “colleges DE”, (DE=descriptor which is a controlled vocabulary term) “college outcomes’ assessment”, DE “higher education”, US, “United States” UK or “United Kingdom”
- Number of results: 202 citations retrieved

Second set

The citations are a result of searches in the databases Academic Search Premier, EconLit, ERIC and Education Abstracts. The search focus aimed to encompass specific subtopics, as suggested by researchers in early May 2014. These themes included articles on particular proxies used to measure learning gain; AHELO; CLA; tests run by ETS; NSS (UK); NSSE (United States); and articles on value added measurement in secondary education in England.

- Search terms: “value added”, “higher education”, AHELO, CLA, ETS, “national student survey”, “higher education”, college? [? = symbol to include college and colleges]
- Number of results: 470 citations retrieved

Third set

The citations are a result of searches in the databases Academic Search Premier, ERIC, and Education Abstracts that focus on relevant articles by authors such as Ramsden and van der Vleuten.

- Search terms included: AU [author] “Van der Vleuten, Cees ”, “value added or learning gain”, AU “Ramsden, Paul”
- Number of results: 30 citations retrieved

Screening the literature

After the searches were retrieved, the expert RAND librarian reviewed articles by title and abstract to screen out irrelevant articles. The inclusion/exclusion criteria for this process can be found below. This resulted in a smaller number of articles being shortlisted:

- Set 1: 23 citations shortlisted (out of 202 citations reviewed)
- Set 2: 53 citations shortlisted (out of 470 citations reviewed)
- Set 3: Five citations shortlisted (out of 30 citations reviewed)

Inclusion/exclusion criteria:

- Publication date range: 1990 to present
- Language: English

Table 8 summarises the inclusion and exclusion criteria used by the librarian.

Table 8: Criteria for inclusion in the literature review

Criterion	Include	Exclude
<i>Quality of the evidence</i>		
Conference abstract	x	
Conference summary	x	
Newspaper article	x	
Opinion piece	x	
Commentary	x	
Bulletin	x	
Announcement		x
Guideline/guidance (without evaluative component)		x
Journal article	x	
Report	x	
Book chapter	x	
Review	x	
Journal editorial	x	
<i>Country</i>		
Focus on but not limited to UK, US, Australia, Ireland, Brazil, Columbia, EU, The Netherlands	x	
<i>Content</i>		
Mere description/narrative of a system without evaluative component (outcome)		x
Evaluative component (outcome) unrelated to topic (i.e. programme evaluation, teacher evaluation)		x
Case study	x	
Framework or theoretical paper on relevant topic		x
Very technical (highly specific models)	x	
Broad modelling, theoretical		x

Reviewing the articles

Further to the searches being conducted, the RAND Europe team received an EndNote file containing 196 documents: 170 had been retrieved, while 26 articles were not available in full text via the RAND library. The team sent emails to the authors to retrieve the papers, and managed to retrieve seven articles. In the end, 75 articles were selected for full review (with a remaining seven articles not being immediately available). The references and abstracts were downloaded into MS Excel and reviewed and mapped onto the study questions. Seven articles were also excluded either because they were duplicates, were too dated to constitute a recent contribution, or touched on issues which were not relevant to the research questions for this project. The advisory group and expert panel members suggested an additional number of articles and documents, and other documents were added.

Study selection and data extraction

Two individuals screened titles and abstracts of studies identified from the literature search. This process was repeated to determine the list of final papers to be included in a review. The results from the search were circulated for feedback to an advisory board of external experts.

Interpreting the findings

The last step of the review process was to interpret the findings in light of the quality level of the body of evidence. Importantly, the degree of consistency across studies was considered. The evidence for each question was presented in the literature review, as were any gaps in the literature.

Appendix B.2: Interview protocol for professional bodies

Background

RAND Europe (an independent, not-for-profit research institute) has been commissioned by the Higher Education Funding Council for England (HEFCE) in July 2014 to produce a study on learning gain in higher education.

- This is part of a broader, joint HEFCE/BIS/Higher Education Academy (HEA) project to explore approaches to the measurement of learning gain. This study aims to critically evaluate a range of assessment methods (including both discipline-based and generic skills testing) for learning gain.
- The first phase of our study consists of a review of the literature to examine key documents on learning gain, value added measurement and student engagement across the UK and internationally. The review also looks into methods in the secondary education sector, which may be applicable to higher education.
- The second phase of the study consists of interviews with professional bodies concerning higher education in England. We aim to discuss the measurement of skills and learning gain in your discipline, but also to ask you about institutions that you feel have implemented good methods in your field, and why you feel these methods are of value.
- Before we start, a couple of short points:
 - The interview should last **60 minutes**.
 - We seek permission from all interviewees to **record the interviews**. As such, may we record this conversation, please? Any citations to be derived from our conversation would be anonymised, with reference made generally to the role of the interviewee and the type of organisation represented. Any such description would be agreed with you in advance of any reporting or publication.

Information on subjects

[**This information is provided solely for the interviewer's use.** It is expected that some of the respondents (particularly in HEIs) will be familiar with the skills students in their field are expected to command by the time they graduate. By contrast, some of the professional bodies may be less familiar with this subject-specific information, in which case the interviewer can briefly outline some of the skills required, based on what is provided below.]

- **Medicine**

- Approximate ratio applicants/applicants accepted 2013: 35%⁷⁶
- Entrance exam: UK Clinical Aptitude Test (UKCAT)/ Biomedical Admissions Test for medicine
- Subject-specific skills for medicine include, according to the Quality Assurance Agency (QAA):
 - Normal structure and function of the major human biological systems/life cycle and how they interrelate
 - Behaviour and relationships between individuals and others
 - Key mechanisms maintaining homeostasis
 - Causes and mechanisms of disease
 - Mechanisms of drugs treatment/non-pharmacological therapies
 - Impairment, disability and handicap, and the principles of rehabilitation
 - Disease surveillance and screening, disease prevention, health promotion, etc.
 - Healthcare planning and prioritisation of service, health economics, etc.
 - Organisation of the NHS
 - Ethical and legal background of medical method
 - Educational principles through which learning takes place
 - Critical thinking
 - Clinical reasoning
 - Insights into scientific research
 - Resourcefulness and creativity
 - Coping with uncertainty
- Generic skills for medicine include, according to the QAA⁷⁷:
 - Use the reflective method and lifelong learning
 - Retrieve, manage, and manipulate information
 - Effectively communicate and argue, both oral and written
 - Manage time and set priorities
 - Effectively manage time and resources and set priorities
 - Apply the principles of scientific research and audit
 - Study topics in depth
 - Deal with uncertainty and work within a changing environment
 - Teach effectively and act as a mentor to others
 - Work effectively within a team
- **History**
 - Approximate ratio applicants/applicants accepted 2013: 87%
 - Subject-specific skills for history include, according to the QAA⁷⁸:

⁷⁶ Please note that this approximation is based on UCAS data (<http://www.ucas.com/data-analysis/data-resources/data-tables/he-subject>), which is not a perfect match per subject (i.e. there is no separate category for history, but only for 'historical and philosophical studies', and accounting is probably located under 'business and administrative studies', etc.).

⁷⁷ Quality Assurance Agency, 2002: 4–5.

⁷⁸ Quality Assurance Agency, 2007a: 4–5.

- Ability to understand how people have existed, acted and thought
- Ability to read and analyse texts and sources critically
- Appreciation of the complexity/diversity of situations, events and mind-sets
- Understanding of problems inherent to historical documents (i.e. standpoint)
- Basic critical skills for testing the validity of documents
- Intellectual independence
- Marshalling arguments
- Generic skills for history include, according to the QAA:
 - Self-discipline, self-direction, independence of mind, and initiative
 - Ability to work with others and have respect for others' reasoned views
 - Ability to gather, organise and deploy evidence, data and information
 - Knowledge of how to identify, find, retrieve, sort and exchange information
 - Analytical and problem-solving ability
 - Structure, coherence, clarity and fluency of oral and written expression
 - Intellectual integrity and maturity, insight and creativity
- **Law**
 - Approximate ratio applicants/applicants accepted 2013: 87%
 - Entrance exam: National Admissions Test for Law (LNAT)
 - Subject-specific skills for law include, according to the QAA⁷⁹:
 - Knowledge of the legal system studied (concepts, values, legal institutions, etc.)
 - Applying knowledge for problem solving
 - Using sources and conducting research
 - Transferable skills and key skills for law include, according to the QAA:
 - Analysis, synthesis, critical judgement and evaluation
 - Autonomy and ability to learn
 - Communication and literacy
 - Numeracy, information technology and teamwork
- **Chemistry**
 - Approximate ratio applicants/applicants accepted 2013: 88%
 - Subject-specific skills for chemistry include, according to the QAA⁸⁰:
 - Chemistry-related cognitive abilities and skills, i.e. abilities and skills relating to intellectual tasks, including problem solving
 - Chemistry-related practical skills, e.g. skills relating to the conduct of laboratory work
 - Generic skills that may be developed in the context of chemistry and are of a general nature and applicable in many other contexts
 - Generic skills for chemistry include, according to the QAA:
 - Communication skills, covering both written and oral communication problem-solving skills, relating to qualitative and quantitative information
 - Numeracy and mathematical skills, including such aspects as error analysis

⁷⁹ Quality Assurance Agency, 2007b: 3.

⁸⁰ Quality Assurance Agency, 2007c: 4–6.

- Order-of-magnitude estimations, correct use of units and modes of data presentation
- Information retrieval skills in relation to primary and secondary information sources, including information retrieval through online computer searches
- IT skills
- Interpersonal skills relating to the ability to interact with other people and to engage in team work
- Time management and organisational skills, as evidenced by the ability to plan and implement efficient and effective modes of working
- Skills needed to undertake appropriate further training of a professional nature
- **Accounting**
 - Approximate ratio applicants/applicants accepted 2013: 84%
 - Subject-specific skills for business and management include, according to the QAA⁸¹:
 - Knowledge of organisations and their management
 - Knowledge of the external environment (economic, political, ethical, etc.)
 - Knowledge of management processes and methods
 - Knowledge of markets, customers, finance, people, operations, IT, business strategy, and such issues as sustainability, globalisation, risk management, etc.
 - Generic skills for business and management include, according to the QAA:
 - Critical thinking, analysis and synthesis
 - Effective problem solving and decision making using quantitative/qualitative data, researching business and management issues
 - Effective oral and written communication
 - Numeracy/quantitative skills
 - IT skills
 - Discipline and time management, initiative and entrepreneurship
 - Interpersonal skills and effective performance within a team
- **Engineering**
 - Approximate ratio applicants/applicants accepted 2013: 84%
 - Subject-specific skills for engineering include, according to the QAA⁸²:
 - Ability to be rational and pragmatic, interested in the practical steps necessary for a concept to become reality
 - Desire to achieve sustainable solutions to problems and have strategies for being creative and innovative and overcoming difficulties by employing knowledge in a flexible manner
 - Be numerate, highly computer literate and capable of attention to detail
 - Be cost- and value-conscious and aware of the social, cultural, environmental, health and safety, and wider professional responsibilities an engineer should display

⁸¹ Quality Assurance Agency, 2007d: 2–4.

⁸² Quality Assurance Agency, 2010: 3.

- Appreciate the international dimension to engineering, commerce and communication
- When faced with an ethical issue, be able to formulate and operate within appropriate codes of conduct
- Generic skills for engineering include, according to QAA:
 - Being professional in one's outlook, capable of team work, an effective communicator, and able to exercise responsibility
- **Exam for generic skills/other exams**
 - Thinking Skills Assessment (TSA) for measuring generic skills for undergraduate admissions
 - Graduate Record Examination (GRE): test used for admission into postgraduate courses in the United States; focuses on numeracy, literacy, and reasoning skills
 - Scholastic Aptitude Test (SAT): test used for admission into university as an undergraduate in the United States; focuses on numeracy, literacy, and reasoning skills

Questions

For this study, we have defined learning gain as discipline, knowledge and skills, as well as broader forms of development (such as 'generic' skills, which include effective communication, team skills, etc.) acquired during studies. Please answer the following questions using the above definition of learning gain.

Skills acquired through tertiary education

[Mention we have looked at subject benchmark statements defined by the Quality Assurance Agency, so answers to the first two questions can be short.]

1. 'Discipline knowledge and skills' may be defined as material or skills that students learn through their classes at university (i.e. taught by a lecturer, often in a formal setting). What do you feel are the key 'discipline knowledge and skills' skills which graduates in your field should have?
2. The OECD defines 'generic skills' as skills which are not subject-specific, such as 'critical thinking, analytical reasoning, problem-solving, and written communication'.⁸³ What do you feel are the main 'generic' skills which graduates in your field should have?

Tools for measuring learning gain in your field

3. Are you familiar with any methods, tests or tools, either within particular higher education institutions or more broadly, for measuring the components of learning gain, such as 'discipline knowledge and skills' or 'generic skills' in your field or beyond?
 - a. Please describe the benefits and drawbacks of using such tools or methods in the context of learning gain.
 - b. If yes, can you provide details of the methods, tools or tests used in your field, and where they are implemented?

⁸³ For further information, please see <http://www.oecd.org/edu/skills-beyond-school/testingstudentanduniversityperformancegloballyoecdshelo.htm>

- c. Are you aware of standard metrics for pre-university admissions in the UK in your field, and whether they have been used to measure learning gain? This can include A-level results, but also subject-specific tests such as the National Admissions Test for Law (LNAT), etc.
 - d. Have specific data such as national surveys been used to measure learning gain in your field?
4. If you are aware of tools or methods for measuring learning gain, do you know what purpose(s) they are used for? This could include informing improvements to learning and teaching, providing information to prospective students, investigating the impact of particular learning and teaching interventions; assisting in international comparison, or their use as part of the quality assurance of learning and teaching.
5. What do you think are the main relevant considerations to be taken into account in developing or importing methods or tools for measuring learning gain in the English higher education context?
6. Are you aware of any data available which could be used for the measurement of learning gain in your field, in the UK or beyond?
 - a. If so, do you know whether it has been used for such purposes?
7. How diverse do you think universities' approaches to measuring learning gain are in the UK (both between departments in your field and between universities)?
 - a. If applicable, may we ask you to provide examples?
 - b. What proportion of the universities you work with are discussing the measurement of learning gain?
 - c. If so, do you know whether they measure learning gain in similar or different ways?

Employment market

- a. Are you aware of any measures of performance/competencies used by employers in your field to assess graduates post-graduation to test both generic and discipline knowledge and skills? Are you aware of any divergence/diversity across employers in testing the generic or 'discipline knowledge and skills' after graduation?
- b. Does your organisation have a role in helping to define/measure what students in your field need to know by the time they graduate? Do you know whether these standards influence applicant selection for university admission?
- c. Do you know whether these standards set progression benchmarks during undergraduate studies?

Further recommendations for the study

8. Is there anything further you would like to add?

9. Can you recommend any experts and practitioners we may talk to about this topic?

10. Which higher education institutions would you recommend we approach to better understand how learning gain is measured in higher education in your field? May we ask you to provide contact details for specific people, please?
 - *We are particularly seeking to obtain insights from institutions which are diverse in terms of their entry criteria, league table rankings, student body, etc.*

Appendix B.3: Interview protocol for higher education institutions

Background

RAND Europe (an independent, not-for-profit research institute) has been commissioned by the Higher Education Funding Council for England (HEFCE) in July 2014 to produce a study on learning gain in higher education. This is part of a broader, joint HEFCE/BIS/Higher Education Academy (HEA) project to explore approaches to the measurement of learning gain.

- This study aims to critically evaluate a range of assessment methods (including both discipline-based and generic skills testing) for learning gain, that we understand as discipline knowledge and skills as well as broader forms of development (such as ‘generic’ skills, which include effective communications, team skills, etc.) acquired over the course of one’s studies.
- The first phase of our study consists of a review of the literature to examine key documents on learning gain, value added measurement and student engagement across England and internationally. The review also looks into methods in the secondary education sector, which may be applicable to higher education.
- The second phase of the study consists of interviews with professional bodies concerning higher education in England, and with universities. We aim to discuss (i) the measurement of skills and learning gain in your discipline or in your institution and (ii) any other tools or methods in use in English higher education institutions or beyond that you may be aware of.
- Before we start, a couple of short points:
 - The interview should last **60 minutes**.
 - We seek permission from all interviewees to **record the interviews**. As such, may we record this conversation, please? Any citations to be derived from our conversation would be anonymised, with reference made generally to the role of the interviewee and the type of organisation represented. Any such description would be agreed with you in advance of any reporting or publication. The recordings will be stored for up to six months, and destroyed after the project is completed.
 - Name of interviewee & position:
 - Interviewer:
 - Date:

Questions

Overview of method

1. Thank you for replying to our call for information/email. We are very interested in finding out how you use [xxx method] to measure learning gain. Could we please have a bit more information on this method; for example, when it started, how you use it and what type of student/departmental or institutional feedback you receive?
2. What do you feel are the discipline-specific knowledge and generic skills and personal development elements that the method you use measures?
 - a. 'Discipline knowledge and skills' may be defined as material or skills that students learn through their classes at university (i.e. taught by a lecturer, often in a formal setting).
 - b. The OECD defines 'generic skills' as skills which are not subject specific, such as 'critical thinking, analytical reasoning, problem-solving, and written communication'.⁸⁴ What do you feel are the main 'generic' skills that the method you use measures?
 - c. 'Personal development' can be characterised by the acquisition of a level of maturity in a student's moral or practical reflection; it is another component of learning gain.
3. What are the advantages and disadvantages of using it to measure learning gain?
4. Which value and specific outputs has this method brought since you started implementing it?

Institutional assessment

5. To your knowledge, how widespread is this method within your institution, or across the sector?
 - a. Why do you think this method is widespread/limited?
 - b. What has stimulated your choice for this method? / What were your motivations?
 - c. Did you benefit from any support to implement this method or tool (i.e. institutional or external grant or other institutional incentive)?
 - d. Has the leadership of your institution made a conscious move to prioritise learning gain in response to/prior to requests from government bodies? Have any discussions taken place in your institution regarding the measurement of learning gain, or has the issue been prioritised by senior staff recently?
 - e. How is this method perceived within your institution?
 - f. Which barriers have you encountered, if any?
6. To your knowledge, which other methods are currently in place in your institution when it comes to learning gain measurement?

PROMPTS: For instance:

⁸⁴ For further information, please see <http://www.oecd.org/edu/skills-beyond-school/testingstudentanduniversityperformancegloballyoecdshelo.htm>

- a. Has data collected by your institution on student performance and/or experience and engagement been analysed by the administrators or by data analysts?
- b. Have the IT systems been modified to enable the collection and analysis of new metrics?
- c. Has external expertise been brought in to consult on how to measure learning gain?
- d. Does your institution use databases of students' records and grades, or data systems, or any other metrics the university may be collecting on its students, to assess learning gain?
- e. Are you aware of any plans for the future to develop your institution's capability when it comes to measuring learning gain?

PROMPTS: For instance, this could be through:

- i. The modification of the IT system
- ii. Additions to the HR team
- iii. Bringing in additional expertise, collecting extra data
- iv. Leadership making learning gain a priority in strategic plans

Sector-wide transferability

7. What do you think would be the benefits and constraints of using this method more broadly across the sector?
8. More generally, what do you think should be taken into account when developing or adapting methods or tools for measuring learning gain from other countries or disciplines in the English higher education context?
9. To which extent do you think universities in England should develop comparable approaches to measure learning gain (both across departments in your field and between universities)?
10. To which extent have you shared this method with colleagues from your institution or others?
11. Would there be any documents you could refer us to, especially regarding the evaluation or review of this method?

Further recommendations for the study

12. Which lessons have you learned or would you like to share?
13. Is there anything further you would like to add that we have not discussed, or that you think we should consider in our study?

Appendix C: Evaluation framework

The research team has also drafted an evaluation framework in order to further define and map the dimensions that interested parties would need to think of when considering using a method to measure learning gain and to test the feasibility of these methods to be scaled up more generally (as illustrated in Tables 9 and 10 at the end of the section).

Evaluation framework to assess the use of the method by a given party

This evaluation framework is presented in Table 9. Because this part of the framework is tailored towards the use by the sector, it has not been filled in by the research team. The ambition of the research team would be that stakeholders and interested parties use the questions in Table 9 in order to further shape their understanding of how and whether to use a given method.

In this respect, the framework asks questions about the purposes that the method would be used to achieve. These purposes were inferred from the terms of reference for this study. The terms of reference referred to the following purposes:

- Seeking to inform improvement to teaching and learning.
- Providing information to prospective students and their advisers.
- Supporting institutions in their enrolment decisions.
- Investigating impact.
- Assisting in institutional and international comparisons.
- Acting as part of the quality assurance of teaching and learning.
- Informing employers about students' performance (called 'support graduate transition to a profession' in the framework).⁸⁵

These purposes are written in a relatively general manner and can be articulated in several practical ways. For example, assisting in institutional comparisons could include discussions of participating institutions (as is the case with UKES) or publication of national league tables (as is the case with the NSS).

Investigating the impact of teaching and learning interventions can take the form of general discussions around an exchange of best practice or lead to the production of precise learning analytics.

⁸⁵ The research team does not assume that a given purpose is more valid than another in the context of the measurement of learning gain, since these purposes depend on the policy objectives that the commissioning bodies seek to achieve.

More generally, each method is more designed for a given purpose rather than for another. A survey such as UKES serves the purpose of encouraging institutions to exchange information in order to enhance student engagement and hence education outcomes. This function is in essence different from the purposes of the NSS, which functions primarily as an accountability and enhancement mechanism, in addition to informing prospective student choice.⁸⁶

The purpose of a standardised test such as MAPP or CLA is to compare the acquisition of generic skills (by comparing institutions in the case of the CLA and students in the case of MAPP).

The purpose of discipline-specific tests currently in existence in England, such as Progress Testing, is to assess students' skills, competencies and content knowledge, and their purpose is to understand how close these students are to being able to understand a specific profession, for example.

Other tests or mapping predictions of pass rates based on grades may help HEIs in their enrolment decisions or help them to assess different modules.

The mixed-method approach, whereby a range of methods was used, was found to support a more tailored approach to teaching and learning *within* an institution.

A second part of Table 9 also includes considerations related to how the method would be used by an institution. This section pays particular attention to the issue of ownership, particularly in the use of data, i.e. its disclosure, level of accountability, data protection rules and sustainability.

Data protection questions assess whether a given method corresponds to general data protection requirements, including informing respondents of the use of their data in a transparent manner, requiring respondents' consent before using data, and abiding by legislation.

The dimensions of disclosure, ownership, accountability and exchange of methods or sustainability are more subjective. For example, whether the results of a given method should be fully disclosed to the public or not is prone to debate. On the one hand, certain public authorities may encourage public disclosure in order to inform student choices. On the other hand, more restricted disclosure (limited to a group of institutions) has yielded positive effects regarding the exchange of best practices. In secondary education, Foley and Goldstein (2012) reported that feeding yearly scores back to schools as one item of information within an inspectorial system enabled these scores to be used alongside other information. The use of value added estimates as a 'screening device' has the potential to avoid many of the harmful side effects of published 'league tables' while still retaining key elements of an accountability device.

Whether a method is owned by a university, public agency or commercial entity (ownership dimension) equally leads to a debate. For example, a commercial ownership may imply that the method is particularly sophisticated, given the competitive environment in which commercial entities operate. But it may be resisted by the sector on the basis of being commercially rather than sector driven, as has been the case for some standardised tests.

A method could also be used as part of a national or institution-wide accountability process. National accountability would tie the method to the broader requirements of value-for-money for taxpayers, for

⁸⁶ HEFCE, 2014a.

example. Institutional accountability could help to enhance quality assurance and teaching and learning within an institution.

Finally, most of the questions around sustainability also depend on the purpose that users of the method seek to achieve. The costs of the method could be borne by students, the government, universities or third parties and/or benefit from national, institutional or other types of support. Who bears the costs and supports the measure indirectly replies to the issue of sustainability, in the sense that it allows the research team to identify broad cost repartition patterns (but does not allow the research team to understand whether the method is financially sustainable and/or cost efficient). Two more objective criteria include whether the method has been endorsed by the senior leadership of the institution (which would facilitate its adoption) and whether the method has been in place for longer than a year.

Table 9: Evaluation framework to assess the (potential) use of the method by the sector

Tick as appropriate		
1. Purpose		
		To inform improvements in teaching and learning
		To provide information to prospective or existing students and their advisers
		To investigate the impact of particular learning and teaching interventions and contextual factors
		To assist in international comparisons
		To assist in institutional comparisons
		To assist in quality assurance of learning and teaching
		To help with students' admissions
		Other
2. Reporting (disclosure, feedback, 'ownership' and accountability)		
Full disclosure	Public	Full disclosure to public
	Stakeholders	Reporting to funding, regulatory and accreditation agencies
Partial disclosure	Institutional	Analysis of measures of learning outcome (not including inference made from qualification-level accreditation)
	Departmental	Reporting to localised academic units, and not beyond these
	Individual	Reporting only to individual teachers or students
No disclosure	No reporting	<i>Ad hoc</i> or absent
Ownership	University	Are the copyrights to the tool or method owned by the university?
	Public Agency	Are the copyrights to the method owned by the government or other public agency?
	Commercial entity	Are the copyrights to the method owned by a commercial entity (third party to the university)?

Accountability and exchange of practices		Is the method used as part of national accountability and exchange of practices mechanisms?
		Is the method used as part of institutional accountability and exchange of practice mechanisms?
		Is the method used as part of other accountability or data-sharing mechanisms?
Data protection		Are respondents informed of how their data will be used using a transparent procedure?
		Are respondents provided with an opportunity to consent to the use of their data?
		Does data protection follow the national legislation in place?
Sustainability		Is the cost of the tool or method borne by students?
		Is the cost of the method borne by universities?
		Is the cost of the method borne by the government or other national agency?
		Is the cost of the method borne by a third party?
		Does the method benefit from institutional material support?
		Does the method benefit from public material support?
		Does the method benefit from other forms of support?
		Has the method been officially endorsed by the senior leadership of the institution?
		Has the measure been used for more than one year?
		Open question: Which other factors have contributed to the sustainability and/or expansion of the tool/method?

Evaluation framework to assess the potential for each method to be scaled up

In order to be scalable, a given method would need to have additional properties. First, the method would need to be transferable. The research team proposes to measure transferability based on whether it is already used across disciplines, institutions or countries, or whether stakeholders and experts think that the method has the potential to be transferable.

Second, the method would need to comply with quality assurance criteria and ideally already be the object of an engagement from peers and/or independent experts (based on review and quality assurance). The process would need to provide sufficient disclosure regarding all corporate and technical interests and processes.

Third, the method would need to be appropriate given the institutional capacity in higher education. For example, an appropriate method would rely on existing academic capacity (rather than require new capacity). Related considerations are whether the measure is available at two periods in time and whether other limitations have been identified during the data analysis phase of the project.

Fourth and finally, the method would need to show evidence of impact, for example, based on evidence of change in teaching and learning methods or institutional management. The framework also asks about the potential consequences of such a method. The experience of the implementation of value added models in secondary education illustrates the unintended consequences of setting up new metrics. Hence the adopted measure would need to pay attention to negative incentives; for example, for HEIs not to enrol certain types of students because they give them low value added (e.g. disabled students, dyslexic students, groups that underachieve).

Evaluating the 'readiness' of the higher education sector for a measure of learning gain

In addition, the sector would need to be 'ready' for a given method, based on whether the method has been recognised by the sector as well as institutional capacity. For example, a method which is known and recognised by the higher education sector may be more easily adopted than an unknown method.

In terms of institutional capacity, institutions would need to have suitable data systems, human resources to analyse the data, communication channels with relevant stakeholders and discussion platforms and appropriate training for academics to help them understand the issues and results arising.

Table 10: Evaluation framework to assess the potential of the method to be adopted by the sector

Based on characteristics of the method	
Transferability	Has the method been identified in the literature?
	Has the method been reported by respondents and interviewees?
	Is the method used across disciplines?
	Is the method used in multiple institutions?
	Is the method used internationally?
	Is there evidence of the tool or method being scalable?
	What is your expert assessment regarding transferability: positive, neutral or negative?
Review and quality assurance	Are the assessment tasks open to peer review?
	Is there evidence that peer review takes place?
	Is the method quality assured?
	Are the analytical processes verified by independent experts and open to academic review?
	Are the reporting processes extensive and sufficient to disclose all corporate and technical interests and processes?
Appropriateness	Does the method build on core academic capacity?
	Does the method require new and additional money (without building core academic capacity)?
	Is the method used in at least two periods of time (individual or cohort level)?
	Have other limitations been identified during the data analysis phase of this project?
Impact	Have some changes been observed in teaching and learning methods?
	Have some changes been observed in 'institutional management'?
	Have some changes been observed in the higher education system?
	Have some changes been observed in the student population?
	Were these changes 'intended'?
	Were there some 'unintended' changes?
	Is evidence of impact available, for example, through evaluation reports?

Based on the 'readiness' of the English sector for a measure of learning gain and on stakeholder perception and institutional capacity	
Stakeholder perception	Is there potential to build this method in UK higher education in robust and efficient ways?
	Has the method been recognised by the English higher education sector?
	Is the outcome of expert research and interviews relatively 'positive' or 'negative' about the applicability of the method?
Institutional capacity	Do HEIs have suitable data systems to administer the method?
	Do HEIs have an institutional research team or other analytical team to process the data?
	Could HEIs communicate this analysis to relevant stakeholders?
	Are the results of the analysis discussed during working groups or other events?
	Does the institution have the appropriate professional training for academics to help them understand the results and the issues arising?

Appendix D: Glossary

Accountability is the obligation on universities, colleges and nominated individuals to explain how well they have met their responsibilities, including the proper use of public funds.⁸⁷

Contextual Value Added measures progress between two educational stages and also takes into account a number of other factors outside the school's control, such as gender, special educational needs, movement between schools and family circumstances (all these factors are known to affect pupils' performance).

Contextual Value Added models make the comparison against other schools more meaningful, for example, when comparing the performance of a school in a leafy suburb against the performance of one in an inner city area – both of which might face quite different challenges.⁸⁸ Contextual Value Added models are similar to contextualised attainment models (CAMs). CAMs measure the influence of schools on student outcomes controlling for family background characteristics in cross-sectional studies. CAMs argue that family background characteristics are adequate substitutes for student prior attainment.⁸⁹

Competence is defined as a proven ability to use knowledge, skills and personal, social and/or methodological abilities, in work or study situations and in professional and personal development. In the context of the European Qualifications Framework, competence is described in terms of responsibility and autonomy. Examples include critical thinking, problem solving, modelling, innovation, creativity, design and investigation, collaboration and team working.⁹⁰

Content knowledge refers to the body of information that teachers teach and that students are expected to learn in a given subject or content area, such as English language, arts, mathematics, science, or social studies. Content knowledge generally refers to the facts, concepts, theories, and principles that are taught and learned, rather than to related skills – such as reading, writing, or researching – that students also learn in academic courses.⁹¹

Further education is for people over compulsory school age (currently 16 in England) which does not take place in a secondary school. Further education courses are generally up to the standard of GCE A level or NVQ Level 3.⁹²

⁸⁷ HEFCE, 2014b.

⁸⁸ Department for Education, 2014b.

⁸⁹ See Lenkeit, 2013; and Goe et al., 2008.

⁹⁰ European Parliament, 2008.

⁹¹ Hidden Curriculum, 2014.

⁹² HEFCE, 2014b.

Higher education provider is a broad term, also shortened to ‘HE provider’, which usually refers to any organisation that delivers higher education.⁹³

Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs.⁹⁴

Learning gain can be defined as the distance travelled by students across two points in time in terms of skills and competencies, content knowledge and personal development. However, this report has shown that different understandings of learning gain existed and learning gain was in turn assimilated to value added or learning more generally.

Learning outcome can be interpreted as learning objective or student achievement. Interpreted as learning objective, a learning outcome is a brief statement that describes what students will be expected to learn by the end of school year, course, unit, lesson, project, or class period. In many cases, learning objectives are the interim academic goals that teachers establish for students who are working towards meeting more comprehensive learning standards, closely related to learning progression and hence learning gain. The term may also be synonymous with *student achievement*, since *achievement* typically implies education-specific results such as improvements in test scores (Hidden Curriculum, 2014).

Likert scale is a widely used scale designed to measure respondents’ attitudes in a survey questionnaire. A Likert scale tests the level of agreement or disagreement of a respondent to a statement, typically on a scale ranging from ‘strongly agree’ to ‘strongly disagree’, and makes the assumption that attitudes can be measured.⁹⁵

Method is a particular procedure for accomplishing or approaching something, especially a systematic or established one.⁹⁶

Methodology is a system of methods used in a particular area of study or activity.⁹⁷

Performance indicators are a set of indicators produced by the Higher Education Statistics Agency (HESA) which measure the performance of HEIs in the UK in a number of areas.⁹⁸

Provider is a short-hand for a broad term, also sometimes written out in full as ‘higher education provider’ or ‘HE provider’, that usually refers to any organisation that delivers higher education.⁹⁹

Proxy is an indicator used to represent the value of something else in a calculation.

Skill is the ability to apply knowledge and use know-how to complete tasks and solve problems. In the context of the European Qualifications Framework, skills are described as cognitive (involving the use of logical, intuitive and creative thinking) or practical (involving manual dexterity and the use of methods,

⁹³ HEFCE, 2014b.

⁹⁴ Learning Analytics and Knowledge, 2011.

⁹⁵ MacLeod, 2008.

⁹⁶ oxforddictionaries.com

⁹⁷ oxforddictionaries.com

⁹⁸ HEFCE, 2014b.

⁹⁹ HEFCE, 2014b.

materials, tools and instruments). Competencies may incorporate a skill, but are more than the skill; they include abilities and behaviour, as well as knowledge, that are fundamental to the use of a skill.¹⁰⁰

Standardised test is any form of test that (i) requires all test takers to answer the same questions, or a selection of questions from common bank of questions, in the same way, and that (ii) is scored in a 'standard' or consistent manner, which makes it possible to compare the relative performance of individual students or groups of students.¹⁰¹

Student engagement is defined various ways. To date, 27 published definitions of student engagement have been reported. Scholarship on this issue originating from the UK (as opposed to the United States) includes a broad definition of student engagement, such as improving the motivation of students to engage in learning and to learn independently, as well as the participation of students in quality enhancement and quality assurance processes, resulting in the improvement of their educational experience.¹⁰² Student engagement also includes connections with structures and processes, as well as 'policies and methods supporting student involvement in shaping their learning experiences', including feedback, student representation or projects to engage students.¹⁰³

North American and Australasian scholars, who produce most of the literature on the topic,¹⁰⁴ connect student engagement with 'the degree to which students approach their studies in a way that contributes towards desired learning outcomes', including critical thinking and self-awareness, engagement with identity, seeing the 'bigger picture', putting theoretical knowledge into method, and personal fulfilment.¹⁰⁵ These then form 'generic skills', which are a component of learning gain. The link between engagement and learning gain is considered critical given the self-directed nature of tertiary studies.¹⁰⁶

Student experience is experience derived from any interaction, course, programme, or other experience related to the student's life in higher education, whether it occurs in traditional academic settings (schools, classrooms) or non-traditional settings (outside-of-school locations, outdoor environments), and whether it includes traditional educational interactions (students learning from teachers and professors) or non-traditional interactions (students learning through games and interactive software applications).¹⁰⁷

Universities are all HEIs, although not all of them choose to apply to be authority-funded.¹⁰⁸ Universities in the UK are diverse, ranging in size, mission and history. Nowadays, the Privy Council and Companies

¹⁰⁰ European Parliament, 2008.

¹⁰¹ Hidden Curriculum, 2014.

¹⁰² Quality Assurance Agency, 2012.

¹⁰³ Callender et al., 2014: 23. See also Trowler and Trowler, 2010: 3.

¹⁰⁴ For a more extensive literature on student engagement in the United States, see Cole et al. (2009); National Survey of Student Engagement (2002); National Survey of Student Engagement (2007). An example of the use of student engagement to measure learning outcome is available at National Survey of Student Engagement (2009).

¹⁰⁵ Callender et al., 2014.

¹⁰⁶ Pace, 1982.

¹⁰⁷ Adapted from Hidden Curriculum, 2014.

¹⁰⁸ HEFCE, 2014b.

House have the power to grant use of the title 'university' to an institution on the advice of BIS. There are certain prerequisites, including the need to have degree-awarding powers.

Validity is a term used to describe a measure that accurately reflects the concept it is intended to measure.¹⁰⁹

Value added is typically measured by the difference between a student's actual test score at a second point in time and his or her predicted test score at a first point in time, and in this respect it is similar to the concept of learning gain.

Value 3 modelling measures value added by comparing a student's actual outcome at Level 3 qualification to the student's estimated outcome (the outcome achieved by students of similar ability nation-wide). The difference between the student's actual and estimated performance gives the student a value added score in the Level 3 qualification. Students' value added is then averaged across schools and colleges at Level 3 qualifications to help identify the schools and colleges that are helping their students make more or less progress than average.¹¹⁰

Vignette is a short description of the use of a method by a particular institution or set of institutions.

¹⁰⁹ Babbie, 2001.

Appendix E: List of acronyms

AHELO: Assessment of Higher Education Learning Outcomes

APTT: Academic Progress Tracking Tool

ALIS: Advanced Level Information System

BIS: Department for Business, Innovation and Skills

CLA: Collegiate Learning Assessment

CVA: Contextual Value Added

GPA: Grade Point Average

HEA: Higher Education Academy

HEFCE: Higher Education Funding Council for England

HESA: Higher Education Statistics Agency

HEI: Higher Education Institute

MAPP: Measure of Academic Proficiency for Progress

NSS: National Student Survey

NSSE: National Survey of Student Engagement

OECD: Organisation for Economic Co-operation and Development

PDP: Personal Development Portfolio

PVC network: network of Pro-Vice Chancellors, Vice-Chancellors and Deputy-Vice-Chancellors

SAT: Scholastic Aptitude Test

UKES: UK Engagement Survey

UUK: Universities UK