Science Student Learning Gain in English Higher Education Institutions: The Development of a Skills Assessment Tool

Rebecca Jane Randles

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

June 2019

Declaration

I declare that this thesis has not been previously accepted in substance for any degree and is not concurrently submitted in candidature for any degree. I state that this thesis is a result of my own independent research except where otherwise stated.

Abstract

Learning gain is a potentially valuable measure of student development. Its importance was highlighted by discussions framing the teaching excellence framework (TEF), which considered the ‘distance travelled’ by a learner as a potential metric. There has been some difficulty surrounding utilising this metric within an English Higher Education context because the Higher Education Funding Council for England and Wales (HEFCE) pilot projects found that using the metrics that are currently being used within the United States is not appropriate. This is due to the need to identify learning gain at a subject level within English Higher Education degrees. With a mixed method approach, this research aimed to create a subject specific learning gain measure which could be used to measure learning gain of students within subject targeted cohorts. A case study methodology was adopted across Liverpool John Moores University (LJMU) with students from the Faculty of Science to focus specifically on one branch of the STEM subjects. STEM subjects were chosen as the targeted cohort due to their importance to functioning society and the recent decline in employable skills, therefore highlighting a need to measure their progress. Preliminary inquiry using interviews was undertaken to understand whether STEM students could be used as the cohort of students and to identify the skill developments to be identified within the assessment tool. This preliminary inquiry led to a focus on science subjects. The assessment tool was then created using published material of literature and learning assessments, in order to identify potential items and to split into subscales the identified areas of engagement, critical thinking and personal development. A pilot was undertaken to check initial wording and understanding of the metric to a cohort of students. Experiences of members of staff were also used to check the validity of the assessment tool. Modifications were made before proceeding to the main part of the research, which was the completion of the assessment tool by a representative cohort of students. Reliability and stability were tested using independent group samples from the targeted cohort using averages, reliability statistics and factor analysis. Validity testing of the measure was undertaken by triangulating interviews with members of staff and qualitative survey data from students with that of the quantitative outcomes of individual data whereby the students were followed over a period of one year to view their learning gain. It was concluded that the assessment tool appeared to be both reliable and valid on its intended targeted cohort of students. It was therefore concluded that it is possible to have a specified self-survey questionnaire which heeds results that are an accurate representation of reality.

Acknowledgments

I want to take this opportunity to thank all those people who made this thesis possible and who kept me going through my PhD Journey. Thank you to my director of studies Prof. Alison Cotgrave who believed in my abilities and whose door was always open to help me with any problems. You were always pushing me to be my best and I thank you for this incredible opportunity. Thank you to Dr. David McIlroy whose guidance on my writing and analysis I was incredibly grateful for. I would also like to thank Dr. Virendra Mistry who took a chance on me and helped me to write my very first article and vastly improve my writing ability. I will always be grateful for your guidance in this project.

Thank you to the staff in the LJMU Graduate School, Faculty of Engineering and Technology, and particularly to the Department of Built Environment. You gave me opportunities to expand my teaching and research skills. You allowed me to complete this research and believed in my ability to do so. I would also like to thank all the staff I have met along the way and who have supported my journey. Thank you to all the participants of this research; the staff and students who took the time to complete the interviews and questionnaires. Without you this research would not have been possible.

I would also like to thank my family and friends, without you all I would have stopped believing in my ability and myself. You were always there for me whenever I needed some time off, and when I needed a bit of a push to keep myself going. I could not have done it without each and every one of you and for that I will be eternally grateful.

I want to give a special mention to my parents for all the support they have given me. You both always asked me about my work and were happy for my achievements. I hope I have made both of you proud. Mum, thank you for all your support in my journey and believing in me. Dad, you were always there to push me to do my best, always believed I could do anything I want to, and made me want to make you proud. You were always interested in what I was doing and supported me. You have been my rock through this whole process and I thank you for that.

This thesis is dedicated to the memory of

Kerry Anne Randles

Stephen James Randles

Thank you for all the memories and motivation you gave me to do well. I am glad that you were both a part of my life, no matter how short. I hope that I have made you proud, to my little sister and grandad; you will always be in my heart.

Publications Arising from this Thesis

Papers

Randles, R., Cotgrave, A. and McIlroy, D. (2018) The development of a methodology to measure undergraduate science student learning gain. *Proceedings of Faculty Research Week at Liverpool John Moores University.*

Randles, R. and Cotgrave, A. (2017) Measuring Student Learning Gain: A Review of Transatlantic Measurements of Assessments in Higher Education. *Innovations in Practice, 11 (1), 50-59.*

Randles, R. and Cotgrave, A. (2017) The Voice of STEM: Using Student Interviews to Discuss Expected Learning Gains. *Proceedings of Faculty Research Week at Liverpool John Moores University.*

Presentations

Randles, R., Cotgrave, A. and McIlroy, D. (2018) *The development of a methodology to measure undergraduate science student learning gain.* Faculty Research Week, Liverpool John Moores University, Liverpool, UK.

Randles, R. and Cotgrave, A. (2017) *Measuring Student Learning Gains: Discovering the Transatlantic Dialogue*. Teaching and Learning Conference, Liverpool John Moores University, Liverpool, UK.

Randles, R. and Cotgrave, A. (2017) *The Voice of STEM: Using Student Interviews to Discuss Expected Learning Gains.* Faculty Research Week, Liverpool John Moores University, Liverpool, UK.

Posters

Randles, R. and Cotgrave, A. (2018) *The Transatlantic Dialogue: Discovering Measures of Student Learning Gain.* Higher Education Academy Annual Conference, Birmingham, UK.

Randles, R. and Cotgrave, A. (2017) *The Voice of STEM: Using Student Interviews to Discuss Expected Learning Gains*. Graduate Conference, Liverpool John Moores University, Liverpool, UK.

Randles, R. and Cotgrave, A. (2017) *Using Student Interviews to Discuss Expected Learning Gains*. Women in Research Conference, Liverpool, UK.

Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

ACT	American College Testing
ALIS	Advanced Level Information System
APTT	Academic Performance Tracking Tool
AUSSE	Australian Survey of Student Engagement
BIS	Department for Business, Innovation and Skills
BME	British, Black and minority ethnic
CAAP	Collegiate Assessment of Academic Proficiency
CAE	Council for Aid to Education
CCC	Calhoun Community College
CIWC	Center for Inquiry at Wabash College
CLA/CLA+	Collegiate Learning Assessment
CSFI	The College Success Factors Index
CVA	Contextual Value Added
DFA	Director of Fair Access
DLHE	Destinations of Leavers from HE Survey
EPP	ETS Proficiency Profile
GPA	Grade Point Average
HE	Higher Education
HEEC	Higher Education Evaluation Center
HEFCE	Higher Education Funding Council for England
HELGA	Higher Educational Learning Gain Analysis Project

LJMU	Liverpool John Moores University
MAPP	Measure of Academic Proficiency and Progress
NMMLGP	National Mixed Methodology Learning Gain Project
NSS	National Student Survey
NSSE	National Survey of Student Engagement
PDP	Personal Development Portfolio
SERU-S	The Student Experience in the Research University Survey
STEM	Science, Technology, Engineering and Mathematics
TEF	Teaching Excellence Framework
UKES	UK Engagement Survey
US	United States
VLE	Virtual Learning Environment
VSA	Voluntary System of Accountability

1. Introduction

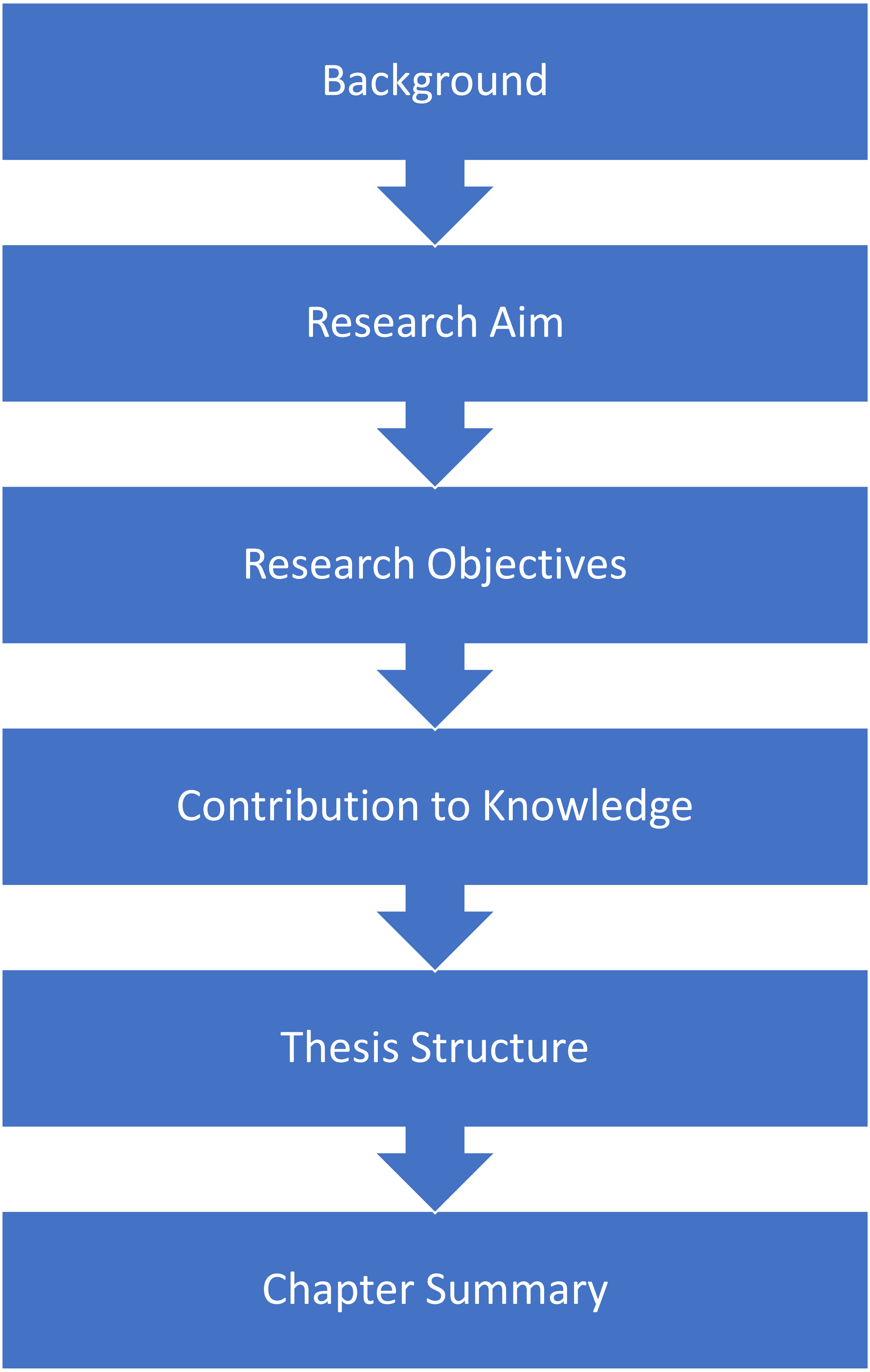

Figure 1.1: Overview of Progression of Sections in Introductory Chapter.

1.1. Background

The need for accountability, quality assurance and value for money within Higher Education (HE) has become an international priority due to increased pressure to demonstrate the knowledge and skills that a student gains during their time at University (Raban and Cairns 2015; OECD 2011a). This “value for money” approach is part of the four primary regulatory objectives that the Office for Students describe as their aim to ensure that English HE delivers positive outcomes for students (Office for Students 2018a). In recent years there has been a continuing 4% increase in the number of students entering HE (HESA 2018a; HESA 2018b). This increase in students begs the question of what the value is in seeking HE and in providing evidence that students have gained important skills from their time at university.

The attempt to measure “gains” has long been discussed in the world of psychometrics. Whether this be a change in ability or attitude, it is something which has been debated as to the difficulty of how it can be measured (Cronbach and Furby 1970). Recently, HE has looked into the possibility of measuring “Learning Gain” which has been defined as the “distance travelled” by a student in their personal development, knowledge or skills (McGrath et al. 2015). Although the concept of learning gain itself is not new, its introduction into the English HE system is a more novel area of study (Nijjar 2018). The introduction of this concept has increased a desire to quantify the impact of student learning and to transfer this impact to reflect that of teaching excellence (Rogaten et al. 2019).

The emergence of the Teaching Excellence Framework (TEF) has further increased the pressure for accountability due to the need to provide evidence of high quality teaching and learning (HEFCE 2017a). This initiative involves the inclusion of Learning Gain as a metric in order to evaluate HE teaching (BIS 2015). Enhancing teaching quality is considered to be a key educational reform policy due to teaching often being seen as second priority to research tasks and resulting in students not reaching their full learning potential (Arum and Roska 2011; OECD 2011a; Wright et al. 2016). Therefore, in order to evaluate teaching quality it is beneficial to understand the student journey and the progress that a student makes during their time in their Undergraduate degree. As the learning that a student gains should be a direct reflection of the teaching that the student has received (HEFCE 2017b). Supporting this, Evans et al (2018) argued that learning gain should not only be used to provide proof of quality but should also be used to inform pedagogy by maximising learning and teaching effectiveness within specific institutional and disciplinary contexts.

There has been much discussion surrounding the current measures within the English HE system. Currently, the measures being utilised stand on the idea that student engagement and satisfaction is a proxy to understanding the benefits of completing an undergraduate degree (Kerr 2018). Engagement, as a concept, has been understood as the relationship between the institution and the student. This concept consists of the student's involvement and effort into obtaining quality outputs from their undergraduate degree (Krause and Coates 2008). Engagement is seen as an important measure of student learning as current measures in England, such as the National Student Survey (NSS), aim to understand student satisfaction and engagement with their undergraduate degree (Surridge 2008). However, these methods generally focus on the outcome or current state of student engagement and satisfaction and do not measure any form of gain or learning from the students, therefore not understanding the direct impact of HE. The understanding of what impact the HE system has on students, has recently been discussed at the University of Winchester which stated that universities should develop critical thinking, enhance employability and immerse students in a discipline (Howson 2017a). This discussion therefore shows a potential focus and direction for the learning gain measures that are currently being developed within an English HE context. Currently, it is estimated that less than 15% of learning gain literature is focused on a English HE context (Evans, Howson, and Forsythe 2018; McGrath et al. 2015).

At this present time, within England, work has commenced surrounding best practice for learning gain measures. Over 70 institutions were involved in 13 collaborative projects to pilot and assess a range of approaches and methods for measuring learning gain. Grades were the primary focus of most projects to measure cognitive gains (McGrath et al. 2015). As well as this areas surrounding student employability and engagement were also explored using items from the UK engagement survey (UKES). Although there was discussion that there was some development in cognitive skills tests within English HE, this was highlighted to be both expensive and time-consuming. As well as this, these skills tests were said to be non-discipline specific and aimed to create a generalised approach to a cognitive learning gain assessment tool (Howson 2017a). The main challenges that were faced within these projects were that of a lack of engagement in the instruments from the students themselves as well as an awareness that the English HE system will not be able to produce a single measure of learning gain due to subject differences (Howson 2017a; Howson 2018).

"*Subject-level differences are emerging across different projects, in terms of how students progress through higher education, how they respond to tests and surveys and how they interpret questions on instruments."* - (Howson 2017a, p. 11)

In support of this, Evans et al (2018) claimed that there will not be one "single solution" to measuring learning gain due to the differences in student characteristics, disciplines and institutional contexts. This paper also argued that learning gain approaches need to be more embedded within the curriculum in order to maximise engagement from students as well as to improve upon curriculum design and delivery. Similar insights into learning gain were also stated from Arico et al (2018) who noted that even measures which are seen as generalisable, such as grades, have differences across individual disciplines. This is further supported by Howson (2018) who found that grades had marking and grading patterns which were highly subject specific in areas such as subjectivity and marking rubrics and profiles.

1.2. Research Problem

Previous research has identified the increasing need for student learning gain measures due to the ever increasing pressure to provide evidence of student learning with the introduction of TEF (McGrath et al. 2015; OECD 2008; Howson 2017b; Bowman 2010a; Nijjar 2018). The research has shown that these types of learning gain measures have been embedded into United States (US) HE culture for a number of years (Arum and Roska 2011; Randles and Cotgrave 2017; Kuh et al. 2015; Ewell 2002). However, the current measures being utilised in England do not measure any learning gains that students have achieved over the course of their degree, they are purely of a reflective nature and ask around areas such as satisfaction and engagement rather than student learning (Callender, Ramsden, and Griggs 2014; HEA 2018; HESA 2018c; HEFCE 2017c). The literature has addressed the need for these measurements within an English HE context through the HEFCE pilot projects that are being conducted as well as the national mixed methodology learning gain project which utilises measurements that are being used over in the US. However, there have been discussions surrounding the difficulties of these projects which include the differences across disciplines (Nijjar 2018; Howson 2017b; Howson 2017a; Howson 2018; Hudson and HEFCE 2016; HEFCE 2016b).

The literature indicates that attempting to bring the measures that are being utilised in the US over to an English HE system would not function effectively within the context of the specified degree programme. The US uses a more generalised degree system whereby there will be general skills that are obtained with a focus on a breadth of knowledge rather than depth (Bishop 2016; International Education Specialist 2016). Within an English context, skills are generally obtained of a specified nature with the way in which a skill is taught differing greatly across different subjects (Ewell 2005; Howson 2017a). Therefore, the literature pointed towards a need for a more specified learning gain assessment tool. This specified approach in which the skills that are measured are specific to the subject, as well as the way in which they are measured, has yet to be researched within the literature. This specified approach will be beneficial as it brings the measurements into an English context which makes the measurements and data much more meaningful to the English HE system. In addition, the current methods which have been tested on students have shown a difficulty in getting students to engage with the instruments (Howson 2017a; Howson 2018). This would point towards a need for an instrument which is more easily accessible and efficient in order to increase student engagement.

1.3. Research Aim

With consideration to the research problem, the aim of this research was identified. The aim of this research was to develop and validate a self-reporting skills assessment tool, that is able to demonstrate student learning gain, with a specific focus on English Science Undergraduate students.

1.4. Research Objectives

To achieve the aim outlined above, the following objectives were identified;

1. To conduct a literature review to explore the current state of learning gain internationally.
2. To explore the developments of learning gain within an English Higher Educational context.
3. To identify the skills that are critical for students to achieve when studying a Science Undergraduate degree within an English Higher Education Institution.
4. To explore and review current measures and literature to define and operationalise the skills which have been identified to belong to Science Undergraduates in England
5. To develop and validate an assessment tool that is able to measure student learning gain with a focus on Undergraduate Science students.

1.5. Contribution to Knowledge

The current research has identified contributions to both theory and practice that the undertaking of this research project has achieved.

1.5.1. Contribution to Theory

Literature has demonstrated the need to develop learning gain measures within an English HE context with work being done to develop and validate these measures on undergraduate students all across the country (Howson 2017a; HEFCE 2016b; Howson 2017b; Howson 2018). Much of this work has been taken from existing assessments and surveys that have been developed from practitioners such as HE providers, academics and educational boards/public bodies. However, the main beneficiaries of HE are that of the students themselves. These students have been identified to be an "untapped resource" of information whereby they should be utilised to create such assessment measures but are yet to be considered (Kuh et al. 2015). The present research begins with an inductive approach whereby students are utilised to give focus and direction to the research project. Therefore, there is a contribution to the theoretical development of learning gain assessment tools. Understanding the personal experiences of these main consumers of HE may be vital in developing the discipline specific measures that are needed within English HE. Therefore, this research is exploring these experiences of the students from a targeted discipline within a specific institutional context paving the way to showing the importance of using the student experiences to develop learning gain measures.

Furthermore, this research is bringing in a different perspective on subject specific skills. Although there is much literature on generic skills of HE (Tremblay, Lalancette and Roseveare 2013; OECD 2018) as well as very discipline specific skills there is little discussion on how these generic skills may differ across the disciplines. This research has taken a small delve into the potential differences in how students are taught and understand skills which are considered to be "generic" to HE such as critical thinking.

In addition, this research is contributing to the philosophical discussion and conversation as to what HE is intended to instill into the students which chose to attend. There is a focus and understanding of the importance of critical thinking with the further understanding of the impact of engagement, motivation and personal development in students learning and growth as a student. This research highlights how these skills can interact with each other as well as understanding the experiences of both staff and students in relation to the outcomes of the assessment tool. This triangulation of this information has yet to be done within an English HE context and in such a targeted approach.

1.5.2. Contribution to Practice

Current pilots into learning gain measures have found variation across disciplines in the way that students learn as well as how they interpret the instrument itself (Howson 2017a). The main outcome of this research is the development of an assessment tool, directly targeted at Science students, which can be used to follow students individually to view their progress in the development of undergraduate skills. Therefore, the understanding of this student journey will allow the institution to be able to not only follow student progress but to utilise the outcomes in order to improve and enhance teaching and the curriculum. In addition to this, the current methods of measuring cognition within English HE are not only not discipline specific but are often grade focused (Howson 2018). The assessment tool which has been created within this research takes a different and unique approach of utilising the multiple choice format. This format is often used in the validated and embedded HE questionnaires which measure satisfaction and engagement outcomes. But, this has yet to be explored in terms of measuring cognitive skills. This method of measuring cognitive skills may also help in the engagement problem in terms of gaining larger samples of students due to the ease and accessibility of having the assessment tool in a multiple choice format that can be distributed via online platforms.

Currently, measuring cognitive skills have relied on already validated and utilised methods that have been taken from the American education system such as the objective measures as the Collegiate Learning Assessment (CLA) or by quantifying methods which are already embedded within English HE such as with the Grade Point Average (GPA) system (Howson 2017b). The development of new methods have been much more focused on attitudes and behavioural outputs rather than that of academic skills. Therefore, this research is focused around the cognitive and academic outputs of HE as well as also considering the key behavioural aspect of student engagement. Having this unique area of focus will be vital in understanding the outputs of HE and whether the students are gaining in these areas.

1.6. Thesis Structure

1. **Introduction:** This current chapter outlines the aims and objectives of the research as well as outlining the background and contribution to knowledge.
2. **Literature Review:** This chapter begins by discussing definitions of student measures. It then discusses the differences between utilising objective vs subjective measures. Finally, the literature review ends by discussing the current measures that are used internationally and bringing this into an English context.
3. **Methodology:** This chapter examines and discusses potential research philosophies, approaches and strategies and justifies the intended methods for this research.
4. **Phase One:** This chapter outlines the research objectives for this phase and the approach chosen for interviews. This was an exploratory research stage in order to identify the specificity of the assessment tool and skills that are intended to be analysed.
5. **Phase Two:** This chapter discusses the process of designing and piloting the assessment tool.
6. **Phase Three:** The reliability testing chapter looks at the stability of the outcomes of the assessment tool from three different samples of the target population. It also analyses Cronbach's lpha and conducts a factor analysis.
7. **Phase Four:** The final, and most important stage, looks at the validity of the assessment tool by following individuals across a period of a year. The outcomes of this are analysed and triangulated with qualitative data from staff and students.
8. **Discussion:** A discussion of the findings takes place in this chapter, it considers each of the objectives and how they have been met by the research.

9. **Conclusions:** This final chapter draws conclusions from the project, identifies the contribution to knowledge, research limitations and implications as well as recommendations for future research.

1.7. Chapter Summary

This chapter outlined the expectations of this thesis. Firstly, by providing a background context that is needed in order to paint a picture of the current HE climate in England, this included the discussion of the importance of learning measurements due to the pressure on HE institutions to provide accountability. Secondly, the chapter outlined the research problem that needs to be addressed within this project. Next, the overall research aim and objectives were stated. This chapter then discussed the contribution to knowledge of this research project and identified the gaps that this project has filled. This included the subject specificity of the assessment tool that has not been previously explored. The importance of this is embedded within English HE due to the one subject focus of English degrees. Finally, this chapter summarised the structure of the thesis.

2. Literature Review

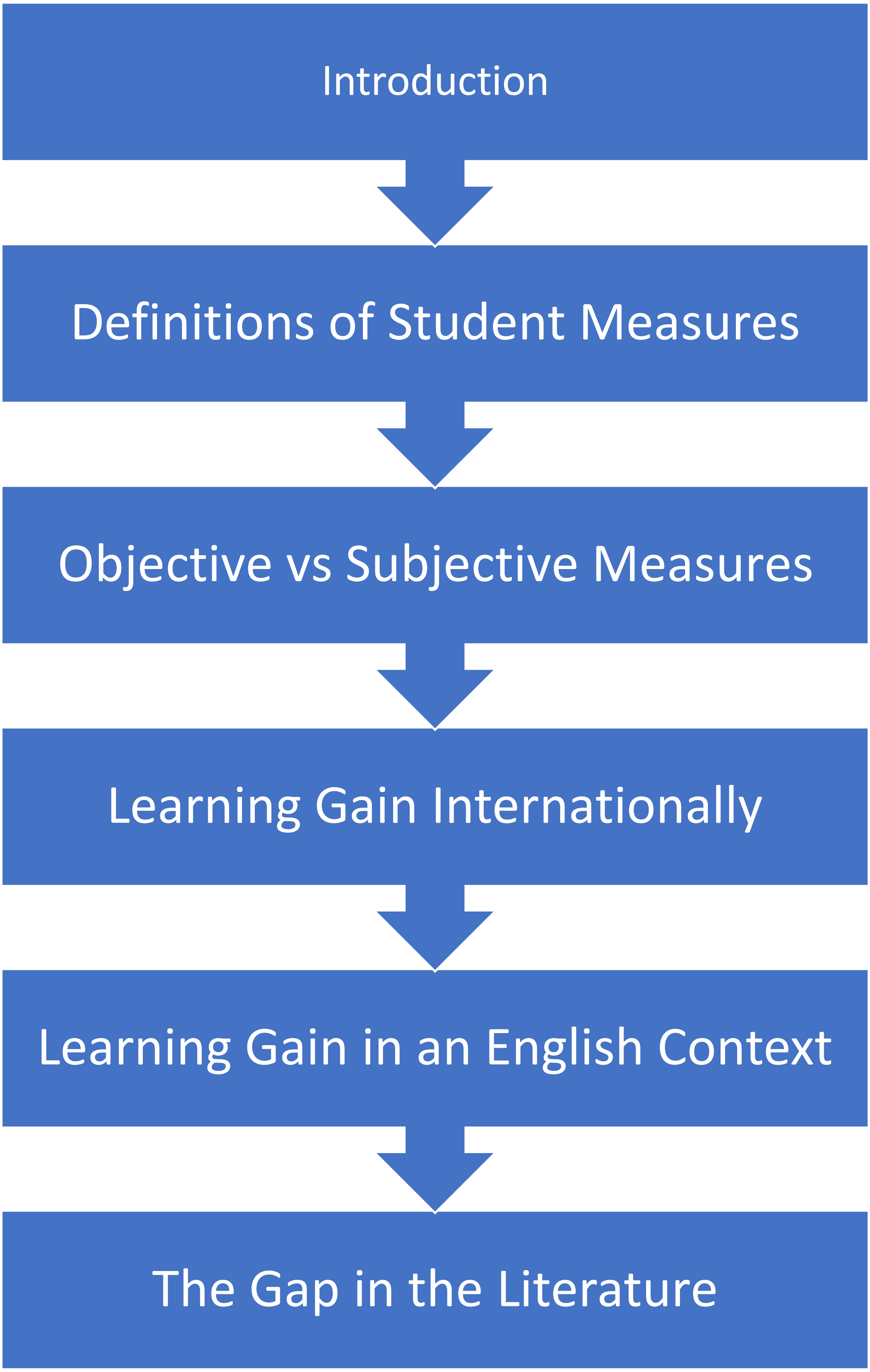

Figure 2.1: Overview of Progression of Sections in Literature Review Chapter.

2.1. Introduction

The purpose of the literature review chapter is to provide an overview of the current knowledge and research conducted within student learning measures, specifically learning gain. This chapter is separated into sections beginning with an explanation of different definitions of terms often used when measuring student learning. This chapter will then discuss comparatively objective and subjective measures of student learning. It will then explore the current views of learning gain firstly on an international level and secondly within an English HE context. Finally, this chapter will then discuss the current gap in the knowledge of learning gain and the difficulties that this area of education has faced within an English context. England was the specific focus rather than the entirety of the UK due to the different HE funding bodies within the countries. England specifically chose to focus upon this development of learning gain that was instigated by the Higher Education Funding Council for England (HEFCE). This initial stage of the research meets the first two objectives of the overall project. This is that of "to conduct a literature review to explore the current state of learning gain internationally" and "to explore the developments of learning gain within an English Higher Educational context". These two objectives will be met by the end of this chapter as learning gain in a national and international context will be discussed.

The literature review process has been outlined in figure 2.1 to show the progression of the sections in this chapter.

2.2. Definitions of Student Measures

Assessments of student outcomes are common practice in HE with a growing international interest in measuring student learning (OECD, 2008). This need has been particularly prominent since HE has become more accessible with an estimation of at least 40% of the population having access to means to be able to attend should they wish to do so (Greatbatch and Holland 2016). As stated by the OECD (2008), opinions have differed on how these assessments should be utilised, with some viewing them as an opportunity to identify the best teaching practices and others as an opportunity to compare HE institutions' output. Regardless of the reasoning, it is important that the assessments are both valid and reliable i.e. that the assessments are measuring what they claim to measure and repeatedly are able to do so with some accuracy (Randles and Cotgrave 2017; Bowman 2010a). However, there has been a worry surrounding student response fatigue, meaning a significant reduction in reliable percentage response rates (G. D. Kuh et al. 2015).

Currently, within the UK, the most widely used student outcome assessments are measures of student engagement, satisfaction and graduate employment such as the NSS, UKES and the Destination of Leavers from Higher Education Survey (DLHE) which was replaced in 2018 by the Graduate Outcomes Survey (HEFCE 2017c; HEA 2017; HESA 2018c). Whilst these assessments may be somewhat useful to HE institutions, they fail to measure any learning gains that students have achieved during the course of their degree (Maskell and Collins 2017). Although, the NSS does contain a small number of questions about student personal development, the questionnaire is implemented once at the end of the degree, asking students to reflect back on their studies (Callender, Ramsden, and Griggs 2014).

However, there are potential problems surrounding measurements of student learning. Of particular note, is a concept known as 'gaming' this refers to institutions targeting specifics groups of students in order to achieve the best outcomes from their student measures (Goldstein et al. 2000). This can be problematic across all the different types of student measures and would suggest a need for sensitivity in creating and implementing measures of student outcomes so institutions do not feel further pressure to perform highly. In addition, comparability can be a big issue, particularly across institutions, as students may have different experiences based upon the focus of their enrolled institution. For example, whether the university itself is research intensive or teaching intensive, which begs the question as to whether one measure can be used within all institutions or whether they need to be institution focused (McGrath et al. 2015). One could also argue that the quality of an institution is not just about the quality of the student learning itself but should also consider areas such as resources and student support systems (Gibbs 2010).

2.2.1. Learning Outcomes

Learning outcomes are defined as the expected outputs that are 'what a learner is expected to know, understand or be able to do at the end of a period of learning' (OECD 2011b) and is often used as a synonym for learning gain within the US HE sector. This ('learning outcomes') is, therefore, different to learning gain as it concentrates on the output level of student development rather than following development over time. There are some fundamental problems with this method of data collection; an output measure that measured student grades would assess an institution based upon students with higher grades. However, it could be the case that the students were already recruited at a high level, it does not show the journey that the student has taken during their degree and what they have gained from their education. Therefore, the outcome of the degree is not a true reflection on the quality and value of an institution (McGrath et al. 2015).

2.2.2. Value Added

Value added measures are regarded as a comparison measure between predicted performance at the outset of studies and the actual performance that is achieved, this differs slightly with learning gain measures as this compares actual performance at differing time points, although they are often used interchangeably. Doran and Lockwood (2006) described value added modelling as a ‘class of longitudinal growth techniques that attempt to identify the unique contributions of schools or teachers to student learning’. England is currently regarded as having international leadership in developing value added measures and models (Tymms 1999). Figure 2.2 shows an illustration of value added. Point A represents the student’s performance in year one, point B represents the student’s expected performance in year two, with point C being the student’s actual performance in year two. Value added is represented by the difference between point B and point C (McGrath et al. 2015).

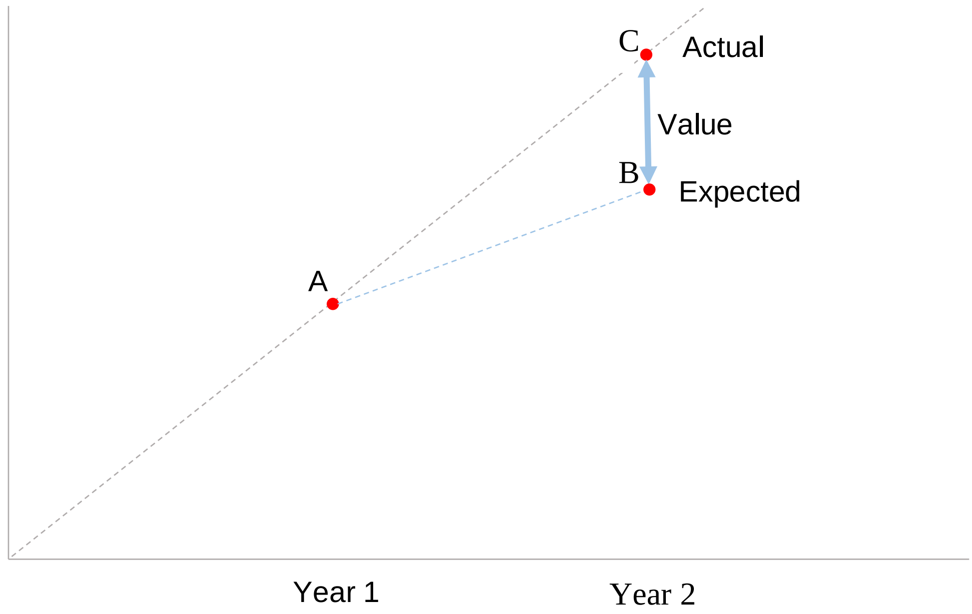

Figure 2.2: Visual Representation of Value Added.

Value added measures can also be achieved by a comparison with other students who achieved similar scores previously (Department for Education 2018a). Similarly, value added models can also be used to measure how a student’s performance is growing towards a particular outcome (Doran and Izumi 2004). This highlights the similarities between value added measures and learning gain as there is an aspect of student development growth to both of these concepts. Some scholars believe value added to include student growth from one year to the next, not just being based on predictive scores suggesting value added is an interchangeable concept for learning gain (Tekwe et al. 2004; Fulcher and Willse 2007; Kim and Lalancette 2013; Liu et al. 2016; Mc Caffrey et al. 2003).

An advancement in value added models included accounting for students' circumstances and prior performance. This led to Contextual Value Added (CVA) models which made comparisons across institutions more representative to the conditions of the institution (Lenkeit 2013; Goe, Bell, and Little 2008).

2.2.3. Learning Gain

Learning gain is currently in its infancy in an English HE context and has not been extensively explored, making this concept a new and novel area of research (McGrath et al. 2015). In 2015, a study by RAND Europe was commissioned to review the current national and international climate of learning gain (HEFCE 2017d). In this report, learning gain is defined as measuring the 'distance travelled on a student's skills, competencies, content knowledge or personal development' (McGrath et al. 2015). This is distinctive in comparison to other potential terms that are often used surrounding student assessments as it is measuring a difference over time. Supporting this, the learning gain progress report (Nijjar 2018) also used a similar definition for learning gain but with the addition of employment readiness. These definitions have been adapted from those suggested by the OECD (2013a; 2013b) and Rogers (2007). Although learning gain is technically defined as a comparison between a minimum of two periods of time, students are often followed across the full period of their studies (McGrath et al. 2015; Nijjar 2018). Due to learning gain being a novel and unexplored area of education in an English context, it has been expressed as a concern within the educational community due to its complexity and lack of understanding across the wider HE sector (Nijjar 2018). Figure 2.3 shows an illustration of the concept of learning gain where learning gain is represented by the differences in point A and point B. Point A is a representation of a student's performance in year one, point B a representation of the performance in year two (McGrath et al. 2015).

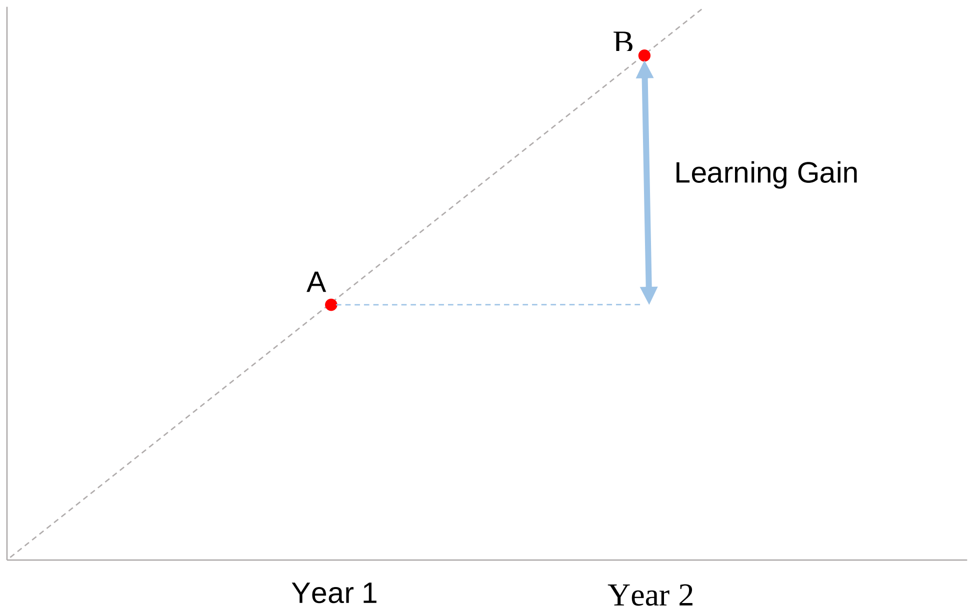

Figure 2.3: Visual Representation of Learning Gain.

The measurement of these attributes is a complex task, the definition of learning gain itself covers a wide breadth of potential areas that could be measured (Howson 2018). Considering what should be measured directly relates to the philosophical questioning of the entire purpose of HE; according to Arum and Roska (2011) the fundamental outcome of HE is the ability to think critically and analytically. This outcome of HE would be considered a "cognitive skill" (Howson, 2018) and is only a small aspect of what learning gain is defined to explore.

Content knowledge is arguably the most significant part of what a student is anticipated to learn during their degree. This relates to the specific knowledge of that particular field and is usually acquired through classroom learning and lectures (Nijjar 2018). The measurement of content knowledge is commonly achieved through the coursework and exams that are assigned for students during their degree and ergo is measured with the result of these usually in a form of a grade (Allen 2005).

Skill development is also another aspect of learning gain that is often measured. Skills are defined as "the abilities or proficiencies to apply cognitive or practical knowledge and use know-how to complete tasks and solve problems" (Nijjar 2018). OECD (2018) discusses a wide range of skills which students should gain from HE such as disciplinary specific knowledge, cognitive skills, social skills and practical skills. Arguably, the skills that are gained from HE would be dependent on the discipline that the student is studying as well as the skills that the course itself emphasises (McGrath et al. 2015). This would also include the skills that students are expected to take onto employment as there is a big need for students who are "work-ready" and have the skills necessary to gain employment and to be successful within it (Nijjar 2018).

The final attribute considered a part of learning gain is students' personal development. This refers to the student's psychological development such as an increase in their confidence or communication (Nijjar 2018; McGrath et al. 2015). Kuh (1995) also discovered, through interviews with students, that during university they believed they had gained leadership skills, interpersonal communication with peers and cognitive complexity which all come under the umbrella of personal development. As learning gain covers a wide breadth within its definition, there have been differences of opinion regarding what matters for measurement (Nijjar 2018).

2.2.4. Implications of Measuring Learning Gain

Upon understanding a clear definition of learning gain, it is also of importance to consider the potential implications of learning gain and the benefits of measuring the construct. There are a number of different angles to view this from and to consider who will benefit from the utilisation of these instruments. This could be from a student perspective, institutional perspective or from an employer's perspective.

Firstly, from a student's perspective, the outcomes of learning gain measures will be of use to both prospective and current students. Prospective students for example, can use the information to make an informed choice on their selection of a university (Howson, 2018). This is of particular importance due to the increase in fees for the undergraduate degrees and the idea of "value for money" as an "investment" into HE (Browne, 2010;Howson, 2018; Nijjar, 2018). Allowing prospective students to view the learning gains of existing students at an institution will provide them with the necessary information to compare which institution is best suited to themselves. However, there have been some arguments surrounding whether prospective students utilise the information that is readily available about institutions such as from satisfaction questionnaires. In a study in the US, it was found that only 18% of prospective adult students use websites, which contain learning outcomes information, to inform their choice of institution with the majority counting on friends, family and institutional websites (Hagelskamp, Schleifer, and DiStasi 2013). Although this may be different for students who are not attending universities as an adult as they could potentially be more informed about the existence of the websites; in addition there may be cultural differences as prospective students in the UK may be more likely to use the websites. However, this kind of implication will only work if the measures of learning gain are comparable across institutions. There are differing opinions as to whether this is possible, due to the different emphasis and structure of courses across institutions (McGrat*h et al*, 2015).

In addition to prospective students, this information will be of use to existing students as they will be able to use learning gain measures to reflect upon their own learning as well as raising the awareness of students as to where their strengths and weaknesses lie in terms of their development (Howson 2018). Some institutions, for example, provide students with personalised reports on their learning and progress which they then pass onto personal tutors who can use the information to encourage students into particular behaviours to increase positive learning outcomes such as the number of independent learning hours (Nijjar 2018). This however, may not be practical and could potentially be time consuming to provide students with individual data, although this is already achieved through giving students individual grades, this may impose further strain on an already heavily stressed teaching faculty (Arum and Roska 2011).

Considering learning gain from an institutional perspective, there are many ways in which this information could be potentially utilised. Enhancement of the classroom and course itself is one way in which this data may be utilised. The data from learning gain measures has the potential to allow teaching staff to tailor their information to the students as well as to enhance learning from a pedagogical perspective. Knowing what current students' strengths and weaknesses are, will be vital in the management of the course itself and the implementation of potential changes to enhance the learning gain outcomes (Howson 2018; Bowman 2010a). Furthermore, the outcomes from the measures can be used a step further in reviewing the entirety of a programme. Not only this, it can be used to evaluate the effectiveness of the programmes themselves and even to reward and recognise members of staff whose students are particularly prominent in their development. In addition, it can provide staff with immediate feedback as to whether a specific module or programme is effective (Nijjar 2018). Finally, the outcomes of the learning gain measures could also be used to enhance student services. This is of particular prominence with the approach to employment and career readiness as this could inform careers services (Howson, 2018; Nijjar, 2018).

Perhaps the highest level of utilisation of the data from these measures is the potential use for employers and the government themselves. For employers, the information may be vital in the recruitment of high functioning graduates as well as to create a varied workforce based on the outcomes of the students' strengths from the data of learning gain measures. The information may also be useful in a quality assessment aspect of the HE institutions in terms of the "value for money" aspect (Bowman, 2010a; Howson, 2018).

However, Mcgrath et al (2015) discovered that HEI's had a unfamiliarity and uncertainty when it came to the concept of learning gain. There was a severe lack of knowledge of the potential implications of learning gain measures as well as a confused understanding, further highlighting the novelty of learning gain in an English HE setting. Although, the institutions did demonstrate a willingness to learn about the concept and its potential worth to institutions. 86% of respondents did provide a definition of learning gain, however, there was a wide diversity of responses further suggesting the novelty of the concept in England as there is a lack of clear definition. Furthermore, this call for information also revealed the usefulness of using learning gain measures to help inform students and to increase transparency in learning outcomes.

2.2.5. Discussion and Conclusions

Learning gain differs from learning outcomes and value added measures as it measures the difference in students' abilities over time by way of comparison. Although some may argue value added can often include this (Mc Caffrey et al. 2003), learning gain purely focuses on this aspect of learning measures to follow the "distance travelled" by students during their undergraduate degree. This is an important distinction as it would suggest that measures of learning gain follow student progress over a period of time with the "gold standard" of learning gain being that of a longitudinal study where individual progress of students is tracked (McGrath et al. 2015).

Implications of learning gain suggest that measures of learning gain can be used to inform student choice as well as to increase transparency in the outcomes of HE. This will allow institutions to emphasise the usefulness of completing an undergraduate degree and to encourage prospective students to attend their specific institution. The measures are useful from an individual perspective in terms of students as well as to employers and stakeholders, therefore being a worthwhile metric in the HE system (Nijjar 2018; Howson 2018). These implications link back to the value for money and pressure for accountability which is currently surrounding HE.

2.3. Objective vs Subjective Measures

Debate has also been sparked surrounding the use of more subjective measures of student learning, for instance self-reports, as opposed to more objective measures, for instance the CLA. Some scholars, such as Gonyea (2005), have noted strong agreement between the use of objective and subjective measures, suggesting that they can be as valid and reliable as each other. However others, such as Pike (1995), have suggested that there are strong differences between objective and subjective measures in the content areas that they assess. There is also the potential of a mixed methods approach, not necessarily of a combination of objective and subjective measures but to provide validation for the output of the measures themselves through the use of qualitative research.

Mcgrath et al (2015) noted several attributes that are needed for a robust method of student learning. Firstly, the measure must include a longitudinal or cross-sectional design, with the longitudinal approach being a much more vigorous method of data collection. Secondly, the measure must be valid i.e. measuring what it is intended to measure. Thirdly, the data must be representative and account for any missing data, usually achieved through a method of student sampling. Finally, the data must be comparable. This can be on a small scale such as between individual performances or can be across disciplines and institutions. The idea being that differences should be able to be seen between students.

2.3.1. Standardised Assessments as an Objective Measure

Standardised assessments are often used as an objective measure of learning gain, usually in the form of a formative or summative assessment. There are two types of standardised assessments; discipline specific and generic skills tests. An example of a discipline specific assessment would be progress testing assessment, which is often used in schools of medicine in England (Freeman et al. 2010; Schuwirth et al. 2010; Chen et al. 2015). A generic skills test is not discipline specific and measures more general skills such as critical thinking, an example being the CLA which is utilised in the US (McGrath et al. 2015).

Standardised measures are seen as more attractive as they rely on techniques to ascertain a student's skillset in an objective way, this allows for easier comparability across groups of students and institutions (Inayah *et al*, 2017). It is also seen as a more valid measure of learning as it is not affected by personal opinion but is a comprehensive measure of that student's skill (OECD 2008). Currently, standardised assessments are mainly used in the US and do not seem to be readily utilised within an English HE context (G. D. Kuh and Ewell 2010).

2.3.2. Limitations of Standardised Measures

When considering the assessment measures, there have been papers that offer a broader sense of the tools, particularly in terms of their appropriateness, applicability and practicality. The University of California, for example, rejected the use of standardised measures to assess student learning gain as it was believed that the tests failed to recognize the diversity, breadth and depth of discipline specific knowledge and learning (Douglass, Thomson, and Zhao 2012). In addition, Banta and Pike (2012) argued that the skills and outcomes that are measured by standardised testing are only a small minority of the learning gain that is taken from completing a HE degree. Furthermore, Brooks (2012) argued that although there are skills that are seen as “generic”, different subjects will be taught these in different ways. For example, a humanities-based subject will require the student to think critically in a different way to that of a science-based subject. This suggests that generic tests would still not be applicable to every subject, this effect would likely be minimised in countries, such as the US, where HE follows a more general approach rather than being subject focused. As standardised measures require time-consuming application and marking, this would mean that rather than a large cohort of students, a smaller sample would be tested, therefore there are issues around representability (McGrat*h et al*, 2015).

2.3.3. Self-Reported Learning Gain as a Subjective Measure

Self-report measures are a subjective way of collecting information about student learning through asking students to report on their current academic ability, the main benefits of using this method of survey being that it requires minimal financing, resources and results can be achieved over a shorter period of time (Bowman 2010b). Gonyea (2005) argued that with the use of further data to back up the self-reported survey, such as institutional records and qualitative interviews, self-reported data can be a valid measure of student learning but should not be used as a substitute for objective measures. This suggests that self-reports should not be solely used to indicate academic ability of students but should be used in conjunction with other methods.

However, self-reports are still often used as a sole measure of student outcomes. The NSS, for example, is a widely used self-report measure of student satisfaction across the UK (HEFCE 2017c) which is often reported as a singular measure of this construct. Although, arguably, student satisfaction is a subjective measure in itself, as it requires the personal experience of the student during their time on their undergraduate degree and therefore a self-report is a more appropriate method of measurement. Nevertheless there are, arguably, subscales from within the NSS which could be subjected to bias and error in judgement by the students as there are questions related to student development (HEFCE 2017c; Surridge 2008; Cheng and Marsh 2010). As there are a limited number of questions which target student development, the NSS and other similar measures such as UKES, have been questioned for validity in being used as a measure of learning gain as they currently stand (McGrath et al. 2015).

2.3.4. Limitations of Self-reports

The validity of self-report measures has been widely discussed. According to Bowman (2010b) and Porter (2013), the problem occurs as students may report that a certain construct has improved when it has not, or they may report no improvement when there really has been. Bowman (2010b) indicated that self-reported measures of learning should be used as a measure of student satisfaction with the education which they are receiving. This is due to research comparing objective and subjective measures whose outcomes indicated that students were not accurately able to report their learning gains both longitudinally and retrospectively. However, this piece of research focused solely on first year university students who may experience development differently to that of the later undergraduate years. In addition, it could be argued that the contents of the objective and subjective measures that were used in this study were fundamentally different. Whilst the objective measure (CAAP) intends to measure cognitive ability, the self-report measure only consisted of four items of which only one related to cognitive ability. This may suggest that it was the content and wording of the self-report questionnaire itself that led to the conclusion of students being unable to self-report their learning.

In addition, Douglass et al (2012) examined students' self-reported learning and compared this to the students' GPA. It was found that higher GPA was correlated to higher levels of self-reported learning gain, suggesting students are able to accurately report their performance over time. Owing to this apparent veracity, Douglass et al (2012) proposed using self-reports on a more regular basis to measure student learning.

Opposing this, Gilbert (2007) found that people are not very good at predicting what makes them happy. It was found that subsequent ratings of happiness did not align with real-world situations. As argued by Bowman (2010b), this would suggest that students will also be unable to make predictions of their performance. However, these examples are fundamentally different to academic ability. Students are constantly assessed and receive feedback on their ability and therefore will have a much more objective awareness of their ability as aligned with reality. In addition, it has been found that self-assessments of knowledge of a subject have highly correlated with objectively tested knowledge, suggesting that people are accurately able to determine their ability (Berdie 1971). Supporting this, Pike (2011) also found that students reported greater gains on outcomes that were directly related to their degree further suggesting a student's awareness of their own academic ability.

However, it could be argued that this may be due to what it known as the "halo effect" in that students may report gains in these areas as their expectations are that they should have gained due to the nature of their degree (Seifert and Asel 2011). In addition, Bowman (2010a) found that there were other external factors that could affect the accuracy of students' self-reports, the biggest difference was found to be between social class groups, suggesting a wide variety of possible influences on reporting learning accurately. However, the study did contain a small sample size of minority groups suggesting that this finding may not be generalisable.

Another potential problem is with the format of self-reports themselves, many self-report methods choose to apply multiple-choice as their method of testing owing to its objective scoring and cost effectiveness. Yet, many researchers such as Nicol (2007) have argued that these kinds of tests do not encourage students to engage in higher order processing, owing to ease of selection, which may result in an inaccurate representation of student learning. Hyytinen et al. (2015) compared the CLA with a multiple-choice test, which both claimed to measure the same constructs but were assessed using different methods; a third of students performed better on the multiple-choice test while a quarter performed better on the CLA. This research also supports the idea that multiple-choice tests simply do not engage a high level of processing.

Bowman (2010b) conducted a study that compared subjective longitudinal assessments with students' self-reports of their learning gain. Human errors in judgement were thought to be the biggest hurdle in self-reports, although it is difficult to say for certain whether these errors were simply due to the construction of the self-report rather than inability to report learning. In addition, it was also found that if self-reports are to be utilised then they should be applied on a longitudinal basis rather than asking students to reflect on their development. Seifert and Asel (2011) concluded that students were able to report on their current situations better than past ones, suggesting learning should not be looked at retrospectively but should be taken as a longitudinal measure. Supporting this, Pike (1995) stated that students are accurately able to estimate their own abilities at a single point in time. This would suggest that a longitudinal method of study would be more suited to self-report surveys as students may find it difficult to reflect backwards on their abilities.

2.3.5. Discussion and Conclusions

In conclusion, standardised assessments appear to be an unpopular method of measuring learning within an English HE context, likely due to the time-consuming element as well as lack of representability from small samples (Banta and Pike 2012; Douglass, Thomson, and Zhao 2012). Self-reported learning gain seems to be more appropriate within the educational context, although there are some potential issues surrounding reliability (Porter 2013; Bowman 2010b); students have frequently shown that they are able to report on their own learning accurately when tested on a longitudinal basis (Pike 1995; Seifert and Asel 2011). This would suggest a longitudinal self-report approach would be the best method to utilise. In addition, the current explorations and developments in assessment tools within an English HE context further suggest a need for a cognitive skills assessment tool. This will be discussed in further detail in section 2.5.

2.4. Learning Gain Internationally

2.4.1. Comparative Education

Measurements of learning gain also face the obstacle of finding a method which is able to be utilised on an international scale, due to the differences in educational contexts across countries. In the US, learning gain measures have been used predominantly in the liberal arts/general education sector, thus providing the argument that their established measures would not translate into an English HE setting due to the more specialized nature of the degree (McGrath et al. 2015). To understand the differences in the US educational system in comparison to England, one must understand how the US educational system is structured.

Access to HE in the US has become a potential problem, with only 29% of people believing that those who are qualified for HE have the opportunity to attend. This figure has decreased from 45% over a period of nine years, suggesting a decreased lack of belief in the educational system for equal access to HE (Immerwahr and Johnson 2012). A distinct difference between the American education system and the English education system is the length of the degree itself. American institutions are split into four year or two year courses whereas most degrees within an English HE system last three years. In 2016, this stood at 3,011 four year universities and 1,616 two year universities (National Center for Educational Statistics 2016). In addition to this, each university can have a specific focus. An example of this is whether the institution is research focused and emphasizing external grant funding or is focused on helping marginally prepared students gain a degree. In addition, an institution can focus on liberal arts, career or technical education (Rippner 2016). Although some may argue that this can be translated into an English educational setting, a comparison of ratings on TEF and research excellence framework suggested that some institutions performed highly in research but poorly in teaching, and vice versa (Morris 2017). This demonstrates the idea that there are institutions within an English context that are either research or teaching driven.

Perhaps the most significant difference between the US educational system and the English educational system is that, the US focuses on a wide breadth of knowledge at degree level whereas the English system is focused on the depth. In the US, the first two years of university consist of general knowledge subjects often referred to as a way to give all students an underlying foundation on which to build upon their knowledge. By the third year, students are expected to choose a “major”, this is the topic in which the student will receive their bachelor degree; there are often requirements students have to meet in order to achieve this “major”, some students even change their mind several times as to what they want it to be. There is also a chance to specialize in a second subject, often referred to as a “minor”(Bishop 2016; International Education Specialist 2016). This shows how the American educational system has quite a wide focus when it comes to a subject area, students do not appear to study one specific subject, even when they choose a “major” they are still able to study another subject. This emphasises how the outcomes in the English HE system will be much more subject focused as they are specializing in one subject rather than several. This also begs the question as to whether measures which are being utilised within the US can be transferred into an English HE system due to the fact that the US is able to create a more general measure of skills.

2.4.2. Transatlantic Dialogue

According to Kuh et al (2015) measuring learning outcomes is not a simple task yet is fundamental to understand the quality and effectiveness of an American HEI. Learning gain has been more embedded in the culture of HE within theUS, this was initially thought to have begun in the 1980s, however roots can be traced to the 1930s with a scholarly inquiry project to measure and understand cognitive gains within an educational setting (Kuh et al. 2015; Ewell 2002). In the 1980s, pressures grew due to the need for evaluation and accountability of student learning outcomes and a dissatisfaction with the increase in costs and decrease of quality (Kuh et al. 2015; Fletcher 2006a; Pike 1995). In 2006, the Commission on the Future of HE in the US, more commonly known as the Spellings Commission, released a report indicating the importance of measuring learning outcomes in HE and to recommend that institutions adopt measures for institutional accountability (Fletcher 2006b). This recommendation led to the creation of the Voluntary System of Accountability (VSA) (Fletcher 2006b).

More recently, the need for assessments and evaluation of student learning was brought into sharper focus with the release of Academically Adrift. This text contained a critical review of the current state of HE and learning gain measures in the US, utilising the CLA to assess students. This research indicated 45% of students demonstrated no significant improvement on a range of skills such as critical thinking, complex reasoning and writing in the first two years of their degree, leading to an increase in the capacity and commitment of universities to assess student learning outcomes (Arum and Roska 2011; Kuh et al. 2015; Brooks 2012). The critical review of the state of American HE led to the creation of the CLA + which addressed some of the issues of the measures that were highlighted in Academically Adrift (CAE 2017a). The importance of measuring outcomes has been emphasised as institutions wanting to understand whether their students have attained the intended outcomes, students wanting assurances for their employability and government boards wanting to be assured of the institutions academic quality; all of which can be achieved through the measuring of student learning outcomes and gain (Kuh et al. 2015).

The assessment method most widely used on university campuses in the US has been the survey method (Kuh et al. 2015; Kuh et al. 2014). Similarly to English HE, the most widely utilised source of student outcomes in the US is that of national student surveys. This includes the National Survey of Student Engagement (NSSE), Community College Survey of Student Engagement (CCSSE) and the Cooperative Institutional Research Project Freshman Survey (CIRP) (Kuh et al. 2015). However, these surveys do not show evidence of student learning but rather of student experience, behaviours and attitudes (Kuh et al. 2015; Pascarella, Seifert, and Blaich 2010). As a wide range of universities utilise the same assessments, this has the advantage that they could potentially be used comparatively across institutions as well as within programmes (Kuh et al. 2015; Edgerton et al. 2003).

2.4.2.1. Assessment Tools

The VSA is an initiative created in the US which reports on three standardized measures of college learning: ETS Proficiency Profile (EPP), CLA and Collegiate Assessment of Academic Proficiency (CAAP), as well as one self-report measure: The Student Experience in the Research University Survey (SERU-S). All of which measure slightly different constructs as shown in Table 2.1 (Steedle, Kugelmass, and Nemeth 2010). According to Liu (2011a) these assessment were chosen from sixteen shortlisted instruments as they were deemed the most adequate in measuring improvement in core skills.

Assessment	Constructs	Creator
Collegiate Assessment of Academic Proficiency (CAAP)	Critical Thinking Writing Mathematics Reading Scientific Reasoning	American College Testing (ACT)
Collegiate Learning Assessment (CLA/CLA +)	Critical Thinking Analytical Reasoning Problem Solving Writing	Council for Aid to Education (CAE)
ETS Proficiency Profile (EPP) formerly known as Measure of Academic Proficiency and Progress (MAPP)	Critical Thinking Writing Mathematics Reading	Educational Testing Service (ETS)
Student Experience in the Research University Survey (SERU-S)	Critical thinking Writing Reading Oral Presentation Quantitative Field Specific Skills	Centre for Studies in Higher Education at the University of California – Berkeley (CSHE)

Table 2.1: Overview of Measures and their Constructs in the US.:

Only 28% of community colleges use one of the VSA standardised assessment measures. Although this figure has increased in recent years, it may suggest that the time-consuming (qualitative) elements are unappealing (Liu and Roohr 2013). In addition, the institutions, which chose to utilise the assessments, did not find the results meaningful in helping to improve their institution. As well as this, the VSA was created for the public to be able to utilise the results but this would seldom occur (Kuh et al. 2015). Advantages of utilising these standard measures are that they are readily available for institutions to purchase knowing that the instruments will be reliable and valid in what they intend to measure (Association of American Colleges and Universities 2014). Although there are a wide range of measures being used within the US, the VSA initiative recognises institutions which successfully integrate learning measures into their institutions and therefore is an important area to investigate to understand the assessments in the US (Voluntary System for Accountability 2018).

2.4.2.1.1. Collegiate Learning Assessment (CLA/CLA+)

The CLA was developed by the Council for Aid to Education (CAE) and originally measured several different academic constructs; critical thinking, analytical reasoning, problem solving and writing (CAE 2017a). The development of the CLA+ led to the measurement of more specific academic constructs (which is included as a 'performance task'): analysis, problem solving, scientific reasoning, quantitative reasoning, critical reading and evaluation, argument critiquing as well as writing mechanics and effectiveness (CAE 2017b). The CLA assesses students' abilities by presenting information, which the students are then asked to assess for quality and argument and make a recommendation based upon the information presented to them. According to Brooks (2012) this makes the assessment much more ecologically valid. With the CLA +, a multiple choice test was added to the performance task which was intended to allow for the assessment of a wide range of content and skills as well as the ability to compare at a student level (CAE 2017b).

The outcomes of such a test can be useful as the results can be exploited in a way that can develop and enhance curriculums. An example of this can be seen with the Richard Stockton College of New Jersey who developed additional measures of critical thinking and high level performance tasks within classrooms to familiarise students with complex thinking after initially finding their students to have low levels of critical thinking skills with the CLA (Kuh et al. 2015).

Some institutions in the US have successfully managed to utilise the CLA to improve upon their student learning outcomes. Barton College in North Carolina, implemented the CLA to their students and found that writing ability was the students' most common downfall, they used this information to inform the creation of a quality enhancement plan where staff members focused on the need for the college to improve the student writing skills (The Council of Independent Colleges 2008). Supporting this, Arizona State University viewed the outcomes of the CLA as a way to measure quality and to prove the effectiveness of their institution (Carnegie Results 2007). However, the CLA has stimulated debate around its practicality; its scoring is labour-intensive often resulting in a small sample as well as potential differences in scoring as there would be with marking any piece of work (Possin 2013; Aloisi and Callaghan 2018a; Hardison and Vilamovska 2009). Furthermore, there has been some debate on whether the constructs measured by the CLA are applicable to all subjects. Brooks (2012) argued that the CLA does not test humanities subjects particularly well as students in these subjects are taught to evaluate information differently. In addition, there is debate on the validity of the constructs of the CLA and whether the claims of measuring "critical thinking" are actually simply measuring reading and writing ability (Aloisi and Callaghan 2018a; Arum and Roska 2011). This all points to some obvious issues with the CLA as a measure of learning gain, a further problem occurs when put into an English HE setting as the skills that are assessed are very generic, whereas students of English HE will have gained subject specific skills.

2.4.2.1.2. Collegiate Assessment of Academic Proficiency (CAAP)

The Collegiate Assessment of Academic Proficiency (CAAP) was created by American College Testing (ACT) and measures five academic constructs; critical thinking, writing, mathematics, reading and scientific reasoning (ACT 2017). Each of the sections takes 40 minutes to complete; all of these are multiple choice except for the writing section. In addition, institutions are permitted to add up to 9 of their own developed multiple choice questions to each module,ment which can have problems with scalability across institutions as there will be some variability to the questions asked should institutions choose to modify the test (DeBoer 2016).

An institution that utilises the CAAP is Calhoun Community College (CCC). The college uses the CAAP to highlight problems and trends; this enables the leaders of the institution to take the necessary steps towards addressing any concerns. For example, in their 2012 report (CCC 2012a), out of a total of 674 students, 28% were found to be scoring below the national mean in writing skills; 55% were below average for critical thinking skills. In the 2014 report (CCC 2012b), out of a total of 486 students, a different picture emerged; 55% were found to be below average in writing skills while 68% were below average in critical thinking skills. In 2016, which had a different focus (CCC 2016), out of a total of 546 students, 48% were found to be below the national average on reading while as high as 63% were below average in mathematical skills.

In addition, Eastern Wyoming College also used the CAAP as their measure to view student progress. They found that in the majority of cases they were above the national average in all areas, Mathematics, science and critical thinking skills were their weakest area and the area which they sought to emphasise and improve upon in the future as they were less than 1% above the national average (Steinbock 2014). However, the sample size that the university chose for the CAAP was only of 62 students, this would question the scalability of the outcomes of this measure; particularly it could potentially be that the students with the higher grades were more likely to agree to complete an assessment task. Furthermore, the University of Nebraska at Kearney utilises the CAAP as an indication of the success of their general studies program. In each of the testing groups it was found that more than 50% of the students scored above the national average (Dubbs 2011). Again, there is a notable problem with the sample size as well as potential sample bias as participation was voluntary and those who voluntarily take part in examinations are more likely to do well on such tests and those with test anxiety/poor performance more likely to avoid such activities.

2.4.2.1.3. ETS Proficiency Profile (EPP)

The EPP, formerly known as the Measure of Academic Proficiency and Progress (MAPP), was created by the Educational Testing Service (ETS) and measures four different constructs; reading, critical thinking, writing and mathematics. It applies 108 multiple choice questions, or 27 questions per construct (ETS 2017), even the writing section consists of multiple choice in the form of grammar and sentence correction, however it does also contain an optional essay testing section which is rated on a score of 1-6 in a typical short-answer essay examination format (DeBoer 2016). As well as the CAAP, the EPP also has the ability to include up to 50 additional questions created by the universities, again creating a problem of potential scalability due to the ability to modify the test (DeBoer 2016).

Liu (2011b) conducted a study looking at the differences in performance between first and final year students using the EPP; 4,373 first year and 1,823 final year students participated. Results indicated that there was a difference between first and final year students; first year students' score average was 110 on critical thinking whereas final year students scored 113. Similarly, on the writing aspect of the test, first year students averaged 113 and final year students 115, suggesting that the measurement is able to distinguish between the two levels of study (Liu 2011b). In addition, it was found that student SAT scores were moderately correlated with performance on the EPP in both of the constructs; in other words, the measure has construct validity. Supporting this, an early study conducted by the ETS showed that scores on the EPP increased as GPA, class level and amount of core curriculum completed increased, this suggests that the assessment has some construct validity (Young 2007; Marr 1995).

In addition, Roohr et al (2016) conducted a longitudinal study consisting of 168 students from a large university in the US using the EPP; the university in question had a retention rate of over 90% from first to final year . Students undertook the test when they first entered the university and were tested again later in their degree. The time between the two testing periods ranged from 4 to 55 months; most students took the second test after more than one year (one third of the students were taking Science, Technology, Engineering or Maths (STEM) subjects as their chosen degree). For the total EPP scores, results became more significant the longer the time between testing, for example when the second test was after three years ($p<.01$), whereas after four or five years ($p<.001$). This was also found to be the case on the reading construct and mathematics, however in critical thinking and writing the difference was only found to be significant after a four or five year gap.

However, the problem with this learning assessment goes back to the idea of multiple choice options not allowing students to engage in higher order thinking skills. This would suggest that this form of assessment could potentially lead to a false representation of student learning (Nicol 2007; Hyytinen et al. 2015; Scouller 1998). Contradicting this, Palmer et al (2007) stated that it is possible to construct a multiple choice examination that engages students in higher order thinking, owing to this it would suggest that the wording and construction of the questions is the most important aspect of the creation of a multiple choice test.

2.4.2.1.4. The Student Experience in the Research University Survey (SERU-S)

The SERU-S is the only self-assessment survey that is part of the VSA initiative. The SERU-S asks students to rate their level of proficiency at two different time points on a six-point scale on a variety of educational constructs such as critical thinking, writing, reading and comprehension skills (Douglass, Thomson, and Zhao 2012). Data from this instrument has been used at three levels; internally, externally and policy informant. Internally the data has been used to initiate campus discussion on strengths and weaknesses as well as to review programmes. Externally, the SERU-S has been used for campus accreditation and part of national accountability regimes, such as the VSA. Finally, the data has also been used to inform policy on information such as international students and research engagement (Center for Studies in Higher Education 2018).

The University of Minnesota is an example of an institution that has implemented and utilised the data of the SERU-S. Due to the format of this survey, it elicits a higher response rate from students with a 41% response rate from the University of Minnesota in 2017. This institution has used this data in a number of ways such as: to evaluate the effectiveness of a welcome week programme to better understand student retention rates, to enhance student services as well as to provide evidence of the effectiveness of a writing-intensive course (University of Minnesota 2018). This shows the value in this assessment and the different ways in which the data can be used, although there may be some potential validity problems due to the use of self-assessment, this measure still appears to be of high value.

2.4.2.1.5. National Survey of Student Engagement's (NSSE)

The NSSE is among the most popular surveys for measuring student engagement in the US, not dissimilar to the UK's NSS (Center for Postsecondary Research 2017). A survey conducted by the National Institute for Learning Outcomes assessment found that 76% of participating institutions measured their student satisfaction and engagement with some form of survey (G. Kuh and Ikenberry 2009). It is unsurprising that this is the case as research has found student engagement to be positively correlated to outcomes such as critical thinking (Laird et al. 2014). The current version of the NSSE reports on four engagement themes; academic challenge, working with peers, experiences with faculty and the campus environment (Center for Postsecondary Research 2017). As this method of data collection is a self-report, one could argue potential validity and reliability problems as students may be unable to provide accurate information (Bowman 2010b; Porter 2013). However, with the topic of the instrument regarding how students use their time with their own opinion of the institution, then one may say that self-reporting is a meaningful source of data (G. D. Kuh 2001). Supporting this, research comparing the NSSE with self-reported learning gain found a positive correlation suggesting that the assessment has strong convergent validity meaning the two constructs assessed (engagement and learning) appear to be correlated, as would be expected (Zilvinskis, Masseria, and Pike 2017).

2.4.2.1.6. The College Success Factors Index (CSFI)

Another measure from the US that is worth mentioning is The College Success Factors Index (CSFI). The CSFI is a self-reporting instrument that has recently been introduced in the US and focuses on ten student success factors grounded in research (Cox and Lemon 2016). Students are tested at the start and end of the academic year to measure any gain. This assessment looks at areas such as task planning, time management and engagement. A study by Cox and Lemon (2016) found that using interventions to improve the teaching of these academic gains improved students' performance across two years and asserted that self-report assessments are effective in measuring the changes in teaching quality and should not be completely disregarded.

2.4.2.2. Longitudinal Research in the United States

2.4.2.2.1. Lumina Longitudinal Study

The Lumina Longitudinal study was a five year long research project, supported by the Lumina Foundation, to investigate the difference between cross-sectional and longitudinal analyses as well as learning gain differences of first year to final year students both within an institution and in comparison to other institutions, using the CLA (Klein 2009). At the beginning of the research, 9,167 first year students completed at least one part of the CLA with 1,330 first year students completing all three sections. 40% of the schools that agreed to take part in the research, met the minimum sample size requirements for all three parts of the CLA(Klein 2009). These schools tested 4,748 at the end of their first year, 2,327 continuing into their second year and 1,675 were followed all the way into their final year (Klein 2009). Results indicated that cross-sectional and longitudinal scores were consistent, the performance task indicated that students gained as much during their first two years as their last two, however all the analytical writing gains seemed to occur during the final two years (Klein 2009).

2.4.2.2.2. Academically Adrift Longitudinal Study

Arum and Roska (2011) suggested that the most important outcome of HE is the ability to think critically. However, students are often poorly prepared for the transition from prior schooling into university due to the differing levels of engagement and independent learning (Arum and Roska 2011). The “academic effort” of students has steadily declined in recent decades dating back to 1961 where 67% of participants originally stated studying more than 20 hours per week, present day only ascertains 20% studying more than 20 hours per week (Babcock and Marks 2010). Not only this, but students have also been found to be choosing courses based on whether they believe they can pass with minimal effort, focused solely on grades as opposed to deep learning of a subject (Johnson 2003). Arum and Roska (2011) argued that universities focus more upon their research output than their teaching quality, it was also argued that due to no significant penalties to students for not engaging with teaching, they are more likely to put minimal effort into their studies. The researchers, therefore, conducted a longitudinal study, using the CLA, to ascertain student engagement, learning and skill development during their time at university.

Interestingly, one of the findings of this study was in student contact with the faculty outside of the classroom. It was found that students entering university with lower academic achievement were more likely to report having no contact with the faculty outside of the classroom with only 5% of students in the high academic achievement group reporting this same experience. This was even apparent after controlling for individual difference factors such as academic preparation and sociodemographic characteristics (Arum and Roska 2011). Furthermore, differences in independent work and contact hours also appeared to vary based on subject. Mathematics and science based subject reported a higher number of contact hours and independent learning (Arum and Roska 2011). However, a perhaps alarming result was that 37% of students also reported spending less than five hours per week preparing for their courses. Arum and Roska (2011) suggested that this may be due to the cultural shift of HE to simply gain a degree but to have a higher focus on the social side as well as managing part-time work. In addition, the results of the CLA scores did not appear to be related to contact with faculty, the more important factor for increasing CLA scores was due to the students' perceived expectations from their faculty.

Student studying alone compared to studying with others was also looked at in this research. It was found that, unsurprisingly, the number of hours studying alone related to gains on the CLA. However, it was found that studying with peers may actually result in a negative impact on the result of the CLA. This would suggest that studying alone would have a better impact on student grades. However, this difference could simply be due to the demographics of the students who prefer to study alone compared to those who study with others. This information could potentially be examined further, as collaborative learning is often studied within a classroom setting rather than outside the classroom. A lack of observation by a faculty member, may result in students "hanging out" rather than engaging in course material.

Furthermore, performance on CLA scores was also found to differ based on the subject focus of the student. Students' performance was higher if they chose to study a science or mathematics based subject, perhaps owing to the idea of STEM based subjects being of higher performance output. However, outputs for engineering and technology were quite low compared to other subjects (Arum and Roska 2011). However, in terms of overall student change in skill development, at least 45% of students were found to not demonstrate any significant gains in critical thinking during their first two years of university.

The authors asserted that this signalled the decline of effective HE teaching in the US and they suggested that this may be due to research focused faculties, who place their teaching as their second priority (Lindsay 2013; Arum and Roska 2011). However, Oppenheimer et al. (2017) suggested that even though the assessment may not have found any learning gain, this may not be due to a lack thereof but rather that the assessment itself is not sensitive enough to pick up small changes. In addition, Arum and Roska (2011) also suggested that this lack of change may be due to the students gaining more subject specific skills rather than generic skill development. Although all students may be taught to "think critically" the way in which they are taught to do so may differ resulting in the CLA being unable to pick up these changes as it appears to be purely based on a scientific method of critical thinking (Brooks 2012).

2.4.2.2.3. The Wabash National Study

The Wabash national study, implemented by the Center for Inquiry in Wabash College (CIWC), utilises a sub-section of the CAAP for its measure of critical thinking (Blaich and Wise 2011). This section is a 32-item instrument that measures skills in clarifying, analysing, evaluating and extending arguments (CIWC 2016). This study was a longitudinal assessment project with the aim of providing information to improve student learning and enhance the educational impact of their studies (CIWC 2016). This research went beyond just using the constructs of the CAAP by utilising a wide variety of scales, such as the student experience survey in order to attempt to achieve the aim of having high-quality data for institutions to promote improvements in student learning (Blaich and Wise 2011). Since the pilot stage of this research in 2005, over 17,000 students from 49 HE institutions have joined the study, with 30 in the new version of the study in 2010 (CIWC 2016).

Students were found to increase their critical thinking skills across the four years as well as on the "need for cognition" scale. However, there was a decline in student academic motivation perhaps suggesting that students have less motivation to complete their course material nearer the end of the degree (University of Rhode Island 2012). The outcomes of this ongoing research have allowed institutions to judge the need to make steps towards helping with student motivation (Blaich and Wise 2011).

However, this study was not without its problems. Although the assessments were a key part of this research, the issue was not with the measures themselves, but rather the difficulty that the institutions have had in identifying and implementing changes (Blaich and Wise 2011; G. D. Kuh et al. 2014). This highlights a problem with the assessments. Although they are able to measure the learning gain outcomes of students, they fail to identify the reasons why a student is at a certain level as well as failing to suggest changes that the institution can put in place to improve student performance (Klein et al. 2007).

2.4.2.3. Learning Outcomes in Canada

Canada identified accountability in education as a potential area for growth in 2005 with the release of their plan to improve upon educational outcomes (Rae 2005; Ontario Government 2005). This new scheme intended to achieve four main goals; what is needed to achieve excellence opportunity and sustainability, setting out goals and strategies that should be undertaken by key stakeholders, recommendations for how and when the changes should take place as well as the knowledge that has been developed over the course of reviewing this information (Rae 2005).

One strategy outlined in this report was to "pursue quality and innovation to make the student experience rewarding and successful". This strategy referred mainly to outcome and benchmark data including advice on performance measures, the collection of benchmark data which was yet to exist in the Canadian HE system as well as monitoring, evaluating and reporting on the quality and performance of students and the HE system. Currently, eight Ontario based universities have piloted the CLA in order to recommend whether this particular measure will be useful and effective in demonstrating quality to key stakeholders in HE – there was a divide as to whether the information that came from this pilot project was useful or could potentially be so (Lennon and Frank 2014). In 2011, Ontario also became involved in determining what standardised measures can be utilised in HE (Tremblay, Lalancette, and Roseveare 2013).

Queens University reported performance information that contained data on student performance in areas such as problem solving, economic reasoning and ethical reasoning. This information was presented to the faculty and informed a change in the curriculum to better improve and meet the outcomes that were expected of the students. This university was also one of six institutions that formed the HEQCO Learning Outcomes Assessment Consortium, which piloted assessments such as the CLA in a Canadian HE context. Although the steps towards learning outcome measures in Canada appears to be slow moving, it is taking the necessary strategies to find measures which are suitable to their HE context (Lennon and Frank 2014).

2.4.3. Other International Perspectives

In 2008, the OECD launched the Assessments of Learning Outcomes in Higher Education (AHELO) project with the aim to deduce whether measures of learning could be developed in such a way that they could be used internationally (Tremblay, Lalancette and Roseveare 2013; H. Braun and Bily 2013; OECD 2015). The results of this study indicated that there were differences across countries when using the same measure of student learning (Tremblay, Lalancette, and Roseveare 2013). This suggests that there may be too much variation in educational policy for there to be an assessment that can be implemented on an international scale, further supporting the idea that these assessments should be scaled down into specificity. Although learning gain is a relatively new concept, several countries have explored measures of student learning.

Engagement and satisfaction appears to be an area where most countries have started their assessment with many countries replicating the use of the NSSE such as Australia, Canada, China, Japan, Mexico, New Zealand and South Africa (Tremblay, Lalancette, and Roseveare 2013). However, it is important to note that this simply reports upon student participation rather than specific learning outcomes, this is not to say the results are not noteworthy as it has opened important discussions surrounding student support (G. Kuh and Ikenberry 2009).

Australia's implementation of learning measures is still within its infancy. In 2009, the Australian government proposed a new government incentive to achieve a diverse HE sector as well as creating highly skilled workers through a higher quality education (Australian Government 2009; Lodge and Bonsanquet 2014). The AHELO Feasibility Study was funded through the OECD and implemented in Australia between 2009 and 2013 in order to see if it is possible to measure student learning (Tremblay, Lalancette, and Roseveare 2013; Edwards and Pearce 2014). This was done through developing, testing and assessing measures of student learning on final year HE students (Coates and Richardson 2012). Focus groups were conducted with students to gain their insights into the drafted assessment items of a particular learning measure (Edwards and Pearce 2014). Students provided rich feedback which resulted in modifications to the testing instruments, further supporting the idea of Kuh et al (2014) that students are a valuable resource in measurement creation. The outcomes of this study were most prominent in the wider dialogue which opened further discussion within Australian HE about the importance and implications of measuring student learning outcomes.

In addition, there have also been discussions surrounding education quality within Chinese society as it has become a growing concern to establish quality assurance, this is hoped to be achieved through the implementation of the Quality Evaluation of Undergraduate Teaching scheme, not dissimilar to the English Teaching Excellence Framework (Xia and Zhong 2014). As stated by Xia and Zhong (2014) a number of problems were being faced by Chinese HE which led to the need for a quality assurance scheme such as institutions not utilizing evaluation results as well as a lack of guidance about benchmarks in HE. This process involves the submission of a Self-Study Report by the institution to the Higher Education Evaluation Center (HEEC). Each institution that submits a report will receive a classification, similar to TEF, to encourage further development in teaching quality. Not only is this scheme in place, but several institutions also utilize The College Student Experiences Questionnaire, which consists of 165 items that gains information on demographics, environment and self-report learning gains. However, most of the current learning outcomes assessments that are being implemented are carried out by non-government organisations. This suggests a need for the Chinese government to become more involved in the assessment of HE quality.

Finally, it is also worth mentioning other European countries who have taken on the need for learning outcome assessments such as Germany. In 2007, the German Federal Ministry of Education and Research funded studies in HE. This led to the creation and implementation of the Modelling and Measuring Competencies in HE Program (Zlatkin-Troitschanskaia, Kuhn, and Toepper 2014). According to Zlatkin-Troitschanskaia et al (2014) the aim was to pursue key objectives such as creating a framework for evaluating HE effectiveness and maintaining quality of the German HE system in comparison to other international systems. With this program, Germany has developed several instruments, which they are in the process of assessing and implementing to HE students to assess their competency (Zlatkin-Troitschanskaia, Kuhn, and Toepper 2014).

2.4.4. Discussion and Conclusions

In conclusion, it is clear that there is a growing international interest in learning gain and outcomes of HE, particularly to provide quality assurance and accountability for student outcomes. The US is the clear leader internationally in having a more well established and long standing initiative as is seen with the VSA (Steedle, Kugelmass, and Nemeth 2010; Liu 2011b). However, these kinds of measures cannot simply be transferred into other countries due to the differences in the educational systems. With the US focusing more on the breadth of knowledge rather than the depth, it would suggest that the measures which are utilised within this context are much more generalised (Steedle, Kugelmass, and Nemeth 2010) with an English HE system being in need of discipline specific measures due to the specified nature of the degree (Howson 2017b). As well as this, the discussion has begun surrounding the differences in how different subjects will be taught generalised skills. This discussion is opening the idea that although there are generic skills which all students gain from HE, the way in which a student will be taught the skill will be dependant on the subject which they choose to study (Brooks 2012). This calls for a need to understand what these differences may be in how students are taught these generic skills and the way in which they are expected to apply them. There is also, therefore, the need to grasp to what level of granularity these differences go in terms of whether they are discipline or subject specific as well as whether there may be some crossover with subjects who share characteristics.

The fundamental weaknesses surrounding the methods currently being utilised internationally lay in both the structure and implementation of these tools. The CLA has had it’s criticisms due to the fact that it is a standardised measure which results in small samples and labour-intensive marking (Possin 2013; Aloisi and Callaghan 2018a; Hardison and Vilamovska 2009). The CAAP addresses some of this labour-intensive marking by reducing the generic skills questions down to that of multiple choice. However, this assessment tool still contains a writing section and is considered to be a more standardised assessment. This would suggest that this measure is also going to result in a small sample due to being done on a voluntarily basis and being a measure which is in need of high-level cognitive thinking. Similar weaknesses can be said for the EPP. All of these measures attempt to create a standardised tool which attempts to test the skills of that of students. These measures may result in small samples as well as a false representation of student knowledge due to students potentially not exerting the maximum effort when conducting an assessment on a voluntary basis of which does not count towards their degree classification. In addition, the subjective nature of the marking of these standardised measures make comparability across time, students, disciplines and potentially institutions to be marginally more difficult as marking is known to have reliability problems (Arico et al. 2018). The SERU-S, however, is a multiple choice self-assessment survey and has been found to elicit a much higher response rate than that of the standardised measures suggesting that students engage more with multiple-choice self-reported measures of learning. As previously discussed in section 2.3, students have been found to be able to accurately assess their own learning when done on a longitudinal basis (Douglass, Thomson, and Zhao 2012; Bowman 2010b). Out of the current measures within the US, there is a limited amount which appear to measure on a longitudinal basis with the exception of the CSFI. As well as this there is a smaller number of measures which are self-reported multiple-choice assessment tools. Those which do hold these characteristics tend to focus on more behavioural areas of student experience such as engagement and time management rather than focusing on generic skills. Therefore, this points towards the need for a multiple choice self-report survey which focuses on the generic skills of HE. This would result in a higher response rate than that of standardised measures, more efficient collection and analysis of results as well as easier comparability due to the higher reliability and more quantitative nature of the results.

Internationally, countries all across the world have recognised the need for learning outcome measures and have begun to implement their own schemes for measuring quality in HE such as China's Quality Evaluation of Undergraduate teaching (Tremblay, Lalancette, and Roseveare 2013; Zlatkin-Troitschanskaia, Kuhn, and Toepper 2014; Lennon and Frank 2014). Most countries seem to have focused their attention on the behavioural aspects of student experience, as was the case with some measures in the US. This included looking at student engagement and satisfaction. However, Australia has begun to develop and test instruments which hope to measure student learning outcomes (Tremblay, Lalancette, and Roseveare 2013). This is also said to be the case in certain European countries (Zlatkin-Troitschanskaia, Kuhn, and Toepper 2014). However, these countries appear to be in the development and early stages of these learning outcome measures and do not appear to take into consideration the discipline specific knowledge and skills development which may occur within HE. In addition, there is little discussion surrounding the utilisation of a self-report measure of generic HE skills.

2.5. Learning Gain in an English Context

There have been recent changes to the HE landscape with the increase in university fees, courses and educational providers across the UK. The change of the funding and diversification of provision has increased pressure to provide justification of investment into HE (BIS 2011). The change to the financing of HE has served a purpose of underlining the importance of showing the value in having a HE degree in terms of the qualification that is acquired and the skills that are gained (McGrath et al. 2015).

2.5.1. Methods of Student Learning Currently used in England

Longitudinal measures of student learning appear to be a rarer occurrence within an English context with more methods opting for a cross-sectional or outcomes approach. Furthermore, measures of student learning in England seem to favour more qualitative based approaches rather than quantitative methods. However, there is still a broad variety of different methods that are utilised within an English HE context (McGrath et al. 2015)

2.5.1.1. Grades

Grades are a commonly used method of measuring an educational outcome. This method of learning gain relies on existing institutional methods, as all HEI's issue grades to their students. However, the key issue is comparability as different institutions will use differing assessment methods and have differing standards, thus the grades only provide a broad comparability between students. As learning gain relies on testing students over periods of time, this can be problematic with grades as they have different standards over time and across disciplines (McGrath et al. 2015). For example, students in the final year of their studies are expected to be able to critically analyse information whereas students in their first year are expected to describe (Bloom et al. 1956) this would indicate that the standards for students increase over time meaning the same grade at different time points will have a different meaning.

Arguably, there are problems with grades as a form of assessment in terms of validity in relation to students' academic achievement. Allen (2005) argued that grades should not simply consist of the results of a piece of work but should include other factors which are related to academic success such as student engagement. However, Randall and Engelhard (2010) found that teachers would often consider other factors when assigning a final grade to a student, this also suggests a validity problem, especially in terms of generalisability and a question as to why this method is so widely used. Another potential problem is that of the variance between grades, particularly if using the standard notations of "2:1 or 2:2" as many students achieve these grades and it would be difficult to distinguish between them (McGrath et al. 2015).

Metrics regarding grades have begun to be piloted within an English HE context. This includes using the GPA system, as adopted in America, which is a standardised grading system currently piloted at a number of institutions in England and being successfully utilised at some such as Oxford Brookes University (Higher Education Academy 2013). Furthermore, the Open University has implemented a Predicted rate pass model which uses grades to issue predictions on students' performance in other areas (Simpson 2006).

2.5.1.2. Surveys

Student surveys are commonly used in HE, as previously discussed, this is in the form of self-reporting by the students. In the UK, the NSS is a large scale survey with the involvement of over 300,000 students and 530 universities across the UK, as of 2017 68% of those eligible responded to the survey (HEFCE 2017e). The UKES on the other hand, is used in a smaller number of institutions. This survey has increased in participation rising from 29 institutions to 42 as of 2017 however, the response rate remains relatively low with as little as a 13%

response rate (Neves 2017; Buckley 2014). The NSS and UKES do not primarily attempt to measure student learning gains, however there are a small number of questions which could be relevant should the questionnaires be implemented several times across the year (McGrath et al. 2015).

A recent review of the NSS in 2014 suggested an inclusion of more questions that address student engagement as well as the addition of questions surrounding academic challenge and a student voice (Callender, Ramsden, and Griggs 2014). The UKES includes a 12 part questionnaire which asks students to what extent they believe their university experience to have contributed to areas such as writing, critical thinking and communication skills (HEA 2017; Neves 2017). However, as previously discussed, as these measures are only administered once a year they cannot be said to measure learning gain as they do not compare the "distance travelled" over time (McGrath et al. 2015). The importance of student engagement measures has been discussed in the US with evidence of a relationship between student engagement and learning gains, suggesting measurement of engagement would be a beneficial factor (Callender, Ramsden, and Griggs 2014; Pascarella, Seifert, and Blaich 2010; Tam 2004). Furthermore, testing of UKES revealed that students were more likely to have a positive learning experience if they were strongly engaged, further supporting the idea that engagement is an important factor of student learning (Buckley 2014). These methods of measurement show a lack of self-report measures in an English HE context, in regards to learning gain, as the currently used measures do not measure students across time points and they lack a depth of questions on student development. Although it could be argued that UKES does discuss skill development, the questions themselves are very closed and limited, not revealing much information in the changes over time (McGrath et al. 2015).

2.5.1.3. Standardised Tests

Standardised measures of student learning can be employed in a generic or specialised format. As previously discussed, standardised tests are more commonly utilised within the US (Randles and Cotgrave 2017; Fletcher 2006a; G. D. Kuh et al. 2015). The Educational Testing

Service (1993) provided a review of over 200 measures used within the US, some of which are intended to measure “distance travelled” though only occur at one time point and therefore, would not exist as a measure of learning gain. Within an English context, standardised measures have been investigated to be utilised as an entry test into HE, BIS looked at the feasibility of implementing this and found the standardised test did not provide any further information than that of the student’s qualifications (Kirkup et al. 2010).

Discipline specific tests appear to be the only standardised measure that is currently in use in English HE. One example is a test most commonly used in medical degrees known as Progress testing. This type of measurement is tested on students four times a year, every year across the degree in order to check student progress in the hope of discouraging students from studying for specific exams to then discard the knowledge (Freeman et al. 2010; Schuwirth et al. 2010; Chen et al. 2015). Although several medical schools have chosen to implement the progress testing tool, each institution has their own test suggesting institutions have large differences in the content of their degrees (McGrath et al. 2015). There is the question as to whether these kinds of tests could be extended into other disciplines and why this has not been attempted.

Measures, which are being utilised in US institutions, are in discussions of how they can be validated to being applied to the English HE context. The CLA is endorsed by several national HE commissions in the US, as well as gaining traction across Europe, suggesting a potential metric that may be able to be applied within England. However, there has been some debate as to whether this measure is specifically designed to be used within the US Higher Educational system. This proposes an argument that the CLA measures skills which are “too general” and does not take into account disciplines which put emphasis on specific skill sets (Banta and Pike 2012; Shermis 2008). As degrees in England focus on one subject, this would suggest that the CLA might be unable to gain traction in this HE context due to the students gaining generic skills that are tailored to their specific degree.

2.5.1.4. Qualitative and Mixed Methods

Qualitative methods have the advantage of gaining rich data and allowing students to reflect on their learning. However, the problem can come with comparability, as the rich detail can be difficult to compare across institutions. An example of the qualitative approach is the Personal Development Portfolio (PDP) which encourages students to think about their skills and discuss their gaps and what they want to achieve the following year. This approach to measuring learning is quite well developed in the English HE sector. However, it may be difficult to compare across institutions and can be quite difficult to use longitudinally, although, usually a PDP is conducted each year of the degree and therefore allows students to reflect back on their previous PDP (McGrath et al. 2015).

A mixed method approach can help with comparability, reliability and validity of methods as they can be used in conjunction with, and in support, of each other; however, this approach to measuring learning is rarely used within English HE. Although it is a highly robust and valid method of measuring learning, it requires high experience and resources to undertake. Two examples of mixed methods approach in England are the Academic Performance Tracking Tool (APTT) and the Advanced Level Information System (ALIS). The APTT monitors module and programme outcomes that are specific to subjects rather than generic skills. It consists of a comparison of grades through the American system of GPA as well as monitoring personal development through the NSS. The ALIS is a measure of grades, surveys, portfolios and other metrics from student records. However, the problem with these methods is that they can take time to produce meaningful results and can often only be done on a smaller scale (McGrath et al. 2015).

2.5.2. The Teaching Excellence Framework (TEF)

A UK government green paper (BIS 2015) identified learning gain as a potential metric in the teaching excellence framework (TEF). Learning gain was chosen as it is expected that it will show evidence of the strengths and weaknesses in teaching practice (BIS 2016a; BIS 2015).

TEF is a new scheme which aims to recognize and reward excellent teaching, this relates back to an argument in US education of universities having no distinct reward or incentive to improve teaching quality, the intent of this scheme is to implement this (HEFCE 2017a; Greatbatch and Holland 2016; HEFCE 2017b; Arum and Roska 2011). The TEF was developed by the Department for Education in England but implemented by the HEFCE and, at the time of writing, is being conducted on voluntary basis. In England, this currently only covers the teaching of undergraduate students with the hopes to eventually be expanded into postgraduate students (HEFCE 2017a). The purpose of TEF being a way of informing potential students about where to study, raising esteem for teaching, recognizing and rewarding excellent teaching as well as better meeting the needs of employers, business, industry and professions (HEFCE 2017b).

The TEF is in part related to quality assessment, as the two form a coherent system but have their own distinctive roles. Quality assessment ensures that HE institutions are delivering educational values to a high standard, which will involve looking at a broader range of areas rather than solely on teaching quality. Quality assessment also tends to ensure that standards and quality are met to a common threshold and often does not recognise achievement beyond this. While the TEF will not only incentivise high quality teaching and allow potential students to make better, more informed choices; but will include a judgement on performance beyond the average. Teaching excellence being defined to include teaching quality, the learning environment, student outcomes and learning gain. It is hoped that therefore, quality assessment and TEF will work coherently to promote, support and reward improvement to student learning (HEFCE 2017b).

The outcomes of TEF will feature on the register of HE providers and be made available in official sources of information for potential students (HEFCE 2017b). Those institutions that are assessed to be below the threshold for teaching quality will be named to the chief executive of HEFCE as a concern. HEFCE will then decide upon whether additional action should be taken, giving further incentive for institutions to raise their teaching quality (HEFCE 2017b).

To ensure institutions don't attempt to skew their results by reducing students from disadvantaged, under-represented or BME groups, the Director of Fair Access (DFA) will assess whether there has been a substantial reduction in said students and how this has affected the TEF results (HEFCE 2017b). This will ensure the outcomes of the TEF results are a valid representation of the institution's teaching quality.

In order to receive a TEF rating, there are several eligibility requirements that an institution must meet. This includes the institution having a suitable amount of student metrics data, the amount of which will determine the duration of the award. In addition, the institution must have an approved access agreement; contain undergraduate students being taught by the provider and deliver education that has specific designation for student support. Finally, the institution will undertake a quality assessment (HEFCE 2017b). Following this the institution will undertake the TEF assessment carried out by a panel consisting of experts in teaching and learning, student representatives, employment and widening participation experts (HEFCE 2017b). Table 2.2 shows the assessment framework for the TEF (HEFCE 2017b).

<table>
<tr><th>Areas of teaching and learning quality</th><th>Teaching Quality</th><th>Learning Environment</th><th>Student Outcomes and Learning Gain</th></tr>
<tr><td>Criteria – Statements against which panellists and assessors will make assessments</td><td>Teaching Quality criteria</td><td>Learning environment criteria</td><td>Student outcomes and Learning Gain Criteria</td></tr>
<tr><td rowspan="4">Evidence</td><td colspan="3">Core Metrics</td></tr>
<tr><td>Teaching on my course (NSS scale 1)
Assessment and feedback (NSS scale 2)</td><td>Academic support (NSS scale 3)
Continuation (HESA)</td><td>Employment/Further study (DLHE)
Highly skilled employment/further study (DLHE)</td></tr>
<tr><td colspan="3">Split metrics</td></tr>
<tr><td colspan="3">Additional evidence and supplementary metrics</td></tr>
<tr><td>Statement of findings</td><td colspan="3">Brief Description of why a particular rating was awarded</td></tr>
<tr><td>Overall outcome TEF rating</td><td colspan="3">The level awarded</td></tr>
</table>

Table 2.2: TEF Assessment Framework.

According to HEFCE (HEFCE 2017b) teaching quality is best understood in terms of the outcomes of student learning. Therefore, not only does the framework consider teaching which provides an appropriate level of challenge, contact and stimulation; but it also includes the learning environment i.e. the more personalised academic experience to maximise retention and progression such as work experience, library access and extra-curricular activities (HEFCE 2017b). Finally, student outcomes and learning gain is also part of the framework as this looks at outcomes related to lifelong skills that the students have achieved as a result of the degree as well as whether the students have gained the attributes necessary to go onto further study and employment (HEFCE 2017b). The assessment criteria are outlined in table 2.3 (HEFCE 2017b).

Reference	Criteria
Teaching Quality	
Student Engagement	Teaching provides effective stimulation, challenge and contact time that encourages students to engage and actively commit to their studies
Valuing teaching	Institutional culture facilitates, recognizes and rewards excellent teaching
Rigour and stretch	Course design, development, standards and assessment are effective in stretching students to develop independence, knowledge, understanding and skills that reflect their full potential
Feedback	Assessment and feedback are used effectively in supporting students' development, progression and attainment
Learning Environment	
Resources	Physical and digital resources are used effectively to aid students' learning and the development of independent study and research skills
Scholarship, Research and Professional Practice	The learning environment is enriched by student exposure to and involvement in provision at the forefront of scholarship, research and/or professional practice
Personalised Learning	Students' academic experiences are tailored to the individual, maximising rates of retention, attainment and progression
Student Outcomes and Learning Gain	
Employment and Further Study	Students achieve their educational and professional goals, in particular progression to further study or highly skilled employment
Employability and Transferable skills	Students acquire knowledge, skills and attributes that are valued by employers and that enhance their personal and/or professional lives
Positive Outcomes for All	Positive outcomes are achieved by its students from all backgrounds, in particular those from disadvantaged backgrounds or those who are at greater risk of not achieving positive outcomes

Table 2.3: TEF Assessment Criteria.

Following this assessment the institution will receive a rating of either bronze, silver or gold (HEFCE 2017b). Panels award an institution a bronze should they deem the institution to meet national quality requirements. Consistently exceeding national quality requirements sees the award of silver. Finally, a gold is awarded if the provision is consistently outstanding and to a high quality (HEFCE 2017b). At the time of writing TEF is currently in the pilot stage and is yet to be implemented across all institutions.

In addition, TEF has also begun to be piloted at a subject level. This means that different subjects will have a different set of benchmarks that institutions are expected to adhere to. There has been discussion as to whether this is the best port of action or whether there should be a set benchmark across all subjects. The same benchmarking factors have been used at subject level as were used at institution level; however, there have also been some potential pilot projects to look at different metrics based upon the subject that is studied. The decision was made to focus on 17 subjects which included their sub-disciplines for example "health and social care" included the sub-disciplines of "social work", "health studies" and "childhood and youth studies (Department for Education 2018b; Office for Students 2018b). This demonstrates that there are differences across subjects at an institution, as the difference in benchmarks suggests a difference in quality standards across different subjects. There has also been the potential argument that the metrics that each subject looks for may be different, this could be a potentially interesting area of study.

The emergence of this framework has increased the pressure to demonstrate the knowledge and skills that a student gains during their time at university (Raban and Cairns 2015). In addition, it has also shown a distinct lack of learning gain measures within an English HE context (OECD 2008). It is hoped that the TEF will give HE institutions an incentive to enhance student learning and to improve upon the overall quality of learning for students (Parker, Cleaver, and El-Hakim 2016). Therefore, an assessment that measures the learning gain of an institution's current student cohort would be particularly useful to aid the frameworks objectives.

2.5.3. Current Learning Gain Developments in English HE

The emergence of this framework led to the creation and implementation of learning gain pilot projects. Learning gain is a relatively new concept within an English HE context and is in its preliminary stages. HEFCE began piloting studies into learning gain in 2015, which consisted of 13 collaborative institutional projects with over 70 HE institutions. These projects combined the cross-sectional and longitudinal approaches to learning gain measures as well as a wide range of methodologies (HEFCE 2016b; Nijjar 2018). These projects have resulted in increased interest and activity in HE institutions in relation to student learning gain (Nijjar 2018). Mcgrath et al (2015) identified several methodological approaches to measuring learning gain which were tested within these pilot projects. These included grades, self-report surveys, standardised tests of both generic and specialised skills, qualitative methods such as portfolios and reflections as well as mixed methods (Nijjar 2018).

2.5.3.1. HEFCE Pilot Project Progress

These pilot projects also identified several "approaches" to learning gain measures. The first approach being that of cognitive gain, this has been defined as what students "think and know", an example of this kind of approach being utilised is the use of the CLA+ which the projects have piloted across a wide range of universities within England (Howson 2018). A further approach identified is that of "soft skills" defined as "affective measures of attitudes and how students feel" this includes constructs which may affect students' cognitive gain such as engagement and motivation (Howson 2018). The final approach is described as "employability and career readiness" which has been achieved through surveys, self-assessments and the participation of students in work experience activities. This approach intends to explore the student's skills in terms of their career adaptability and mobility.

As well as this, the pilot projects were categorised as either "telescope" or "microscope" projects (Nijjar 2018; Howson 2017b). Telescope projects defined as analysing large amounts of data from a large cohort of students such as learning analytics and registration data, both of which are useful in identifying patterns in progress and attainment. Whereas, microscope projects focus on collecting data from specific groups of students, usually over a period of time, such as targeting a specific discipline or subject (Howson 2017b; Nijjar 2018).

The pilot projects considered the potential scalability of the tested measures. An initial evaluation suggests that some of these approaches may have the potential for scalability at subject level across HE institutions. However, some of the approaches potentially need to be embedded locally and used as a benchmark (Howson 2018).

The pilot projects have also split their measures into three distinct categories of affective, behavioural and cognitive. Table 2.4 shows the measures that are considered to be part of each category.

Affective	Behavioural	Cognitive
- Self-efficacy - Well-being - Resilience - Disposition to learning - Satisfaction	- Engagement - Work placements - Co-curricular activities - Skills self-assessment - Employability enhancing experiences - Study abroad - VLE engagement - Learning analytics	- Grades - General cognitive gain (e.g. problem solving) - Disciplinary cognitive gain - Critical reasoning - Situational judgement - Research methods

Table 2.4: Measures explored in the HEFCE Pilot Projects

The cognitive approach to learning gain involves two projects. One of these projects is the piloting and evaluation of an instrument that is intended to measure student cognitive, meta-cognitive, affective and socio-communitive gain. This assessment is being tested for its validity, transferability and scalability as it assumes that all cognitive gains will be the same across disciplines. This evaluation is done through student perspective, questionnaires and tests of the instrument itself (Howson 2018). Further to this, the pilot projects are also evaluating the use of assessments which are already utilized in the US such as the CLA+, this is currently being tested in Birmingham City University and the University of Reading in order to consider whether this instrument is appropriate in an English HE setting (Howson 2018). Finally, the projects are also collecting institutional data on the trajectories of student grades. This is broken down into module marks, GPA and the final degree classification.

These projects have not been without their challenges. The projects found a distinct lack of engagement from students as well as a difficulty surrounding financial costs of staff time. In addition, the challenges surrounding the evaluation of learning gain have included a lack of clear purpose, methodological complexity and a difficulty surrounding the breadth of the definition of learning gain itself leading to a less clear focus (Howson 2018).

2.5.3.2. Projects Arising from the HEFCE Pilot Projects

The pilot projects that were run by HEFCE led to the creation of the National Mixed Methodology Learning Gain Project (NMMLGP), which incorporated the methods used in the Wabash National study that was conducted in the US. This study took a longitudinal approach to learning gain with an expected 31,000 students to take part in an online assessment for each year of their studies. This project was administered across a selected ten institutions combining a critical thinking and problem-solving test designed to detect a student’s ability to detect reasoning errors, apply principles, summarise conclusions and to assess impact of evidence (Nijjar 2018). The study also contains self-reflective questions which explored academic motivation, attitudes to literacy and diversity as well as student engagement (HEFCE 2016a; Howson 2018; Nijjar 2018). The purpose of this study was to understand the requirements needed to test across multiple institutions and disciplines, and is due for completion in 2020 (Nijjar 2018).

Complementary to this, HEFCE also announced the Higher Educational Learning Gain Analysis Project (HELGA) which aimed to compare known measures of learning gain by applying them to large administrative datasets to combat issues surrounding learning gain, capacity building and networking events (HEFCE 2017d; Howson 2018). The aim of this piece of research was to test the validity of measures of learning gain as well as to evaluate these techniques and their effectiveness between individual data, subject data and institutional data (Nijjar 2018).

2.5.3.3. Other Developments in Learning gain within English HE

The outcomes of the HEFCE pilot projects have sparked discussion and research into learning gain, particularly beginning to explore more discipline specific approaches and how these are best conducted (Evans, Howson, and Forsythe 2018). As well as this, that of behavioural aspects of learning gain, different approaches to measuring critical thinking and more affective measures such as self-efficacy have also been explored (Evans, Howson, and Forsythe 2018; Vermunt, Ilie, and Vignoles 2018; Forsythe and Jellicoe 2018). Learning gain has also been utilised to demonstrate differences in students in institutional initiatives such as widening participation (Cameron, Wharton, and Scally 2018). In addition, measures which have been developed and utilised in other countries have also begun to be scrutinised and tested within an English HE context to test for suitability (Aloisi and Callaghan 2018a). Baume (2018), however, argued that learning gain will be a difficult concept to embed within English HE due to the reluctance of many institutions to that of change which attempts to quantify and measure any form of educational concept.

Aloisi and Callaghan (2018a) explored the use of the CLA and its attempted integration into an English HE context. They argued that the CLA can be used to capture academic abilities in the liberal arts yet are unclear as to whether the skills that the CLA claims to be measuring can be identified as that of critical thinking, as the measure claims. This article also highlighted the lack of reliability in using this measure in a longitudinal manner, as would likely be the most appropriate use when wanting to measure learning gain. This article is therefore, further highlighting the need for measures which are specifically developed within an English HE context, as those which have been developed within the US have different contextual and utilisation differences that make them unsuitable for that of measuring learning gain within an English HE context. This was also supported by Evans, Howson and Forsythe (2018) whose exploration of the current national and international context of learning gain measures concluded that a more discipline specific approach may be needed due to the differences on the emphasis in skills, knowledge and the way in which the students are taught. Furthermore, they also recommended an approach to learning gain that sways away from that of the more objective measures but towards that of a student self-report of their own learning. These recommendations further point towards the need for a development of a discipline-specific self-report learning gain assessment tool.

Some projects have begun to explore focusing on a specific discipline when measuring learning gain. STEM subjects appear to be a popular choice in terms of focus, the differences between gender, ethnicity, prior academic achievement as well as high retention rates have been listed as the rationale behind the focus on this particular group of students (Rogaten and Rienties 2018). Scalise, Douskey and Stacy (2018) conducted a study testing STEM students on areas such as problem solving and critical thinking using standardised homework based tasks at three points in time; before the course, mid-course and after course completion. This study also split the participants into two groups of conceptual support and conceptual support with differentiated instruction to view which had the greater learning outcomes. It was found that the students who received conceptual support outperformed the other group as well as showing a 43% reduction in those at risk in their studies. This research focused on the outcomes of the assessment rather than focusing on the development of the approach to be more discipline specific. Similarly, Rogaten and Rienties (2018) examined the effects of the socio-demographic factors of STEM students on their learning gain. This was again, focusing on the outcomes of exploring learning gain rather than that of the methodological considerations to ensure that the concept is being measured appropriately.

In addition to this, one of the biggest problems that came out of the HEFCE pilot projects was that of getting students to engage with the measures (Howson 2019). Speight, Crawford and Haddelsey (2018) looked into various approaches to attempt to improve student engagement. It was found that the situational judgement test and self-reporting measure were the most effective in gaining a larger response from that of students. However, this research was unable to state as to whether the students were actually more engaged with the assessments themselves. Nevertheless, this would suggest that an assessment which is much easier and efficient to complete would result in higher proportions of students who engage with the assessment, therefore having higher and more valid sample sizes.

2.5.4. Discussion and Conclusions

The current measures within an English HE context seem to be favourable of more traditional and cross-sectional methods of measuring student learning (McGrath et al. 2015). Grades are still the most widely used and cited method of student learning, despite the flaws that indicate that grades have both reliability and validity problems (Randall and Engelhard 2010; Allen 2005). This would call for a shift towards a new method of measuring student learning, one which is able to take into account specific skills rather than attempting to quantify the overall learning of a student when each discipline and institution will have large contextual differences. Although surveys are widely used within an educational context, these seem to focus on more behavioural aspects and “proxy” measures of student learning such as engagement and satisfaction (Neves 2017; Buckley 2014; HEFCE 2018). Engagement has been listed as a vital aspect of student learning (Neves and Stoakes 2018) yet, it has been stated that this should not be measured alone as a measure of student learning but should be measured alongside other skills (Howson 2018). This would suggest that the surveys currently being utilised within English HE are missing the measurement of the skills that a student develops during their undergraduate degree and focus solely on the behavioural and affective aspects of the student journey. This gap was thought to be able to be filled by utilising the generalised standardised tests that were used in US institutions. However, it has been noted that these measures are “too general” and do not take into account disciplines that put emphasis on specific skill sets (Banta and Pike 2012; Shermis 2008). This is further highlighting the need for a discipline specific approach to the development of an assessment tool as this has yet to be explored thoroughly within an English HE context.

TEF also brought to light the need to measure the quality of teaching within HE. This framework has further put pressure to show accountability for student learning when studying an undergraduate degree. The recognition of differences across disciplines has begun to be addressed through this framework as there has been pilots for a subject level TEF which will have different benchmarks dependant on the subject (Department for Education 2018b; Office for Students 2018b). These developments are further highlighting the differences across subjects as the variations in benchmarking would suggest differences of quality and expectations across the subjects in areas such as engagement and employability. The inclusion of learning gain within this framework led to the creation and implementation of the HEFCE pilot projects (Nijjar 2018; HEFCE 2017d). These projects explored affective, behavioural and cognitive skills to be able to measure student learning gain. These projects found that student engagement was low as well as differences across subjects (Howson 2018; Howson 2017b; Howson 2019). This would suggest a need to have a more efficient and accessible assessment to enable students to be more engaged, as well as a need for a more discipline specific approach.

Learning gain research has begun to develop further within an English HE concept due to the HEFCE pilot projects promoting discussion and interest in the area. The recent developments have resulted in learning gain being utilised in institutional initiatives (Cameron, Wharton, and Scally 2018), the testing of existing generalised approaches (Aloisi and Callaghan 2018a) as well as a more discipline focus when measuring learning gain (Scalise, Douskey, and Stacy 2018; Rogaten and Rienties 2018). However, much of this research is focusing on the outcomes of the assessment or measure rather than that of the methodological aspect of measuring the concept itself. This would suggest that although there has been some developments within learning gain in an English HE context, there has been little progress in understanding and developing discipline specific tools as well as there still being a need for this approach. There seems to be a focus on the product of the assessment tools rather than the development of them. In addition, there seems to be an awareness of the most efficient and appropriate methods such as that of the self-report survey (Howson 2019; Speight, Crawford, and Haddelsey 2018). Yet, there is little advancement in actually developing such tools.

2.6. The Gap in the Literature

The review of the literature into learning gain assessment tools within an English HE context concluded that although developments in learning gain have begun, these tend to focus on the outcomes of a learning gain measure rather than the development of a suitable assessment tool (Evans, Howson, and Forsythe 2018). Furthermore, there is also a lack of focus on a more discipline specific approach, despite the understanding that this is what is needed to further development in learning gain approaches (Banta and Pike 2012; Shermis 2008; Howson 2019). In addition to this, the use of self-report methods of measuring student learning appear to be consistently mentioned within the literature as an efficient approach (Speight, Crawford, and Haddelsey 2018; Douglass, Thomson, and Zhao 2012; Pike 1995). Yet, there is little development and discussion surrounding utilising this method in order to measure student learning gain, particularly when looking at cognitive skills (Howson 2019).

The literature noted that there is no “gold standard” for measuring learning gain and that the concept is complex with potentially many methodological considerations and approaches that could be taken (Baume 2018; Evans, Howson, and Forsythe 2018; Vermunt, Ilie, and Vignoles 2018). Therefore, this research will close the gap in the knowledge by aiming to develop a learning gain assessment tool which is discipline specific and takes the approach of that of a self-report survey method. As this is an area of learning gain which has been identified to be needed but has yet to be explored. As previously discussed, Kuh et al (2015) argued that students should be a part of the process of developing a learning assessment tool due to them being the “consumers” of the end product and who are often overlooked.

The rationale behind choosing that of science students stems from the need to focus on students from the same discipline who will face similar teaching, classroom and assessment styles, expectations of cognitive skills such as a high emphasis on scientific reasoning and critical thinking (National Audit Office 2018; Siekmann and Korbel 2016). As well as similar emphasis on the employability skills that are within the discipline, believing to be at an all time low particularly in the Biological Sciences (BIS 2016b).

3. Research Methodology

Figure 3.1: Overview of Progression of Sections of Research Methodology Chapter.

3.1. Introduction

The purpose of this chapter is to provide an overview and justification for the intended methodology of this research project. Firstly, the chapter will explore philosophical assumptions, the potential research approaches as well as the research strategy for this project. The selected approach for this project was a mixed methods design, which consists of both qualitative and quantitative techniques. The objectives, approach and strategy for each of the individual phases are also discussed within this chapter.

Saunders, Lewis and Thornhill (2016) proposed the idea of the "research onion" which indicates the steps which one should take when conducting research and justifying why that particular approach or method has been chosen over the other potential options. The outer layers of the "research onion" are often overlooked and ignored; these will be explored within this chapter. As figure 3.2 shows; the outer layer of the onion is the consideration of a research philosophy. This includes the approach one may take such as Ontology as well as the stance within that such as Positivism. The researcher must then consider an approach to the research i.e. whether the approach will be inductive or deductive. Following this comes the methodological aspects of the research. This includes the methodology itself as to whether it is qualitative, quantitative or mixed methods, as well as the strategy the researcher intends to take, e.g. experiment, survey or case study. In addition to this, the researcher must also consider whether their research should be of a longitudinal or cross-sectional nature. The final stage being the actual data collection and analysis.

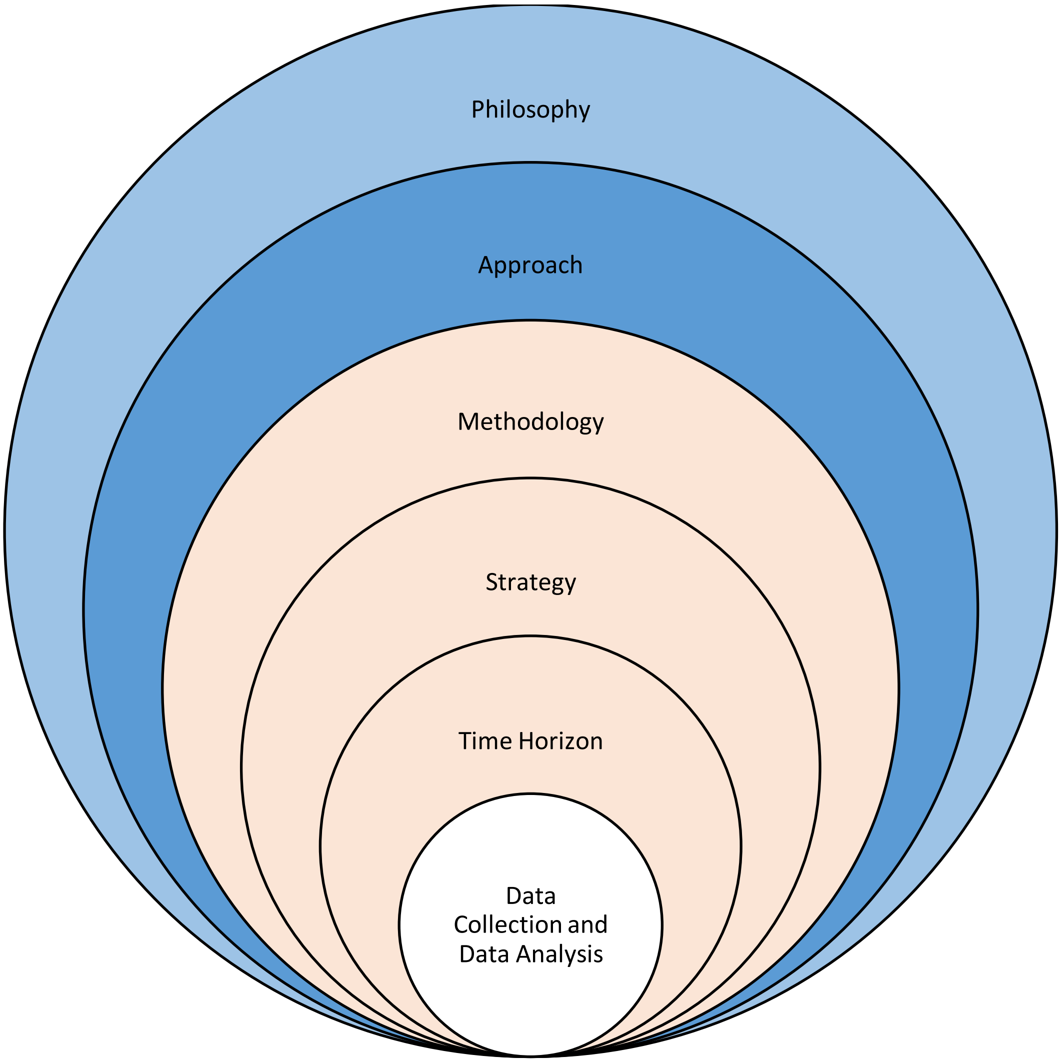

Figure 3.2: The Research Onion

The literature review provided an awareness of what should be considered when creating a measure of student learning. However, it has shown that there has been a lack of subject specific measures within an English context. Another untapped resource is that of a student perspective as to what they believe they have gained and what should be included within these measures. They are the best equipped to understand the skills which they are gaining across the course of their degree. In addition to this, as institutions differ, there is grounding in the idea of being more focused on a particular institution and subject, to have an internal understanding of the level that the students are at. Whether STEM was too broad a cohort to create a measure for, was firstly explored. Before considering this, we must consider the philosophical underpinnings of the methodological approach that was undertaken in this research project, as indicated by the “research onion”.

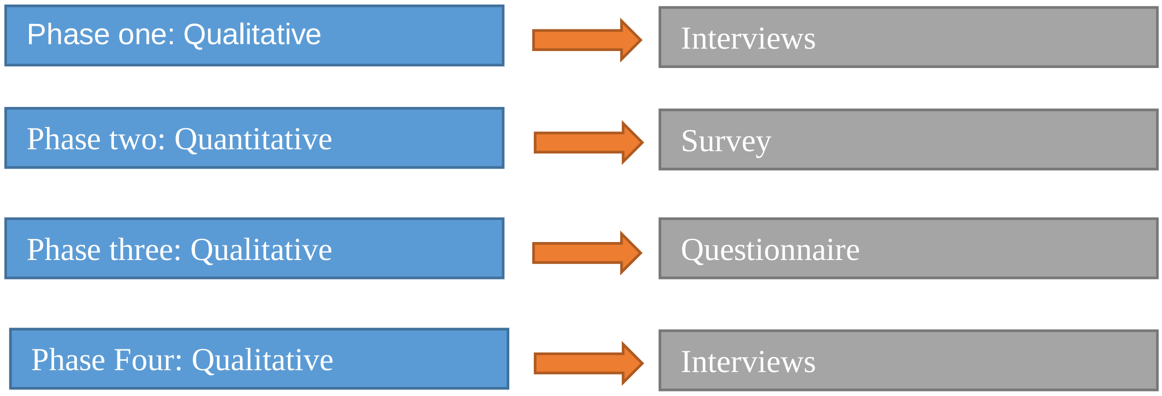

Figure 3.3: Breakdown of the methodological approach in the research project.

3.2. Research Philosophy

Several scholars agree that knowledge is a result of different research paradigms and dimensions which allow research to be conducted and ultimately, knowledge is furthered (Patton 2001; Lincoln and Guba 1985). Kuhn (1970) described the idea of ‘paradigm shifts’ i.e. what is considered scientific “truth” changes over time resulting in a ‘paradigm shift’ and change of beliefs. Research philosophy refers to the beliefs of the researcher in how the research should be conducted as well as the researcher’s beliefs and assumptions of developing knowledge (Saunders 2016). This is not dissimilar to what has been previously mentioned as these beliefs may change dependent on the chosen stance of the researcher. These philosophical stances can be broken down into assumptions about human knowledge and how it can be communicated (epistemology), assumptions about the very nature of what is reality (ontology) as well as the extent to which the researcher’s own values influence their research (axiology). The stance that the researcher chooses to take will ultimately form the research questions and hypotheses (Saunders 2016; Burrell and Morgan 1979).

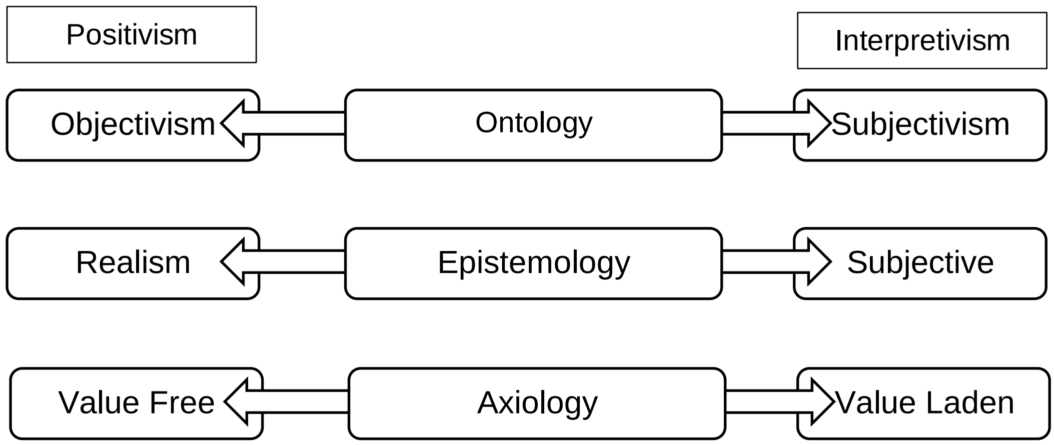

Figure 3.4: Research Philosophies and their paradigms.

The development of a reflexivity and introspection of one's own beliefs is important to be able to examine and critically analyse one's own beliefs with the same scrutiny as the biases of others (Saunders 2016). Familiarising oneself with these philosophies and providing a justification as to which is being held by this particular research project, is a significant step in the research process. A number of different researchers cannot seem to come to an agreed definition on several of these paradigms and how they relate to each other, causing some difficulties in choosing an appropriate philosophical stance (Mason 2014). Each strand of research philosophy will have a positivist approach and an interpretivist approach. These approaches have advantages and disadvantages which should be carefully considered as can be seen in Table 3.1.

	Positivism	Interpretivism
Advantages	• Collection of a large amount of data • Clear theoretical focus • More control of the research process • Easily comparable data	• Gives an understanding of the how and why • Researcher can adapt to changes when they occur • Good approach for understanding social processes • Allows for complexity and contextual factors
Disadvantages	• Inflexible • Weak at understanding social processes • Often does not discover meanings and the "why"	• Data collection can be time consuming • Data analysis is challenging • Researcher has to be aware of the uncertainty of patterns • Generally perceived as less credible

Table 3.1: Pros and Cons of the Positivist vs. Interpretivist Approach (Raddon 2010).

3.2.1. Ontological Considerations

Ontology refers to the reality of the research and shapes the way in which the researcher sees and studies their research variables (Saunders 2016; Panke 2018). The types of questions, which are asked by an ontological stance, are questions such as; what is the nature of reality and what the world is like. An example of this would be 'what is it like being in a managerial position' (Saunders 2016). There are two sides of this debate; some believe that research consists of one definitive truth that is made of a distinct set of laws. The other side of the debate is that of relativity i.e. there are many subjective truths that are open to interpretation (Mason 2014). These positions are often referred to as objectivism and subjectivism (Bryman 2012). Both these positions need to be considered before one can be chosen for the research project.

3.2.1.1. Objectivism

Objectivism takes on the assumption of general scientific principles, it is considered 'classic' scientific research (Bryman 2012). This approach to Ontology states that reality consists of solid unchanging ‘things’ that are measurable and observable (Burrell and Morgan 1979). Objectivism considers that research is external to ourselves; social entities are similar to that of physical entities and cannot be influenced therefore, there is also the idea that there are some structures that although present, cannot yet be articulated (Saunders 2016; Bryman 2012; Burrell and Morgan 1979). Essentially, it can be thought of as an organisation that has rules, regulations and standardised procedures that cannot be changed. The objectivists seek to keep their research free of values and opinions with a focus on factual information so as not to bias their findings, there is therefore a detachment from their own morals and values throughout the research process (Saunders 2016). An example of this objectivist paradigm would be looking at bad behaviour in schools, this would be done by agreeing on definitions of bad behaviour (e.g. shouting, answering back) and counting the number of times, this occurs. This is simply looking at what is happening and recording this in an objective way. The problem with this kind of stance is that it does not show the reasoning behind what is happening, in particular the thoughts and opinions of the participants involved, this is completely disregarded in this paradigm of ontology.

3.2.1.2. Subjectivism

Subjectivism is essentially the opposite end of the spectrum to objectivism. This paradigm argues that social phenomena are produced through social interaction and are in a constant changing state of numerous truths and interpretations (Bryman 2012). These ideas are based on perceptions and interpretations rather than a set of “rules”, an extreme view of which is Nominalism; which states that social structures are creations of the researchers and other individuals and that there is no underlying reality beyond what people attribute to it (Saunders 2016). These ideas come from the notion that two individuals may perceive the same situation entirely differently, therefore the “reality” of a situation is dependent on that individual. In addition, if adopting this paradigm, the researcher must research the phenomena in extreme detail and consider influences such as geographical and socio-cultural contexts in order to understand the “reality” of an individual, often resulting in the study of different opinions and narratives that can help to explain these different realities (Saunders 2016). Using the same example that was used in objectivism, subjectivism would argue that there are no agreed definitions for bad behaviour but it would be the perceptions of the staff as to which child is difficult and which is simply lively.

3.2.2. Epistemological Considerations

Epistemology is concerned with knowledge and how we know what we know. It is also concerned with how we communicate that knowledge to other people and what knowledge is considered as valid (Saunders 2016; Burrell and Morgan 1979). This area of philosophy may ask questions such as "What constitutes good quality data?", it is concerned with the reliability and validity of the research and knowledge that is being studied (Saunders 2016). Epistemology is also directly related to ontology as once we perceive reality in a certain way, this will have an effect on the knowledge of the world (Mason 2014). It is also concerned with the limitations of knowledge and can be regarded as the study of criteria for what the researcher believes constitutes knowledge (Research Methodology 2018).

3.2.2.1. Realism

Realism is research which seeks to explain and predict results by searching for regularities and causal relationships between variables, essentially the more "scientific" method of research (Burrell and Morgan 1979). Polit and Beck (2008) associated realism with natural sciences, however Bryman (1984) associated realism with social research that applies natural sciences. Although there are some different views surrounding this, there appears to be an agreement that this branch of epistemology is the more objective approach. There are arguments as to whether a hypothesis can actually be deemed as true or can only be falsified i.e. proven wrong. Even with this argument, the falsification of a hypothesis is still adding to existing knowledge by eliminating potential theories (Burrell and Morgan 1979). The main idea of realism is that social sciences should treat their research in the same way as any other science would, therefore being more objective. It focuses on causality and reduces phenomena to their simplest elements, making it a quite reductionist approach (Research Methodology 2018). Realists generally use existing theories to develop a hypothesis which will then be tested to either be confirmed or falsified (Saunders 2016). In addition, realists try to remain impartial to their research so as not to bias their findings, this is done from attempting to undertake research without considering your own values and morals (Crotty 1998).

3.2.2.2. Subjective

A subjective view is similar to that of subjectivism, it accepts that the world views can constantly change and shift. There are no objective facts out there but rather the results of the research are constructed (P. Mason 2014). The subjective view focuses on the details of the situation, the meanings and motivations behind the actions, rather than an objective measure of the said actions (Research Methodology 2018). This approach places emphasis on getting the opinions of the individuals who are directly involved in the activities of the research, it rejects the point of view of the “observer” and believes the only way to understand behaviour is by occupying the frame of reference of the participant in the action (Burrell and Morgan 1979). This therefore, focuses on narratives, stories, perceptions and interpretations of a situation and takes new understandings and worldviews as a contribution to research (Saunders 2016). This approach emerged in early twentieth-century Europe and has several strands such as hermeneutics, phenomenology and symbolic interactionism (Crotty 1998). In the opposite way to positivism, this approach argues that social research cannot be studied in the same way as other sciences, those who take this approach do not believe that there are general “laws” that can be generalised to everyone but rather that the complexity of humanity is lost if you try to force generalisations. Subjectivists are all about collecting rich detailed information, from participants, about their individual circumstances (Saunders 2016).

3.2.2.3. Pragmatism

This approach to epistemology is that which is not method focused. It does not consider knowledge to be purely scientific nor does it consider it to be entirely subjective. This approach is somewhat of a middle ground which purely focuses on finding the most appropriate and fitting methodology for finding the answers to the research question. Pragmatist researchers look at both the what and how, with an ideal to seek an explanation for any findings. They focus on both the internal and external world i.e. the views of individuals as well as the scientific observations (Creswell 2014). The focus of this approach is that of problem solving, it is a highly reflexive approach as the outcomes of this form of approach are very value-driven (Saunders, Lewis, and Thornhill 2016)

3.2.3. Axiological Considerations

Axiology is concerned with judgements of value and specifically the role of the researcher's own values on all stages of the research process, it primarily focuses on the aims of the research and to clarify whether you are trying to explain, predict or understand the phenomena (Research Methodology 2018). Axiology refers to the values and ethics of the research process, it looks at questions as to how we, as researchers, deal with values. It is argued that researchers show their axiological values by making a judgement on why they deem one topic to be more important than another. An example of this would be if you choose to focus your research data collection in the form of interviews, this would suggest that you consider personal interaction of more value than a quantitative questionnaire (Saunders, Lewis, and Thornhill 2016)

3.2.4. Philosophical Stance of this Research Project

The ontological stance of this research project is that of objectivism. This research project assumes that social phenomena can be measured and observed in the same way as any other scientific research. With the aim of this research being to create an objective measure of student learning, this aligns with this philosophical view due to the nature of wanting to objectively measure student skills. The epistemological stance that this research has taken is most closely related to that of pragmatism. Pragmatists usually take the mixed methods approach in that they often seek to find the "how" with the qualitative method and use the most appropriate methods for the specific research project. This view is appropriate for this research as the quantitative nature of the assessment tool will require validation; this can be sought from qualitative data that could reveal whether the assessment tool aligns with reality. This eliminates any potential limitations of creating an assessment tool to be an objective measure as it will have support from a different method.

3.3. Research Approach

All research begins with a question which the researcher intends to answer through data collection and inference of the outcomes. However, there are two approaches that one can take when seeking answers to a research question. One being the inductive approach, which is considered a “bottom-up” approach and the other a deductive approach, considered “top-down”. These approaches represent whether the researcher begins with a theory that they intend to revise or add knowledge to; or whether the researcher simply aims to collect data and to form a theory based upon that data collection alone (Panke 2018). In addition to this, there is also a third approach known as abduction which is a crossover between the two approaches and goes back and forth between using theory and developing one’s own theory (Saunders, Lewis, and Thornhill 2016).

	Deduction	Induction	Abduction
Logic	When the premises are true, the conclusion must also be true.	Known premises are used to generate untested conclusions	Known premises are used to generate testable conclusions
Generalisability	Generalising from the general to the specific	Generalising from the specific to the general	Generalising from interactions between the specific and the general
Use of Data	Data collection is used to evaluate propositions or hypotheses related to an existing theory	Data collection is used to explore phenomenon, identify themes and patterns and create a conceptual framework	Data collection is used to explore a phenomenon, identify themes and patterns, locate these in a conceptual framework and test this through subsequent data collection
Theory	Theory falsification or verification	Theory generation and building	Theory generation or modification; incorporating existing theory where appropriate to build a new theory or modify existing theory

Table 3.2: Deduction, Induction and abduction assumptions (Saunders, Lewis, and Thornhill 2016)

Each of these approaches has a set of assumptions and beliefs about logic, generalisability, use of data and approach to theory. These can be seen in Table 3.2 as viewed in Saunders, Lewis and Thornhill (2016).

3.3.1. Inductive Approach

The inductive approach to research is the method most commonly associated with qualitative research. Those who take this approach argue that participants should be seen as human beings with individual perspectives and views of the world rather than as research objects. This approach often takes on a small sample of participants as opposed to a larger sample which is often used in the deductive approach (Saunders, Lewis, and Thornhill 2016). This approach is often used when the research in question has not been researched before or is a brand new phenomenon that has no previous knowledge to be taken from (Panke 2018). Those who follow the inductive approach tend to be quite critical of deduction, arguing that it reduces phenomena to objective measures and does not consider reasoning behind the cause-effect of the research (Saunders, Lewis, and Thornhill 2016). Figure 3.5 shows the stages in the inductive approach.

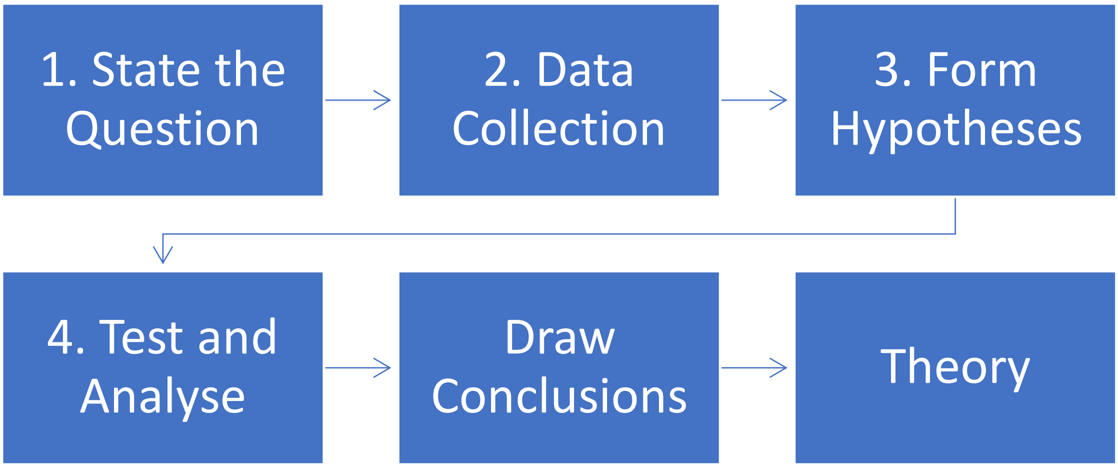

Figure 3.5: Process of the Inductive Approach (Bryman 2012).

However, Creswell (2014) argues that no research can be purely based on data as there will be some prior conceptual knowledge from the researcher that leads them to decide to conduct that particular research in the first place. There is therefore the argument as to whether any research could purely be considered “inductive”. Although, there is the argument that phenomenology does exactly that by gathering rich detail on a particular phenomenon without prior experience of the topic.

3.3.2. Deductive Approach

The deductive approach to research is what many people would consider the most common method of scientific research. This approach takes existing laws and theory and subjects them to rigorous tests to either validate or reject the current knowledge (Saunders, Lewis, and Thornhill 2016). Figure 3.6 shows the process by which a deductive approach to research is taken. It shows that researchers will develop a hypothesis based upon existing knowledge which they will then test through their own data collection (Bryman 2012). The results of which can be used to revise an existing theory or knowledge.

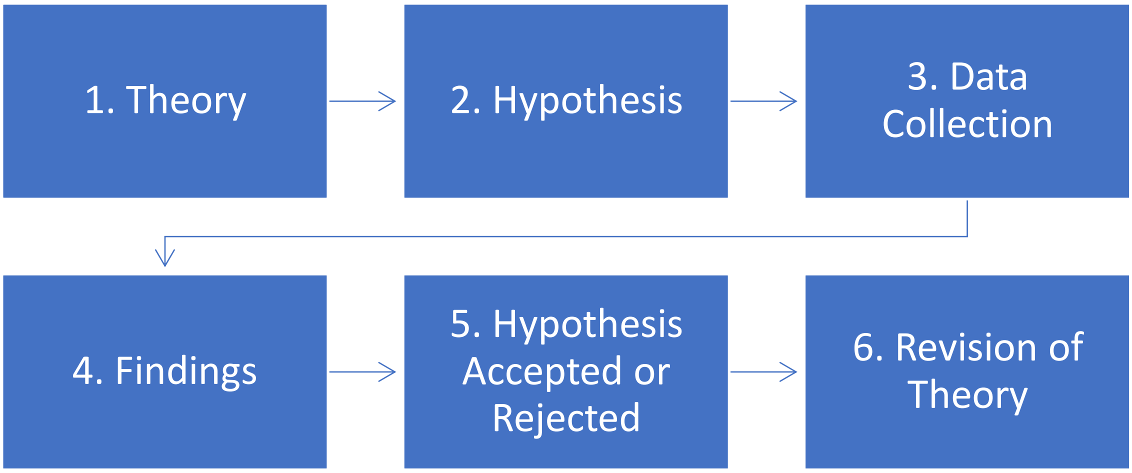

Figure 3.6: Process of the Deductive Approach (Bryman 2012).

According to Bryman (2012), a skilful researcher is able to operationalise their hypothesis by indicating the appropriate methods that need to be undertaken in order to test the hypothesis. However, the limitation of this approach is that it is very reductionist in nature i.e. it reduces the most complex of problems to simple elements so that they can be objectively measured, often quantitatively. In addition, as this approach tends to generalise from a sample of participants, it can often be tricky as to whether it can be generalised to the entire population or simply to the target population. For example, if one was to study views of working at a particular company, and only one branch was studied, this could then not be generalised to all branches within that company as it may differ from the others (Saunders, Lewis, and Thornhill 2016).

3.3.3. Abduction Approach

Whilst the inductive and deductive approaches have a clear linear process of going from theory to data (deduction) or data to theory (induction); the abductive approach is an approach that combines the two by going back and forth between theory and data (Suddaby 2006). Abduction is viewed as a useful approach for researchers due to the fact that it is not governed by rules of a process that must be followed but rather constantly reviews the process, going back and forth, to allow for the generation of new knowledge (Reichertz 1995). This approach is one of introspection and reflexivity that the researcher must adopt in order to be successful at their research, arguably this is one of the most important skills that any researcher can have, to be able to criticise oneself (Saunders 2016).

3.3.4. Summary of Research Approaches Used

This research adopted the abduction approach by alternating between theory and findings. This research deduced some theory from the literature review and background search but also induced theory from the qualitative data such as from the preliminary interviews and staff interviews. This allows for the researcher to be more flexible in their approach to looking at the data and to take a more critical approach to the outcomes of each stage of the project as it is occurring. Suddaby (2006) argued that no research project can be purely inductive or deductive and requires the use of both approaches. In support of this, Saunders, Lewis and Thornhill (2016) suggested that it would be logical to use both of the approaches in conjunction with each other and it is often beneficial to the researcher, therefore the abduction approach is the most appropriate method.

3.4. Research Strategy and Methodology

Choosing the appropriate research methodology is an important stage in the research process as it informs the methods that you choose later on as well as the kinds of hypotheses that you can explore; the options being quantitative, qualitative and mixed method (Panke 2018). In simple terms, quantitative research is considered to be research that is concerned with numerical data and that can be analysed statistically. Whereas qualitative research is concerned with non-numeric data such as interviews, video clips etc. However, it is not always this straightforward; often a primarily quantitative survey will require further information that the participant will then have to write, which then is providing qualitative data. Conversely, qualitative data may also be used to inform the creation of a quantitative learning gain assessment tool. These methods are often used in conjunction with each other, often known as a mixed method approach, and are not without their advantages and disadvantages, as is summarized in Table 3.3 (Saunders, Lewis, and Thornhill 2016).

	Quantitative	Qualitative
Pros	• Large-N Projects • Allows for statistical analysis to analyse patterns • Several hypotheses can be examined at once • Can generalise to bigger populations	• Can conduct an in-depth analysis and shed light on the "why" • Move beyond a purely correlational view of variables • Due to smaller number, often collects newer, novel information
Cons	• Reductionist • Cannot provide detailed insight into cause-effect	• Small-N Projects • Limitations to generalization • Potential problems with replicability with analysis

Table 3.3: Advantages and Disadvantages if Quantitative and Qualitative Method (Panke 2018).

Built environment research has been known to draw upon different disciplines such as psychology, engineering and management. Research within this area has mainly focused upon quantitative research strategies but has recently begun to incorporate more qualitative approaches in order to stay in line with the "real world" application of built environment research (Amaratunga et al. 2002; Tucker, Mulliner, and Wilson 2017; Wilson and Cotgrave 2016). Creswell (2014) stated that qualitative and quantitative methodologies should not be seen as polar opposites but should be viewed as lying on a continuum with mixed methodology in the centre. Although some research tends to be more on one end of the continuum than the other, it is critical to understand where on this continuum this current research sits.

3.4.1. Quantitative

The quantitative approach to research is most often associated with positivism due to its mostly objective and scientific nature as well as the deductive approach; focusing on data to test a theory. However, this may not always be the case, there is not a black and white way to conduct a quantitative research project; there may sometimes be a philosophical approach which requires the researcher to adopt interpretivism, and sometimes there may be a case where the researcher is required to take an inductive approach (Saunders, Lewis, and Thornhill 2016; Bryman 2012). Quantitative research often relies heavily on coming to a hypothesis based on existing theory, however sometimes the hypothesis will not be specified and this is inferred from the nature of the theory and the concerns that the researcher has surrounding this (Bryman 2012). Creswell (2014) described the quantitative method as a means for "testing objective theories by examining the relationship among variables". It was also discussed that the variables would be measured to produce a numerical outcome that could then be statistically analysed. Furthermore, it was also argued that those who take this approach to research have set assumptions about testing theories in a deductive approach, reducing bias, controlling for extraneous variables, and being able to generalise and replicate the findings. In addition to this, when sampling for participants, quantitative methods focus on selecting a sample which will be representative to the target population. This can be done through probability or non-probability sampling, essentially the difference being whether all the people in the target population have an equal chance of being selected to participate (Bryman 2012).

Mujis (2004) argued that there are four types of research question for which the quantitative method is most suited;

- When we require a numerical answer such as a "how many" research question, this makes it obvious that qualitative methods would not be suitable.
- When we want to study a numerical change such as an increase/decrease in a variable or phenomenon.
- When wanting to understand what factors predict a certain phenomenon.
- When testing hypotheses.

In the case of this research, as the focus is upon learning gain, this is looking at whether there is a numerical change in a student's skill set over a period of time, therefore suggesting that a quantitative approach to measuring said skills is an appropriate fit. However, there are also instances to be aware of where quantitative research is not deemed to be an appropriate methodology. These cases include when wanting to explore a problem in great depth and to understand the underlying mechanisms behind a phenomenon. As well as this, quantitative methods can be difficult when attempting to develop new theory, although not impossible, the qualitative method is better suited to this (Muijs 2004).

3.4.1.1. Strategy

The most common strategies for inquiry that are used in the quantitative approach are those of survey research, experimental research and observational research; in table 3.4 the suitability of the strategy for this research is explored. (Creswell 2014).

		Suitability for this Research
Survey Research	Pros	• Could be used in both a cross-sectional and longitudinal nature • Allows for easily comparable results on skills and seeing changes over time that may even be miniscule through statistical analysis. • Easy accessibility for students; more likely to engage with project. • Could also incorporate qualitative open questions; or follow up with purely qualitative survey.
	Cons	• Self-survey problems as to whether students are accurate at knowing their skills; however, could be reduced by taking the longitudinal approach. • Difficult to know whether the measurement is accurate to students' actual skill or just what they believe (i.e. could be under or overestimated).
Experimental Research	Pros	• Can allow for a high degree of control over the research. • Would measure skills directly.
	Cons	• Could be difficult to get students to engage without an incentive – even then students have often still not engaged. • Lacks ecological validity, can be artificial • Difficult to capture perceptions of students and staff on skills gained during degree • Difficult to capture student engagement with the course
Observational Research	Pros	• Is observing students in natural environment; high ecological validity • Engagement would be captured naturalistically and won't be construed due to students over or under estimating
	Cons	• Difficult to measure skills based on an observation; how would this be operationalized? • Lack of student and staff perceptions • May be difficult to compare across students • Time consuming and difficult for the type of project

Table 3.4: Suitability of Survey, Experimental and Observational Research for this Project.

3.4.1.1.1. Survey Research

Survey research includes the use of questionnaires and structured interviews in order to find numerical trends or patterns from a sample of participants which will then be intended to be generalised to the entire population, questionnaires usually consist of multiple choice questions and Likert scale questions in order to quantify a participant's attitude or performance (Creswell 2014). A common misconception of survey research is that it is an easy form of research, however, it can be difficult to design a well constructed survey. It can often be difficult to collect data on all the variables that the researcher is interested in. In survey research, research objectives are of particular importance as they can help to guide the researcher into creating the appropriate questions. In addition to this, survey research can often try and answer a hypothesis but can also simply be a descriptive survey (Muijs 2004). The advantages and disadvantages of each method of survey research has been summarised in Table 3.5.

	Pros	Cons
Pencil and Paper Questionnaire	• Familiarity with the method by the participants • Participants able to complete at own convenience • Allows participants time to think of responses	• Often low response rates • Can be time-consuming to follow up • Time consuming for data entry, especially with large number of participants
Telephone Interviews	• Allows researcher to continue until the target sample size has been met • Better suited to some random sampling methods • Allows for direct input into computer system; saving time • Can make the questionnaire adaptive based on responses	• Bias can occur as some potential participants will not be listed in directories • Participants may find them intrusive and refuse participation • Little time for respondents to think of answers • Participants could be unavailable to take calls
Face to Face Interviews	• Can be adaptive and able to change questions based on responses	• Can be intrusive and induce non-cooperation • Place of interview may introduce bias dependent on who is at the location at the specific time chosen • Data input is time consuming
Online/Email Questionnaires	• Answers can easily be transferred into a statistical analysis program • Can be made to be adaptive • Participants find it easy access to do from phones or computers • Participants are able to think of answers to questions • Participants can complete in their own time • Known for having higher response rates due to ease	• Potentially can induce bias if wanting to target an older population; may not have access to such technology

Table 3.5: Methods of Survey Research and Their Advantages and Disadvantages

When finding a sample for survey research, as is important for all research, one should not try to bias the sample towards a particular group but rather should have equal opportunities for all within the target population. One important thing to note with the creation of survey tools is that the data itself relies upon the instrument being of a high quality i.e. that it is measuring what it intends to measure. This suggests that further analysis would be needed on the specific reliability and validity of the instrument itself (Muijs 2004).

The advantage of survey research is that it is possible to study a wide range of research questions using survey research. Also, it is arguably not as artificial as an experimental design as it is describing a situation or studying variables either directly or by proxy. Therefore, with survey research having higher ecological validity it is arguably easier to generalise findings to real-life situations. Furthermore, the survey method is very efficient in gathering large amounts of data with minimal effort and cost, which is also beneficial to generalising findings due to having a larger sample, representative of a target population. This is beneficial for this particular research project as the researcher requires the collection of as much data as possible from this particular cohort of students. Additionally, it withholds the ethical consideration of anonymity as it is easy to keep participants' data completely confidential in the majority of the methods, face to face interviews can be more tricky but can still lend themselves to confidentiality. Finally, and perhaps the most important advantage in terms of this research, the survey method allows for easy comparability across participants, as previously mentioned, this holds importance as there is a need to be able to compare longitudinally and between participants on the skills that they have gained, this particular method would allow for this (Muijs 2004). However, there are disadvantages to this method, although it does have higher ecological validity, it lacks an ability for the researcher to control the environment which can often mean that the researcher cannot deduce cause and effect. However, longitudinal studies and appropriate statistical modelling can sometimes allow the data itself to reach a verdict on cause and effect without the need for control. As with most quantitative methods, there is also the limitation of a lack of depth to the phenomenon that is being studied, this is why survey methods are often used in conjunction with qualitative methods (Muijs 2004).

3.4.2. Qualitative

Qualitative research is often used to discover a participant's attitude, trends or opinions on a certain phenomenon. This form of methodology is heavily reliant on the participant to discuss their experiences in a way which helps the researcher to interpret outcomes in order to answer their research question. This method goes into incredible depth and detail with participants and often is used to discover the meaning and “why” behind phenomena; essentially to discover cause and effect (Jackson, Drummond, and Camara 2007). Shank (2002) defined qualitative research as “ a form of systematic empirical inquiry into meaning”. Ospina (2004) discussed this meaning as following a procedure which is grounded in “the world of experience” to try to understand “how others make sense of their own experiences”. Qualitative research often takes the inductive approach to research and seeks to form a theory based on the data, it also takes the more interpretivist philosophical approach to data meaning that researchers are expected to interpret the outcomes of the research, a contrasting method to that of quantitative research (Bryman 2012). This method is often used when there is a small number of hypotheses that the researcher wants to explore in depth; it is difficult to control for several hypotheses and therefore a quantitative method is more suitable within those circumstances (Panke 2018). Qualitative data can take two different forms; one of which involves one data collection technique, such as interviews, and the analysis of this, known as the mono method qualitative study. It can also take the form of several data collection techniques such as both interviews and focus groups with corresponding analysis, known as the multi-method qualitative study (Saunders, Lewis, and Thornhill 2016). Patton (2002) noted that a substantial difference between qualitative and quantitative methods can be seen within the sampling techniques. With the quantitative method the focus is on a representative sample, as previously discussed, however, with qualitative methods the focus is on selecting smaller samples with purpose.

Ospina (2004) describes several instances where one may decide to use qualitative research rather than quantitative. Firstly, it is often used when a new area of research is born i.e. an area which has not been studied before. Furthermore, if one aims to add richer detail to an area of existing knowledge which had previously been predominantly quantitative. And finally, when one has a need to “understand a phenomenon” from the perspective of the participants themselves without the influence of the researcher. There are several advantages to conducting qualitative research. This includes the ability to follow new ideas when they occur within the process as well as being able to explore new phenomena effectively, this is due to the fact that the collection of qualitative data is mainly non-procedural allowing for easy change whereas quantitative methods follow a much more standardised procedure (Saunders, Lewis, and Thornhill 2016). There is also an increased opportunity to develop new ideas and theories, as is evident from the overwhelming use of the inductive approach in qualitative research (Ospina 2004). However, this method is not without its limitations, it is a difficult method to replicate due to the subjectivity of the interpretation of the results. It also has a big problem with generalization suggesting that qualitative methods alone would not suit this particular research project due to the need to be able to compare and generalise to the student cohort.

3.4.2.1. Strategy

There are several strategies of inquiry that are used in qualitative research some of which are; ethnography, grounded theory, case study research, phenomenological research, narrative research and thematic analysis. These strategies are explored in Table 3.6 for their suitability for this particular research project (Creswell 2014; Bryman 2012).

		Suitability for this Research
Ethnography	Pros	• High ecological validity as requires a lot of observation; could mean higher accuracy to determine student engagement • Could be useful in validating quantitative data • Does often include interviews as well as observation for further clarification and individual views/opinions
	Cons	• Difficulty with observation as don't have access to all potential student places where engagement may be missed • Skill measure cannot be observed, or at least not with ease • Difficulty with comparability; how are we going to see the changes in the students over time, how would that translate into comparing students with text
Grounded Theory	Pros	• Can develop a theory of information with no prior background influence; purely focusing on the experiences of the students in that particular institution and of that particular cohort (in this case science students) • Easy to see the process behind the creation of a theory through notes and memos; in-depth qualitative method.
	Cons	• Difficult to not be influenced by prior knowledge • Already been pilot projects in this area which gives the researcher an idea of expectations; therefore may be difficult to remain unbiased • Time-consuming and may be something which researcher may not be able to do well; is more experience needed for this method.
Case Study Research	Pros	• Appropriate for collecting in depth views of an individual; could potentially be used to gain information from staff and students to align with the survey • Allows for flexibility to explore views more naturally; can edit and change questions and flow as interviews take place
	Cons	• Focuses on a very minute number of individuals due to the amount of depth that is needed; could potentially also be a pro as this means that you can get more information from those willing to engage however, lacks generalisability
Phenomenological Research	Pros	• Can give original experiences of student skills; how they perceive to have gained during their degree • Allows members of staff to explore their own expectations of skills without influence
	Cons	• Can make it difficult to compare across different interviews
Narrative Research	Pros	• Used often in educational research • Extensive procedures to follow; suggests some reliability in the method
	Cons	• Requires students to "tell a story" of their experience at the institution; requiring retrospective thinking which research shows students are often very poor at, especially over longer periods. • Relies heavily on a student's ability to give highly detailed information which can be difficult to ascertain from students, especially if not engaged in the research project

Table 3.6: Suitability of Qualitative Research Strategies for this Project

3.4.2.1.1. Case Study Research

Case study research is the in-depth data collection of a community, school, family, organization, event or individual through a variety of different methods over a period of time (Creswell 2014; Bryman 2012). Although this definition has been provided, there is actually much debate over what exactly a case study is. It is often used to describe research that does not appear to fit into other categories but is most often used when describing qualitative research that explores an area in-depth from different perspectives (Moriarty 2011). The case study approach is referred to as a "naturalistic" design, meaning that the researcher does not exert control or manipulation over the variables that are being studied (Crowe et al. 2011) Stake (1995) discussed three different types of case study; intrinsic, instrumental and collective. The intrinsic case study is the study of a case where the subject itself is the interest of the study. The instrumental case study is the study of a case to provide insight into an issue, redraw generalisations or built theory. The collective case study is the study of multiple cases simultaneously or sequentially to gain a broader insight into a particular phenomenon. According to Rebolj (2017), case studies can also be categorized by their "time dimension". This includes retrospective case studies which collect data on a phenomenon which has happened in the past. Snapshot studies; which examine one particular period of time such as an event or particular day. Diachronic studies; which are essentially studying changes over time and are a form of longitudinal study. Nested studies; data is compared across multiple cases in order to give a broader picture. Parallel studies; data is collected across multiple cases and studies simultaneously and finally, sequential studies; where the cases happen consecutively and are studied so.

Rebolj (2017) described case studies as being "strong precisely where quantitative studies are weaker". Case study research is high in conceptual validity; essentially meaning that when trying to measure variables which are difficult to measure, researchers must find an equivalent phenomenon that is expressed in a different way. As this requires contextual factors, this would therefore be difficult to achieve through quantitative methods, but is very common through case study research. In addition to this, case study research is highly suitable to derive new hypotheses due to the fact that it is exploring new areas and taking into account high complexity factors. Furthermore, as with most qualitative research, case studies are able to explore cause and effect, essentially delving into the reasons why behind a phenomenon and not simply describing the relationship. In addition to this, case studies will often use multiple data sources such as documentation, interviews and observation. This allows for high credibility in the data due to the different types and the potential validation from one source to another (Rebolj 2017; Baxter and Jack 2008; Yin 2002; Yazan 2015). However, limitations of the case study method lie in the reliability and validity of the method. Generalisability is a big problem when using case study research as the majority of case studies focus on one single case; there is therefore a worry as to how this can be representative, therefore some researchers believe case studies should only be used as an exploratory part of research and not the main research method. Furthermore, case studies can be time consuming as they often employ a longitudinal time frame, making it sometimes difficult to conduct. In addition to this, there are no standardised “rules” for a case study, therefore meaning that the method can differ greatly from one researcher to another (Bryman 2012; Moriarty 2011; Yin 2002). There is an aspect of case study research to this project due to the fact that the focus is on one single institution as well as a focused cohort of students. Although this would then argue that there are generalizability problems, there is still value to this method in that a measure is being created and validated that is specifically targeted at the science cohort of students within this particular institution. Due to the vast difference in courses across institutions, the skills which are tested and those which are found to change over time, may differ. Therefore, it would suggest that there is a case study aspect to this research.

3.4.2.1.2. Phenomenological Research

Phenomenological research is a strategy of inquiry where the researcher looks to understand human experiences from the view of the participants, without the influence of any of the researcher’s own experiences. This method of research is also considered to be a philosophical stance and therefore is often used in conjunction with other qualitative research methods. This method is conducted via studying a small number of participants through extensive engagement to develop relationships and patterns among the experiences (Creswell 2014). Pure phenomenological research is seen as a descriptive qualitative method rather than explanatory, it often overlaps with ethnographic research, however, it focuses on the experiences and perceptions of individuals from their very own perspectives (Lester 1999). Padilla-Diaz (2015) described different types of phenomenology. Firstly, descriptive or hermeneutical phenomenology which is the study of participants' own experiences and the interpretation of this. Secondly, eidetic or transcendental phenomenology which "analyses essences perceived by consciousness with regard to individual experiences". Finally, egological, genetic or constitutional phenomenology which is "the analysis of the self as a conscious entity". The most common method of data collection in phenomenology is that of the interview method, this allows for exploration of participants' experiences with the flexibility to expand upon the discussion. Furthermore, the analysis in phenomenology is that of the identification of common meanings within the data. The first stage of which is known as "horizontalization" meaning the researcher identifies a list of quotes that are of equal value to identify the relevant topics that the participants are communicating (Padilla-diaz 2015).

Phenomenological research allows for a systematic approach to highlight and explain experiences of individuals, to deeply explore these perspectives and experiences, which can be beneficial for certain research questions (Khan 2014; Lester 1999). However, there is yet again the concern of generalisability due to the small sample size of phenomenology. (Lester 1999). This method of research could arguably be an approach adopted by this project due to the idea of it having a more descriptive approach to it, this method could allow the researcher to describe the experiences of the students and staff in relation to the skills that are gained in order to validate the quantitative method.

3.4.3. Mixed Methods

Mixed methods research is seen as the combination of both qualitative and quantitative methods of collecting data (Halcomb and Hickman 2015). Bryman (2012) discussed seven different justifications for conducting mixed methods research:

- **Triangulation**; where data methods are combined in order to "triangulate findings".
- **Offset**; when the researcher uses the methodologies to offset the weaknesses of each other.
- **Completeness**; which refers to the belief that the use of both methods brings a more comprehensive account of the research area.
- **Process**; when quantitative research provides an account of the structures, qualitative research can indicate the processes behind this.
- **Research Questions**; the methods are often combined due to the need to answer several different research questions.
- **Explanation**; they may also be used to provide an explanation of the other.
- **Unexpected Results**; they may also be used in combination when one method produces surprising results where the other may be able to offer an explanation for this.

For the purpose of this research project, triangulation appears to be the most appropriate method and reasoning to employ a mixed method approach. Bryman (2012) described triangulation as "the results of an investigation employing a method associated with one research strategy are cross-checked against the results of using a method associated with the other research strategy". Triangulation is seen as a sensible approach to conducting research due to the fact that it is gathering information from different sources and methods which work effectively in conjunction with each other (Almalki 2016). Turner and Turner (2009) discussed four different kinds of triangulation, these included;

- Data triangulation; gathering data from multiple sources
- Methodological triangulation; gathering data from multiple methodologies

- Theory triangulation; Using multiple theoretical frameworks
- Investigator triangulation; Using several researchers to corroborate analysis, most commonly used in data analysis of qualitative methods.

Triangulation has been described to be used for "completeness" i.e. to gain an in-depth understanding of a phenomenon from several different methods. As well as for "confirmation" i.e. to validate and confirm the outcomes of one method with the outcomes of the other (Adami and Kiger 2005).

3.4.4. Chosen Strategy and Methodology

After reviewing all the potential strategies and methodologies that could be employed in research, the methodology and strategies were chosen with a justification as to why they are appropriate for this particular project.

A mixed method approach was chosen for this research project. The strategies have been explained below including the mixed method approach and which qualitative and quantitative strategies will be used. For this project, data and methodological triangulation will be used to help inform the outcomes of this research. Different data sources and multiple methods are used in order to collect information which includes; systematic literature review, semi-structured interviews, qualitative surveys and quantitative learning gain assessment tool. Initially the assessment tool was created and implemented. This was then followed by interviews with staff and distribution of a qualitative survey to students. Triangulation was used as a "confirmation" method in which the outcomes of the qualitative research were used to verify the outcomes of the quantitative method. Thematic analysis was conducted on the qualitative data in order to triangulate with the findings of the quantitative assessment tool. This allowed for potential weaknesses in the quantitative data to be corroborated and to inform the outcomes.

3.4.4.1. Case Study Research

Due to this project being conducted in one institution and targeted at one particular cohort of students; this would suggest that this project is a case study. This is due to the fact that the research is conducting in-depth data collection and analysis on one particular case. Specifically, this case study would be indicative of an instrumental case study due to the focus on a particular institution as well as the fundamental aspects of an instrumental case study being of that to use the focus to facilitate understanding of another phenomenon. In this case, the case study of looking into the institution of LJMU will help to facilitate the understanding of learning gain measures and to better further the knowledge of utilising learning gain assessment tools within Higher Education Institutions. In terms of the time dimension of this particular case study, as this is viewing the changes in student learning gain over a peiod of time, this would suggest that this particular case study belongs to that of a Diachronic study which studies changes longitudinally. In this project, the case study is being used to collect both quantitative and qualitative data and is therefore not limited to one particular method. This is beneficial due to the fact that previous research has suggested that as institutions differ so greatly, questionnaires should be specifically targeted to the institution's specific courses and the skills which they emphasise. Institutional Context and the Science Cohort.

As previously discussed, there is great difficulty surrounding the ability to create a generalisable learning gain assessment tool which can be used across institutions. This can be due to a number of factors such as different specialisms of universities; such as whether they are a medical school, well known for psychology etc., as well as the facilities, course details and emphasis. In addition to this, there have also been found to be differences across subjects in the way in which they respond to surveys, how they interpret questions and whether they are likely to engage. An example of this being that those studying engineering courses have been found to be the most reluctant to take part in research (Williams et al. 2007; Wiers-jenssen, Stensaker, and Grogaard 2010). Originally, the Liverpool Mechanics Institute, Liverpool John Moores University (LJMU) was founded in 1823 and currently offers over 250 degree courses and houses a community of 25,000 students. The University currently consists of five faculties; Faculty of Arts Professional and Social studies, Faculty of Health Education and Community, Faculty of Engineering and Technology, Faculty of Science and Liverpool Business School.Currently, for the academic year 2017/18 the cohort of LJMU students consists of 54.13% Female and 45.87% Male. Below is a breakdown of the entire STEM student cohort by the subject studied;

- 15.02% studying Biological Sciences
- 3.38% studying Physical sciences
- 1.05% studying Mathematical sciences
- 9.23% studying Engineering and Technology
- 15.82% studying subjects related to medicine

This has seen a rise in all of these subjects since 2014/15 academic year. Overall, as can be seen from the subject breakdown it appears that 44.5% of students study a STEM related subject. The second largest number of students appear to be studying biological sciences, with the largest being subjects related to medicine. This helps to give an overall view of the students in the targeted cohort. It would appear that those specifically studying a science related subject are 34.22% of students which is a large cohort of the university. This made for an appropriate target upon the decision to narrow the scope of the students. Although there may be problems in engaging this cohort, having a larger pool of students is beneficial to attempt to gain a larger snapshot rather than focusing solely on one particular science such as biological.

3.4.4.2. Survey Research

For this research, it appeared that survey research would be most appropriate, specifically the online/email method. This would allow for statistical analysis on numerical changes over a period of time as well as easy comparability across the cohort of students. The disadvantage of not being able to target certain populations did not apply as the target population is that of students who have regular access to emails and technology as part of their education. In addition, this method allows easy access for students suggesting they are more likely to engage as well as the potential to incorporate questions which elaborate on student experiences. This, therefore, is seen as an appropriate choice, however, purely using the quantitative approach is reductionist and can be difficult to validate whether the numerical outputs are indeed in line with the expectations and outcomes that the students and staff discuss.

3.4.4.2.1. Learning Gain Assessment Tool Design

The creation of a measurement which will be specifically tailored to the institution and the cohort of students was undertaken. This was due to the fact that there are many differences in skill emphasis, skills that are taught and expectations across both institutions and different subjects. The preliminary inquiry of STEM students informed the assessment tool of what skills needed to be aimed for, as well as a search of the literature, and also narrowed the scope down to focus on the science student cohort. It has previously been recognised that the students themselves are an “untapped resource” that should be used to inform the creation of such assessment tools (Kuh et al. 2015). Questions were divided into three categories of; Engagement, Critical Thinking and Communication. Wiers-Jenssen et al (2010) discussed the importance of including engagement measures when attempting to measure student learning and skills. This is due to the fact that this may provide context in the outcomes of the skill measurement part of the tool. This was also suggested to be an important factor in the preliminary inquiry interviews. The engagement section of the tool was informed by the already published and widely used NSS as well as the version used in the US; NSSE. The final two areas were informed from literature and the preliminary inquiry. Chirk (2006) discussed some important notes when designing questionnaires; the questions need to be in a clear and logical order, if of a multiple choice format there should be options for the participant to add their own thoughts should their option not be there, though this would not apply to a Likert scale. There is also a need to avoid double-barrelled and ambiguous questions to ensure you are getting the answers that you seek. In order to check for these kinds of problems a pilot of the assessment tool was utilised to be completed by students as well as experienced members of staff for comments. This allowed a check that the questions made sense to be undertaken and could also facilitate an initial reliability test using Cronbachs Alpha in SPSS. This pilot stage led to some changes in the assessment tool, which will be discussed in more depth in chapter 5.

3.4.4.3. Semi-structured Interviews

Semi-structured interviews are a popular method of constructing and conducting an interview for research purposes. The reason being that the researcher is able to construct questions prior to the session which they can ask of all interviewees, but can also prompt participants and encourage further discussion should the answer be too brief or off-topic (Mathers, Fox, and Hunn 2002). Alshenqeeti (2014) noted that participants may often not give a full answer to a question. Therefore, by taking this approach to the interviews it gives the researcher the flexibility to attempt to gain further insight into participants' answers but whilst also having a set standardised structure which will be consistent across all interviews. Interviews were used on students in the preliminary inquiry where students were asked openly to discuss what skills they thought they have gained during their degree. These interviews consisted of a small number of seven students from different STEM subjects in order to gain a snapshot into the types of skills that are discussed. This open question allowed for the answer not to be biased, but also allowed the researcher to prompt the participant should they appear to not give very in-depth answers. The questions that were asked of the members of staff were much more targeted to seek whether they aligned with the outcomes of the assessment tool. They asked about specific skills and the expectations that the staff have of these. Six members of staff from across different science subjects were spoken to.

3.4.4.4. Qualitative Survey

Due to the difficulty in encouraging students to engage with an interview unless they believe it to be beneficial to themselves, a qualitative survey approach was taken in order to validate the outcomes of the learning gain assessment tool. Students were asked the same questions as members of staff in the semi-structured interview however, due to the survey being conducted online it released the ability to be able to prompt students for richer and more detailed answers. This meant that some answers in this survey lacked the necessary depth needed to validate the outcomes, however this still provided beneficial information from across several different science related subjects.

3.4.5. Overview of Stages of the Research Project

Before delving into the more detailed information on the individual stages of the research project, it is useful to provide a visual representation of the stages to better understand the process that was undertaken. Before beginning the research, a literature review was conducted to be able to understand the current state of learning gain both nationally and internationally to be able to identify the gap in the knowledge. It was identified that there has been little developments into the survey research of a discipline-specific assessment tool. The focus of the current research within an English HE context appears to be on that of the outcomes of the learning gain measures rather than the development of a suitable assessment tool. Figure 3.7 provides a visual representation of all of the stages of the research project.

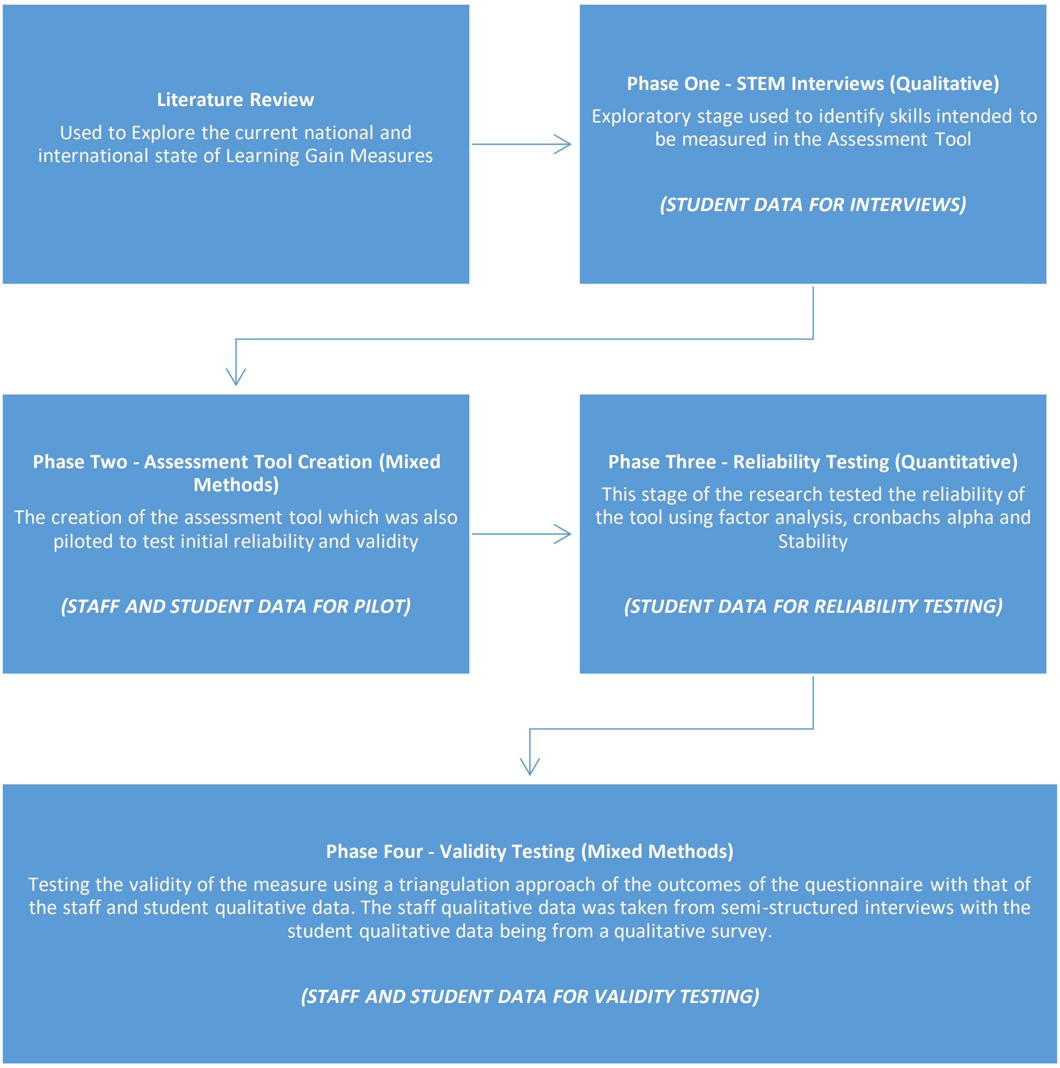

Figure 3.7: Overview of the Stages of the Research Project.

Phase one was an exploratory stage of the research whereby STEM students were interviewed using semi-structured interviews to be able to explore the skills that are believed to be critical to these subjects. Differences led to a more focused approach onto Science students. The following three stages of the research project involved the creation and testing of the assessment tool. The tool was tested for reliability by looking at the stability of the measure across three different groups of students as well as using factor analysis and cronbachs alpha. The validity of the tool was tested using a triangulation method with that of qualitative data from both staff and students.

3.5. Phase One Strategy and Methodology

In order to understand how specific the subject needed to be for the assessment tool creation as well as to inform the skills which will be measured; a preliminary inquiry stage informs the first phase of the research. This phase was undertaken using the qualitative approach by conducting semi-structured interviews on STEM students. This stage of the research is an exploratory stage in which the researcher intended to identify whether the STEM student cohort is a suitable target population for a subject specific learning gain assessment tool. This is a particularly important phase of the creation of the assessment tool as it determines a focus of the research.

3.5.1. Research Objectives for Phase One

This part of the research informs the overall research objective of “To identify the skills that are critical for students to achieve when studying a Science Undergraduate degree”. At this current stage, the researcher intended to focus on the entire STEM cohort as the target population for the assessment tool. This part of the research allowed the researcher to understand whether the focus needed to be narrowed. Objectives were identified in order to achieve the outcomes of this stage:

- Identify the skills that STEM students consider to be critical to success in their undergraduate degree.
- View the differences in the STEM student cohort on their skills and align with previous literature.
- Define and select the skills which will be utilised in the assessment tool in order to gain an insight into the specific subject’s learning gains.

The outcomes of this preliminary enquiry stage led to a narrowing of the research down to focusing solely on Science subjects rather than that of the entire STEM cohort. This was due to differences with the subjects that were identified through the interviews and literature.

3.5.2. Semi-Structured Interview

Due to the need to explore a phenomenon with STEM students and wanting to utilise their own opinions and knowledge, the semi-structured interview was chosen. Blandford (2013) stated that interviews are "best suited for understanding people's perceptions and experiences". This further supports the idea that this method is appropriate for the exploration of students' skills. Arthur and Nazroo (2003) discussed the importance of creating an "interview guide" or "interview schedule" when conducting semi-structured interviews. They discussed the need for constantly having the research purpose in mind to allow for asking questions which will give outcomes that are of interest. The "guide" should also include any prods or prompts that may have to be included should the participant not be giving the right information or going into the right amount of depth. This is of particular importance when it comes to interviewing students who have a tendency to answer the questions with minimal information due to believing that the information they may give is obvious or not of interest. This highlights the importance of preparing for methods to prompt students (Becker 1956).

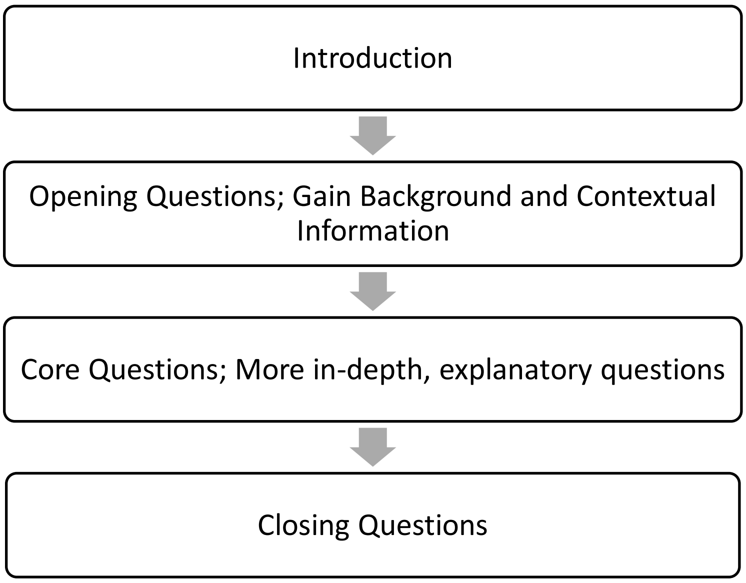

Figure 3.8: Stages of Discussion in Interviews to Help with Guide Creation (Arthur and Nazroo 2003)

Blandford (2013) suggested that a particularly useful technique for getting information from participants is to ask for examples, this can help with recall of information. Harrell and Bradley

(2009) discussed three different kinds of questions that can be used. This included descriptive questions, which are simply used to ask students to describe things, which can often help with areas to probe further. In addition, structural questions are used to help researchers determine relations. Finally, contrasting questions help with the understanding of meaning. Furthermore, there are things to avoid when designing questions in order to get the most beneficial outcomes from the participants (Harrell and Bradley 2009)., this includes;

- avoiding double-barrelled questions
- leading questions
- double negatives
- technical terms
- vague questions

Alshenqeeti (2014) suggested that the atmosphere is also of high importance when conducting interviews; this includes the environment and the interviewer themselves. The interviewee is more likely to open up when they feel more comfortable.

An advantage of this kind of interview is that of allowing the researcher to prepare the questions ahead of time, this will make sure that the questions are focused on the research topic. In addition to this, it allows for the participants to express their own views and leaves it open to their own opinions. This is of particular importance for this research project as the aim is to gain the skills that will be addressed in the tool, purely from the students themselves to make it specific to this cohort. Furthermore, due to the fact that every interview will have some structure, this can make for much easier comparability between them and therefore, makes the data much more reliable and valid as the same questions and prompts will occur, particularly with a well written interview guide (Cohen and Crabtree 2006). The structure that this kind of interview gives has also been used as a diagnostic tool due to the higher validity, suggesting that when utilised properly the semi-structured interview can be advantageous and as standardised as quantitative research (Segal et al. 2006).

3.5.3. Designing the Interview Schedule

As previously discussed, the interview schedule began with an introduction. This introduction explained the nature of the study to the participants; that the interviewer simply wanted the students' opinions on the skills that they believe they have gained during the course of their degree. It was explained to the participants that the data will be recorded but kept strictly confidential and they were told that there is no right or wrong answer. This part of the interview schedule is of particular importance as it gives the students expectations and an understanding of the purpose of the interview. Even though participants are given an information sheet and consent form, it is still beneficial to repeat these important messages (Arthur and Nazroo 2003). Leech (2002) discussed the importance of question order. It was suggested that "easy" questions should be put at the start of the interview, ones which respondents will be able to answer without much thought in order to gain a rapport with the interviewer. With semi-structured interviews, there has been debate as to whether one should ask contextual/personal questions first or last. Leech (2002) believes that personal questions should be asked last because if the personal questions are asked first, it gives the impression that the interview is about the personal response of the participant. However, they did also state that should the question be about the participant's own personal opinion or beliefs, then this would be a good place to begin. As this was the case for this research project, this was chosen as the starting point.

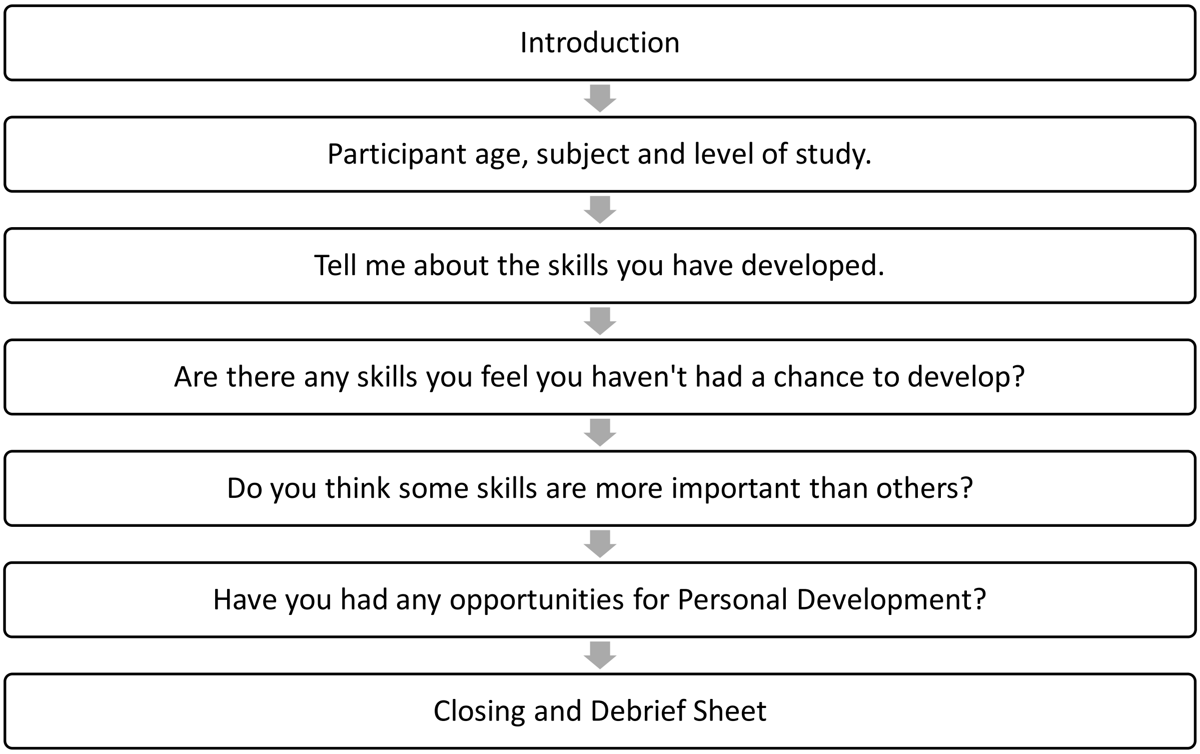

Figure 3.9: Flowchart of Main Questions of Interview with STEM Subjects

Questions were left open to participants, they were asked in a way so as not to bias the answers to participants but to gain the information necessary. Prompts were also used in the interview schedule which would help participants to further develop their answer. This could be by asking participants for an example of when they used a particular skill, asking participants about their particular exams or assignments in order to gauge skills they have learnt as well as about their subject itself and the knowledge they have gained, these are referred to as “example questions” (Leech 2002; Spradley 1979). Personal development was asked as a separate question due to the potential that participants may focus on skills that are purely academic. Should participants have mentioned personal development during their answer to the first question, this was then not omitted. If participants were struggling to answer a question, or were not giving sufficient detail, the Spradley (1979) “grand tour question” technique was adopted. This is essential when a participant is asked to describe something which is incredibly familiar to them, in this case it would be to describe a typical day or a typical assignment for their course. This then can lead to prompting questions to ask students about the skills they need in order to complete the said tasks and assignments. Leech (2002) described this kind of question as the “single best question I know of for a semi structured interview”. See Appendix 1 and 2 for Participant Information Sheet, Consent Form, Debriefing Form and Example Interview Transcript.

3.5.4. Interview Procedure

Participants were invited to interview in private dedicated rooms on the premises of the university. Prior to the start of the interview, participants were given an information sheet which they were encouraged to read as well as a consent form they were required to sign which informed the participants that the interviews would be recorded and transcribed. Interviews were recorded using a Dictaphone and averaged 30 minutes in length. A semi-structured interview schedule was used by the researcher that contained a list of questions as well as prompts, as previously discussed, the researcher was able to adapt this schedule when necessary and some prompts were added to use in later interviews. All interviews began with the question of “Can you start by describing some of the skills you have developed or expect to develop during your undergraduate degree” this was used so as not to bias participants to discuss specific skills and left it open to the participants to lead the interview. Questions were also asked on participants’ personal development as well as their independent learning and engagement with their degree. Interviews were led by the participants’ responses and therefore differed in the order and phrasing of questions. Additional questions were also asked of participants should they have not given sufficient information or depth to their answer. Recurring themes were identified in the analysis process and whilst there was overlap in these themes, the importance of each of these themes differed from each participant and even more so across different subjects.

3.5.5. Analytical Strategy

A thematic analysis was deemed to be the most appropriate for the qualitative data. The same day that the interviews had been conducted, they were transcribed, which had the advantage of the researcher being close to the data and able to consider initial analysis of the themes that were beginning to emerge. The same interviewer was present for all interviews allowing for the questions from previous interviews to be taken in the same direction. Braun and Clarke(2008) describe six phases of thematic analysis, these stages were followed in this phase of the data analysis. Firstly, transcripts were printed and read to get a sense of each interview as a whole and to have the researcher familiar with the themes that were being discussed in each interview, this allowed the research to begin to understand the emerging themes. Notes were taken at this stage of the research of skills that the participants were mentioning. Following this, key words and phrases were highlighted using Nvivo i.e. any information that students mentioned about their learning outcomes that they had specifically gained from their degree, any other areas of interest were also highlighted, such as aspects that students believed affected their ability to gain from their degree. This allowed for the generation of initial codes; areas began to emerge such as "analysis". The next stage involved grouping these initial codes into themes; these themes were the basis of the skills that would be used in the assessment tool. These emerged such as "Critical Thinking and Problem Solving" as there was some overlap in these two areas. Furthermore, the themes were then reviewed to check whether they were working with the data, this also resulted in a thematic "map" which can be seen in the results section. Finally, the themes were defined as to what should be included within them, this is also detailed in the results section. As these themes emerged from the data, it therefore reflected the view of the participants and the "student voice" making this preliminary inquiry a more distinctive approach. See Appendix 3 for a Coding Report for STEM Interviews.

3.6. Phase Two Strategy and Methodology

This stage of the research has been informed by the preliminary inquiry stage which looked into the skills perceived to have been gained by STEM students at LJMU. The questions that this stage of the research intends to answer is whether the created assessment tool appears to have statistical reliability and face validity in relation to science students. The overall research question that this tool intends to answer is whether science students gain skills over their degree. This is broken down into validating and testing the reliability of the assessment tool for this specific cohort as well as following the students individually over time.

3.6.1. Research Objectives for Phase Two

This stage of the research meets the fourth and fifth objective of the research project of "to explore and review current measures and literature to define and operationalize the skills which have been identified" as well as, "to develop and validate an assessment tool that is able to measure student learning gain within a Science student context", specifically focused on the development aspect of the final objective. Breaking down Phase two into the following objectives allows a more focused approach;

- Design an assessment tool that is tailored to science student skills in order to be able to measure a change in these skills over time.
- Conduct a preliminary assessment of the assessment tool through a small pilot study to assess the reliability of the questions using Cronbach's Alpha coefficient statistic.
- Test the validity of the tool through the opinion of experienced researchers in the field of questionnaire creation in order to check for face validity.

It is expected that the pilot stage of the assessment tool will trigger some potential changes to the instrument before continuing onto phase three and four. The following stages will test the tool for reliability and validity on a much larger cohort of students, as well as testing the instrument for the intended use of following students over a period of time. This will be discussed further in chapter 6 and 7.

3.6.2. Considerations for Design of the Assessment Tool

Mujis (2004) discussed that designing a questionnaire has similar stages to those of an experimental quantitative design but with some slight differences in the way in which the stages are undertaken. Seven stages were identified;

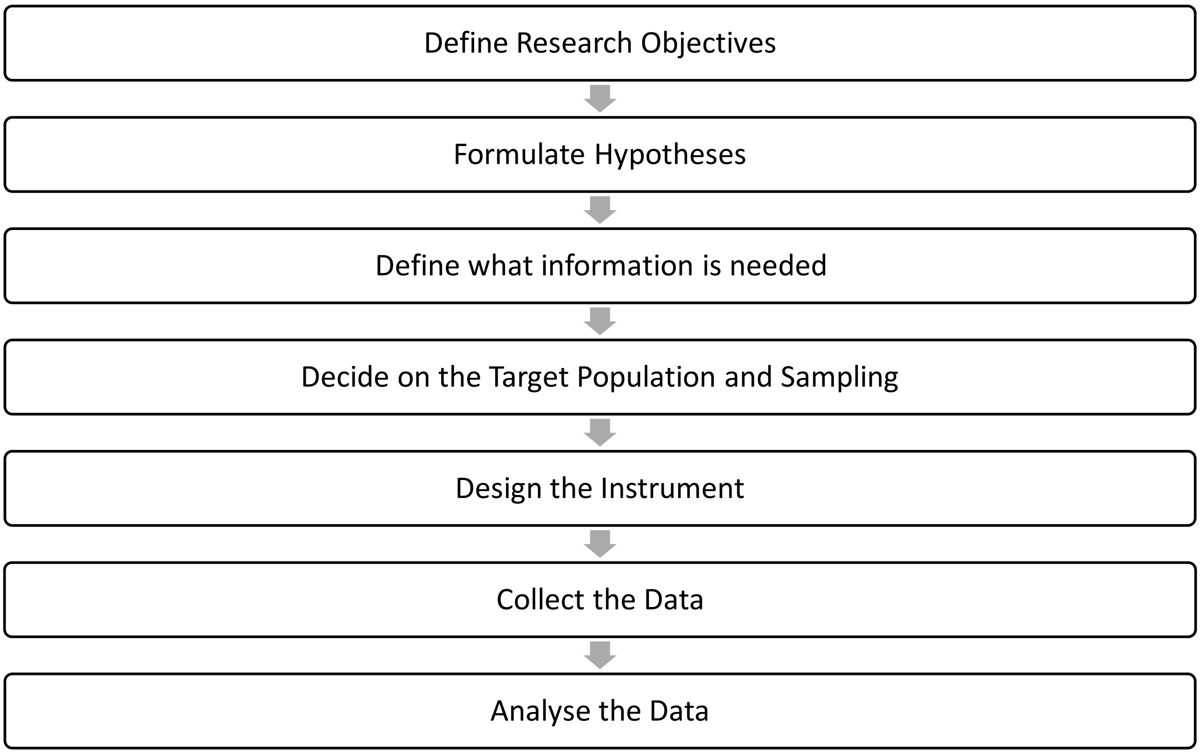

Figure 3.10: Stages for Designing and Implementing Survey Research

Rowley (2011) discussed the importance of considering different types of questions and how they fit into answering the research questions. Generally, types of questions are broadly categorised as open or closed questions. Meaning whether the participant is able to freely respond (open) or is faced with a multiple-choice question (closed). Adams and Cox (2008) discussed four main types of questions:

- Factual questions requiring a yes/no response
- Complex factual questions requiring some interpretation or analysis
- Opinion and Attitudinal questions requiring deeper concentration
- Open Ended Questions requiring participants full concentration

However, closed questions can take many different forms. This can include a Likert scale question which asks participants to what extent they agree with a statement either on a five point or seven point scale. As well as a ranking question which may ask participants to rank statements on a scale of most to least important to them (Rowley 2011). Closed questions are quick for participants to answer which may lead to a higher response rate as well as being easier to analyse using statistical analysis. Open questions allow for more in-depth insight into participants' opinions but can often be difficult to analyse, especially with an issue of comparability (Rowley 2011; Adams and Cox 2008; Muijs 2004).

Furthermore, there is importance with the length of the questionnaire as well as the structure. Having too long a questionnaire can decrease response rate or make it more likely that participants will not finish the entirety of the questionnaire, having too short will not gather the necessary information. Closed questions can also help with the length and time it takes to complete a questionnaire as it allows for a quick response. In addition to this, wording of questions is also considered to be an important consideration. Being careful not to ask a question where some participants may not understand the wording (particularly being wary of those where English is not a first language) as well as not asking too many questions within one, causing confusion as to which is intended to be answered (Adams and Cox 2008). Supporting this, Oppenheim (1972) stated to ensure that all questions are worded in a neutral tone so not to ask a leading question which may pressure some participants into believing that they should answer in a specific way such as "Are you against giving too much power to trade unions?"; as this may make participants believe that the researchers stance is that of against, which may make them more likely to answer in such a way even if it does not reflect their own beliefs.

As well as question structure and wording, the researcher should be aware of potential errors that may occur. Oppenheim (1972) noted some potential reasons for sampling errors that it is important to be aware of in an attempt to control for such things;

- Faults in the survey design
- Errors due to non-response
- Bias in questionnaire design or wording
- Unreliability or lack of validity
- Respondent misunderstanding

Supporting this, Gillham (2004) noted similar errors to this with the addition of problems with motivating respondents to participate unless they see a personal benefit, difficulty in checking validity of responses as well as participants finding it easier to talk than they do to write. In order to control for such errors, often a pilot project is conducted with the intended questionnaire. This can allow the researcher to eradicate these potential errors as they are able to catch them prior to sending the questionnaire out for the main experiment. This is of particular importance in the cases of reliability and validity. In terms of non-response, this may be particularly challenging when following students on an individual level, though this can be improved with simple reminder emails. Furthermore, the topic of this particular questionnaire may be more motivating to students as they may seek to view the outcomes of the questionnaire as it is following student progress over time, something which they may potentially be interested to see. In addition, the problem of writing over talking is less of a potential error in this case due to the questionnaire taking a multiple choice Likert scale approach. All of these errors can be easily controlled and taken into account through the implementation of a pilot project and awareness of these errors.

3.6.3. Target Population

The intended target population for this assessment tool is that of undergraduate students that are studying a science-based subject at LJMU. Undergraduates were chosen as the intended targets due to the vast differences between undergraduate and postgraduate degrees in terms of teaching and expectations. Furthermore, science students were chosen due to the preliminary inquiry. Originally, the intended target was that of STEM students, however, the preliminary inquiry highlighted vast differences across the cohort in terms of the way in which skills are taught and therefore, the scope was narrowed to that of science students. These students were chosen over the rest of STEM due to the targeted institution. LJMU has a large cohort of science students (34.22%), therefore it would be beneficial to focus the scope to that of a large discipline area in the institution. As previously discussed, this institution was of particular focus and therefore this project is considered as a case study as it would be unclear whether students from other institutions would report the same skills that are developed. Therefore, this tool is targeted specifically at one discipline in one specific institution. This would therefore be useful for comparability for internal use, but perhaps would not be useful for comparability across other institutions unless the skills were found to be of a similar nature.

As the science students were chosen to be the target population, it was not only important for the skills to be discipline specific but also for the questions that were asked. As previously discussed, science students in this particular institution, discuss critical thinking as an evaluative process. This would suggest that the questions for this part of the assessment tool need to be worded in such a way that probes into their use of evaluation. Furthermore, science students discussed their personal development more in terms of developing communication and presentation skills rather than group work. Therefore, it would be appropriate to discuss working with others in ways that do not just include group work. The wording and specific nature of the tool is what makes this project a novelty. There has yet to be a creation of a learning gain assessment tool that is this focused in its approach, therefore it is important to ensure that the questions are doing what is intended. See Appendix 4 for the Questionnaire in the Online Tool as well as Appendix 5 and 6 for Example comments from the pilot of the questionnaire.

3.7. Phase Three Research Approach and Strategy

This stage of the project involves reliability testing of the questionnaire. This is the intention of checking the stability of the questionnaire by testing the instrument on different groups at different periods of time. There is an expectation that there may be a slight change in the skills when testing over time as there should be an increase in skills. However, there should be a stability in the spread of the data and potentially the means. In addition to this, statistical tests were also run to check the reliability of the questionnaire. The tests used were that of cronbachs alpha as well as confirmatory factor analysis.

3.7.1. Research Objectives for Phase Three

The final two stages of the research project meet that of the final objective of the research project. Splitting this objective into small stages for this phase, the intended objectives of this stage of the research are;

- Conduct descriptive statistics on the three samples in order to compare means and standard deviations.
- Conduct a Cronbach's Alpha statistics to test the internal consistency of the measure
- Conduct Confirmatory Factor Analysis to test the fit of the questions with the different subscales.

Reliability is important for questionnaire creation as it shows that the questionnaire is consistent in its measurements. The next stage of the research will look into the validity of the questionnaire, perhaps the most important stage, as it checks that the questionnaire is measuring that which it claims to measure.

3.7.2. Reliability

Reliability is considered as the consistency and repeatability of the research. Measures should be able to discriminate between different individuals who perform on different occasions between different conditions. Alternatively, reliability could also be considered as the stability of the measurement over a course of time (Drost 2004). One of the most common measures of reliability is a statistical test known as the reliability coefficient. However, there are many other methods of testing reliability such as; test-retest reliability, alternative forms, split-half approach, interrater reliability and internal consistency (Rosenthal and Rosnow 2008).

An example of test-retest reliability is to administer the test or research to participants, to then have them repeat the same test at a later date in the hope to gain similar results. However, there are some limitations of this as leaving too short a time between the tests can mean participants may remember the first test and have what is known as "practice effects". Alternatively, too long between the tests can mean that there will be changes due to maturation of participants which may lead to different responses. For example, if the test was looking at participants' writing skills and they were currently in education; you would expect this to improve over time and therefore difficult to tell whether this is in fact reliable (Rosenthal and Rosnow 2008). Internal consistency refers to whether the scales are consistent across participants. This can be easily tested through statistical analysis with the Cronbachs Alpha statistic in SPSS, mentioned previously as the reliability coefficient. In order for the measure to be considered "high" value the statistic should be 0.7 or above. This method of statistical analysis will also show the statistic with the removal of some questions, as it may be one question which is dragging the internal consistency down. This easily allows the researcher to provide evidence for this type of reliability (Bryman 2012).

Generally, as reliability is concerned with replicability, the use of a survey quantitative method holds this general assumption, as the same instrument can be used in different situations, this can be tested by simply repeating the instrument on different groups of people, this will verify the test-retest reliability of the instrument. This was achieved by sending the quantitative survey out to several student levels at different time periods. In addition to this, reliability is also concerned with stability and this could generally be tested by looking at the outcomes of the instrument over time and whether the averages stay the same. Furthermore, Cronbachs Alpha will be run both at the pilot stage of the assessment tool and on both stages that the assessment tool is sent out to further validate the reliability of the tool. However, it would be difficult to tell whether the outcomes of this instrument are validly aligning with reality.

3.7.3. Procedure

The design of the questionnaire was that of a multiple choice Likert scale format. This structure allows for quantitative outcomes and therefore the ability to conduct statistical analysis. The phase of the research intended to look at reliability and stability of the data. This was achieved through looking at the data on three different samples from the target population.

Emails were sent out to undergraduate science students at LJMU to ask them to take part in the questionnaire. This questionnaire was sent out at three different time points across the year, students who had previously participated were asked not to participate again in order to get three different groups of participants. The questionnaire began with the participant information sheet and consent form in which participants were asked to agree to participate. The questionnaire consisted of 61 items of which 5 items were demographic items, chapter five outlines the creation of the questionnaire. Following this, a debriefing page thanked participants for participation and outlined the objectives of the research.

3.8. Phase Four Research Approach and Strategy

This stage of the research is the final and perhaps most important stage. This phase intends to look at the validity of the questionnaire by triangulating individual quantitative outcomes with qualitative data from staff and students. This triangulation will help to validate the outcomes and determine whether the questionnaire appears to be measuring the intended scales correctly. In the quantitative stage of the data, the data will be explored in order to view the outcomes and to see whether distinct differences can be seen. This will then be compared with the qualitative data.

3.8.1. Research Objectives for Phase Four

As previously discussed, the final two phases of the research project come from meeting the final aim of research, particularly in implementing and checking the validity and reliability of the questionnaire. This stage focuses on the validity and was broken down into the following objectives;

- Distribute the questionnaire to a sample of students who will be followed individually over a period of a year in order to identify learning gains.
- Analyse data from the questionnaire to give an overall picture of the outcomes of the questionnaire which can be looked at overall and by each level. This can also be compared across different subjects.
- Interview members of staff to understand their expectations and experiences with students on these constructs.
- Analyse the interview data in order to triangulate the outcomes with the quantitative data.
- Distribute a qualitative survey to students to understand their perspective of what is expected of them and their experiences of their course.
- Analyse the survey data to also triangulate with the quantitative data.

This final phase of the data will indicate whether the questionnaire is both reliable and valid. Phase three has indicated the questionnaire to have high reliability, however the validity of the questionnaire needs to also be confirmed.

3.8.2. Validity

Validity is defined as how accurate the research/measure is to gauge a concept and its alignment with reality. As with reliability, there are also several types of validity such as; face validity, concurrent validity, predictive validity, construct validity and convergent validity. Face validity is the minimum expectation of any research project, whereby it tests that on "the face of it" the measure appears to be accurately measuring the concept. This is often done through the simple task of asking others, particularly well-established experts, whether they believe this to be true. Convergent validity is considered to be gauging the validity of a measure by seeing whether the same outcome occurs from a different method. This can be done by comparing an assessment tool with the outcomes of observation, discussions or interviews.

The problem of being unknown if the research is aligning with reality that comes from testing reliability is reduced by the introduction of qualitative research. This will show what is expected of the students from staff, what the students believe they have improved upon themselves and whether these outcomes align well enough with the outcomes of the quantitative survey. Face validity was considered for this project as the pilot stage of the assessment tool involved asking those who are experienced in questionnaire creation, whether the assessment tool appeared to be asking the questions in a way that reflected the concept. Finally, convergent validity is also considered here as the outcomes of the quantitative measure are not considered alone but rather are to be compared with those of qualitative methods. This will help to validate the tool itself and to see whether the outcomes are accurate when compared to the reality.

3.8.3. Implementation of the Assessment Tool

The questionnaire was in a Likert scale format to be able to achieve quantitative data outcomes. This stage of the project followed individual data over the period of a year. This stage of the research intended to test whether the questionnaire is able to pick up changes in individuals' learning. Not only this, but it tests that the questionnaire is able to see differences between the student levels. This quantitative data will also be used to be triangulated with the qualitative data. This is a particularly important stage of the research as it validates the questionnaire.

Emails were sent out to undergraduate science students asking them to participate in the questionnaire. At the start of the questionnaire, participants were asked to give consent to the questionnaire via ticking boxes. The questionnaire consisted of 61 items of which 5 items were demographic items, chapter five outlines the creation of the questionnaire. Participants were also given a debriefing page after completing the questionnaire. At the very end of the questionnaire participants were asked whether they were comfortable with being followed over the rest of the year by repeating the completion of the questionnaire two more times; halfway through the year and at the end of the year. If participants were happy to do this, they were asked to leave their email in order to be followed up. These emails were kept in Bristol Online Surveys, which is password protected, and only used in order to distribute the questionnaire. Participants were given 4 weeks to complete the questionnaire at each distribution stage. In the second and third stage, participants were reminded four times to complete the questionnaire using the built-in reminder tool in Bristol Online Surveys. They were reminded once a week over the four weeks, in order to try and reduce participant drop off over the final two stages. See Appendix 7 for an example of raw data.

3.8.4. Staff Qualitative Data

Members of staff from the faculty of science were contacted for interview. This was decided on to discuss the expectations and experiences that staff had of their students. This will help to validate the outcomes of the quantitative data as the qualitative data can be triangulated with the quantitative data. Although in the previous section, the subscales were discussed as a total, in order to triangulate findings, quantitative data from individual questions may be considered in order to get a better picture of whether the outcomes are validated.

Participants were six members of staff from the faculty of science in LJMU. An email was sent out to all the faculty staff members asking them to participate in a short interview where they were to discuss the expectations and experiences that they had of their students' learning and skill development. This research adopted a theoretical sampling approach, participants were chosen as they were staff of the target population of the students that were of interest and therefore had the ability to answer the research question. A reminder email was sent out twice, in order to try and gain a larger sample size. However, due to the busy schedules of members of staff and a lack of incentive it was difficult to gain a larger sample. Although, as this part of the research is a triangulation and confirmatory basis of the data, this smaller sample may still provide some insights into the validity of the quantitative questionnaire.

Participants were informed that the interviews were going to be surrounding their expectations and experiences with students regarding their skills and learning gain during their time at the university. A semi-structured format was taken in which a set of questions were taken into the interview but the researcher was free to give additional prompts and the participant was able to steer the direction of the interview. They were informed that the session would be recorded and transcribed but all names would be anonymised and would not be identifiable.

3.8.4.1. Interview Procedure

Members of staff were interviewed at their own convenience in a quiet room, usually their own office, at the university. Prior to starting the interview, participants were given an information sheet and consent form which they were asked to carefully consider and to sign the consent form if they were happy to proceed. The semi-structured interview took on average 25 minutes in length. The semi-structured interview was split into sections of; general discussion, critical thinking, engagement and personal development. This allowed for some unbiased discussion at the beginning of the interview to then be brought into more specific questions relating to the questionnaire. All participants started with the question "What skills would you say students have developed this year?". There was room to prompt the participant into discussing the differences between the year groups if they did not do this automatically. When asking about critical thinking, questions were first asked indirectly about the construct such as asking about whether research had been conducted, this was to get into the flow that would lead to the discussion of critical thinking, whereby participants were asked "Has any work been done on critical thinking?" For the engagement aspect, participants were asked questions about how attendance changes throughout the year, whether students get involved in their lessons etc. Finally, with personal development, participants were asked about student group work, presentations etc. These questions were directly related to the questions that were asked in the quantitative questionnaire in order to see if the outcomes aligned together. The interviews were led by the participants and the order in which the questions were asked may differ depending on what was being discussed. Additional questions and prompts may also occur to gain further information when needed. See Appendix 8 for the Interview Schedule for Staff Interviews.

3.8.4.2. Analytical Strategy

As the outcomes of this qualitative data are to be utilised to triangulate with the outcomes of the quantitative study, thematic analysis was deemed to be most appropriate. This analysis looks for themes within the data, which is needed in order to be able to triangulate the findings. Upon completion of the interview, they were transcribed verbatim on the same day. Once all interviews were completed, each interview was printed to look for initial themes. These interviews were read over and highlighted with some initial notes to understand what potential themes may emerge, any areas of interest which showed information about what students were learning during their degree were highlighted as well as any expectations that members of staff had of the students. Upon familiarisation with the data, the transcripts were coded using Nvivo. Firstly, the transcripts were combed for initial themes; these began to emerge such as “independent learning”. Upon gaining the initial themes, these were then split into main themes such as “Engagement”, these themes emerged and were expected due to the focus of the questions. The main themes were expected to be related to the outcomes of the questions. These themes were then reviewed to check that they have come from the data and not had a biased view due to the need to validate the outcomes of the questionnaire. These stages of thematic analysis align with Braun and Clarke (2008) and the stages of thematic analysis that they proposed. See Appendix 9 for an Example of an Interview Transcript from Staff.

3.8.5. Student Qualitative Data

In order to further validate and triangulate the quantitative outcomes of the questionnaire, hearing from the students themselves would be valuable. Originally, interviews were the chosen method for this part of the research. However, students did not respond to requests to participate in such a way. Therefore, the questions which were originally going to be asked in interviews were transferred into a qualitative survey in which students were given open questions about their current experiences on their degree course.

3.8.5.1. Procedure

Participants were recruited via email using the university systems. They were asked whether they would take 30 minutes of their time to participate in a qualitative survey which asks students about their course and the experiences and expectations that the students had thus so far. The approach to participant selection was a theoretical sampling approach as subjects were asked to only participate if they were studying a science-based subject at LJMU. The recruitment email was sent out twice in order to gain a larger sample of participants and a wider representation from across the faculty of science. Participation was stopped as the outcomes of the questionnaire began to overlap and therefore it was decided that no further participants were needed in order to gain an insight into the student experience. Students were asked to provide as much detail as they could for each of the questions, this was to attempt to gain rich qualitative data rather than students writing minimal information. Before beginning the survey, students were presented with an information sheet and asked to complete an online consent form which consisted of ticking boxes. Upon completion of the survey, a debriefing sheet appeared. Participants were assured that they would remain anonymous and no data could be linked to them individually. See Appendix 10 for an example of survey data from students.

3.9. Data Analysis

As this research project adopted a mixed method approach, considerations were taken as to the data analysis which needed to be conducted dependent on whether the outcomes were quantitative or qualitative in nature. The programmes that were to be used in aiding the data analysis were also considered. Table 3.7 shows an overview of the approaches that were used at each phase of the research project. This section will then discuss the approaches taken to data analysis and the programmes used.

Phase	Approach	Strategy	Analysis
Phase one	Qualitative	Interviews	SPSS was used to analyse the descriptive statistics of the participant demographics. Thematic Analysis was utilised using Nvivo to analyse the interviews.
Phase Two	Mixed	Questionnaire (Assessment Tool) and Qualitative Survey.	Quantitative data was analysed in the pilot stage of the project. This included Descriptive statistics of the student and staff participats of the pilot project as well as using Cronbachs Alpha reliability statistic in SPSS. Qualitative data was also used in this stage. Staff and Students comments were analysed using thematic analysis in Nvivo.
Phase Three	Quantitative	Questionnaire (Assessment Tool)	SPSS and an extension of this programme known as AMOS was used to analyse Quantitative data. SPSS was used to analyse the participant demographics and the outcomes of the questionnaire to check for stability of the means. AMOS was used to conduct the factor analysis on the questionnaire.
Phase four	Mixed	Questionnaire (Assessment Tool), Staff Interviews, Student Qualitative Survey.	The outcomes of the questionnaire were analysed using SPSS as were the participant demographics for the questionnaire outcomes, Staff Interviews and Student Qualitative survey. In this phase of the research both descriptive and inferential statistics were used. The staff

			interviews and student qualitative survey were analysed using Thematic analysis in Nvivo.

Table 3.7: Overview of Approaches and Analysis used at Each Phase of the Research.

3.9.1. Quantitative Analysis

Both descriptive and inferential statistics were explored in the quantitative analysis. Descriptive statistics simply describe the characteristics of the data by summarising the distributions and patterns. The most common methods used in descriptive statistics is that of the mean and the standard deviation. These allow to see an overview of the averages of the scores as well as the spread (Field 2017). Inferential statistics allow the research to draw inferences from the data by completing bivariate or multivariate analysis. The general level of statistical significance which is most commonly used in research is that of 0.05, which will also be adopted within this research project (Bryman 2012).

There are, however, many different tests of statistical significance that can be conducted. In order to identify which test to utilise, a test of normality must first be conducted. The most common test or normality is that of the Kolmogorov-Smimov test. If the data is identified to be normally distributed, parametric tests can be conducted. If the data is not normally distributed then non-parametric tests will be conducted (Bryman 2012). Factor analysis is another statistical test which is intended to be utilised in order to confirm whether the questions in each of the individual subscales belong to that of which they have been assigned to. This is known as confirmatory factor analysis as opposed to utilising exploratory factor analysis which will identify which questions should be grouped together (Field 2017). Factor analysis also comes with it’s own test of suitability known as Bartlett’s test of Sphericity and the Kaiser-Meyer-Okin measure.

3.9.1.1. SPSS for Quantitative Analysis

The most widely used computer software for analysing quantitative data, particularly for research purposes is that of the Statistical Package for the Social Sciences (SPSS). There have been many different versions of this software as well as books on the utilisation of it

(Bryman 2012). SPSS can be used both for descriptive statistical analysis, inferential statistical analysis and significance testing. This software is essential in the analysis of quantitative data as it is able to perform a vast multitude of functions which the researcher can use to statistically analyse the data. This will primarily be used on the outcomes of the assessment tool which consists of multiple choice questions to allow the answers to be given a numerical value. This information can be analysed descriptively in order to give information such as student demographics, averages and standard deviations for each subsection as well as descriptive differences across the different subjects from within science. This information can also be tested for statistical significance. Statistical significance was run on both the repeated measures of the assessment tool and the following of individual students. This will allow the researcher to see whether the measure is both stable (reliable) and valid in terms of statistical significance. See Appendix 7 for an example of raw data.

3.9.2. Qualitative Data Analysis

Thematic analysis will be utilised within this project as it appears to be the best approach for a project which requires some form of interpretation of the data. This will allow for easy comparability between the qualitative data and the quantitative data due to the fact that the data will be split into themes and therefore will be much simpler to compare the results. Interviews and qualitative surveys will be used to achieve this with both staff and students.

Sometimes also known as the “inductive analysis” this approach to qualitative research is one which simply conducts interviews, focus groups, or whatever form of data collection they see fit; to then find categories or themes from within the data. This often leads to the creation of a model or a visualisation of the outcomes from the analysis (Thomas 2006). There is a debate within thematic analysis as to whether “themes” are seen as the same as “codes” and has previously not been considered to be a separate method of analysis but rather a method employed within other qualitative approaches, such as grounded theory. Thematic analysis is seen as a process in which the researcher reads over the transcripts to identify “higher order themes” which are then divided into “subthemes” or “lower order themes”(Alhojailan 2012).

Bryman (2012) identified a theme to be “a category identified by the analyst through the data that relates to the research focus”. There can often be confusion as to how to go about identifying themes. Ryan and Bernard (2003) suggested looking for occurrences such as repetitions, metaphors and analogies, similarities and differences and transitions to difference topics.

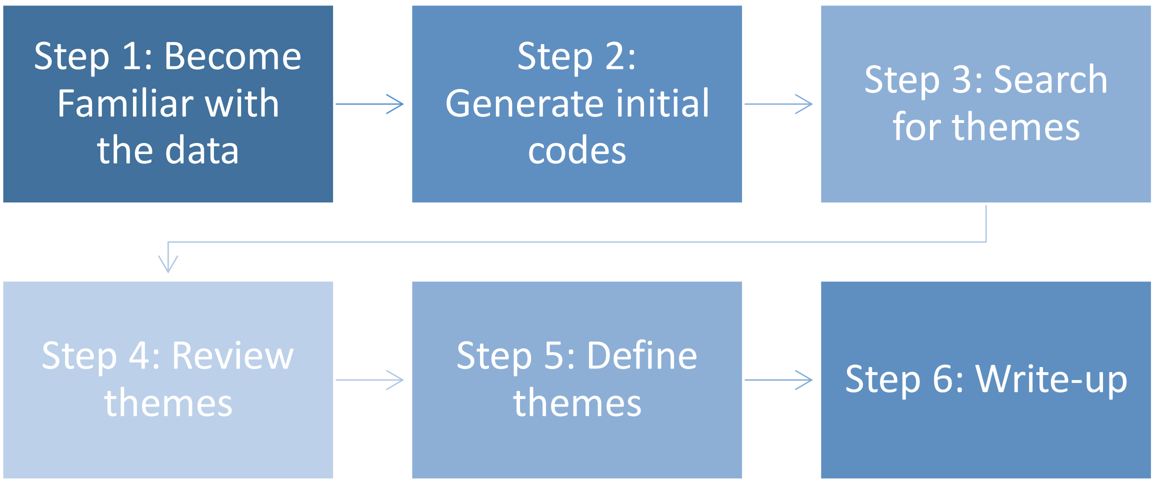

Figure 3.11: Six Phase Framework for doing Thematic Analysis (Braun et al. 2008)

Braun et al (2008) discussed six stages that form a framework for conducting thematic analysis. The first stage is that of familiarisation; which is described as the “immersion” of the data, usually achieved through repetition of reading through the data. The emphasis here being that the researcher may already have some initial ideas and thoughts on the potential codes and themes prior to actually conducting the analysis. Some researchers may also use this time to write memos or notes with their initial ideas. Secondly, the researcher must then generate initial codes; this essentially means the marking of interesting data that is usually very specific. It is suggested to complete each stage through the entirety of the dataset rather than completing all six stages on one participant. After completing the generation of initial codes, there then comes the grouping of these codes, essentially looking for the potential themes and collating them into these to create a broader dataset which will be useful when creating the visual representation of the data. Some researchers may find that their initial codes turn into a theme or subtheme of the data which then have codes within them. The next stage involves the reviewing of these themes by which the researcher considers the patterns and validity of the generated themes and codes; whether they fit together and form a pattern or whether they need to be reconsidered. The fifth stage consists of defining the themes themselves; this is particularly important for repetition and reliability of the data. The researcher needs to identify what is meant by each of the themes and what codes should and should not be included. Finally, comes the stage of writing up and visually representing the analysis.

Thematic analysis is a particularly flexible method of qualitative analysis as it does not lay roots in one particular epistemological or theoretical perspective. In addition to this, this method is seen as the most simple method of qualitative analysis with very easy to follow stages and process that most quantitative researchers will be used to; and is often seen as the first form of qualitative analysis that should be learnt as a stepping stone towards others. Furthermore, not only is the method easy to follow from a researcher's perspective but it is also easily understandable to the majority of the public (Maguire and Delahunt 2017; V. Braun and Clarke 2008; Javadi and Zarea 2016). However, there is a potential for researchers to focus simply on the summarisation of the data with the interview questions as their main themes; this can fall into the trap of simply organising the data rather than analysing and interpreting what has been discussed. It is therefore important for this to be kept in mind when using this kind of analysis. In addition to this, there are also potential problems with the bias that could occur from the researcher, a theme could be produced which is not backed up by the data but rather from the preconceived idea of the outcomes that the researcher hopes for, therefore this makes the reviewing stage of thematic analysis of particular importance. Furthermore, as with most methods of qualitative analysis, there is a worry surrounding generalisability and reliability due to the small sample size that is often used as well as the analysis relying on the interpretation of the researcher (Maguire and Delahunt 2017; Javadi and Zarea 2016). This form of analysis and methodology seems appropriate for this particular research project due to the need for interpretation of qualitative data from staff and students;

this will allow for an easy comparison between the outcomes of quantitative data and qualitative data due to the categorisation of the themes.

3.9.2.1. Nvivo for Qualitative Analysis

Computer-assisted qualitative data analysis (CAQDAS) is a fairly new innovative method of analysing qualitative data. Most programs that can be found use a method of coding and retrieval, essentially organising the themes to allow for easy access rather than attempting to code the data by hand. However, CAQDAS does not automatically analyse the data for you, like quantitative programs are able to, they simply assist in making the analysis process much easier (Bryman 2012). Nvivo is the CAQDAS program of choice in order to thematically code the data from the interviews with the members of staff as well as the qualitative surveys from the students. This allowed for easy retrieval for write up through the use of coding reports which allow the researcher to see the quotes that belong to each code.

3.9.3. Overview of the Methods of Analysis at each Stage

Table 3.8 shows a visual representation of the data collection method, the method of analysis and the purpose/justification as to why this particular stage of the research was employed. This begins with the literature review stage of the research and follows the stages into the initial exploratory stage as well as the development and testing of the assessment tool.

Data Collection Method	Method of Analysis	Purpose
Literature Review	Strategic search of the literature with critical analysis.	To identify gaps in the literature and to see what has and has not worked previously.
Preliminary inquiry with STEM students (Interviews)	Thematic analysis using Nvivo. Used to identify the skills that STEM students perceive to have gained during the course of their degree.	To identify how specific the cohort of students needs to be for the creation of the assessment tool; do STEM students discuss the same skills or is there differences?
Learning Gain Assessment Tool Creation (Questionnaire and Qualitative Survey)	An assessment tool was created from the literature as well as some questions taken from existing measurements such as the NSS. Pilot Stage: Statistical analysis with SPSS was conducted in order to test for reliability across questions. A survey was also sent out to knowledgeable staff members who identified any potential validity problems.	To test for initial reliability and validity problems with the assessment tool before proceeding to main testing on target cohort of students.
Reliability Testing (Questionnaire)	Statistical analysis using SPSS of both descriptive statistics and inferential analysis.	To test the reliability and stability of the measure across different levels and groups of students. Should the mean and spread appear to be the same this would indicate stability. Cronbachs alpha and Confirmatory Factor Analysis are also utilised.
Validity Testing (Questionnaire, Interviews and Qualitative Survey)	Statistical analysis using SPSS with both descriptive and inferential statistics.	To test for validity of the measure by following students individually over time. This will allow to see differences in students and whether the measure is able to show difference in learning gain.

Table 3.8: Data Analysis at Each Stage and Purpose.

3.10. Ethical Considerations

A set of ethical principles should also be considered in order to protect the participants of research studies. Researchers are generally expected to protect the "rights and dignity" of their participants whilst also withholding a set of ethical standards. According to the British

Psychological Society (2014) these principles include; risk, deception, consent, confidentiality and debriefing. During this research there have been efforts made to follow these guidelines during the entire research process and at every individual stage. Risk refers to physical or psychological harm of the participant. This includes the protection of personal data, relationships and potential risks to social status. During this research, as it was purely based on participants' educational experiences and learning, there was no risk of harm to the participant on the questions that they were being asked during both quantitative and qualitative stages. Efforts were made to protect participants' individual information that will be discussed in more detail further on. Deception refers to falsely giving information to a participant, this can be on the specific nature of the study itself, researchers are expected to give detailed information on the justification and reasoning behind the research. In some methodological cases, there may be a need to withhold some information; however, this should then be included in a debriefing at the end of the study. This research did not withhold the true nature of the study; it was clear from the start that it was the intention to follow student progress on their learning and to gain an insight into what they believe they have gained. It was also made clear to the members of staff that the information they gave was being used to validate the outcomes of the assessment tool.

In addition, participant consent to the research study is perhaps one of the more important principles. This can be in the form of written consent or implied consent as in the case of completion of a assessment tool. All questionnaires/surveys included a consent form before the questionnaire began, any interviews were asked to give written consent prior to starting the interview, this included consent to be recorded and transcribed. Participants who took part in the individual data collection stage voluntarily gave their email address to be contacted further. Consent also considers participant withdrawal both during and after the study has been conducted. During this project participants were made aware that they could withdraw at any time during, but could not on completion due to anonymity of the data. Finally, confidentiality refers to the anonymisation of the data so that individuals cannot be identified.

During the individual data collection, there was some identifiable data as participants chose to give their email to be contacted. However, these emails were kept under a password-protected website and were only used to send out the follow-up questionnaires that the students had consented to. This research was also reviewed and approved by Liverpool John Moores Universities' research ethics committee who deemed the research ethically sound and able to be conducted.

3.11. Chapter Summary

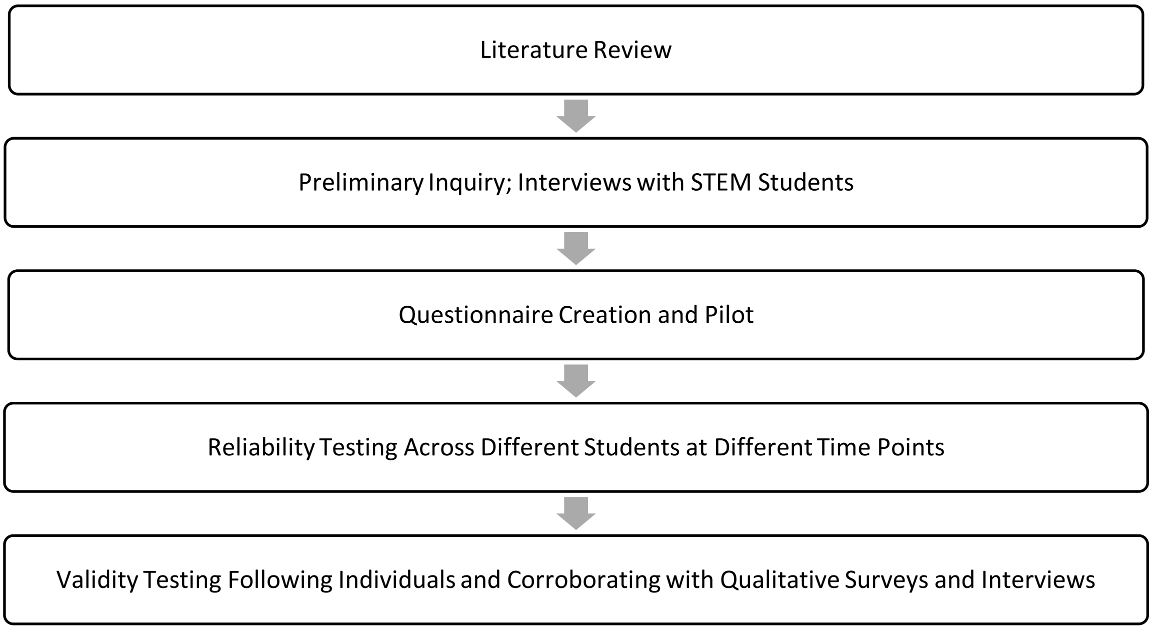

Figure 3.12: Overview of Stages and Methodologies of Research Project

To conclude this chapter, a mixed methodology research approach was chosen in order to be able to triangulate the results. The first stage of the research consisted of a qualitative preliminary inquiry with STEM students in order to understand the skills which this cohort believes to have gained and whether there is any overlap or disagreement in what these skills are. The next stage of the research consisted of a quantitative approach which was the creation and piloting of the learning gain assessment tool. Following this, in order to test for reliability, the tool was then implemented several times across different groups of students and at different time periods, allowing for tests of Cronbachs Alpha and stability of the instrument. Finally, to test for validity, the instrument was implemented again following students on an individual level to test for changes in learning gain. As well as this, qualitative surveys were given to students to corroborate these outcomes as well as qualitative semi-structured interviews with members of staff. These results were used in triangulation in order to prove the reliability and validity of the created instrument.

4. Phase One – STEM Interviews

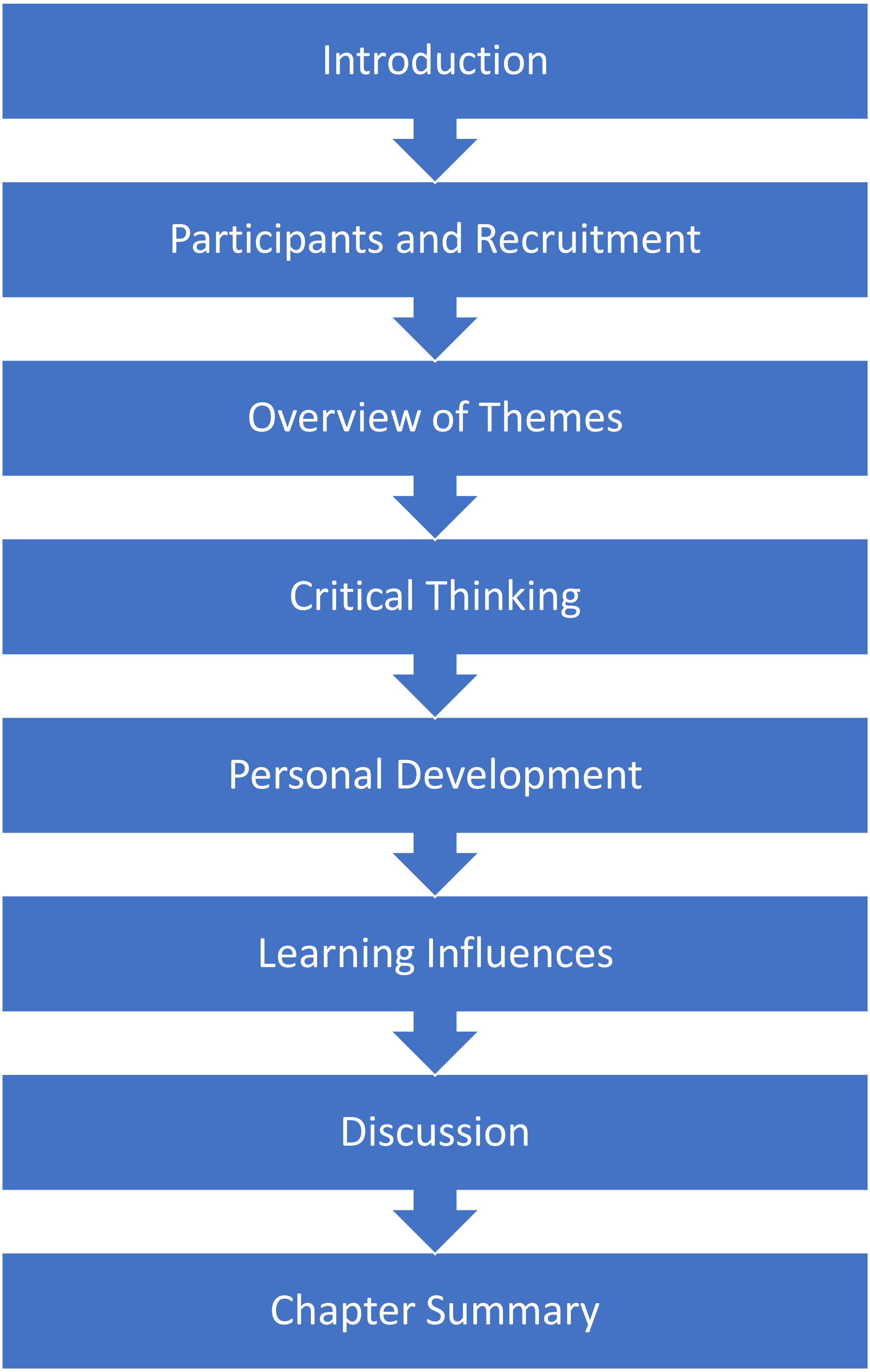

Figure 4.1: Overview of the Sections in the Phase One Chapter.

4.1. Introduction

In order to understand how specific the subject needed to be for the assessment tool creation as well as to inform the skills which will be measured; a preliminary inquiry stage informs the first phase of the research. This phase was undertaken using the qualitative approach. This stage of the research is an exploratory stage in which the researcher intended to identify whether the STEM student cohort is a sutiable target population for a subject specific learning gain assessment tool. This is a particularly important phase of the creation of the assessment tool as it determines a focus of the research.

Prior to this preliminary inquiry stage, a literature review was conducted which can be seen in Chapter 2. This stage was particularly important as it allowed the researcher to understand what had previously been researched and the potential problems that were encountered. The literature review indicated that the tool should focus on a specific discipline within an institution due to the differences across subjects and lack of a standardised curriculum in HE. Therefore, a more targeted approach to the tool was decided upon whereby the skills measured will be taken from the point of view of the students who are currently studying the subject. This current phase focused on looking at STEM students and the skills which they believe they have gained during their degree. STEM students were intended to be the original focus of the learning gain assessment tool.

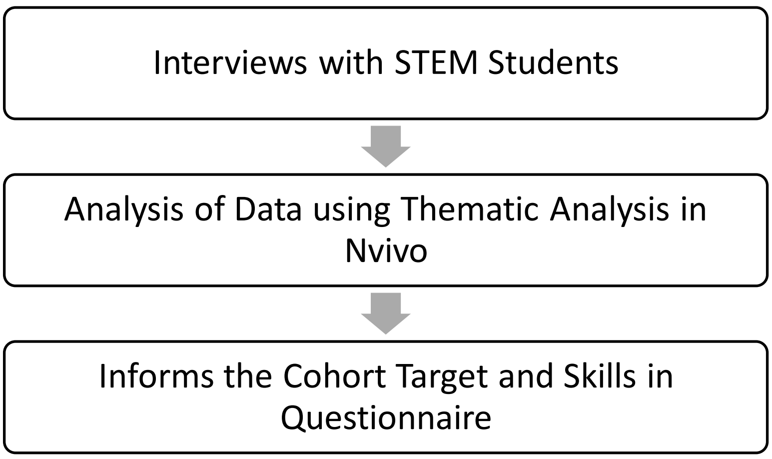

Figure 4.2: Stages and Justification of Phase One

The outcomes of this data lead to a narrowed focus. This meant that the assessment tool would be specifically targeted at science students rather than the entire STEM cohort. Science was chosen to be the targeted subject due to the large percentage of students studying science-based subjects at LJMU (34.22%) compared to the number of students studying engineering and technology (9.23%) and Mathematics (1.05%). The outcomes of this data were also used to inform the skills which will be measured in the learning gain assessment tool.

4.2. Participants and Recruitment

Qualitative research has a basis in the assumption that a larger diversity of students can lead to a better understanding of a research topic (Mason 2002). STEM subjects students were chosen to be interviewed as this was intended to be the original focus of the learning gain assessment tool. This research adopted a theoretical sampling approach by selecting participants based on their potential contribution to answering the research question. Participant selection was stopped when themes tended to overlap and a representation from each area of STEM was achieved. Due to the time constraints and student availability of this small-scale study, there was some difficulty in the extent to which the sample could be further explored. However, themes still began to emerge from the selected participants and created an interesting picture. Students were recruited via email using the LJMU email system; they were asked to only contact the researcher should they be from a STEM based subject, to arrange an appropriate time and day for the participant. The recruitment email was sent out a total of three times to attempt to recruit participants, the researcher also attended lectures of students to ask them to take part in the interviews in the hope it would recruit further participants. There was great difficulty in getting students to engage with interviews, the researcher attempted to gain at least one representative from each area of STEM to try and help with comparability. Table 4.1 provides a breakdown of participant demographics for this phase of the research project.

	Age	Gender	Subject	Year
1	24	Female	Mathematics	Final Year
2	23	Male	Civil Engineering	Final Year
3	22	Male	Computer Science	Second Year
4	37	Male	IT Multimedia Computing	Second Year
5	21	Female	Geography	Final Year
6	20	Female	Psychology	Final Year
7	19	Female	Biology	First Year

Table 4.1: Breakdown of Participants of STEM Interviews.

Participants were informed that the aim of the study was to discuss the learning outcomes of their degree to date and their expectations of what they will have gained by the end of the degree. A semi-structured interview approach was taken in that a set of questions was taken into the interview but participants were able to freely answer and lead the interview in the direction that they wished to do so. Participants were assured that their anonymity would remain intact throughout the entirety of the interview and that names would be changed to preserve confidentiality, as required by the university's code of ethics. This research was approved by the university's ethics committee.

Participants were seven undergraduate students who are currently studying an undergraduate STEM subject. Participants' ages averaged 23.71 with a standard deviation of 6.10. Four of these participants were female and three were male. Three participants were from a science-based subject, two from technology, one from engineering and one from mathematics. Four of the participants were in the final year of their degree meaning they were able to discuss their degree as a whole; two were in their second year and one in their first year. This allowed for different perspectives on learning and development.

4.3. Overview of Themes

The interviews generated a number of themes about learning outcomes. This research has the originality of a perspective from the students themselves, the outcomes have been linked with the current literature. This preliminary inquiry stage is part of ongoing research which will lead to the creation and implementation of the skills-based assessment tool. Below the themes have been split into subheadings with an explanation as to where the specific theme emerged from. The themes were split into three main themes of Critical Thinking, Personal Development and Learning Influences. These themes came directly from the data of the students and their discussions. There were some differences between the STEM students as to how they discussed these themes. The numbers indicated in brackets show how many quotes were found to contain the subtheme.

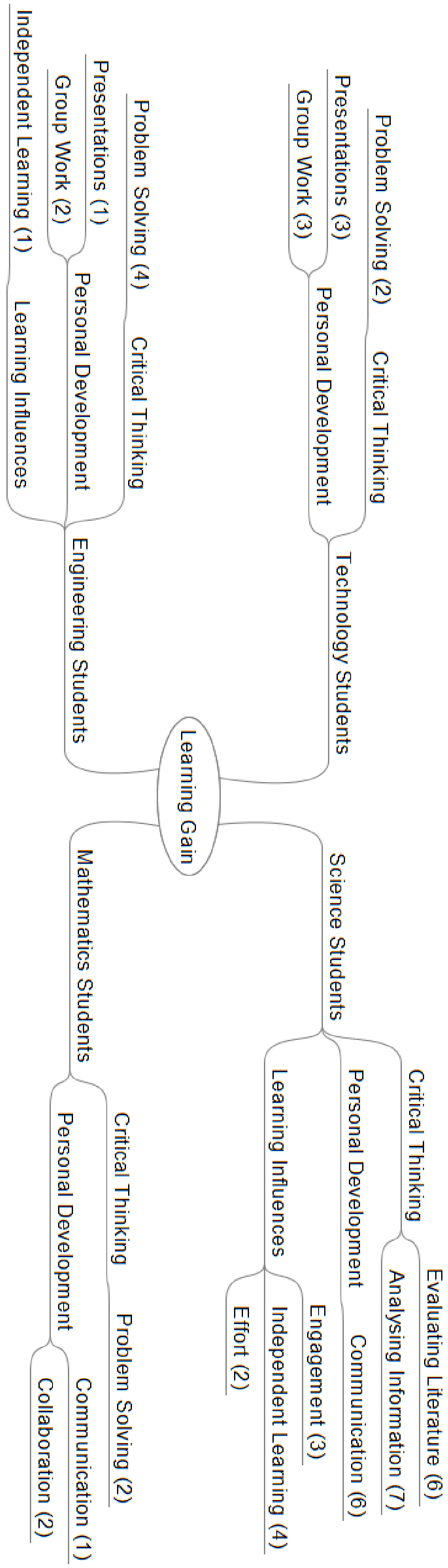

Figure 4.3: Thematic Map of Skills Discussed by STEM Students.

4.4. Critical Thinking

The first question asked of all interviewees was to identify the skills they believed they have developed during their undergraduate degree, this was presented as an open question so as not to bias the students into giving a particular answer. Students unanimously mentioned some form of analysis or evaluation that was expected of them. Additionally, all of the interviewees mentioned critical thinking as a skill that had been developed during their undergraduate degree. Critical thinking would often overlap with problem solving and students would mention the two in conjunction with each other, often interchangeably, this was particularly prominent in the technology, engineering and mathematics students. Sternberg (1986) defined critical thinking as the 'mental processes, strategies and representations people use to solve problems'. This definition suggests problem solving as a specific aspect of critical thinking itself. In support of this, Robbins (2011) discusses 'analytical thinking' as a key aspect to be able to problem solve, this suggests that an individual cannot have one without the other. It was decided to merge these two codes within the same theme as they were discussing the analytical side of their degree.

"Critical analysis is key. I'd say the ability to problem solve in STEM fields is arguably the most important asset you have." (Participant 3)

The importance of critical thinking appeared to be emphasised more in participants that were in their final year of study who stated that critical thinking was a 'big focus' of level 6. This suggests that when implementing a measure of critical thinking, this should significantly improve during the final year due to the shift in focus. The science-based subjects in this study placed greater emphasis on critical thinking in terms of 'evaluating' literature and being able to 'analyse' information rather than simply accepting the information. This is in line with the research conducted by Moore (2013) who found aspects of critical thinking to involve the ability to be sceptical with information as well as to be able to make a judgement broadly as whether or not the information presented is of high or low quality. This, therefore, supports the idea put forward by students in the science-based subjects who discussed evaluating their information. This evidence suggests that the science-based students in this sample are taught about thinking critically more directly than students from the other STEM subjects as they were able to understand what was meant by “critical thinking” and to provide examples of this when prompted.

“You’ve got to look at information that’s been given to you and then say whether you think that’s right or not. Information that when you look at papers, you then analyse whether you think what information they are saying compares to information B and what you think compared to that, which I probably didn’t do before I came to University, I just read it and thought “yes that’s right”.” (Participant 5)

Participants who weren’t from science-based subjects tended to discuss problem solving as a separate entity and understood this concept more than the idea of thinking critically. Participant 4 understood critical thinking as solving problems saying that he would ‘solve it himself’. Considering this, it would suggest that even though STEM subjects supposedly overlap in their skills, in this particular sample each discipline appears to have a different emphasis on certain skills.

“Problem solving would definitely be one [of the skills gained]”. (Participant 1)

This ties in with the debate put forward by Lai (2011) of whether or not critical thinking skills are domain specific, in the case of this research it would suggest that this is the case. This would further suggest that it may be of benefit to narrow the scope of the research as there are differences in the “analysis” aspect of the degree and the way in which it is taught across disciplines. Supporting this, scholars have also argued that critical thinking is domain specific due to differences in what constitutes valid evidence, arguments and standards (Bailin 2002). However, there are some scholars which argue that critical thinking is a general skill that is the same across all disciplines (Gelder 2005). Additionally, there are also some scholars who believe that there are both general and domain-specific aspects of critical thinking. Ennis (1989) stated that different domains place greater emphasis on different aspects of critical thinking such as mathematical subjects placing emphasis on deductive proof. This would suggest that there may be differences across subjects in the aspects of critical thinking that are taught and emphasised across subjects. In this case it appears to be evaluation that is emphasised for the science based subjects.

In addition to this, the science-based participants tended to discuss that the skills which they had gained were not only based on their subject but based upon the modules that they had chosen. Participant 6 stated that there was only one module where she felt as though she had to problem solve. Furthermore, Participant 5 stated that she did not problem solve at all due to the modules that she had chosen in her final year, however, some of her peers had chosen 'lab work' which meant that problem solving was a much bigger issue. This suggests that there may even be individual differences dependent on the modules which the participants have chosen. This would suggest that a larger sample size would be needed in order to see these differences, this aligns with the previous discussion that LJMU has a large cohort of students who study science based subjects, this then may suggest that this cohort is the best option for narrowing the scope. In addition, it could also be argued that the students could potentially choose modules which limit their critical thinking skills, such as modules that are not problem-based, and therefore there may be certain modules which are key in building these processes.

"I don't really [problem solve] because I've picked subjects where most of the modules, and subjects that I chose, are not really problem based. But, some of my friends do lab work which goes wrong and then they have to change the whole method. Whereas I don't really do that." (Participant 5)

Overall, critical thinking appeared to be a very important aspect that was discussed by the participants. However, it appeared that the way in which the different disciplines were taught to think critically varied; as science-based subjects focused on analysing literature whereas the rest of the subjects appeared to focus more on the problem-solving aspect of critical analysis.

4.5. Personal Development

Another aspect which was mentioned by all students was personal development. This included areas such as communication skills, time management and writing skills. A number of participants discussed that their confidence had grown dramatically since starting their degree particularly in 'public speaking' and 'presenting'. All of the science-based participants stated this but only one of the other STEM based participants, which was mathematics. This is showing how different the degree journey can be across different disciplines, though it could be argued that this is on an individual differences basis, the shared experience of the science-based participants would suggest otherwise. Although, several of the participants discussed presenting as a part of their degree. Participant 2 and 3 both stated that presentations were 'key' in their discipline. Participant 5 stated that she was 'quite confident' but was aware that this was not always the case with others as she knew people who would find presentations 'petrifying'. Communication in general was also discussed as something which had been improved.

"We had to do a presentation so it was more setting a presentation out. It wasn't written communication but was more verbal communication." (Participant 7)

Furthermore, participants also discussed their experiences of working with other people and how this has helped with their skills. Participant 1 discussed the difficulty of having to write a report collaboratively as 'little things' began to matter as people would differ in how they would 'write things'. Although the participant did also state that she thought the experience would be useful 'in work'. In contrast to this, participant 3 stated that although he was able to do group projects, the way in which the students are expected to communicate is not 'how you would do it in an industry standard'. As these participants are from two different subjects within STEM, it highlights the potential differences even from within this knowledge base that is so often grouped together. Furthermore, science-based participants reported that group work was not a big feature of their degree even though they had expected that it would be.

"I thought it was going to be much more involved – a lot of group work – but it's not really been like that." (Participant 6)

The Wakeham review of STEM degrees (2016b) stated that employability skills of STEM subjects have been declining in that this has led to poor employment outcomes, particularly in the science based subjects. With this exploratory stage appearing to show differences in the emphasis on the skills in personal development between that of science subjects and the rest of the STEM cohort, it would suggest that these skills may be more focused on the development of communication in other ways rather than presentations and group work. This would suggest a further need to narrow the scope down to a more specific discipline due to the differences across the subjects.

Some participants also expressed their dislike for group work even though they were aware that it is important to develop 'team skills' there was a general feeling that having a mark based on a group project was 'unfair' due to a lot of factors that could potentially cause the mark to be lower out of the control of the individual. This is in support of Bourner, Hughes, and Bourner (2001) who found that the most common complaints of student experiences of group work were things which students could not control for example having an unmotivated group member.

Not all participants, however, had a dislike for group work with participant 4 stating that he was able to 'learn' from other people via 'working in teams'. This suggests that group work is only beneficial dependent on the individual themselves rather than the specific subject. This is supported by Almond (2009) who found that students with higher individual marks did less well in group work whereas students with lower individual marks did better in group work. This further supports the idea that this form of skill development is based on the individual, this may also be due to the lack of opportunities to participate in group work that appear to be in the science-based subjects.

Becoming an independent learner and being able to manage time effectively was another theme which appeared throughout several of the interviews. Time management appeared to be seen as more important in the final year, likely due to having to manage an individual research project. Students also expressed that they felt university education was a lot more self-reliant as there are times when students ‘have to go and actually do work.’

Finally, writing skill development appeared to differ between the subjects. Science-based participants discussed feelings of being able to write ‘more cohesively’ as well as the ability to ‘write reports’ This was of particular importance to participant 5 who stated that her writing skills were ‘absolutely terrible’ prior to starting her degree and highlighted how her degree had helped her with different ‘styles of writing’.

“I think one thing that comes from science subject is the ability to write reports.” (Participant 5)

In contrast to this, participants from other STEM subjects felt as though the ‘written aspect’ of their courses was very minimal. Participant 2 even went as far to say as he believed that the lack of writing in STEM subjects is what draws some people to do them if their ‘written skills aren’t that good’. This suggests that the participant believes that STEM related subjects are more ‘mathematics and physics’ based and therefore would not develop writing skills as such. Supporting this, participant 1, who is studying a mathematics degree, also said that they had not ‘written that much’ during the course of their degree and therefore did not feel as though her writing skills had developed much. However, the participant was aware of the importance of developing these skills ‘generally for life’ and indicated that she had attempted to do so in her own time and during her final year project.

“I don’t think my writing has really improved but I think that’s just because of the nature of my course.” (Participant 1)

In conclusion, the science-based participants appeared to develop personal skills differently from the participants from other STEM subjects. Their degrees appear to involve less group work and more writing. All subjects did agree upon learning to study more independently and to manage time more efficiently.

4.6. Learning Influences

An interesting outcome of the interviews was the discussion of influences on learning. Participant 5 felt particularly strongly about the engagement and effort which students put into their studying. She felt it was important for students to take up ‘opportunities’ which are available to them that will help to develop their skills. Kuh et al. (2008) found that student engagement was related to academic outcomes, and this supports the idea that participant 5 was alluding to. This suggests that students have an awareness that although there are some expectations of what a HE institution can offer, there is only so much before a student needs to be motivated enough to drive their own skill development during the course of their degree. Therefore, if student engagement is low and their skill development is poor then it would be likely that this would not be due to the institution but due to the student’s lack of motivation to engage with the programme as well as the amount of effort that the student puts in to their degree. This would also suggest a link to motivation and what may potentially cause students to engage, and what may cause them to disengage with their degree.

“I think there is the opportunity to develop their skills but, it’s whether you take those opportunities. Which I think, a lot of people don’t and I also think it is a lot of the same people who do the same thing. I think if you put yourself forward for something then you will benefit from it but if you don’t then you won’t. If you turn up to a group piece of work and sit in silence then it won’t work.” (Participant 5)

In relation to a student's engagement with the course, participants were also aware of the importance of their independent learning. There was a particular emphasis on 'taking control of your own work' as well as 'taking your initiative'. Participants discussed about how their time spent in the library increased over the course of their degree as they realised it was important to work more independently. There was also a feeling of university culture expecting you to be more independent as participant 1 stated they're not 'holding your hand' as is done in lower levels of education. This is supported by previous research which suggests that some students find the transition from college to university a difficult one due to the fact that there is a greater expectation of students taking control of their own learning and understanding the need to take in extra information outside of their contact hours (Christie, Barron, and D'Annunzio-Green 2013).

"So, per year, my time spent in the library has increased and understanding why I have to be there, if that makes sense, not just turning up because I feel like that is what I should be doing. It's more that, I have to be there." (Participant 2)

In conclusion, whilst skills are gained during their time at university there can also be influences on this skill development that are outside the control of the university itself . Participants suggested they need to learn to become more independent and engage with their own learning in order to develop the required skills. Furthermore, there was discussion surrounding the motivation and effort that the student puts into their degree that will directly influence the outcomes that the student gains.

4.7. Discussion

The results of this preliminary inquiry have suggested that there may be potential differences across subjects in the ways in which the students are taught. The science-based participants in this small sample appeared to consider critical thinking in an evaluative sense and considered this in relation to analysing information. Whereas the rest of the participants discussed critical thinking as a form of problem solving. Although this is a small sample, the discussions appeared to be consistent across the technology, engineering and mathematics students who appeared to have more in common than the science-based participants did. The science participants reported less group work and presentation skills, suggesting they may be more ill prepared in this sense than the technology, engineering and mathematics students. It is of course unclear as to whether this information can be generalised to the entirety of STEM at the institution, however, this small snapshot has already begun to demonstrate the differences across the disciplines.

The outcomes of this exploratory stage also align with literature on domain specificity in skills. Literature suggested that critical thinking skills have a domain-specific aspect to them in that the way in which the skills are taught and the aspects of which are emphasised differ across subjects (Lai 2011; Ennis 1989; Bailin 2002). This can be seen in the differences in how science students discuss critical thinking compared to the rest of the STEM cohort; they place greater emphasis on evaluation of sources where as the rest of STEM place emphasis on problem solving. In addition, for the personal development aspect, science based students appear to have less emphasis on the more traditional methods of improving their communication skills such as presentation skills yet this still seems to play a small part in their degrees. This aligns with the more recent literature on STEM skills which states that their employability is declining, specifically in that of science based subjects (BIS 2016b).

This research intended to question whether there was enough overlap within the STEM subjects for their skills to be grouped as one. It is entirely possible that due to the small sample size, the differences within the discussions are simply due to individual differences i.e. the individual experiences of each of the participants. However, the two participants from the technology-based subjects both spoke about their learning outcomes in similar ways and in line with the engineering and mathematics participants. All of these participants discussed a lack of a writing aspect within their course as well as an emphasis on problem solving. In addition, all of the science-based participants had an agreement on what they had achieved during their degree. The participants discussed that they were often expected to analyse information and to not accept it at face value. They were also in agreement that there was a lack of group and communication related opportunities on their course. This suggests that this has provided an appropriate snapshot of the disciplines. Further study would be needed with a wider range of students, larger sample and across different institutions to see if this data is generalisable across the cohort.

The main benefit of this stage was to inform the future development of this entire research project. This information has led the main research project to be focused solely on science-based subjects rather than the entirety of the STEM cohort as was originally expected. As although it is clear there is a need for a more subject-specific measure (Brooks 2012; Howson 2017), it was not clear exactly how specific this needs to be. As STEM subjects are often grouped as having "STEM skills" suggesting their skills overlap, this research would suggest that the way in which they are taught to utilise these skills may be different. The focus on science-based subjects has also suggested that it would be beneficial to look at the different science subjects in the main research project. This would be to see whether there are any differences in the emphasis on different skills; as this project suggests they are taught in similar ways however, the emphasis on the skills still may be different across different science subjects. Due to the institutional context of having 34% of the student population being from a science-based subject, it would also make sense to take the target population as this rather large cohort rather than one of the other subjects in STEM which make up smaller parts of the student population. Therefore, for the assessment tool, it would be beneficial to collect data on the individual sub-discipline of science that the student is studying to allow an insight into the differences within these subjects.

In terms of the skills that were chosen for the tool; due to the narrowing of focus onto the science students, critical thinking was chosen as one subscale for the tool. This was chosen instead of analysis due to the way in which the students discussed this aspect as the area in which they believe they have gained the most. This also aligns with previous literature which states that universities generally argue that their goal for HE is to instill into their students the ability to think critically (Arum and Roska 2011). The data from these interviews were used to inform some of the questions that were asked in the assessment tool, this is discussed in further detail in chapter 5. In addition to this, personal development was also considered a subscale; this included areas such as communication, presenting and working with others. Although group work has been suggested to not generally be a part of assignments, some interviewees still discussed the idea of meeting with peers and therefore this would be the focus of the questions rather than on group work. Finally, as pointed out by several interviewees, the engagement, independent learning and motivation of students is an important area on which to collect data when attempting to measure these skills. This is due to the fact that it will be able to show whether the change in learning gain is due to students' engagement. It is something which could potentially be considered a control for the outcomes of this measure, and is definitely a subscale which will be beneficial to collect and compare with the outcomes.

4.8. Chapter Summary

This chapter was a preliminary inquiry stage of the research that informed the researcher of the focus of the student cohort for the assessment tool as well as the skills that need to be included. Originally, the project was to focus on the STEM student cohort, however this preliminary stage led to a change in focus. Interestingly, the outcomes of this preliminary inquiry stage also informed the researcher of the need to include engagement within the assessment tool. Particularly if there is a decline in a student's skills as this could then be due to a decline in the student's engagement with the course rather than the failure of the institution to provide the necessary information.

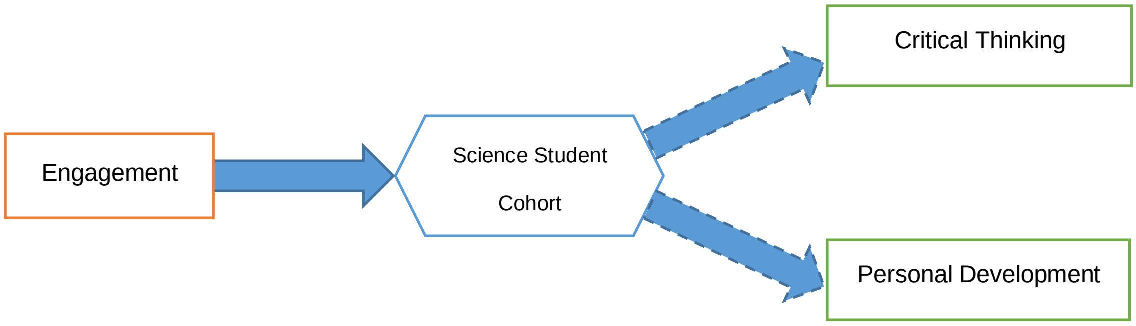

Figure 4.4: Skills Identified for Targeted Student Cohort

The way in which the science cohort discussed critical thinking compared to the rest of STEM was of particular interest as they discussed evaluation and analysis of information rather than the ability to problem solve. This makes the tool very targeted to the cohort and shows the need for a specified measure which has not been done previously. This tool will not only target the necessary skills, but will also target the skills in the way in which they are taught for this specific cohort of students. This makes the ability to see the students' learning gain much more appropriate as it can be seen in relation to what would be expected of the students in this disciplinary area.

5. Phase Two – Assessment Tool Creation and Pilot

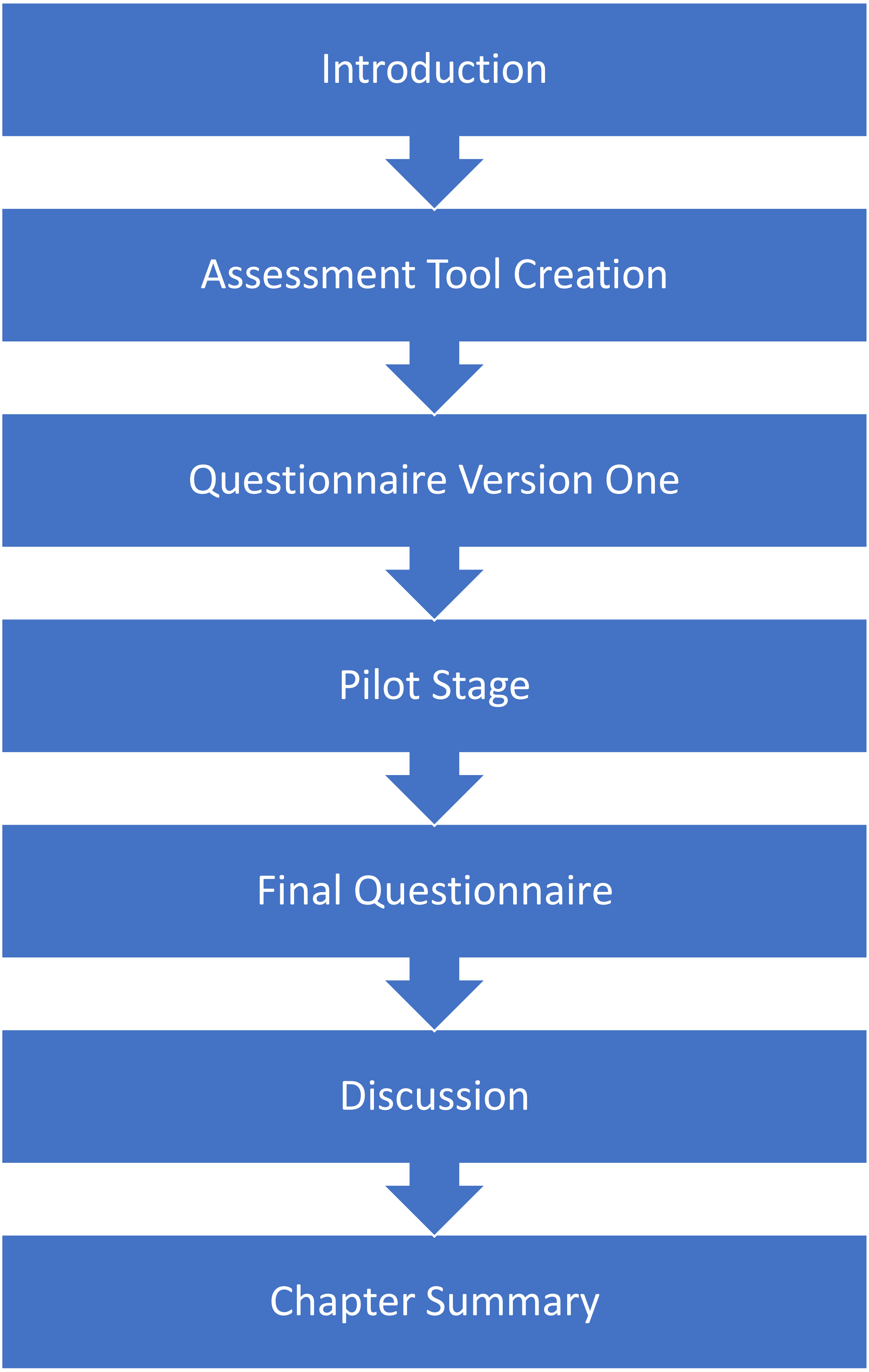

Figure 5.1: Overview of Chapter Progression for Phase Two.

5.1. Introduction

This project was originally intended to be targeted at the STEM student cohort. Upon the preliminary inquiry of this project, it was decided to narrow the focus of the assessment tool to that of science students due to it being the largest cohort in the targeted institution. This phase of the project involved the creation and pilot of the assessment tool. The preliminary inquiry phase informed the skills which were to be used in the tool. Critical thinking and personal development were selected to be the main focus of learning gain in science students. Engagement was also considered to be a measure in this learning gain assessment tool, this is due to the fact that it can affect the outcomes of the skills and therefore is an area which needs to be monitored as a potential control.

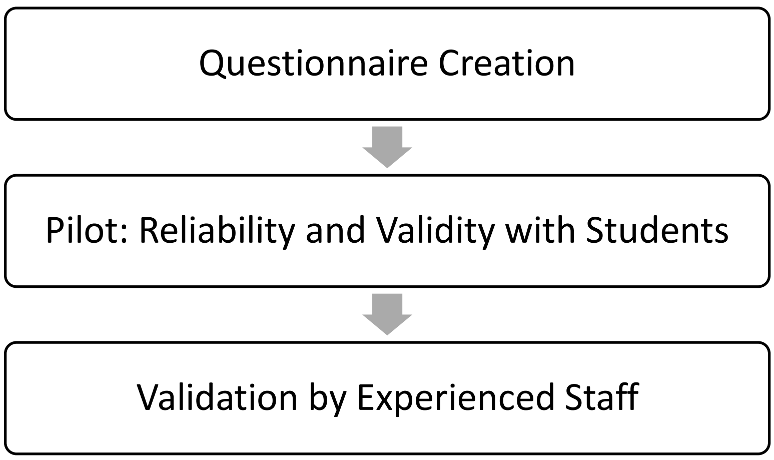

Figure 5.2: Stages and Justification of Phase Two

Following the creation of the assessment tool, an initial pilot stage was taken. This pilot stage allowed for a test of reliability using Cronbach's Alpha by giving the questionnaire to students. In addition, it also allowed for validity as students were able to leave comments on the assessment tool regarding the wording and understanding of the questions. Furthermore, the learning gain assessment tool was sent to members of staff who are experienced in questionnaire creation. This allowed for an initial look into the face validity of the assessment tool; where wording and legitimacy of the questions were considered. Following this pilot stage, some changes were made to the tool which will be discussed later in this chapter.

5.2. Assessment Tool Creation

The assessment tool predominantly begins with questions on the demographics of the participant in relation to the outcomes (Rowley 2011). For this particular project, it was of interest to gather data from participants such as gender, age, subject of study, level of study and mode of study. This gives some potential comparison statistics to show differences between these demographics which may align with or disregard expectations. The majority of these questions were given as closed, multiple choice options. However, subject of study was left open for participants to write the name of their course, this was then later coded in SPSS to make for easier analysis. Age was also left open for participants to put their exact age rather than multiple choice options to make for more accurate averages.

The intended design of the tool was that of a closed, multiple-choice self-report questionnaire. Previous research has found that students are accurately able to report their own learning, likely due to the fact that they are getting constant feedback on their work (Pike 2011; Douglass, Thomson, and Zhao 2012). Closed questionnaires are beneficial as they allow statistical analysis, which is easier for comparability at even minute levels. This is of importance in this project as when following student learning gain, there may only be very small changes which would be difficult to pick up from open qualitative questionnaires. In addition to this, having a multiple choice questionnaire keeps the focus of the answers on the topic at hand. Open questions can sometimes lead to irrelevant or extraneous responses. Not only this but it makes the question clearer to the participant as they can see the options of responses and what kind of answer the researcher is looking for. Moreover, closed questions are much quicker and easier for the respondent to answer. This may be beneficial in this case as the tool will be distributed several times and would become tedious to the participant should they have to answer open questions. This may mean that there is a higher rate of responses as well as a smaller rate of retention (Rea and Parker 2014).

There is also consideration for the types of data questions that could be asked in the assessment tool, this is important not only to recognize whether the question is appropriate but also for data analysis (Brace 2013):

- Nominal data; Discrete categories that can have a number applied but that have no value (Male/Female)
- Ordinal data; Usually found in ranking scales (Most used to Least Used)
- Interval data; numerical data that has equal distance between each point (Likert scales)
- Ratio data; similar to interval data but has an absolute zero (Age)

Interval data has been described as most commonly used for measuring attitudes and perceptions. As this project is intending to measure student skills in the form of self-report, this in itself is a perception that is held by the student. Therefore, it would suggest that interval data is most appropriate. A commonly used method of interval data is that of the Likert scale. This can be on a five point, seven or even nine point scale (Rea and Parker 2014). However, giving too many options can make it difficult for the respondent to pick an appropriate answer. Therefore, a five point scale is chosen here as not to give too many options. In addition to this, many questionnaires that intend to ask students about their learning and engagement, such as the NSS, often take the format of a statement followed by a five point Likert scale. This suggests that this may be the most appropriate method for this particular project.

In addition to this, Oppenheim (1972) discussed the three different survey methods of cross-sectional design, factorial design and longitudinal design. In this particular project, longitudinal design is the method which is being undertaken as the students' skills want to be followed to see the "distance travelled" on the skills. Oppenheim (1972) notes that there should always be a method of controlling for extraneous factors which may explain the differences in the outcomes. In this case, the additional subscale of student engagement was included as this may affect the outcome of the skill development i.e. should the student not be engaging with the course it would be expected that their skills would not increase as much as a student who is heavily engaged. This may be important as it may provide an explanation for any potential decreases in the skill set.

5.2.1. Engagement

Student engagement was decided to be the first subscale of the tool as it will make for easy introductory questions that students are able to answer as well as being able to get students into the topic at hand of thinking about their learning. A search was conducted of validated questionnaires which looked into student satisfaction, engagement and independent learning. The questionnaires that were looked at and utilised for questions in this phase were; UKES NSS, NSSE and finally the Australian Survey of Student Engagement (AUSSE). As these questionnaires have been used widely across institutions in their countries of UK, USA and Austrailia, they were referenced for potential questions that could be asked of students about the engagement they undertake on their course.

Assessment Tool Item	Reference
Attendance	(ACER 2018; Center for Postsecondary Research 2018)
Preparation	(Center for Postsecondary Research 2018; ACER 2018)
Participation	(HEFCE 2017c; ACER 2018)
Contribution	(Center for Postsecondary Research 2018; ACER 2018)
Discussion of Grades	(HEFCE 2017c; Center for Postsecondary Research 2018; HEA 2018; ACER 2018)
Reviewing Notes	(Center for Postsecondary Research 2018)
Reading Beyond Course Material	(Center for Postsecondary Research 2018; ACER 2018)
Last Minute Assignments	(Center for Postsecondary Research 2018; ACER 2018)
Number of Hours Independent Learning	(HEA 2018; Center for Postsecondary Research 2018)

Table 5.1: Assessment Tool Items and References for Engagement Subscale

Students were asked to what extent they agreed with a statement on a five point Likert scale. This aligned with the validated assessment tool format as they all took this approach to asking students about their engagement. As these areas were identified to be the engagement that would directly affect student learning, there were sometimes several questions on the same area. For example, the question of attendance was split into attendance at lectures and attendance at workshops. This is due to the well known idea that students are more likely to attend smaller, classroom like lessons than lectures (Schmidt et al. 2015). Preparation and participation were simply asked in statements as to “I prepare for lectures/workshops”, with participation asking about the student’s contribution to class discussions. Discussion of grades was asked of students in relation to discussing with a member of academic staff. Reviewing notes and reading beyond the course was again, simply asked of students in a statement. All of these statements had a five point scale which ranged from “All of the time” to “Never”. Last minute assignments was used as something which was a more negative question as to attempt to ensure students were paying attention to the questions, as this would be a reverse scored item. Number of hours spent independent learning was split into multiple choice categories, the least being 0 hours, the most being “More than 30” which had eight options in total.

5.2.2. Critical Thinking

The next subscale of this tool was that of critical thinking. Critical thinking is considered to be the main outcome of undergraduate degrees. This skill is seen as a higher level skill that students should be taught and be able to execute on completion of their degree (Arum and Roska 2011). It is clear that this is the case as all students in the preliminary inquiry discussed critical thinking. However, science students discussed critical thinking in terms of analysis and evaluation, suggesting that they are taught this skill in a specific way. Although this gives a direction to aim with the questions, it can be difficult to understand how critical thinking can be measured, what questions should be asked to determine this in students. Many questionnaires can often simply ask whether the student has been expected to think critically on their course such as in the UKES where students are asked at the end of the survey how much the student experience has contributed to “thinking critically and analytically”. However, the start of UKES, and the majority of these surveys, do split critical thinking up into several questions in order to gain an insight into the student's critical thinking. Surveys were consulted such as UKES, NSS, NSSE and AUSSE. As well as this, literature was also searched to look at what is considered to be "critical thinking". Due to the aim of the assessment tool to be focused solely on science students, questions were only included that focused on analysis and evaluation rather than problem solving. This is supported by Lai (2011) who discussed the importance of domain specificity when learning and applying critical thinking skills. Although this is well known in the educational literature, the questionnaires which aim to measure critical thinking still do this with a general approach, this tool aims to focus the questions based on science students.

Assessment Tool Item	Reference
Argument Quality	(Moore 2013)
Application of facts, theories and models	(ACER 2018; Center for Postsecondary Research 2018; HEA 2018; Moore 2013)
Analysing	(Halpern 1999; ACER 2018; Center for Postsecondary Research 2018; Moore 2013; Ennis 1985; HEA 2018)
Evaluating	(Center for Postsecondary Research 2018; HEA 2018; Moore 2013)
Source Validity and Reliability	(Moore 2013)
Challenge Assumptions	(Moore 2013; Ennis 1989)
Usefulness of Sources	(Moore 2013)
Making a Judgement	(ACER 2018; HEA 2018; HEFCE 2017c; Moore 2013; Facione 1990; Halpern 1999)
Forming Ideas	(Center for Postsecondary Research 2018; HEA 2018; Moore 2013)
Understanding and Exploring Sources	(HEFCE 2017c; Moore 2013)
Justification for Conclusions	(Center for Postsecondary Research 2018; HEA 2018; Moore 2013; Halpern 1999)
Reflection	(ACER 2018; Center for Postsecondary Research 2018; HEA 2018; Moore 2013)

Table 5.2: Assessment Tool Items and References for Critical Thinking Subscale

All questions in this subscale of the assessment tool consisted of statements such as “I am able to determine the quality of an argument” where the participant had to indicate on a five point Likert scale, ranging from “Strongly Agree” to “Strongly Disagree”, to what extent they believed the statement applied to them. As with the previous subscale, some areas were split into several questions. For example, forming ideas was split into asking students about building upon existing ideas as well as constructing completely new and novel ideas. This will distinguish between the different levels of higher order thinking as outlined in Blooms Taxonomy (Bloom 1956). Blooms taxonomy indicates that the highest level of learning is creating ideas, however there are two strands of this with creating new information that has never been studied before as well as building on ideas to make something new, with the creating of new information considered to be the higher and more difficult skill.

5.2.3. Personal Development

The final subscale of the tool is personal development. As personal development is often a key part of existing educational questionnaires, these were again consulted for potential items. These consisted of the UKES, NSS, NSSE and AUSSE. However, the items that were picked were specifically chosen as they were mentioned by the science students in the preliminary inquiry. Group work was sometimes mentioned but this was something that was indicated to be lacking in the science degree. Therefore, instead of specifically mentioning working with others as a group, questions were asked about working with students on assignments as a preparation technique or when asking other students for help.

Assessment Tool Item	Reference
Speaking	(ACER 2018; Center for Postsecondary Research 2018)
Explanation of Course Material	(ACER 2018; Center for Postsecondary Research 2018; HEA 2018)
Presenting	(ACER 2018; Center for Postsecondary Research 2018)
Working with Others	(ACER 2018; Center for Postsecondary Research 2018; HEA 2018; HEFCE 2017c)
Preparing for Assessments with Others	(ACER 2018; Center for Postsecondary Research 2018; HEA 2018)
Asking Other Students for Help	(ACER 2018; Center for Postsecondary Research 2018; HEA 2018)

Table 5.3: Assessment Tool Items and References for Personal Development Subscale

In this section, all questions were still asked in the form of a statement and a five point Likert scale. However, some questions answered consisted of a “Strongly Agree” to “Strongly Disagree” format and some of “Very Often” to “Never” this is due to the nature of the questions and the format not making sense. For example, “I am able to speak clearly and effectively about course material” made more sense to have an agreement scale as this is a personal skill and asking the participant to what extent they believe they have gained this skill. However, some skills required the participants to indicate the frequency of something, such as “I have prepared for exams/assessments by discussing the material with other students". This was more of interest to see whether the frequency changes over a period of time. For working with others, participants were asked about whether they have worked with students on projects (which can be in the form of group work, or simply by working on an assignment), whether they believe that they work effectively with others, frequency of preparing with other students and finally asking for help from other students. These questions give an indication of the student's team work skills and how this may change over time, this had to be considered in ways other than that of group work due to the nature of a science degree.

5.3. Draft of the Assessment Tool

Below is the first version of the questionnaire which was taken to pilot stage, this started with demographics questions. All of the questions in the engagement and personal development section were already validated questions that were taken from existing surveys. The critical thinking section, however, were questions that were new and created based off of the literature that stated the aspects of critical thinking.

Demographics

Gender

Male	Female

Age

Subject

Science	Technology	Engineering	Mathematics	Art	Other

Engagement

I attend all of the lectures on my course

All of the time	Most of the Time	Some of the Time	Rarely	Never

I attend all of the workshops/seminars on my course

All of the time	Most of the Time	Some of the Time	Rarely	Never

I prepare for my lectures/workshops

All of the time	Most of the Time	Some of the Time	Rarely	Never

I ask questions in class.

All of the time	Most of the Time	Some of the Time	Rarely	Never

I contribute to class discussions

All of the time	Most of the Time	Some of the Time	Rarely	Never

I have discussed my grades or assignments with a member of staff

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

I review my notes after I have attended a lecture/workshop

All of the time	Most of the Time	Some of the Time	Rarely	Never

I read beyond the course material

All of the time	Most of the Time	Some of the Time	Rarely	Never

I leave my assessments and exam revision until last minute

All of the time	Most of the Time	Some of the Time	Rarely	Never

Roughly, how many hours per week do you spend doing work outside of class? This includes extra reading or working on assignments.

0	1-5	6-10	11-15	16-20	21-25	26-30	30+

Critical Thinking

I am able to determine the quality of an argument

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to apply facts, theories or models to practical problems or new situations

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to analyse an idea or theory in depth

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to effectively evaluate a source

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to determine whether a source is valid and reliable

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

When reading sources, I try to challenge the assumptions that are made within the text

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to distinguish between sources which are useful and not useful

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

After reading many sources, I am able to come to my own conclusions

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to build upon existing ideas to come up with something original

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to construct my own ideas in regards to my subject

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

After reading a source, I would be able to explain the basic meaning of the text.

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

Upon further inspection of a source, I would be able to determine the underlying message and motives of a text

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

When judging a text, I am able to provide justification for my answer using evidence

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

After coming to a conclusion on a text, I am able to reflect on my own decision and critique the judgements I have made

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

Personal Development

I am able to speak clearly and effectively about course material

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I have explained course material to another student

Very Often	Often	Sometimes	Rarely	Never

I have given a presentation on my course

Very Often	Often	Sometimes	Rarely	Never

I have worked with other students on projects

Very Often	Often	Sometimes	Rarely	Never

I can work effectively with other students

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I have prepared for exams/assessments by discussing the material with other students

Very Often	Often	Sometimes	Rarely	Never

I have asked another student for help with course material

Very Often	Often	Sometimes	Rarely	Never

5.4. Pilot Stage

When creating an assessment tool, it is important to conduct an initial pilot stage whereby a smaller number of participants are taken in order to test for reliability and validity. The number of participants for this stage is normally around twenty to forty respondents, the researcher is not concerned with a more controlled sampling procedure but simply attempting to test on the target population (Rea and Parker 2014). Often reliability is tested with the Cronbach's Alpha which is a statistical test of reliability, as discussed in chapter 3. Brace (2013) discussed several questions that should be considered when piloting a questionnaire.

Reliability

- Do the questions sound right?
- Do the participants understand the questions?
- Does the questionnaire retain attention? i.e. not too long
- Does it flow properly?

Validity

- Can respondents answer the question?
- Are the options provided sufficient?
- Do the questions and answers match with what is being studied?

As previously discussed, face validity is often simply done by asking those who are experienced in the creation of questionnaires, as well as the area, to review the assessment tool and give their thoughts. Therefore, this phase of the project will consist of piloting the tool to students to test reliability, with the option for students to comment on the wording and understanding of the questions. As well as to members of staff to test face validity whereby they are able to comment on the questions and whether there should be any additional questions.

5.4.1. Student Pilot

The tool was set up online using Bristol Online Surveys. This tool allows for easy creation of the tool as well as controls. It is particularly helpful as it can cut off the respondent number which is useful for pilot projects. The assessment tool was sent out to students and was set at a cut off of 30 participants using the online tool. Participants were asked to answer the questions within the assessment tool so that Cronbach's Alpha analysis could be utilised. At the end of each subscale section, there was an open question for students to be able to write in should they not understand certain questions or be confused by the wording. This allowed to test for the validity of the questions.

5.4.1.1. Methodology

5.4.1.1.1. Participants

The assessment tool was sent to all students studying at LJMU, this meant that there were also students of Masters and PhD within this population as well as those from different disciplines. Although this was not the intended target population, this was to ensure that some of the questions were not asking too high of a level of students as well as to make sure that the questions were targeted to science students. The Masters and PhD students would then be able to comment should they believe that the questions asked were more at their level rather than that of undergraduate. And students who did not belong to the science-based discipline would be able to comment on whether or not they felt the questions applied to them. This was an attempt to validate that the questions were in fact targeted. This pilot stage ended up with 30 students which consisted of 36.7% male and 63.3% female. The average age of participants was 21.36 years. Students were also asked about their discipline, PhD and MPhil students were asked to record the main subject of their degree. Out of the 30 students, there was a total of 3 postgraduate students and 27 undergraduate students. Of the 3 students who were postgraduate in nature, two listed their subject to be “other” and one listed it to be science.

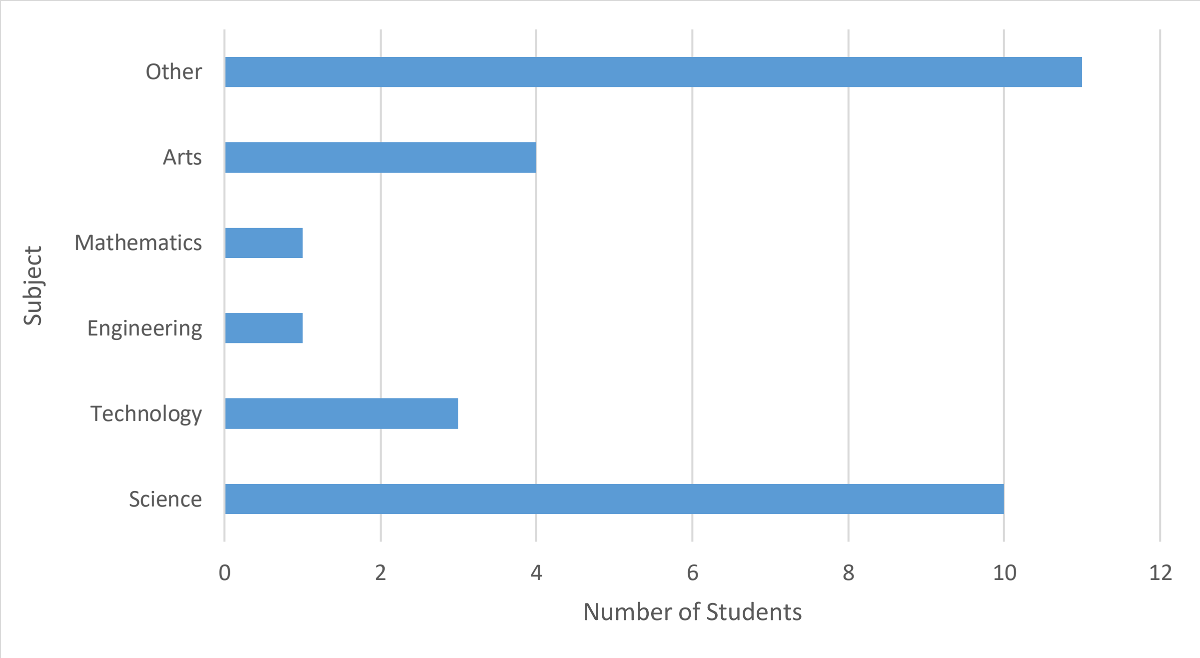

Figure 5.3: Breakdown of Subjects in Pilot Stage of Assessment Tool.

The majority of participants in the pilot project came from a science-based discipline or felt they did not belong to any of the categories and selected "Other". This was useful as it will show whether or not the tool is targeted to science students and that they understand the questions.

5.4.1.1.2. Procedure

The tool was created in Bristol Online Surveys where the cut-off point for participants was set to 30. Participants were contacted via the university email system. They were asked to participate in the assessment tool as well as to leave comments when there were questions where they believed the wording to be confusing, when they did not understand a question, or when the questions did not apply to them. The first two pages of the tool consisted of a participant information sheet that gave participants information on the purpose of the assessment tool and what the results will be utilised for. Following this, participants were also asked to give consent to participating, this simply consisted of ticking boxes to agree to consent. The assessment tool would take approximately 20 minutes to complete dependent on the detail participants put into the open questions. Finally, on competition, participants were presented with a debriefing sheet that thanked them for their participation in the study.

5.4.1.2. Results

The engagement subscale consisted of 10 items ($\alpha = 0.773$), the critical thinking subscale consisted of 14 items ($\alpha = 0.943$) and finally, the personal development subscale consisted of 7 items ($\alpha = 0.499$). Both the engagement and critical thinking subscales appeared to be highly reliable. However, the personal development subscale was of low reliability. Even with the removal of some questions, the reliability statistic is still not considered high. This would suggest that this subscale needs additional questions, or changing of the questions in order for it to become more reliable.

In terms of the comments from students, 24 of the students chose not to leave any comments in the optional textboxes. This would suggest that the majority of the students had no trouble with the tool in terms of wording, layout or understanding. However, 3 students commented that more than one answer could be selected. This would mean that some students may accidentally select several answers to the question. In addition to this, one student commented on the wording of one of the questions which stated “I attend all of my lectures” where the options of the Likert ranged from “All of the time” to “Never”. The student commented that this may be confusing as you cannot attend “all” lectures “some of the time”. Therefore, the wording of this question needs to be reconsidered. Furthermore, 2 students mentioned that the format of the tool was quite confusing. The Likert scale of the tool was laid out to go across the page, students noted that when completing this on a mobile phone this can make it difficult to see all options and suggested that this should go down the page rather than across. Finally, one student noted that a question which asked students about leaving assignments till last minute was quite confusing and misplaced, although this had been put in as a way to have a reverse score within the tool, it may be that this question is simply confusing to participants and should perhaps be excluded.

Although the majority of participants of this inquiry stage did not report any problems with the tool, the different subjects did show that the tool may be fit for purpose. Students who were not from a science-based subject would often check “Neither agree nor disagree” on several of the critical thinking-based questions. This could suggest that the questions do not apply to them and therefore they are unable to answer the question properly. In addition, some students even answered “disagree” to these questions suggesting that they had not been taught this particular skill. The students who answered in this way, were not from science-based subjects. Science-based subjects were always found to be more on the agreeable side of the skills. As this sample does include higher levels than the target population, it would potentially be expected that students from lower levels in their science degree will likely start off “disagree” on some statements as it potentially could not have been taught yet and for this to increase over time. Though, this sample appears to include more higher level students due to the average age range being generally higher. In addition to this, the postgraduate science student had a significantly higher score in critical thinking than the undergraduate science students, suggesting that the critical thinking is able to distinguish between the levels of students.

5.4.2. Staff Validation

The staff pilot assessment tool was also set up online using Bristol Online Surveys. This was sent to members of staff in psychology as well as the supervisory team. This phase simply consisted of asking members of staff about the wording, validity and whether any questions needed to be added or taken away.

5.4.2.1. Methodology

5.4.2.1.1. Participants

The tool was sent out to staff at LJMU in order to receive comment on the validity of the assessment tool. Seven members of staff responded to this assessment tool which was left as an open question survey after viewing the assessment tool. The majority of staff that responded came from the Psychology department of university, there was also a member of staff from the mathematics department and a member of staff from the built environment, the other five members of staff were from Psychology.

5.4.2.1.2. Procedure

The survey was given via Bristol Online Surveys where staff were given a copy of the assessment tool and asked to answer questions following this. The open questions were given at the end of viewing each subscale of the assessment tool so the staff were easily able to comment and view the assessment tool. There were four questions asked at the end of each subscale;

- Are there any items which you believe should be excluded?
- Are there any items which you believe should be reworded?
- Are there any items which the question type or layout is not appropriate?
- Are there any additional items which you believe should be considered?

This allowed for staff members to give their opinion on the assessment tool whilst also prompting them with what should be considered. The beginning of the assessment tool included a participant information sheet to inform members of staff of the purpose of the research and the reason why these questions were being asked. Following this, a consent form would be shown whereby participants were asked to tick a box to confirm their consent. On completion of the assessment tool, a debrief was shown. Although these opinions were sought of members of staff, this does not necessarily mean that all suggestions would be taken. This part of the pilot stage was used as a checking process and for the potential to add any ideas into the assessment tool. There was particular interest in the need for questions for the personal development subscale due to the fact that the Cronbach's Alpha was low in the student pilot.

5.4.2.2. Results

5.4.2.2.1. Engagement

Participant one commented on the engagement part of the assessment tool by saying that one of the questions should be reworded in order to be able to have the same Likert scale output as the rest of the questions and to keep consistency. This participant also suggested the addition of questions which ask about making the use of learning materials as well as a question of students' motivation to attend. Supporting this, participant two also suggested adding questions which indicate what students believe to be important. This could be an interesting outcome as it may help to understand behaviour rather than just monitoring it. Participant three suggested that there was a wording problem where one of the questions referred to "work" it was suggested to write this as "academic work" as students may read this as paid or voluntary work rather than academic. Similarly, Participant four suggested that students may be confused when referring to "class" as the behaviour may be different depending on whether it is a lecture or a workshop environment. This participant also suggested consideration of learning gain in more informal group settings, although this form of learning tends to be covered in the personal development section. Moreover, participant five suggested splitting up questions which contain "grades and assignments" into two separate questions as, again, the behaviour may be different dependent on what it is, this was also suggested by participant six and seven. Furthermore, participant five also suggested allowing students to simply write the number of hours that they study independently rather than giving multiple choice options as the analysis would be "more useful". In addition to this, participant six suggested that questions on how students use feedback may be useful here. These outcomes would suggest that the potentially biggest problem with this section is wording and using questions as one that should potentially be split.

5.4.2.2.2. Critical Thinking

Participant one commented that questions would only apply to science students as there is a wide variety of disciplines and therefore there would be some questions which would not apply. This is a useful comment as it suggests that the critical thinking questions are aimed at science students and not at students of all subjects, as was intended. Participant four suggested being potentially aware that students will always answer these questions in relation to their own subject, despite only some questions asking them to do so. However, this is the intention of the project due to the fact that it is focused on science-based subjects and therefore the wording is not a concern here. Participant seven suggested the addition of linking evidence and theoretical positions. This section appeared to be the least problematic of the assessment tool and requires little modification.

5.4.2.2.3. Personal Development

Participant one suggested the addition of questions in this section asking students whether they learn more from students than teaching staff as well as asking whether they avoid group work. This may be useful as it will show whether science subjects tend to shy away from projects that involve team work. Furthermore, participant three pointed out that one of the questions was asking a double-barrelled question as it asks about student preparation and discussion with other students and these should be split into two to not cause confusion. Furthermore, participant four suggested that perhaps asking students whether they value the skills they have gained towards their chosen career may be something of interest as it shows development towards an intended career path. In addition, participant six suggested that communication with members of staff may be useful here, this is key in personal development as it may further the student’s communication skills. There appear to be some ideas for additional questions that could be added to show both action and skill-based questions of the participant. This may make this subscale much more reliable.

5.4.3. Modifications

Overall, the layout of the questionnaire was changed so that the Likert scale answers to the questions run down the page instead of across. This was done to make the answering of the assessment tool easier and more seamless on a mobile device. In addition, the questions were set so that participants were only able to select one answer so as not to accidentally select more than one.

5.4.3.1.1. Engagement

Wording of questions was considered to be changed based on the comments received in the staff pilot. For example, "I attend all the lectures on my course" was changed to "I attend the lectures on my course". There was also a consideration of splitting up questions that may potentially have changed behaviour dependant on the circumstances. For example, a question was initially "I prepare for lectures/workshops." This was then split into "I prepare for my lectures" and "I prepare for my workshops". This will help to distinguish whether engagement in students changes dependent on the circumstances of the lesson. The question of feedback was also added into this section. It was asked as to whether students have discussed this feedback with a member of staff, and whether they actually utilise their feedback in future assignments. It was decided to be added as this use of feedback may be important in the increase of skills. In addition to this, as there was a student who was confused by the wording of the question to "last minute" this was changed and split into two questions of "I am organised when it comes to completing my assignments" and "I am organised with revision during exam season". Although there was some indication of working with others outside of the classroom in the personal development section, there were additional questions added to the engagement section which asked students about their engagement with learning specifically with other students such as "During exams, I revise with fellow students." This allowed for the addition of more specific questions in relation to engagement. Finally, there was an additional question about the use of learning materials on Canvas as well as numbers of hours spent independent learning; it was decided to be changed to be an open, free text question whereby students are able to input the exact number of hours they believe they spend learning outside of their classroom hours. This was decided upon due to the fact that it would allow students to be more accurate in the amount of time they believe they spend independently learning and therefore leading to a more specific statistical analysis.

This subscale changed from a total of 10 items to a total of 25 items.

5.4.3.1.2. Critical Thinking

As this subscale was seen to be the least problematic and had the fewest suggestions of changes or comments; the only change that this section endured was that of an additional question which was suggested in the staff stage of the pilot. This question asked students about linking evidence to a theory. This was suggested by a participant as an important aspect of critical thinking.

This subscale changed from a total of 14 items to a total of 15 items.

5.4.3.1.3. Personal Development

This subscale of the assessment tool had the biggest number of changes due to the findings of a low reliability scale in the initial pilot phase. This meant that there was a strong need for suggestions of questions for this scale. Originally, this part of the assessment tool solely focused on asking about skills, however with the staff pilot stage there was a suggestion to include actions within this section as a way to further view whether there has been an acquisition of skills. For example, “I avoid group work where possible” has been a question which has been added into this section. The addition of this question will help to see the actions of the students as well as the opportunities that are presented on the course from questions such as “I have given a presentation on my course”. This section had the suggestion of splitting up exams and assignments into separate questions when asking about discussing material with other students. This allows us to see some potential distinctions between the two as some students may work with others on exam material but not assessments or vice versa. Moreover, there was also a suggestion to include the value of such skills towards a chosen career and everyday life. The addition of these questions also helps to indicate the skill set of the student as it will see whether they believe that their personal development has been important and worthwhile. This may be a good indication of the improvements that the student has made as they may see the value in gaining this skill set more towards the time when they are considering career options. Finally, there was a final suggestion of students' communication with members of staff. This will help to show whether students are engaging and communicating their needs with members of staff and utilising their personal development skill.

This subscale changed from a total of 7 items to a total of 17 items.

5.5. Discussion

This section showed the development of the assessment tool and its subscales. Following this, a pilot project of the assessment tool led to modifications which involved the addition of questions as well as the rewording and splitting of potential double-barrelled questions. The results of the preliminary inquiry led to the decision to use critical thinking and personal development as the skills that would be followed in this learning gain assessment tool. These skills were specifically chosen as they were deemed to be skills specific to HE learning as well as specific to science students (Arum and Roska 2011). Additionally, engagement was also considered to be a subscale as it is a form of control which could affect the outcomes of the skills. Several validated HE questionnaires from several different countries were consulted for questions that could be included in this assessment tool. This included the NSS, NSSE, AUSSE and UKES. Literature was also consulted for the critical thinking aspect of the assessment tool. However, the critical thinking subscale focused in on the evaluative and analytical process of critical thinking rather than problem solving. This made sure that the assessment tool was specifically tailored to science students.

This stage of the research is the main developmental stage of the subject specific learning gain assessment tool. This stage will lead to more thorough testing of the assessment tool which intends to test the reliability and validity of the instrument. This assessment tool intended to use questions that came from validated, already in use, measures of student learning. However, this particular project wanted to keep the focus of the critical thinking and personal development skills purely relevant to the science-based subjects. It appears, from the outcomes of both the staff and student pilot that this has been achieved. Of particular note is in the staff pilot where one member of staff noted that some of the critical thinking questions would only work for science students and not for any others. However, as the pilot project included students not just from science and undergraduate programmes, this project needs to be tested solely on the target population. The next stages of this research project should include reliability and validity testing. Reliability testing will be achieved through testing on different groups of students from the target populations several times across the year. This will help to see whether the assessment tool is stable across different groups of participants as well as at different time points. It would of course be expected that there may be some minute changes, however, the spread and averages of the scores should not change too dramatically. The validity testing stage will be achieved through following students on an individual level across a year, this can be any level of undergraduate as long as they are captured at several different time points.

5.6. Finalised Assessment Tool

The final version of the assessment tool is seen below. Questions which were added or changed have been highlighted for ease. The modifications made were previously explained in further detail.

Demographics

Gender

Male	Female	Other	Prefer not to Say

Age

Please specify the name of your course below.

Level of Study

Level 3	Level 4	Level 5	Level 6

Mode of Study

Full Time	Part Time

Engagement

I attend the lectures on my course

All of the time	Most of the Time	Some of the Time	Rarely	Never

I attend the workshops/seminars on my course

All of the time	Most of the Time	Some of the Time	Rarely	Never

My main motivation to attend lectures and workshops is

Interest	Obtain better grades	Gain skills for employment	Satisfy Course Requirements	Other

I prepare for my lectures

All of the time	Most of the Time	Some of the Time	Rarely	Never

I prepare for my workshops/seminars

All of the time	Most of the Time	Some of the Time	Rarely	Never

I ask questions during lectures.

All of the time	Most of the Time	Some of the Time	Rarely	Never

I ask questions during workshops and seminars.

All of the time	Most of the Time	Some of the Time	Rarely	Never

I contribute to class discussions

All of the time	Most of the Time	Some of the Time	Rarely	Never

I have discussed my grades with a member of staff

Very Often	Often	Sometimes	Rarely	Never

I have discussed my assignments with a member of staff

Very Often	Often	Sometimes	Rarely	Never

I have discussed my feedback with a member of staff

Very Often	Often	Sometimes	Rarely	Never

When I receive feedback on assignments, I will use this to help with future assignments

Very	Often	Sometimes	Rarely	Never

I review my notes after I have attended a lecture/workshop

All of the time	Most of the Time	Some of the Time	Rarely	Never

I read beyond the course material

All of the time	Most of the Time	Some of the Time	Rarely	Never

I am organised when it comes to completing my assignments

All of the time	Most of the Time	Some of the Time	Rarely	Never

I am organised with revision during exam season

All of the time	Most of the Time	Some of the Time	Rarely	Never

During exams, I revise with fellow students

All of the time	Most of the Time	Some of the Time	Rarely	Never

I meet with fellow students to discuss course material as a group

All of the time	Most of the Time	Some of the Time	Rarely	Never

I discuss course material with fellow students

All of the time	Most of the Time	Some of the Time	Rarely	Never

I make use of learning materials on Canvas. This includes accessing module guides, the discussion board and lecture slides.

All of the time	Most of the Time	Some of the Time	Rarely	Never

Roughly, how many hours per week do you spend completing academic work outside of class? This includes extra reading or working on assignments

Attending lectures on my course is important to succeed at university

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

Attending workshops/seminars on my course is important to succeed at university

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

Reading beyond course material is important to succeed at university

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

Preparing for lectures and workshops is important to succeed at university

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

Critical Thinking

I am able to determine the quality of an argument

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to apply facts, theories or models to practical problems or new situations

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to effectively link evidence to a theory

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to analyse an idea or theory in depth

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to effectively evaluate a source

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to determine whether a source is valid and reliable

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

When reading sources, I try to challenge the assumptions that are made within the text

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to distinguish between sources which are useful and not useful

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

After reading many sources, I am able to come to my own conclusions

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to build upon existing ideas within my subject area to come up with something original

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I am able to construct my own ideas in regards to my subject

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

After reading a source, I would be able to explain the basic meaning of the text.

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

Upon further inspection of a source, I would be able to determine the underlying message and motives of a text

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

When judging a text, I am able to provide justification for my answer using evidence

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

After coming to a conclusion on a text, I am able to reflect on my own decision and critique the judgements I have made

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

Personal Development

I am able to speak clearly and effectively about course material

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I have explained course material to another student

Very Often	Often	Sometimes	Rarely	Never

I learn more from fellow students than from teaching staff

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I have given a presentation on my course

Very Often	Often	Sometimes	Rarely	Never

I have worked with other students on projects

Very Often	Often	Sometimes	Rarely	Never

I can work effectively with other students

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I avoid group work where possible

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I have prepared for exams by discussing the material with other students

Very Often	Often	Sometimes	Rarely	Never

I have prepared for assessments by discussing the material with other students

Very Often	Often	Sometimes	Rarely	Never

I have asked another student for help with course material

Very Often	Often	Sometimes	Rarely	Never

The communication skills that I have/will gain from university will be valuable towards my chosen career.

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

The communication skills that I have/will gain from university will be valuable for my everyday life.

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

The ability to work with others that I have/will gain from university will be valuable towards my chosen career.

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

The ability to work with others that I have/will gain from university will be valuable during my everyday life.

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree

I discuss academic problems with members of staff.

Very Often	Often	Sometimes	Rarely	Never

I discuss personal problems with members of staff

Very Often	Often	Sometimes	Rarely	Never

I have discussed career options with a member of staff

Very Often	Often	Sometimes	Rarely	Never

5.7. Chapter Summary

This chapter was the stage of the research where the assessment tool was created to be targeted specifically at science undergraduate students. Originally, the pilot project of the assessment tool with students found that the reliability of the personal development subscale was low and needed to be changed. The other two subscales were of high reliability. The pilot stage with staff revealed some changes to questions in the engagement subscale, as well as suggestions for additional questions and splitting of double-barrelled questions. The critical thinking subscale was the least problematic with just one suggestion of an additional question. This pilot stage with staff was particularly useful for the personal development subscale as it increased the number of questions from 7 to 17. The next stages of the project involve more thorough reliability and validity testing of the assessment tool on the intended target population.

6. Phase Three – Reliability Testing

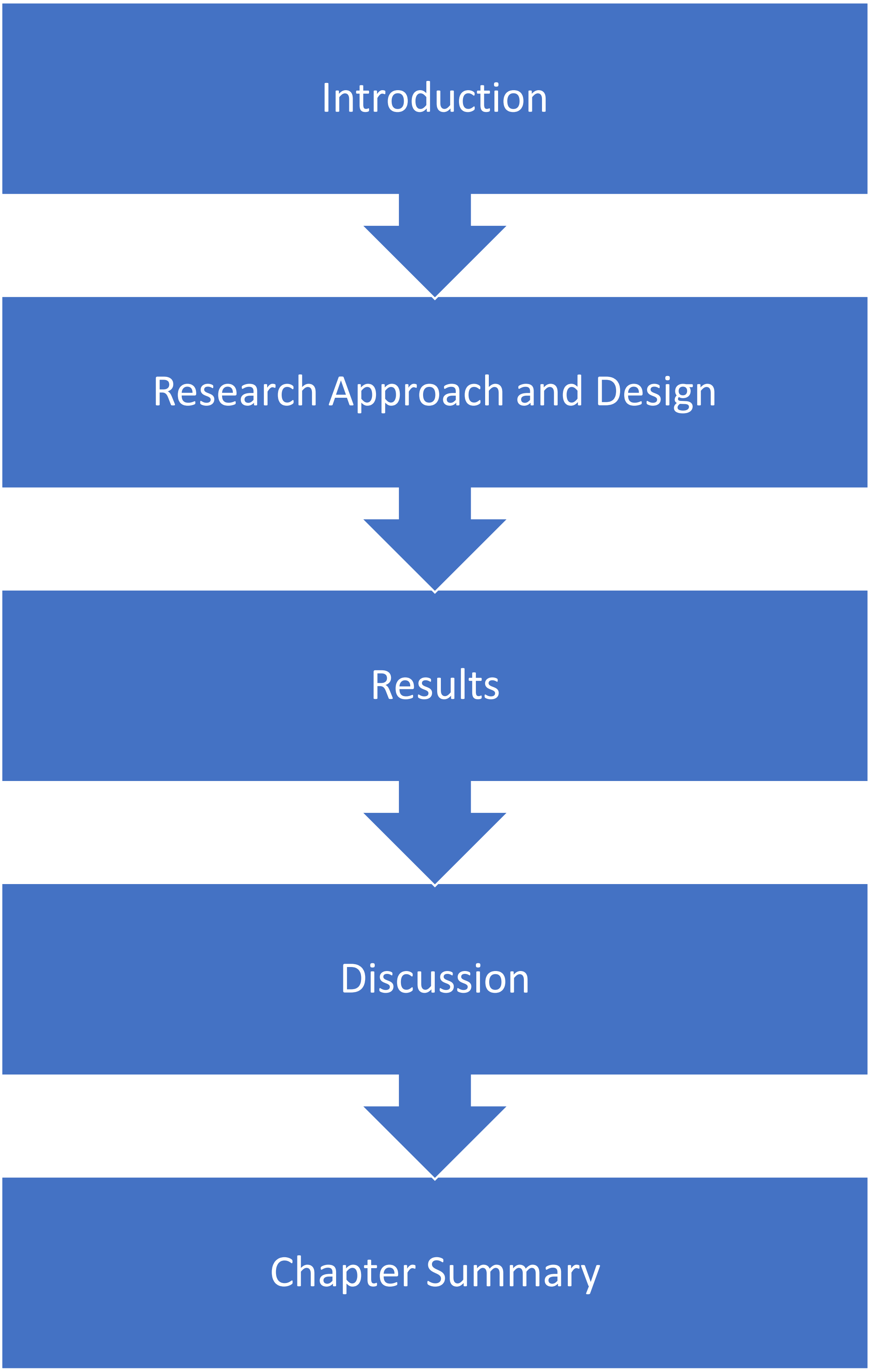

Figure 6.1: Overview of Chapter Progression for Phase Three.

6.1. Introduction

Once assessment tool had been created and piloted, it was important to delve deeper into the reliability and validity testing of the instrument. The time points that are chosen for these stages are the beginning of the year (October), halfway through the year (January) and the end of the year (May). These time points are chosen as this is likely to be when there will be changes in the skills and we will be able to track them which is particularly important in the validity testing stage.

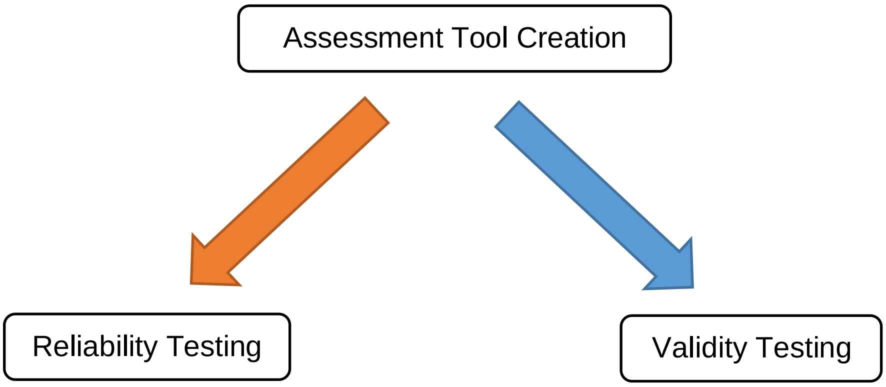

Figure 6.2: Showing Next Two Stages Following Assessment Tool Creation with Reliability Testing Highlighted

This stage, Phase three, of the project involves reliability testing of the assessment tool. This is the intention of checking the stability of the assessment tool by testing the instrument on different groups at different periods of time. There is an expectation that there may be a slight change in the skills when testing over time as there should be an increase in skills. However, there should be a stability in the spread of the data and potentially the means.

6.2. Participants

Participants consisted of undergraduate science students studying at LJMU. Across all three time points there was a total of 384 participants when the data was collected there was a total of 3909 science students studying at the institution, this translates to a response rate of 9.8%. The breakdown of gender overall was 74.15% female, 24.81% male, 0.52% other and 0.52%

prefer not to say. The average age of overall participants was 20.81 with a standard deviation of 3.63. Overall, there were 27 level 6 students (third year), 161 level 5 students (second year) and 196 level 4 students (first year). In addition, 6 of the participants were part time students. A total of 21 different courses participated in this phase of the testing. Below is a table which shows the split of participants and colour coding of the school to which they belong. The subjects are split into three different schools; natural sciences and psychology (blue), sport and exercise science (grey) and pharmacy and biomolecular science (orange).

Subject	Number of Participants
Criminology and Psychology	8
Psychology	65
Forensic Psychology and Criminal Justice	16
Criminology	4
Sport Psychology	16
Animal Behaviour	24
Geography	15
Biomedical Science	29
Zoology	31
Forensic Science	30
Biochemistry	11
Sport and Exercise Science	33
Wildlife Conservation	10
Pharmacy	27
Biology	15
Policing Studies and Forensic Psychology	3
Science and Football	9
Pharmaceutical Science	18
Forensic Anthropology	12
Chemistry	6
Healthcare	2

Table 6.1: Breakdown of Participants by Subject in Reliability Testing

Table 6.1 shows the largest school in the faculty of science is that of natural sciences and psychology. Some of these subjects also belong to the school of law such as the criminology, policing studies and forensic psychology and criminal justice courses. However, they are all bachelors of science undergraduate degrees. Below is a breakdown of these subjects by school.

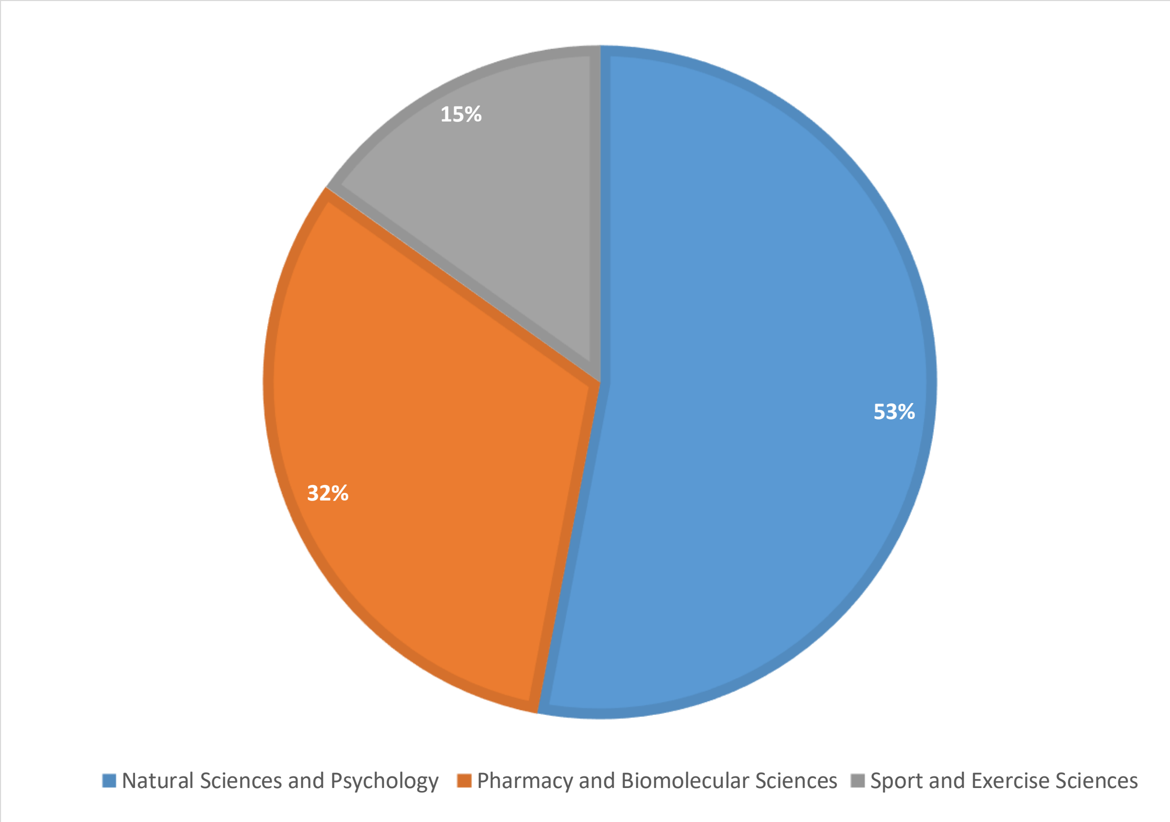

Figure 6.3: Breakdown of Participants by School

The largest number of participants come from the natural sciences and psychology school (53%) and the smallest from sport and exercise science (15%). This breakdown of the participants from the schools makes logical sense in relation to the number of courses in each of the schools. This suggests there may be a good representation of each of the schools in this reliability stage. This is also true for each individual distribution. The first distribution of participants consisted of 185 participants of which 55.14% belong to natural sciences and psychology, 31.81% belong to Pharmacy and Biomolecular Sciences and 13.05% from Sport and Exercise Sciences. The second distribution of participants consisted of 100 participants broken down into 53% belonging to natural sciences and psychology, 31% belong to pharmacy and biomolecular sciences, 16% belonging to sport and exercise science. The final distribution of participants consisted of 99 participants broken down into 46.46% from natural sciences and psychology, 34.34% from pharmacy and biomolecular sciences and 19.20% from sport and exercise science. Although these percentages slightly differ at each distribution stage, it would appear that the ratio of the subjects shows a good distribution of the subjects.

6.3. Results

6.3.1. Data Screening

Before beginning the data analysis process, it is important to inspect the data for errors prior to doing any data analysis. This can involve checking raw data, identifying outliers and dealing with missing data. Solving these issues can also make the data more reliable and valid prior to testing.

6.3.1.1. Missing Data

Missing data can cause problems when conducting certain analyses. This can also be a problem when attempting to do a total score for subscales (as is the case with this assessment tool) as missing data can cause a total score to be skewed. Overall, there was found to be 6 missing pieces of information. This would suggest that this will not be problematic for the outcome of the data as it is a minute number. This may simply be due to participants not wanting to answer the question or to accidentally skipping the question if they were completing it on a mobile device.

6.3.1.2. Unengaged Responses and Outliers

There were no outliers in the data. The data was visually reviewed in order to check for potential unengagement. There was no indication of this, an example would be answering with the same answer for several questions.

6.3.1.3. Factor Analysis Suitability

When conducting factor analysis it is important to check whether the data is suitable to be able to conduct the said analysis. Bartlett's test of Sphericity and the Kaiser-Meyer-Okin measure were used in order to determine the suitability of the data. For factor analysis, all three samples were used as one piece of data. Bartlett's test of sphericity, which tests the overall significance of all the correlations, was significant (χ^2 (1540) = 9655.90, p<.001), suggesting that it is appropriate to use factor analysis on this dataset. Furthermore, the Kaiser-Meyer-Olkin measure indicated the strengths of relationships among variables to be high (KMO = .85), therefore, this suggests that factor analysis is suitable to conduct on this dataset.

6.3.2. Subscale Stability

The three subscale areas were given a “total” variable. Each variable consisted of items relating directly to the subscale. The total for engagement consisted of 19 items. This total did not include the items which asked about motivation to attend lectures, what is importance for success and also the hours spent independent learning. The maximum total a participant could get for this item is 95. The total for critical thinking consisted of 15 items. The maximum total a participant could achieve is 75. Finally, personal development consisted of 17 items. The maximum total a participant could achieve is 85.

	Mean	Standard Deviation
Engagement		
Distribution One	61.81	9.43
Distribution Two	61.77	10.66
Distribution Three	62.61	9.86
Critical Thinking		
Distribution One	57.26	6.13
Distribution Two	58.14	6.40
Distribution Three	58.53	7.72
Personal Development		
Distribution One	56.65	8.17
Distribution Two	57.20	8.84
Distribution Three	58.23	8.64

Table 6.2: Descriptive Statistics for Subscale Totals.

For the engagement subscale, the means and standard deviations do not appear to differ greatly. With Distribution One (M = 61.81, SD = 9.43), Distribution Two (M = 61.77, SD = 10.660 and Distribution Three (M = 62.61, SD = 9.86). This would suggest, as the group of participants are different at each distribution, that there is a stability to the assessment tool. These three different samples of students from the target population are able to stay stable. As the standard deviations are also similar, this would also suggest that the spread of the scores appear to also be stable in the engagement subscale.

For the critical thinking subscale, stability of each group also appeared to be stable as the means and standard deviations did not greatly differ. With Distribution One (M = 57.26, SD = 6.13), Distribution Two (M = 58.14, SD = 6.40) and Distribution Three (M = 58.53, SD = 7.72). There is some small variation in the distributions, though this would be expected due to individual differences. Though as they don't appear to differ greatly this would suggest that there is good stability in the subscale.

Finally, the personal development subscale did not appear to vary greatly between the three groups. However, it did appear to increase. This may be due to the fact that the groups were taken from different time periods and therefore the personal development that is undertaken has increased due to the demands of the degree. However, the standard deviation did not appear to change suggesting the spread of scores is completely stable. With Distribution One (M = 56.65, SD = 8.17), Distribution Two (M = 57.20, SD = 8.84) and Distribution Three (M = 58.23, SD = 8.64).

Overall, it appears that each of these subscales holds a stability with their means and standard deviations. This suggests that on different samples of participants from the same target population, there is the same average from the sample of participants and same spread of scores. This suggests that the assessment tool is able to account for individual differences due to the fact that the scores have a spread, but also suggests that the assessment tool is stable for use within the target population.

6.3.3. Subscale Reliability

In order to test the reliability of the assessment tool, all three samples of data were combined so as to use all 384 participants to gain a better insight into the reliability of the data. Cronbach's Alpha was chosen to analyse the data. This was conducted on each of the individual subscales, as well as the questions which were not included within the subscales. The questions not included in the subscales were used for an understanding of the reasons behind the potential movement in the subscales such as; motivation to attend lectures importance of attending lectures and number of independent learning hours.

	Cronbach's Alpha Statistic
Critical Thinking Subscale	α = .881
Engagement Subscale	α = .839
Personal Development Subscale	α = .812
Motivation/Importance Questions	α = .713
Overall Assessment tool	α = .908

Table 6.3: Reliability Statistics for the Subscales

Table 6.3 shows that the reliability statistics show high reliability in all of the subscales. The Personal Development subscale is of particular importance as it had a low reliability in the pilot of the assessment tool (α = .499) before adjustments were made. The entirety of the tool has particularly high reliability, suggesting that the questions have high internal consistency i.e. how closely related the set of items are as a group. In addition, there was no significant change to Cronbach's Alpha if any of the items were to be deleted.

6.3.4. Factor Analysis

Factor analysis is the statistical technique for identifying which observed variables can be grouped into "factors" which are thought to measure the same thing (Field 2017). This analysis is of particular use when creating an assessment tool as it can shed light on whether the items that are within the same "factor" correlate together well. This analysis works best on large datasets and therefore, all three samples were merged together for this analysis. There are two types of factor analysis that can be conducted; exploratory factor analysis and confirmatory factor analysis. Exploratory factor analysis is the method by which the data is intended to be explored in order to find the “factors” or groups of items. Confirmatory factor analysis, on the other hand, is when one has a preconceived idea of what the “factors” of the assessment tool are and intends to confirm that this is the case. For this project, the latter was deemed to be more appropriate. This factor analysis will be used to confirm that the items belong to the subscales that measure; engagement, critical thinking and personal development. Strong reliability is also indicative for the suitability of factor analysis.

	Estimate	P Value (*** = <.001)
Engagement		
Lecture Attendance	1.000	***
Workshop Attendance	1.192	***
Lecture Preparation	1.714	***
Workshop Preparation	1.770	***
Questions Lectures	1.601	***
Questions Workshops	1.677	***
Contribution Discussions	1.750	***
Discussed Grade with Staff	1.697	***
Discussed Assignment with Staff	1.779	***
Discussed Feedback with Staff	2.077	***
Using Feedback	1.293	***
Reviewing Notes	1.852	***
Read Beyond Course Material	1.302	***
Organisation with Assignments	1.477	***
Organisation with Exams	1.518	***
Revising with Students	0.899	***
Meet with Students about Course	1.253	***
Discuss Course Material with Students	0.919	***
Learning Materials on Canvas	0.706	***
Critical Thinking		
Argument Quality	1.000	***
Application of Facts, Theories and Models	1.047	***
Link Evidence to Theory	1.134	***
Analyse an idea or Theory	1.353	***
Effectively Evaluate a Source	1.227	***
Source Validity	1.226	***

Challenging Assumptions	1.385	***
Distinguishing Useful and Not Useful Sources	0.927	***
Coming to own Conclusions	1.060	***
Build on Existing Ideas	1.366	***
Constructing Own Ideas	1.152	***
Explain the Basic Meaning of Text	0.998	***
Determine Underlying Message and Motives of Text	1.034	***
Justification for Answer	1.037	***
Reflection on Own Judgements	1.150	***
Personal Development		
Speak Clearly and Effectively	1.000	***
Explained Course Material to Another Student	0.942	***
Learn more from Fellow Students than Staff	0.983	***
Given a Presentation	1.027	.001
Worked with Other Students	1.470	***
Work Effectively with Other Students	1.710	***
Avoid Group Work	1.747	***
Prepared for exams with Other Students	1.685	***
Prepared for Assignments with Other Students	1.411	***
Asked Another Student for Help	1.436	***
Communication skills Valuable to Career	2.647	***
Communication skills Valuable to Everyday	2.645	***
Working with Others Valuable to Career	2.960	***
Working with Other Valuable to Everyday	2.938	***
Discuss Academic Problems with Staff	1.222	***
Discuss Personal Problems with Staff	0.653	***
Discuss Career Options with Staff	1.028	***

Table 6.4: Indicator Variables, their estimated loadings and Significance Value to Assigned Factors

The root mean square error approximation (RMSEA) suggests that the model may be a good fit (RMSEA=.088) as it is suggested that below the value of 1 indicates a good fit. In addition to this, all indicator variables loaded significantly on to their assigned factors (p>.001) suggesting that all variables fit well into their assigned factor. Therefore, this would suggest that the current indicator variables are a good fit to the factors to which they have been assigned.

6.4. Discussion

This section showed the reliability of the assessment tool. The assessment tool was distributed at three different time points in order to gain three different samples of students. The results of these samples were compared in order to test the stability of the measure. It was expected that the means for each of the subscales would be comparable to each other as well as the standard deviations. This was found to be the case suggesting that the tool is stable. In addition to this, Cronbach's Alpha statistic was conducted in order to test the internal consistency of the measures. This was found to be high for all three of the subscales suggesting a good internal consistency with the tool. Finally, confirmatory factor analysis was also run. This was run in an attempt to test the reliability and validity of the model. Essentially, this was checking whether the questions (indicator variables) were a good fit for the subscales (factors). It was found that there was a fairly good fit with the model, in turn it was also found that all indicator variables significantly loaded onto the factors. This suggests that the questions which are being asked appear to be useful for understanding the outcomes of the subscales, therefore revealing good reliability and validity of the model.

This stage of the research provided an insight into the measure. This has shown that the measure has high reliability meaning that the repetition of the measure on the same individual will yield similar results. Also, it has shown that all of the questions appear to be asking a question which all participants are able to interpret and understand in the same way. This can be understood from the Cronbach's Alpha statistic. Finally, it has also shown the appropriateness of the subscales. The next stage of the research is perhaps the most vital part; testing the validity. This part of the research will follow participants across a period of a year on an individual basis in order to see whether the assessment tool is able to yield results which can be compared. In addition to this, the results of this data will also be triangulated with qualitative data from staff and students in order to validate them.

6.5. Chapter Summary

This chapter was the stage in which the reliability of the assessment tool was tested. This has shown that the assessment tool has got good internal consistency, stability and is a good model of fit for the questions that were chosen for the subscales. The means and standard deviations across the three samples were stable across all three of the subscales. In addition, the Cronbach's Alpha statistic was high for all of the subscales. Finally, all indicator variables loaded significantly on to their assigned factors. This provides evidence for good reliability of the assessment tool. The next stage of the research will look at the validity of the assessment tool and whether the outcomes of the measure match up to staff expectations and discussions with students.

7. Phase Four – Validity Testing

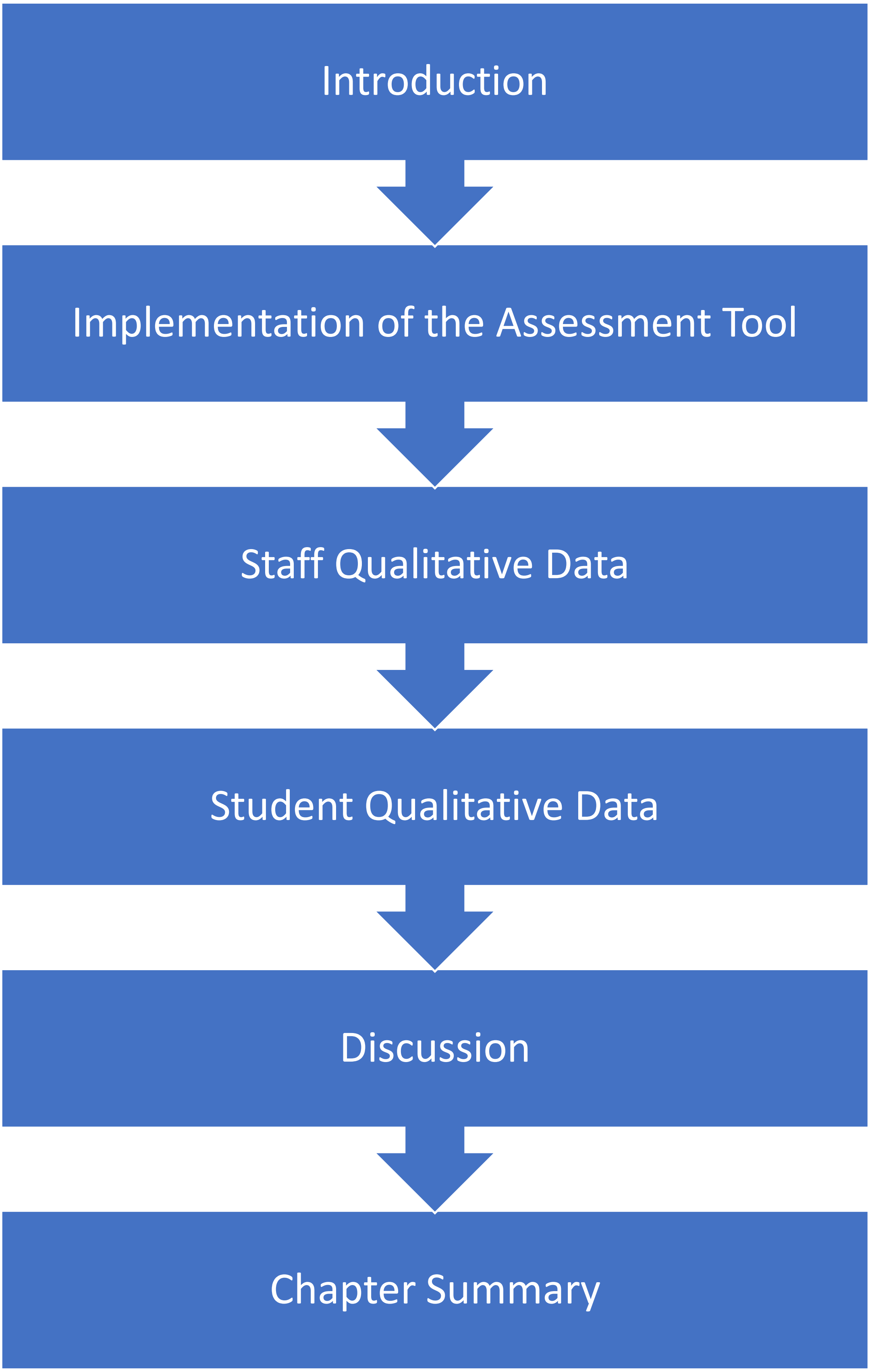

Figure 7.1: Overview of Chapter Progression for Phase Four

7.1. Introduction

Following on from confirming the reliability of the assessment tool, there also comes the important task of confirming the validity. As previously stated, validity can be determined in a number of ways. This particular project's method is that of triangulation by which the outcomes of qualitative data are triangulated with the outcomes of the Quantitative Data.

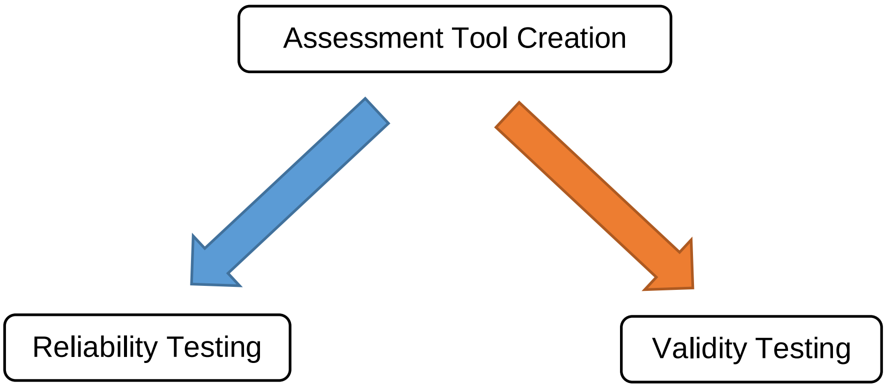

Figure 7.2: Showing Next Two Stages Following Assessment Tool Creation with Validity Testing Highlighted

In this phase of the project, the assessment tool was given to undergraduate science students whose progress was followed on an individual basis over the period of one year. The students were measured at the start, middle and end of the year in an attempt to view changes and follow the "distance travelled" by the individual. This data alone may be useful, as it may be able to show the ability to track minute changes in an individual's learning. However, this data will also be confirmed by the collection of qualitative data. Qualitative data was collected from both staff and students. The staff data was a discussion of expectations of students and the general outline of the courses; there was a hope that this would be able to show differences between courses and confirm any changes. In addition, the survey qualitative data which was taken from students, asked about the chances that they have had to develop their skills on their course. This will also be used to help to confirm any changes that are seen from the assessment tool data. In addition to this, the assessment tool data will be looked at to determine whether the tool is able to pick up significant differences across different time points and different levels.

7.2. Implementation of the Assessment Tool

The first stage of testing the validity of the assessment tool was to implement it onto the target population in the manner of which it was intended. This means following students in a longitudinal manner with the assessment tool whereby the students will be asked to complete the tool at three points in time. At the first point that the assessment tool was sent out, participants were able to voluntarily give a contact email to the researcher so that the assessment tool could be repeated in the middle and end of the year. The tool was sent via institutional emails.

7.2.1. Data Screening for the Assessment Tool

7.2.1.1. Duplicates and Missing Data

The emails of the participants were searched for duplicates. This was done as often participants can forget that they have already completed the assessment tool and may complete it a second time when reminder emails are sent. This was done by using the "Identify Duplicate Cases" analysis in SPSS. There were found to be 6 duplicate cases, these were removed from the dataset. In addition to this, missing data was also considered. Retention rates of participants dropped at each time point. Participants started at 168, dropped to 86 at time two and 78 at time three. Interestingly, these were not necessarily the same participants. Some participants chose to engage with the assessment tool at time three but not at time two. This may be due to forgetting or other commitments at the time. Those with missing data, i.e. chose not to answer a certain question, were not included in the subscale totals but were still utilised to understand the differences on individual questions in which they had provided an answer.

7.2.1.1.1. *Unengaged Responses and Outliers*

The data was also screened for participants who appeared to be disengaged. This referred to those who appeared to be answering the multiple choice questions by selecting the same answer multiple times. This was found not to be the case and it appeared that disengaged participants simply did not respond at all three time points. There were no alarming outliers for this dataset.

7.2.2. Participants of the Assessment Tool

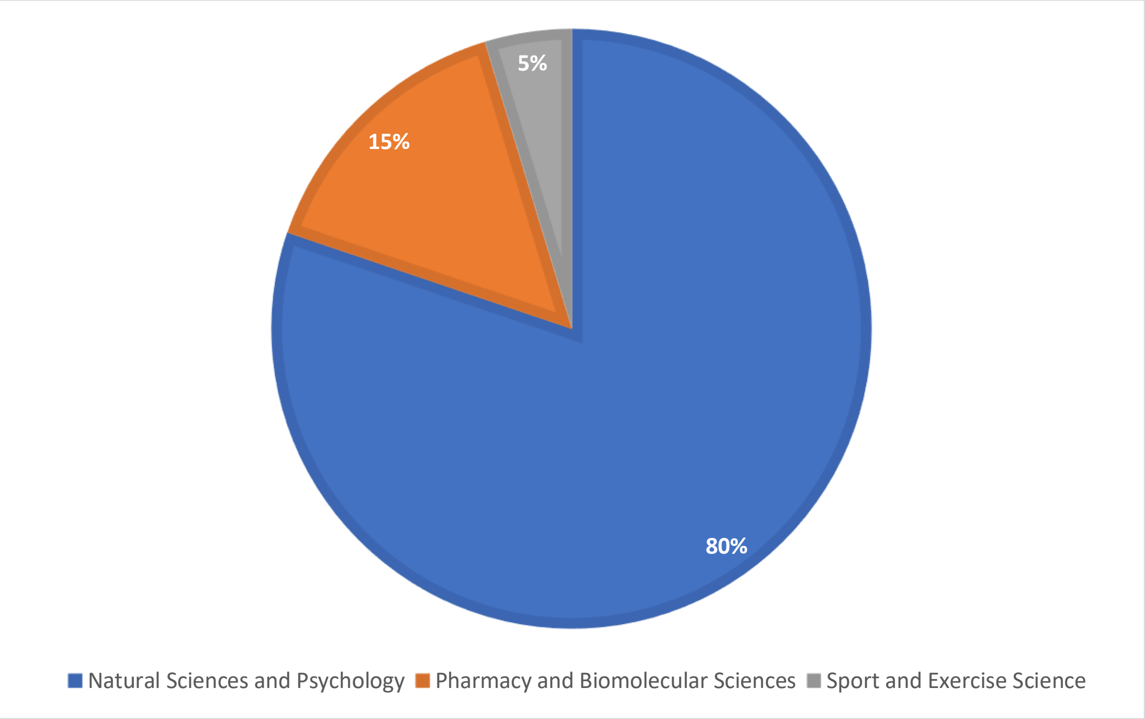

Figure 7.3: Breakdown of Subjects by School for Individual Data in First Distribution

Participants consisted of science undergraduate students in LJMU. At the first stage of the research project a total of 192 students participated, this was at the time where there were a total of 3849 science students studying at the institution making the response rate 5%. However, only 179 of those participants agreed to be tracked for a further two times across the year, bringing the response rate down to 4.7%. Upon screening for duplicates, it was found to only have 168 participants, meaning the response rate went down to 4.4% for the final number of students at the first iteration of the assessment tool. Of those 168, 86 participated in the second round of data collection and dropped to 78 in the final round of data collection. Overall, 18 different subjects participated. In the first round of data collection, this was split into the majority coming from Psychology (53.6%) which is likely due to the assessment tool being put onto the psychology student research participation system. This system is an institutional based points system whereby the students need to participate in research and gain a number of points as part of module credits, this happens in the first year of the degree and explains the high number of Level 4 students and Psychology students. Psychology was then followed by Animal behaviour (6.8%) and Biology (6.8%). The split of gender was 75.5% Female and 24% Male with 0.5% indicating other. In addition to this, the split of levels was the majority coming from Level 4/First Year at 68.8%, followed by Level 6/Final Year at 15.1%, Level 5/Second Year at 9.4% and finally, Level 3/Foundation Year at 6.8%. Again, this is likely due to the assessment tool being put onto the psychology student research participation system which is primarily used by Level 4/First Year students. The average age of participants at this stage was 21.35 with a standard deviation of 7.25. Only one participant was a part-time student, the rest were full-time.

7.2.3. Subscale Total Results

To get an overall view of the outcomes of the assessment tool, the means and standard deviations were computed for the total subscales of; engagement, critical thinking and personal development. Although this may be affected by different levels, it can begin to paint a picture of the outcomes which will later be discussed by level.

	Engagement Total (Max = 95)		Critical Thinking Total (Max = 75)		Personal Development total (Max = 85)	
	Mean	SD	Mean	SD	Mean	SD
Time Point One	63.89	8.74	57.79	6.98	55.28	8.57
Time Point Two	61.59	9.22	58.72	7.55	55.21	10.70
Time Point Three	62.48	8.90	59.14	7.46	56.29	8.63

Table 7.1: Means and Standard Deviations for Overall Outcomes of Subscales.

Interestingly, it appears that engagement had slightly decreased from time one (M=63.89, SD = 8.74) to time two (M=61.59, SD = 9.22). To then slightly increase at the final time point (M=62.48, SD = 8.90). This change could be suggested to be due to students becoming increasingly less engaged throughout the year to then becoming engaged again before their exams. The potential reason for this change will be discussed later in the qualitative data section of this chapter. The total for critical thinking appears to increase across all three time periods. This suggests that critical thinking ability increases over the course of a degree. This would be perhaps the most important outcome as undergraduate degrees are viewed as having the ability to increase critical thinking skills (Arum and Roska 2011). Finally, personal development appears to slightly decrease from time one (M=55.28,SD=8.57) to time two (M=55.21,SD=1-.70). However, it does increase at the final timepoint (M=56.29, SD=8.63).

To test for significance in these time points, assumptions test must first be conducted in order to determine the test which needs to be utilised. Normality tests were run using the Shapiro-Wilk test which found that all time points and subscales were normally distributed ($P<.05$) except for critical thinking at time point two and three. This would suggest that the test which is run for significance needs to allow for non-normally distributed data. Should all the data have been normally distributed, a repeated measures ANOVA would have been utilised, however the non-parametric equivalent to this is that of the Friedman Test. This test also assumes the data from the dependent variable is ordinal data or continuous. The differences across time points was found to be significant on student critical thinking (P>.001) and personal development (P>.05). However, there was not a significant difference on engagement (P<.05). The lack of significance on engagement may be largely due to differences across year groups. Therefore, it would be beneficial to compare means across the different years on each of the subscales.

		Engagement Total (Max = 95)		Critical Thinking Total (Max = 75)		Personal Development total (Max = 85)	
		Mean	SD	Mean	SD	Mean	SD
Time Point One	Lv3	65.00	7.37	58.62	5.00	54.38	5.26
	Lv4	62.57	9.05	57.79	7.11	53.93	8.79
	Lv5	64.89	6.51	55.78	5.81	55.67	7.28
	Lv6	68.74	7.57	58.71	7.85	61.57	6.85
Time Point Two	Lv3	63.57	9.88	63.71	7.30	56.00	11.54
	Lv4	60.56	8.87	58.38	7.89	53.90	8.93
	Lv5	62.71	10.39	56.57	3.46	52.00	13.39
	Lv6	66.91	7.82	60.55	9.18	65.82	8.45
Time Point Three	Lv3	66.71	8.48	59.00	6.83	56.86	7.20
	Lv4	60.88	8.37	59.31	8.43	55.26	8.52
	Lv5	64.83	14.66	59.43	8.08	52.57	12.34
	Lv6	64.54	6.77	59.69	6.21	63.77	4.95

Table 7.2: Means and Standard Deviations for each Subscale Split by Level.

For engagement, it appears that Level 6/Final Year begins as the most engaged year group (M=68.74,SD=7.57). However, the engagement drops down over time. This drop over time appears to be the same for all levels except for Level 5/Second Year who increase in engagement from time point two (M=62.71,SD=10.39) to time point three (M=64.83, SD=14.66). However, these differences were found not to be significant (P<.05). Although we can see that there are changes here, this may be due to the retention of the sample of students over time or the need for a larger sample. For critical thinking, interestingly, Level 3/Foundation Year appear to start at higher levels of critical thinking, on a par with those of Level 6/Final Year. All of the year groups appear to increase in their critical thinking over time except for Level 6 who have a minute drop from time two (M=60.55,SD=9.18) to time three (M=59.69,SD=6.21). This may be due to retention rates, as those who were higher in critical thinking may not have participated in the final time point distribution. The most significant difference across the time points in critical thinking was that of level 4/First Year (P>.001). Suggesting that Level 4/First Year is the year in which the biggest gains occur for critical thinking. Finally, for personal development, for Level 3/Foundation Year and Level 4/First Year, there appeared to be an increase in personal development over time. The difference over time for Level4/First Year was found to be significant (P>.05). However, Level 5/Second Year appeared to decrease over time, suggesting perhaps there was less chance for personal development over the year. Level 6/Final year, appeared to increase from time point one (M=61.57,SD=6.85) to time point two (M=65.82,SD=8.45) to then decrease at time point three (M=63.77, SD=4.95). The significant differences that have been found for Level 4/First Year and not the other levels may be due to the fact that the majority of this sample is made up from that year group (68.8%) and therefore significant differences may have been shown in other levels should there have been a larger sample.

7.2.4. Motivational and Importance Results

The assessment tool also contained questions which asked about student motivations, importance and number of hours that they spend independent learning. These questions intended to understand potential reasons why there may be some changes across the time periods.

7.2.4.1. Motivation to Attend Lectures/Workshops

It would be first beneficial to view the reasons which motivate students to attend lectures at the different time points. Students were given four options as reasons which motivate them to attend lectures, they were also given an “other” option where they were able to input their own alternative answer. This is a question of interest as it can show if the motivations of students change over time.

	Interest in the Subject	To Obtain a Better Grade	To Gain Skills for Employment	To Satisfy Course Requirements	Other
Time Point One	35.1%	47.1%	8.1%	8.6%	1.1%
Time Point Two	26.7%	57%	4.7%	11.6%	0%
Time Point Three	26%	58.4%	5.2%	10.4%	0%

Table 7.3: Changes in Student Motivations to Attend Lectures and Workshops Over Time.

It would appear that at each time point “To obtain a better grade” is the biggest reason that motivates students to attend lectures and workshops. As well as this, it appears to increase over time with a decrease in all the other reasons. This would suggest that students are motivated by the outcomes of their assignments and exams i.e. to get a better grade. This may suggest that students may be less likely to attend or engage with something should they believe that it will not directly help with obtaining a better grade. Therefore, to increase engagement it may be beneficial to emphasise the importance of all lectures/workshops. It was found that there was a significant difference between all three time points (P>.001), this suggests that the motivations of students change over time, near the end of their degree is the most prominent time that students are motivated for a better grade; likely due to the fact that this is usually when exams occur. The only time point where students selected “other” as a reason was time point one. These reasons were from two students who both gave the same reason; that they have paid for university and it wouldn’t make sense to skip them. This relates back to the literature review chapter, about students and their need of a ”value for money” aspect of university (Howson 2017b). As the most important year for percentage toward final grade is Level 6/Final Year, it goes some way to explain why this year group is the most highly engaged as the final year is the most important year. However, it is interesting that engagement appears to decrease over time as the motivation increases to obtain a better grade over time. This question may be answered with the qualitative data.

7.2.4.2. Number of Hours Spent Independent Learning

Students were also asked to estimate the number of hours they spend per week "completing academic work outside of class?", this includes extra reading or working on assignments. This was to gain an understanding of how many hours students spend independent learning each week. This has been split into Levels as it may provide more meaningful data as to whether this changes as the work becomes more meaningful and difficult.

		Mean	SD
Time Point One	Level 3	8.54	3.76
	Level 4	14.64	11.87
	Level 5	10.56	6.85
	Level 6	22.18	11.74
Time Point Two	Level 3	11.00	6.33
	Level 4	12.73	7.25
	Level 5	11.29	8.69
	Level 6	16.73	10.84
Time Point Three	Level 3	10.43	5.86
	Level 4	16.24	11.97
	Level 5	15.14	12.71
	Level 6	26.38	20.85

Table 7.4: Independent Learning Hours Across Each Time Point, Split by Level.

There does not appear to be a logical pattern to the changes in the independent learning, which could be suggestive of an inaccurate estimate of the time spent studying. There are, however, quite large standard deviations for a lot of the levels suggesting quite a large variance in individual differences. Level 3 students appear to increase from time one (M=8.54,SD=3.76) to time two (M=11.00,SD=6.33) to then decrease at time three (M=10.43,SD=5.86). Interestingly for Level 4, there is a decrease from time one (M=14.64, SD=11.87) to time two (M=12.73,SD=7.25) to then increase at time three (M=16.24,SD=11.97). This may be due to the fact that the end of the year is usually when exams occur and therefore higher levels of independent learning occur due to revision for exams. Level 5 is the only year group who consistently increased in their independent learning time, however this was not significant. Finally, Level 6 decreased at time point two but significantly increased at time point three. Again, further suggesting that at the end of the year independent learning time increases potentially due to revision.

7.2.4.3. Importance to Succeed

Participants were also asked on a scale of Strongly Agree (5) to Strongly Disagree (1) about how important certain aspects of university life are to succeed at university. This included asking about attendance (split into lectures and workshops), reading beyond course material and preparing for lectures/workshops. The higher the score the higher the agreement, the lower the score the lower the agreement.

	Attending Lectures on my course		Attending workshops/seminars on my course		Reading beyond course material		Preparing for lectures and workshops	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Time Point One	4.70	0.53	4.63	0.58	4.19	0.80	4.03	0.87
Time Point Two	4.40	0.83	4.41	0.80	3.98	0.88	3.72	0.97
Time Point Three	4.36	0.68	4.21	0.86	3.96	0.84	3.58	0.99

Table 7.5: Students Opinions on Importance to Succeed at University Across the Time Points.

Interestingly, all of the questions decreased in agreement over the different time points. This may suggest that as students progress in their degree they find that they are still able to succeed without doing these things. Furthermore, it appears that all of the different indicators are fairly high in agreement that they are important to succeed at university, with attending lectures being the highest. This provides an interesting perspective as engagement has been found to lower over time. The difference in importance for attending lectures was found to be significant (P>.001), as well as for attending workshops, reading beyond course material and preparation (P>.05). This suggests that the attitudes towards these constructs change over time and students place less importance on each one as time goes on. However, this is an overall subscale count and therefore it may be beneficial to look at the individual questions that related to the importance questions.

	Attending Lectures on my course		**Attending workshops on my course**		**Reading beyond course material**		**Preparing for lectures**		**Preparing for Workshops**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Time Point One	4.58	0.55	4.70	0.57	3.40	0.94	3.73	0.96	3.65	0.98
Time Point Two	4.23	0.73	4.30	0.87	3.27	1.02	3.38	0.97	3.38	1.09
Time Point Three	4.09	0.76	4.12	0.91	3.24	0.91	3.06	1.07	3.08	1.18

Table 7.6: Students answers on the corresponding questions to importance across the time points.

As can be seen from Table 7.6, each of these constructs declines over time. This aligns with the importance constructs as it appears that as students decline in areas such as attendance, they find they are still able to succeed and therefore place less importance on this area to succeed. The difference in attending lectures is statistically significant (P>.001) this suggests that over time, students attend lectures less. This was also the case for attending workshops (P>.001). Which further supports the idea that students may attend less as they see no consequences to their success, leading them to place less importance on attending. Both preparation questions were also found to be statistically significant (P>.001) further suggesting that the decline may have led students to place less importance on doing this action for success. However, reading beyond course material was found not to be significant (P<.05).

7.3. Staff Qualitative Data

Following the quantitative data, qualitative data will be utilised to be able to confirm and triangulate the outcomes. It was decided to get the perspective of both staff and students by asking both of these cohorts about their experiences and expectations of the learning environment and outcomes.

7.3.1. Participants

Staff were invited to interview via email and were interviewed on the premises of the institution. These interviews were semi-structured whereby there was a list of set questions that were asked and taken into the interviews but additional questions may be asked should it be needed. There was a total of six members of staff that were interviews. 33.3% of participants were male, with 66.6% female. Two of the participants came from Psychology, one from Sport Psychology/Science and Football, Two from Zoology/Animal behaviour one of which also taught Biology and Finally, one from forensic psychology and criminal justice. Table 7.7 provides an overview of participant demographics for this stage.

	Subjects Taught	Role	Gender
1	Science and Football, Sport Psychology	Programme Leader	Male
2	Psychology	Lecturer	Female
3	Forensic Psychology and Criminal Justice	Programme Leader	Female
4	Biology, Zoology, Animal Behaviour	Senior Lecturer	Male
5	Psychology	Programme Leader	Female
6	Zoology, Animal Behaviour	Senior Lecturer	Female

Table 7.7: Participant Summary for Staff Interviews.

The length of the interviews averaged 35 minutes across all six of them. They were each recorded using a dictophone and transcribed within 24 hours of completing the interview. Each of the interviews were analysed using thematic analysis, as was discussed in detail in the methodology chapter.

7.3.2. Analysis

The first question asked of all interviewees was about what skills the students have developed during this year. This was left as an open question to begin a discussion about skills with the participant that was not biased towards the outcomes that the researcher expected. Interestingly, three out of the six members of staff indicated critical thinking as the most important skill that should be acquired from a degree. However, the three members of staff who did not cite critical thinking as the most important skill, still discussed the acquisition of it and the importance of this as the degree programme advances. This may suggest that when asked for personal opinion this will be subject to individual differences and shouldn't necessarily be taken as fact. However, the opinions are still of interest to understand the experiences of staff. There were found to be some slight differences across the subjects however there were a lot of similarities especially in terms of engagement. The results will first be discussed and then triangulated with the quantitative data. A thematic map of the analysis can be seen below.

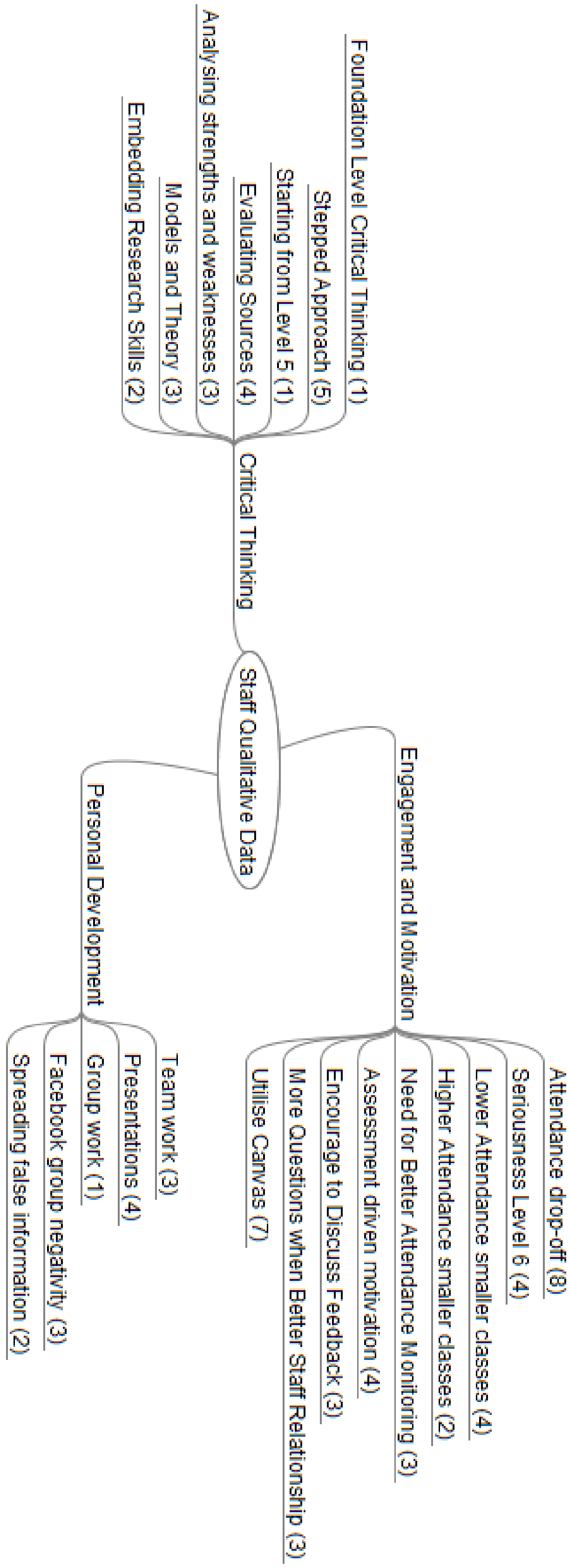

Figure 7.4: Thematic Analysis Map of the Staff Qualitative Interviews

7.3.2.1. Engagement and Motivation

Participants were asked about student attendance as a means to open up a conversation about student engagement. All participants stated that attendance tends to "drop off" as the year progresses. However, several participants did state that it tends to get better when students go to Level 6 as they are more aware of the "seriousness" of the year as it counts towards the majority of their final grade. In addition, Level 6 is the level in which a lot of the programmes become more specialist or give students options of modules to choose from, meaning that by this level students are studying the areas which they want to study. Furthermore, three out of the six participants stated that smaller classes actually had a lower attendance than larger classes such as lectures. However, Participant 1 and 4 stated that the smaller classes had better attendance as you're "noticed more" if you do not attend.

"Generally throughout the year, it tails off as the year goes on, but it does seem to get better at third year." (Participant 2)

There was also a discussion around the differences across students and the lack of consequences for non-attendance. Participant 3 discussed how the attendance on the forensic psychology and criminal justice programme is fairly high compared to a lot of other subjects within the science faculty as the students are "chased up" if they don't attend timetabled lectures/seminars/workshops etc. Participant 6 suggested that this lack of attendance should be taken more seriously in that there are serious consequences for non-attendance as occurs in some other science subjects, it was noted that those in pharmaceutical sciences are unable to graduate unless they show that they have attended their timetabled sessions. Several participants noted that the lack of attendance is a "serious problem" with Participant 5 stating that there was less attendance in smaller classes, where they did not "understand the reasons" behind why this may occur. As often, other participants would report that attendance in smaller classes was generally higher due to students "getting noticed more if you don't go".

"I think that we need a system whereby their financial bodies are informed if they drop below a certain percent without medical evidence". (Participant 6)

Participant 4 stated that this attendance gets worse in Level 5 and 6 because the students "know the system and know what they can get away with". This suggest that attendance should be worse as the years go on, however, this is in direct contrast with some of the other participants who believe attendance to be better once students reach Level 6. In addition, a lot of the participants identified the students to be assessment driven as their motivation to attend any timetabled sessions. There were some participants that stated that "interest" can sometimes be a motivator however, the majority of the staff members believed that this was only the case for a minority of students. Participant 1 also stated that it may be when they know a staff member well, they are much more likely to attend and engage with the material. There appeared to be a general agreement across the participants that students are very assessment focused.

"I think it's whether or not it's linked to an assessment. Although that's not ideal, we'd like them to come to all of the lectures and learn as much as possible whilst they're here but students are very assessment focused. If they can see a link between the lecture and assessment then their attendance tends to be higher" (Participant 5)

As well as this, several participants also discussed about students coming to them about feedback. In the sport programme, feedback online is given very minimally and students are encouraged to have meetings with their tutor in order to discuss feedback. This was different to other programmes where the majority of feedback is given online, but some students would often still seek guidance in order to understand the feedback in "language that they get". However, the psychology courses stated that they "encourage" students to bring their feedback at the "individual meetings" they have every semester. Participant 4 stated that "relatively few" students will come to discuss feedback, suggesting that students either are happy with the feedback they receive or do not utilise it. In addition, there was also the mention of students engaging in lectures and seminars by asking questions. Participant 1 stated that questions were more likely to occur if there is a relationship between the staff member and student. The participant stated that the student is “more likely to speak to the staff member and ask for support or help”. This suggested that engagement may be due to how comfortable the students feels with the staff member.

“I’m yet to find the pattern of what makes them engage. Last year it was very much Level 4’s were engaging and Level 5’s weren’t. But, this year, Level 5’s aren’t engaging, who are last year’s Level 4’s, and the Level 4’s are. So maybe it’s just that they can’t be bothered by Level 5.” (Participant 4)

However, Participant 4 stated that he found difficulty in understanding what it is that makes students engage. Although, participant 3 stated that this engagement could be helped through the use of discussion boards. These discussion boards are essentially online forums, within the virtual learning environment (VLE), where students are able to ask questions to members of staff and other students can also view the answers to the questions. Finally, all participants stated that they utilise Canvas (the virtual learning environment) by putting resources there for students to use. Resources that participants mentioned included; lecture slides, assignment outlines, module guides, marking and feedback and reading lists. Participant 6 had a strong view on the use of Canvas and how putting all the information on there can hinder students. She stated that students have a perception that they can “get by with what’s on Canvas”. She stated that she believed this was the reason for students tail off in engagement and attendance. Potentially suggesting that as students’ attendance tails off, their Canvas use may increase.

In summary, it appears that overall, students’ attendance should decrease as the year progresses but increase when the students enter their final year. Furthermore, there appears to be a split between whether there is better attendance in smaller classes compared to larger classes. This may be based on the subject and the student’s perception as to what is important. In addition to this, it appears that it is difficult to understand what gets students to engage in lectures in terms of asking questions. There appears to be a general consensus that Level 4 students are the most engaged, perhaps due to enthusiasm. Finally, there was a worry that as Canvas contains a lot of information this may be the reason for the decrease in engagement. However, there was also an agreement that students will attend if they believe something to be related to their assessment.

7.3.2.2. Critical Thinking

Participants were asked about how they embed critical thinking into the students and whether they have formal sessions on this skill. All participants stated that they include critical thinking into their degree programme, however, the way in which they did it and from what level would often differ. Interestingly, Participant 6 stated that she introduces critical thinking right from foundation level students as she believes that "they have got to university, they are critical thinkers". This participant was the only participant who did not have a "stepped approach" to introducing critical thinking. The majority of participants would discuss beginning Level 4 with discussing "strengths and weaknesses" with students to build this further as the degree progresses. The psychology subjects discussed how in Level 4 modules they introduce critical thinking by giving the students two media reports and a journal article to discuss the differences in how the research is portrayed. However, for the most part critical thinking is not technically assessed in students until Level 5 and 6, the most important year being Level 6. This would expect that the majority of students will increase over time in their critical thinking skill. In addition to this, Participant 6 also stated that their Level 3 students did overall better at their degree than someone who comes into the degree at Level 4. This is because critical thinking skills are introduced straight from Level 3.

"My Level 4 module, anything I tell them I then say "so why is this not true" and they respond with "you just told us it's a seminal study", and I would say "yes it is, but what might not be true about it". And start to critique it; whether that's the ethics, the methodology, the assumptions they've made based on the data, what doesn't fit with other things that they know. I want them to know from an early stage that they're allowed to think for themselves. Because they very

much have the attitude that if it's in a book it must be perfect and right, but that's not the case." (Participant 3)

In contrast, participant 1 stated that they did not do any critical thinking at level 4 as they felt that students were not "prepared". They did, however, state that they would introduce some critical thinking concepts, suggesting that they begin to lay down groundwork for the Level 4 students to build onto in their later years. This participant stated that level 4 was about understanding research that they are asked to read and not the "value" of that research. Participant 1 stated that this was not properly introduced in Level 5 and 6.

"We don't feel the students are necessarily prepared for that. Staff will talk them through some of the concepts of critical thinking towards the latter end of level 4. But, a lot of what we do in the initial stage is more getting students to look at research and what is being said. Not necessarily value." (Participant 1)

In addition to this, all participants stated that there were a lot of models and theory to the degree that the students had to "understand" at level 4 and to then "link to evidence" towards level 6. Models and theory have been suggested to be key in critical thinking. Participant 3 stated that it "opens their mind" and "changes the way they think about things" as often the models and theories come with "critique". This would suggest that this is an important aspect of furthering critical thinking in students.

"Theories are attached to most of the modules and this is what we hope the students will come to understand, especially the relationship between theory and evidence. How to use evidence to support or challenge a theory" (Participant 5).

In relation to critical thinking, a number of participants also discussed embedding research into students. Not only from the view of reading research but from conducting their own research projects. It appears that all participants stated that within their subjects, research skills are taught straight from Level 4 (or Foundation year) in order to prepare students for their dissertation research project that they have to complete in Level 6. These research skills feed into the students' critical thinking as they are expected to "think critically about the sources that they are reading" as stated by Participant 2. Furthermore, students are also "encouraged" to use journal articles as their main form of sources, some subjects stated that they move from a "book based system at Level 4 to specialising in Level 5 and 6 with articles" as stated by participant 1. This would suggest that students critical thinking skills should particularly increase in Level 5 and 6 as this would be the years in which these skills are more prominently introduced and students are more used to critiquing the literature.

In summary, it appears that the majority of subjects begin to introduce critical thinking and the skills surrounding this, such as research skills, in Level 4. However, the expectation for students to actually utilise these skills and to be assessed on their critique does not appear to come into play until Level 5 and 6. This would suggest that although students' critical thinking may go up slightly in Level 4, their critical thinking should significantly increase in Level 5 and 6. Furthermore, there may be some differences in this across subjects as it appears the animal focused subjects attempt to bring in the critical thinking aspect as early as possible whereas the sport related subjects tend to view Level 4's as ill prepared for this.

7.3.2.3. Personal Development

Participants were also asked about aspects of personal development. This included discussing areas such as; presentation skills, group work and communication. Discussion of group work tended to be more focused towards developing students to work as a team. Participant one, for instance, discussed how students will do "analysis together before writing up separately" in Level 4. This was also the case for Participant 2 and 5 who stated that students gather data together but write up separately. Participant 6 discussed how they had combined the group work with presentation skills whereby students "research" an area "longitudinally" that they then have to "present in front of the class". Conversely, participant 4 stated that he "won't touch" group work as there have been "too many complaints" due to the effect other students can have upon individual grades. However, participant 4 still stated that he believed group work to be important as "when you go to a work place you are going to be working with other people". For presentation skills, all participants stated that students would complete a presentation at some point across the course of their degree. It appeared that these two types of work, group work and presentation skills, tended to be favoured more in Level 4. However, there were some exceptions to this as Participant 1 stated that the presentations were more towards "the final year". Two participants also mentioned the struggles of doing presentations as participant 4 stated that there are exceptions which "make presentations too difficult for everybody" such as those who have "anxiety, stress or just don't like public speaking". Participant 3 stated that presentations usually begin as a group so "people aren't stood up on their own". This can mean that personal development in these areas can be difficult as students can find presenting to be daunting.

"They tend to do really well in the presentation, what they think is that they're going to do terribly because they're shaking and stuttering and sweating. But we aren't looking at that, we are looking at whether they have done the work." (Participant 3)

In addition to this, there was also a narrative surrounding the positives and negatives of having students discuss assignments with each other. Some participants discussed problems that they have encountered surrounding "Facebook groups" whereby students have spread "inaccurate rumours" – as was stated by participant 5 – which can be "unhelpful". However, participant 5 also stated that there are students who will not discuss with others as they are "isolated" from the rest of their cohort, but "probably still go on Facebook." Supporting this, participant 3 discussed an experience whereby students have argued over a due date of an assignment, causing her to "come in to 42 emails all asking the same question". This suggests that students often rely heavily on the discussions that they have with each other.

"It [the Facebook chat] went on for a week before the deadline and nobody emailed me to just clarify. They were all getting worked up in a frenzy in this Facebook chat and not a single student emailed me to just say, "when is the assignment due?"" (Participant 4)

In summary, it appears that although group work and presenting occur in the degree programmes, these can be problematic and tend to occur at different time points in the degree dependent on the subject. In addition, the narrative that the participants have created surrounding student discussions potentially suggest that students will listen to each other over the members of staff and potentially may even learn more via discussions with their fellow students.

7.3.3. Triangulation with Quantitative Data

Following the analysis of the staff qualitative data. It is now important to triangulate these outcomes with the outcomes of the assessment tool. In order to validate these outcomes, there may be a need to look at individual questions more in depth rather than the overall total of the subscales. It may also mean that the data would need to be split between levels and subjects for more meaningful analysis that can be compared with the qualitative outcomes of the staff interviews.

7.3.3.1. Engagement and Motivation

In the staff qualitative data it was found that students appeared to decrease in their attendance as the year progresses. There were two questions on attendance that were asked of students in the assessment tool. These were split into asking about attendance at larger classes i.e. lectures and smaller classes i.e. workshops and seminars. In order to align the outcomes of the interviews, one would expect that the attendance of students will decrease over time.

	Attendance in Lectures		Attendance in Workshops	
	Mean	SD	Mean	SD
Time Point One	4.58	0.55	4.70	0.57
Time Point Two	4.23	0.73	4.30	0.87
Time Point Three	4.09	0.76	4.12	0.91

Table 7.8: Descriptive Statistics for Attendance at Each Time Point.

Table 7.8 shows that the attendance in both lectures and workshops decreases over time. Students begin to agree less with the idea that they attend their lectures and workshops, suggesting that students attend less as the year progresses. Not only this, but both of these drops in attendance were found to be statistically significant (P>.001) suggesting that students' attendance at lectures and workshops significantly decreases across the year. In addition to this, there was found to be a split between whether student attendance was better in smaller classes than larger classes. The psychology subjects found that their smaller classes were potentially worse in their attendance whereas the rest of the subjects found that the attendance in smaller classes was better. There is also the expectation that students from pharmaceutical sciences will have better overall attendance as they are unable to complete their degree if they do not attend, as stated by one of the participants in the staff qualitative stage. In order to view whether these outcomes align together, the students were split into subjects of; Psychology, Natural Sciences, Animal Sciences, Sport Sciences and Pharmacy and Biomolecular Sciences. This was chosen to get an overall idea of the differences in these groups of subjects and due to low representations from some individual subjects, they were decided to be grouped for more meaningful analysis.

		Attendance Lectures		Attendance Workshops	
		Mean	SD	Mean	SD
Psychology	Time Point One	4.47	0.56	4.71	0.56
	Time Point Two	4.11	0.62	4.36	0.80
	Time Point Three	3.80	0.85	3.97	0.85
Natural Sciences	Time Point One	4.77	0.43	4.63	0.72
	Time Point Two	4.47	0.64	4.47	0.83
	Time Point Three	4.25	0.68	4.38	0.81
Animal Sciences	Time Point One	4.57	0.59	4.65	0.57
	Time Point Two	4.00	0.82	4.43	0.54
	Time Point Three	4.17	0.75	4.00	1.10
Sport Sciences	Time Point One	4.44	0.73	4.44	0.73
	Time Point Two	4.33	0.58	4.33	0.58
	Time Point Three	4.00	0.00	4.00	1.41
Pharmacy and Biomolecular Sciences	Time Point One	4.91	0.29	4.91	0.29
	Time Point Two	4.50	0.52	4.08	0.79
	Time Point Three	4.40	0.63	4.20	0.78

Table 7.9: Attendance in Students Across all Three Time Points Split by Subject.

When looking at lectures, it can be seen that the majority of the subjects decrease over time on both lectures and workshops. However, interestingly, animal sciences appears to increase again at time point three for lectures and pharmacy and biomolecular sciences appears to increase at time point three for workshops. For psychology and pharmacy and biomolecular sciences the differences across time points were significant for lectures ($P>.05$). However, the differences for the rest of the subjects were not significant, although changes can be seen by looking at the means. This was also the same for the workshops, where psychology and pharmacy and biomolecular sciences were also significant ($P>.001$). For looking at the differences between workshops and lectures; there appear to be some slight differences across subjects. Interestingly, students reported that they attend workshops more than lectures in both psychology and animal sciences. Although this aligns with the staff qualitative outcomes for animal sciences, it is in direct contrast to the outcomes for psychology. This may suggest that either students inaccurately report their attendance or the students who attend workshops are more engaged and therefore more likely to complete the assessment tool. Interestingly, natural sciences appeared to report less attendance in workshops than lectures, again in opposition to the staff qualitative outcomes, which is of interest. However, the participant who reported that students attend smaller classes does not actually teach those classes himself and reported it off what his expectations would be rather than experiences, which may explain the differences. However, the only subject which was significantly different between workshops and lecture attendance was that of psychology. Suggesting that students are more likely to attend a workshop than a lecture. As previously suggested, the lack of alignment with the staff qualitative outcomes may be due to students who are more engaged being more likely to voluntarily complete the assessment tool.

Furthermore, there appears to be a general consensus that Level 4 students are the most engaged, perhaps due to enthusiasm. This difference has previously been discussed within the quantitative outcomes stage. It was found that the changes across time with all of the students in their overall engagement was not significant, this suggests that student engagement does not decline enough for it to be a significant change. In addition to this, from looking at the means across the different levels, it would suggest that students actually become more engaged the higher the level. However, there may be some subject differences. Although, participants also stated that level 6 students become more engaged perhaps due to the “seriousness” of the final year, which is corroborated by the quantitative analysis. However, it may be meaningful to look at whether the differences across levels are significant. It has been found that at time one the differences across engagement were significant ($P>.001$) this suggests that, as seen in the summary section, the engagement significantly increases dependent on the level. This may suggest that students become more engaged the higher the level. This is interesting as the members of staff stated that participants become less engaged as they knew “what they could get away with” however, some participants reported level 6 were highly engaged.

However, this may be related to the discussion of Canvas. There was a worry that as Canvas contains a lot of information this may be the reason for the decrease in engagement in lectures. However, this may be the reason that students’ total engagement appears to increase. It may be that their utilisation of Canvas significantly increases and this is where they are highly engaged. Therefore, it is of interest to look into this question that is asked of the students.

		Use of Canvas (VLE)	
		Mean	SD
Time Point One	Level 3	3.92	1.26
	Level 4	4.44	0.78
	Level 5	4.72	0.46
	Level 6	4.50	0.64
Time Point Two	Level 3	4.29	0.95
	Level 4	4.44	0.77
	Level 5	4.29	0.76
	Level 6	4.27	0.65
Time Point Three	Level 3	4.57	0.54
	Level 4	4.43	0.59
	Level 5	4.14	0.69
	Level 6	4.15	0.90

Table 7.10: Student Use of Canvas Split by Level Across Time Points.

Interestingly, all levels appeared to decrease over the time periods apart from Level 3, who increased over time and level 4 who appeared to stay the same over time. This is an unexpected outcome and does not appear to align with the expectations and experiences of the members of staff. This may suggest that students’ engagement simply decreases as they are able to still do well in their degree with the utilisation of Canvas and therefore do not see the benefit in attending when they are able to succeed without. Overall, the engagement with

Canvas is fairly high, all around the mark of 4 out of a possible 5 points. It may suggest that in order to gain more meaningful results in the future that the assessment tool should potentially have a 7 point scale to attempt to gain smaller differences. However, there was also an agreement that students will attend if they believe something to be related to their assessment. This aligns with the outcomes of motivation that were discussed earlier in the quantitative summary section as the majority of students stated that the main reason that motivated them to attend was to get a better grade (starting at 47.1% at time one and increasing to 58% at time three). Finally, it would be beneficial to look at students' individual data in order to see if the engagement behaves in a way which is expected from the staff qualitative outcomes. A random sample of participants were selected. Participants were only included if they had data from all three time points. It was found that 52 cases had information at all three time points. A random sample of 5 students (10%) was taken by using a random number generator to select participants in order to look at individual journeys in engagement.

			Engagement (Max 95)		
Subject	Level	Gender	Time Point One	Time Point Two	Time Point Three
Animal Behaviour	4	F	59	58	60
Geography	6	F	66	55	58
Psychology	4	F	64	59	54
Biology	3	F	64	70	75
Animal Behaviour	4	F	62	52	52

Table 7.11: Random Sample of Individual Data for Engagement

As can be seen from Table 7.11, two participants' engagement decreased from time point one to time point two but then increased again at time point three. This would suggest that perhaps students become more engaged towards the end of the year as this is likely to be when students have their exams. Furthermore, it appears the level 6 student within the sample are more engaged than the other students at the start of the year, which aligns with the staff qualitative outcomes that students become more engaged in final year due to the "serious" nature of this level. One participant, from level 4 psychology, consistently decreased in their engagement over time, again aligning with the staff qualitative outcomes that students' engagement drops off as the year progresses. Interestingly, the level 3 biology student consistently increases in their engagement as the year progresses. This may suggest that the assessment tool is able to pick up individual differences with students. The members of staff will be likely to discuss the overall average of students rather than the individuals that are exceptions to the rule, and therefore this is showing that there will be some students who do not follow the norm.

In summary, it appears that some of the staff qualitative outcomes aligned with the quantitative outcomes whilst others did not. The decrease in attendance that was reported by members of staff was easily corroborated with the quantitative data. In addition to this, the differences in workshops and lectures also aligned with the discussions with staff apart from the psychology subject. This may be because students who choose to voluntarily take part in these kinds of assessment tool are generally more engaged and therefore may not be a good representation of the overall engagement of the cohort, as without an incentive the low engagement students are less likely to participate. Yet, a pattern of decreased engagement can still be seen. However, it was estimated, by some staff members, that the decrease could be likely due to students "getting by" using only Canvas. This may suggest an increase in the use of Canvas, however this could not be seen. When looking at individual data, there was some alignment with the staff qualitative outcomes however there were also individuals that did not align with these averages, suggesting that the assessment tool is able to show individual differences. It appears that the majority of what was found from the staff qualitative data could be seen in some patterns with the quantitative outcomes suggesting some validation for the measure. However, as these outcomes are from the experiences of the staff members it will also be of interest to hear from a student perspective.

7.3.3.2. Critical Thinking

Staff qualitative data revealed that the majority of subjects begin to introduce critical thinking all the way from level 4, but do not expect these skills to be utilised until level 5 and 6. However, there were some notable subject differences as the animal sciences stated that they have the expectation that students will utilise critical thinking right from Level 3 as this is a required skill for university. Psychology subjects also mentioned that they begin to introduce critical thinking strategies straight into level 4 tutorials and into a research skill module. In contrast to this, sport sciences suggested that Level 4 students are not prepared to take on such a skill from that level. In the quantitative data summary, there were some interesting differences across the different levels. However, in the majority of cases it appeared that generally critical thinking would increase from time point one to time point two, with some increasing again at time point three and some decreasing. This decrease may be likely due to retention rates and potentially students who were improving may have not taken part in the final stage of the data collection.

		Critical Thinking Total (Max = 75)	
		Mean	SD
Time Point One	Lv3	58.62	5.00
	Lv4	57.79	7.11
	Lv5	55.78	5.81
	Lv6	58.71	7.85
Time Point Two	Lv3	63.71	7.30
	Lv4	58.38	7.89
	Lv5	56.57	3.46
	Lv6	60.55	9.18
Time Point Three	Lv3	59.00	6.83
	Lv4	59.31	8.43
	Lv5	59.43	8.08
	Lv6	59.69	6.21

Table 7.12: Overall Summary of Critical Thinking Across Time Points

However, as there have been differences noted from the staff qualitative data, it may also be meaningful to check the differences across the different subject groups with critical thinking as this may affect the learning gain for this particular skill. It would be expected that the animal sciences and psychology start off at a higher level than other subjects and increase over time. It would also be expected, should the results corroborate with the staff qualitative outcomes, that the sport sciences will start off lower than the rest of the subjects as it was stated that this was not introduced at Level 4. There was some missing data for students at time points and therefore only the data that is available will be included, as well as if there was no participants for that level for that particular subject, data was also not included if only one participant fulfilled that demographic as it would not result in a fair comparison. Although this may be difficult to view due to retention rates, the data will still allow for some idea as to whether this is indicative of the staff qualitative analyses.

			Critical Thinking (Max 75)	
			Mean	SD
Psychology	Level 3	Time Point One	59.33	3.05
	Level 4	Time Point One	57.56	6.69
		Time Point Two	59.09	8.37
		Time Point Three	59.46	9.16
	Level 5	Time Point One	61.00	7.21
	Level 6	Time Point Two	64.33	9.29
Natural Sciences	Level 3	Time Point One	57.80	7.50
		Time Point Two	62.75	7.27
		Time Point Three	61.67	8.15
	Level 4	Time Point One	58.93	7.48
		Time Point Two	58.50	4.23
		Time Point Three	60.14	7.29
	Level 6	Time Point One	61.90	9.32
		Time Point Two	57.40	11.15
		Time Point Three	58.67	8.11

Animal Sciences	Level 3	Time Point One	59.50	0.71
	Level 4	Time Point One	55.60	6.87
		Time Point Two	55.25	5.74
		Time Point Three	55.00	7.94
	Level 5	Time Point One	52.40	6.80
	Level 6	Time Point One	55.00	9.45
		Time Point Three	60.67	7.51
Sport Sciences	Level 4	Time Point One	61.50	2.12
	Level 5	Time Point One	54.33	2.08
		Time Point Two	52.00	1.41
	Level 6	Time Point One	58.00	5.35
Pharmacy and Biomolecular Sciences	Level 3	Time Point One	58.67	4.93
		Time Point Two	68.00	9.90
		Time Point Three	56.33	7.23
	Level 4	Time Point One	59.42	10.18
		Time Point Two	56.50	9.85
		Time Point Three	59.83	7.91
	Level 5	Time Point One	56.50	5.01
		Time Point Two	58.67	1.15
		Time Point Three	56.20	4.92
	Level 6	Time Point One	51.50	3.54

Table 7.13: Critical Thinking Descriptive Statistics Split by Subject Groups and Level

It appears that there is no consistent pattern with critical thinking when looking at the totals as split by subject and year. For animal sciences, however, the level 3 students appeared to have higher critical thinking at time point one than the rest of the years and time points except from time point three for level 6. This aligns with the staff qualitative outcomes as the staff member stated that Level 3 students do overall better than those who come in at Level 4. This is shown by their critical thinking being higher than the Level 4 students. The differences across the levels were also found to be statistically significant (P>.001). Interestingly, for sport science, their critical thinking decreased over time to then slightly increase again at Level 6. This may be related to the retention rates or to the students decrease in engagement. For Psychology students, although they decrease from Level 3 to Level 4, after this the students consistently increase over time and across different levels suggesting that this aligns with the outcomes of the staff qualitative data. The difference between Level 3 and 4 may be due to the addition of students who have come from college who may drag down the average of the students as they would not have previously come across critical thinking, as mentioned by the member of staff from the animal sciences. There also appear to be some decreases in natural sciences and pharmacy. Again, this may be due to retention rates or decrease in engagement. However, it could also be indicative that students are unable to grasp the concepts of critical thinking and bringing that into their work. This will become more clear when discussing with the students themselves. In order to get a better understanding as to whether critical thinking is a journey which is more individual, it may be beneficial to look at an individual's journey rather than collecting the averages. A random sample of students were taken to view individual data. Students who have information available from all three time points were used. It was found that 52 cases had information at all three time points. A random sample of 5 students (10%) was taken using a random number generator.

			Critical Thinking (Max 75)		
Subject	Level	Gender	Time Point One	Time Point Two	Time Point Three
Pharmacy	6	F	49	55	59
Psychology	6	F	58	60	61
Psychology	4	F	49	59	59
Biomedical Science	4	F	48	51	56
Pharmaceutical Science	3	F	60	61	61

Table 7.14: Random Sample of Individual Data for Critical Thinking

When looking at a sample of individuals' progress, it appears that students generally increase in their critical thinking skills over time. The only participant whose skills do not increase is that of the Pharmaceutical science level 3 student. This may align with the staff qualitative outcomes as the majority of the staff members that were interviewed stated that they did not introduce, and expect students to use, critical thinking skills until level 5 and 6. However, interestingly, this student did appear to have a higher overall critical thinking skill than the rest of the students from different subjects. This could be due to the nature of the programme as Level 3 is a foundation year and as was stated by a staff member, most participants do better on overall skills when coming in at a foundation year rather than from college.

In summary, it appears that there is no consistent pattern when it comes to critical thinking. The students may increase or decrease over time and different levels. This may be positive as it shows the students' individual differences. When following individual data instead of averages it was found that the sample generally increased in their critical thinking skills over time which aligned with the staff qualitative data. Overall, it appears that the assessment tool is valid in its measurements as it often aligns with the staff qualitative outcomes however, it does show a sensitivity to individual differences and to changes over time.

7.3.3.3. Personal Development

Personal development was a more difficult and varied area when it came to the staff qualitative outcomes. Although group work and presenting skills are developed in the degree programmes, the time that these occur can be completely different depending on the subject. This would therefore suggest that it would be beneficial to view the quantitative outcomes in relation to the individual subjects to be able to triangulate the outcomes with the staff qualitative measures. The majority of these skills appeared to occur within Level 3 and 4, suggesting that gains should be seen straight from the beginning of the levels and increasing the higher the level. Three questions were asked of students in relation to presentation skills. For group work related questions, students were asked five questions. For the purpose of this triangulation, the time points were not considered and an average was taken across the three time points. This was done as the staff qualitative outcomes indicated that the differences were due to the amount in different levels rather than across the year. If data was missing, or only one student completed that area and level, it was not included in the table.

		Presenting (Max 45)		Group Work (Max 75)	
		Mean	SD	Mean	SD
Psychology	Level 4	30.82	6.67	40.77	9.63
	Level 6	37.50	0.71	63.00	4.24
Natural Sciences	Level 3	30.33	5.51	48.00	6.00
	Level 4	29.40	4.72	49.80	13.37
	Level 6	34.25	5.06	60.25	5.06
Animal Sciences	Level 4	28.67	2.08	57.67	10.01
Pharmacy and Biomolecular Sciences	Level 3	34.00	1.41	51.50	3.54
	Level 4	28.00	3.39	38.00	6.44
	Level 5	32.00	5.20	42.67	10.41

Table 7.15: Descriptive Statistics for Presentation Skills and Group Work Split by Subject and Level.

It would appear that the majority of the subjects showed similar averages regardless of the subject that the student was studying. This is supported by statistical significance as none of the differences across levels across the subjects were seen as significant ($P>.005$). This suggests that all subjects appear to introduce these presenting and group work skills at similar levels and have similar gains in the outcomes of those skills. This is supported by the staff qualitative outcomes as the majority of subjects stated that they would begin these skills from Level 3 and 4 and continue to develop them across the degree. However, the sport science subjects were not represented here due to a small sample of only one participant which had inputted across all time points for these questions. This would mean that it is difficult to know whether the sport sciences align with the staff qualitative outcomes, as this participant stated that these skills were mostly developed in Level 6. In addition to this, the staff qualitative outcomes also potentially indicated that students will favour learning from their fellow students rather than from members of staff. Questions were asked of students about their time spent with other students preparing for assignments as well as being asked about whether they learn more from fellow students than staff. It would be beneficial to look at the quantitative outcomes of these questions to view whether they align with the staff qualitative data.

	Learn More From Students		Asked Another Student for Help		Prepared For Exams with Other Students		Prepared for Assessments with Other Students	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Time Point One	2.66	0.91	3.01	1.03	2.91	1.10	3.26	1.08
Time Point Two	2.55	1.05	2.85	1.13	2.90	1.28	3.26	1.12
Time Point Three	2.50	1.02	2.60	1.02	2.72	1.10	3.24	1.02

Table 7.16: Descriptive Statistics for Students Development with Other Students.

Interestingly, it would appear that students' involvement with other students decreases over time. This may suggest that students begin to see the benefits of working alone over working with a group. In addition to this, over time students also waiver in their opinion that they learn more from students than staff. This may be because, as staff qualitative outcomes have suggested, there have been problems of students getting the wrong information, spreading and causing problems. This would align with the staff qualitative outcomes as it would appear that students begin to favour steering away from other students, perhaps due to the fact that students can give false information. However, none of these differences were statistically significant ($P>.005$) suggesting that students do not significantly decrease in these behaviours over time.

In summary, it would appear that students presenting and group work skills align with the staff qualitative outcomes. It was found that students across the different subjects appeared to be on similar levels when in the same year group. As the differences were not significant, this further supports this as it shows that students' personal development appears to be the same across the subjects, as was reported by the staff qualitative outcomes. Finally, when looking at how students interact with each other, it also appears that this tends to slightly decrease, perhaps due to misinformation that was reported from staff qualitative outcomes. However, this decrease was not statistically significant and therefore, suggests students do not differ greatly in this behaviour over time. Further information is therefore needed in order to help explain these potential differences, This will be therefore beneficial to discuss with students themselves.

7.4. Student Qualitative Data

The current qualitative outcomes that have been collected focus on the perspective of staff members whereby their expectations and experiences of students were discussed. An overview of the participants and their demographics can be seen in Table 7.17.

	Age	Gender	Subject	Year
1	20	F	Zoology	Final Year
2	23	F	Biomedical Science	First Year
3	19	M	Sport and Exercise Science	Second Year
4	28	M	Biomedical Science	Second Year
5	21	F	Biology	Final Year
6	22	M	Geography	Final Year
7	22	F	Wildlife Conservation	Final Year
8	24	M	Pharmacy	Final Year
9	20	F	Biochemistry	Second Year
10	21	F	Biomedical Science	First Year
11	27	F	Animal Behaviour	First Year
12	21	F	Wildlife Conservation	Second Year
13	19	F	Sport Psychology	First Year
14	20	F	Biology	Second Year
15	21	M	Geography	Second Year
16	20	F	Forensic Science	First Year
17	20	F	Sport and Exercise Science	Second Year
18	33	F	Healthcare	First Year
19	24	F	Sport Psychology	Second Year
20	19	F	Sport and Exercise Science	First Year
21	20	F	Biomedical Science	Foundation Year
22	21	F	Forensic Science	Second Year
23	18	F	Biology	First Year
24	24	F	Animal Behaviour	Second Year
25	21	M	Science and Football	First Year
26	20	M	Science and Football	Second Year
27	26	M	Sport and Exercise Science	Second Year
28	19	F	Zoology	Second Year
29	20	M	Pharmacy	First Year
30	19	F	Sport and Exercise Science	First Year
31	25	F	Biomedical Science	First Year

Table 7.17:Participant Demographics from Student Qualitative Survey Data.

Participants consisted of 31 undergraduate science students with an average age of 21.84 and standard deviation of 3.26. 70.97% of participants were female and 29.03% of participants were male. Furthermore, the sample consisted of 16.13% of students in their final year, 41.94% of students in their second year, 38.70% in their first year and 3.23% in foundation year. Subjects have been colour coded in Table 7.17 to the school which the participants belong to, this can also be seen from the figure below.

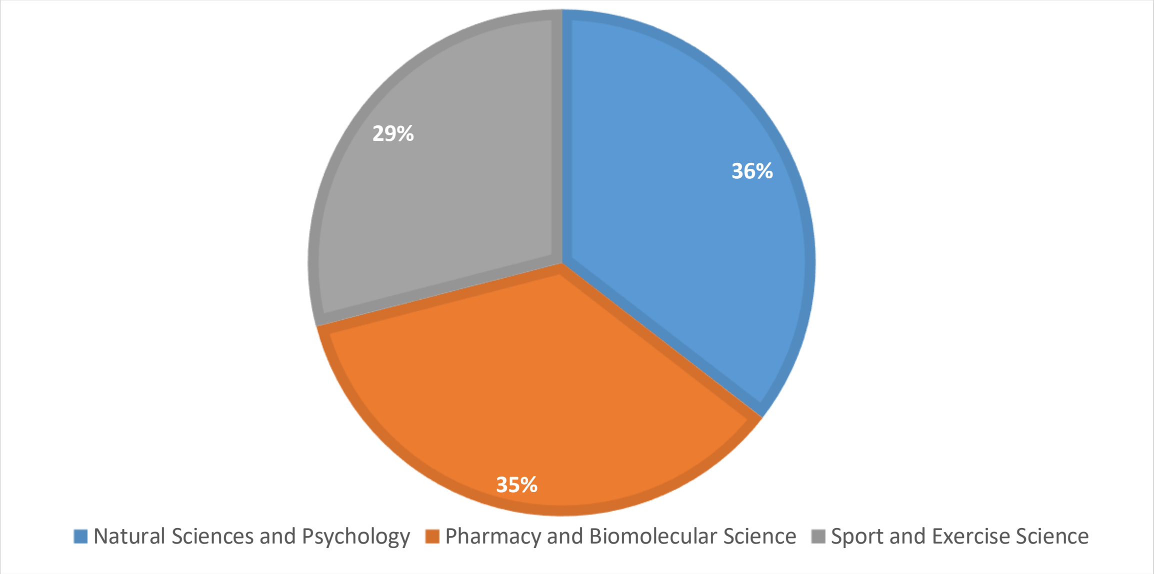

Figure 7.5: Breakdown of Participants by School for Qualitative Survey Data.

The number of participants from each of the different schools was roughly equal. Meaning that there was a fair representation from all of the different schools, within the faculty of science, for this stage of the research. Participants were asked questions which were similar to those of the staff interviews. This included beginning quite openly by asking students to discuss their degree, what it involves and the skills that they are gaining from this. This survey would then lead into more specific questions about critical thinking, engagement and personal development as was done with the members of staff. This data was also analysed with thematic analysis. The steps for this kind of analysis were taken from Braun and Clarke (2008) as was previously discussed in other qualitative chapters.

7.4.1. Analysis

There was a wide variation in how students chose to answer their questions. Some students would go into heavily detailed, in-depth answers whereas others would write a single word or potentially a sentence. This meant that there was some data which was much more valuable than others. However, as there was quite a substantial sample of students with a representation from each school, there was still data available that could be used to align with the quantitative outcomes and triangulate with the qualitative staff outcomes. In this analysis section, an overall picture will be given on what the students discussed in their surveys. This data will then be triangulated with the quantitative data and the qualitative data from the members of staff in order to gain a full understanding into the validity of the assessment tool. Interestingly, it appeared that there was a vast amount of individual difference from students in all areas of engagement, critical thinking and personal development. This appears to support the quantitative data as there were many individual differences between students. Although the staff qualitative data discussed the general consensus of the subjects, there would still be students who were anomalies to this. As previously discussed, it is likely that these anomalies may be more likely to engage in the assessment tool as they are more likely to be highly engaged students. This would explain potential differences between the staff qualitative data and the quantitative outcomes. In this analysis section, the qualitative data from the students will be discussed. This data was split into the same themes in order to make for easier comparison and triangulation.

7.4.1.1. Engagement and Motivation

Students were asked about their lecture and workshop attendance. They were asked about how this attendance had changed across the year, and if they were higher than first year how it compared across levels. They were also asked, if there had been any change, why this was, as well as their motivations for attending. Of the 31 students, 20 of them stated that their attendance had stayed the same during their time at university. 15 of these 20 stated that this was due to the fact that they “attend all lectures” and therefore, their attendance had not changed as they were already highly engaged. The reasons that students would often give for this would be that it helps with getting “higher grades”, some of which stated that it would be a “waste” to not attend lectures as they help to “understand material more”. There was also a general consensus that attendance was important as the work would “count towards the grade”. Participant 10 stated that they “knew, and still know, how important they [lectures/workshops] are” to their degree. Furthermore, Participant 24 stated that “lecturers have always stated that higher attendance strongly correlates to higher grades.” This is showing that students who are highly engaged with their degree tend to view lectures as more important to their overall degree classification. This would be a potential area for triangulation with that of quantitative outcomes as the assessment tool asked students about what they viewed as important. Furthermore, the fact that the majority of the students answering this survey appeared to be highly engaged, may suggest that the outcomes of the quantitative assessment tool may reflect this, as previously stated, students who are highly engaged are more likely to voluntarily engage in research.

“I attend all lectures so I attend lectures the same amount. Every lecture is different than the last as it may be a different lecturer or the same one with a different topic but it does interlink in some way and can help understand the topic more. It is helpful if you attend the lecture as if the lecture notes mainly have pictures / diagrams, then you can write more of the topic.” – *Participant 12.*

However, there were also 9 students in this cohort who stated that their attendance had actually decreased. Participant 6 discussed the idea of having to play catch up due to a “pile-up of work” which meant that their attendance therefore suffered. Similarly, participant 11 stated that they found attending lectures difficult due to being a single mother and having to sort out childcare. This would suggest that some students whose attendance appears to suffer may not be due to a lack of engagement with the course but due to exceptional circumstances. However, the rest of the students who stated that their attendance had decreased appeared to have a similar reasoning behind this. A lot of students stated that they found using “online materials” and “reading the slides” in their own time to be “more useful” than attending the timetabled sessions themselves. There was also an idea that as attendance is “not recorded” it is easy to “skip lectures” and to catch up by watching “recorded lectures” online. In addition to this, students in this mindset would state that they often found lecturers “unengaging” and therefore were unmotivated to attend. This may go some way to explaining why some students' attendance may change, it may be more likely to be due to the lecturer themselves and their delivery, rather than anything to do with the content.

"I do not find the lecturers or the lectures useful, I get more done if I stay at home and study myself." – Participant 22.

Finally, in terms of attendance, two students actually stated that their attendance had increased. It is clear that an increase in attendance may be more of a rare occurrence, and students will often stay consistently highly engaged during their degree, or their attendance decreased. The two students who mentioned that their engagement increased gave very similar reasons. These two students were a Level 5 student and a Level 6 student, who both stated that they attended more as their grades now "counted towards their degree".

"I attend lectures more because it matters the most now, grade wise" – Participant 7.

When students were asked about the topic of answering and asking questions when in lectures and workshops; the majority (74.19%) of students stated that they stayed the same throughout the year in the amount that they asked and answered questions. This would mean that if the student was quite confident, they would generally always answer and ask questions and continue to do so, and if a student was quiet and shy they would not answer or ask any questions and would also stay that way throughout their degree. The majority of students would state that they did not "answer questions" for fear of "getting them wrong". This shows the need to make students feel comfortable and may also explain some individual differences between student engagement, as they may only engage when they feel comfortable with the lecturer.

"I don't think I've ever answered a question in a lecture or workshop unless picked upon. Similarly to most people I think, the environment doesn't really make you feel comfortable in potentially getting the answer wrong" – Participant 26

However, there were 8 students who stated that the amount that they answered questions did change over time; 3 of whom stated that they answered more and 5 of whom that they answered less. The students who stated that they answered questions more stated that it was due to an increase in “confidence” which appears to be in relation to their personal development. Therefore, it may mean that if students are given opportunities to develop their confidence, they may also be able to engage more with the lectures. In addition, the students who stated they answered questions less often had different reasons for doing so. Participant 4 stated that because all the “content was new” they were afraid of answering incorrectly and therefore answered questions less. There was also a personal experience that was described by participant 12, who stated they were “moved a year down” so found it was easier to keep to themselves. In addition to this, participant 25 stated that a change in the “format of lectures” has meant that the “amount of direct questions being asked” has changed and therefore led to fewer questions being asked. Finally, another participant stated that it was often difficult to answer questions due to the way in which lecturers “answer it themselves”. This may suggest that students’ reasons for being less engaged often differ based on individual experiences, there does not appear to be one particular reason.

“I often find that lecturers don’t always allow people to think or discuss a question and they will usually answer it themselves after a couple seconds of silence.” – Participant 30.

As well as this, students’ independent learning hours appeared to overall, increase, with 64.52% of students stating that this was the case. The majority of students would state that this was due to “deadlines” and “work increase” with some even stating that the “increased difficulty” and “importance” of the year caused them to increase. Furthermore, some students stated that their “disappointment” in the outcomes of their grades from previous years had caused them to “put more effort” into their current year. Participant 5 stated that it was the “research project” in the final year that caused an increase of hours as the assignment was worth a high amount of their grades. However, there were 9 students who stated that their work had stayed the same. These students were students who had always put in a large number of hours or had found the number of hours which allowed them to “achieve the grades in which they wanted to achieve” and therefore felt no need to increase this. However, there were 2 students who stated that their independent hours had decreased one of which was due to stress and one due to working full time. This further highlights the fact that the students who engage with these assessment tools are often highly engaged, and for those who are not it is likely to be due to extenuating circumstances. This also highlights the individual experiences of students during their degree.

"I decided that, for myself, I needed to work harder as the modules were less suited to my interests and what I understand so I need to work more so that I can get a better understanding." – Participant 10.

Furthermore, students were also asked about their use of Canvas, the virtual online learning environment. 61.29% of students stated that their use of Canvas stayed the same over time, the majority of students stated that using Canvas is "important to use and check every day", as students would often "use it to print lecture slides". As well as this, participant 18 stated that they use Canvas to "review notes prior to sessions". Furthermore, 10 students stated that they used Canvas more due to "exam periods" which made them use it more as they "find content beneficial". However, there were 2 students who said that they used Canvas less as they "knew what was expected of them" and therefore did not need to use it as often. In addition to this, when students were asked about preparation there was an equal split between doing more preparation and doing the same. The students who stated that they did the same amount would either state that they "only read when necessary" or would say that they "always take notes on what lecturers expand on". This would suggest that students tend to stick to what they find works for them. In addition, the students who reported doing more preparation usually stated that it was due to a need to do "further reading" from a lack of "understanding". However, there were 5 students who found that they prepared less. This was split between students stating that they were "too busy" or that they found it "wasn't necessary" as they were able to "record the lectures". This suggested that generally students would start as they meant to go on and would stick to the amount of preparation that they would do at the start of the year.

"Extra reading is relatively constant, doing more around the subjects I find more complex. I write rough notes during lectures and then write them up neat including any extra info obtained from the extra reading, note taking does increase near exam time" – Participant 2.

In summary, it appears that the students who are engaging with the project are overall quite highly engaged students. Students' attendance generally appeared to have stayed the same or to have increased, with a lower number of students decreasing in their attendance. Furthermore, when it came to answering and asking questions, students appeared to generally stay the same throughout their degree. In addition, students also stated that the number of independent hours had increased due to a bigger workload. When it came to utilising Canvas, students tended to stay the same in their use, suggesting that they would find the amount of time which worked for them. Finally, with preparation, their seemed to be an equal split with the same amount and more preparation being done, suggesting that students perhaps find the amount that works best for their grades and schedule.

7.4.1.2. Critical Thinking

Student critical thinking was also discussed whereby students were asked about whether they were taught about the subject and if so, how. 3 of the students did not answer questions about critical thinking, the reasons why are unknown but may be likely due to not understanding the concept. 39.29% of students who did respond to the question stated that they had not been taught about critical thinking. Participant 22 stated that although they knew they "had to use it" they had not "been taught about it very well". 70% of the students who claimed they had not been taught about critical thinking were level 4, first year, students. This suggests that the fact that students had not been taught about critical thinking is due to the level which they are at.

"I think at level 4 we are just learning facts and not done much critical thinking at all apart from in our poster presentation." – Participant 18.

Students who stated that they had been taught critical thinking would mention that they would learn this through "tutorials" and sometimes in "essays". Students appeared to be aware that it was “expected” when they had to “review sources”. Participant 15 mentioned that they were told not to take things on “face value”. Several students also stated that critical thinking had “increased throughout the year” and in particular had “changed since level 5”. This was further supported as participant 26 stated that “the majority of critical thinking has been taught and expected” in second year rather than first year. This is suggesting that the higher years is when critical thinking comes more into play, aligning with previous qualitative research and quantitative outcomes.

“We have been taught critical thinking in preparation for writing essays along with which sources to use and which to avoid using. This has increased throughout the year as we have been further introduced to the proper ways of writing essays and reports at this level.” – Participant 23.

In summary, critical thinking appears to occur mostly into the second and final year of the student’s degree. Students who were in the first year of their degree stated that they had not been taught about the concept as of yet. Furthermore, students discussed how they were often taught in tutorials and were told not to take things on face value. It appears that these outcomes align with the previously found outcomes of the staff qualitative data and student quantitative data, this will be discussed in more detail when triangulating the data.

7.4.1.3. Personal Development

Finally, students were also asked about areas of personal development such as presentation skills, group work and discussion of exams and assessment with other students. As was found from the staff qualitative data, it appears that the majority of the presentations being conducted were done in groups meaning that the benefits of presentation skills and group work are being simultaneously tested. Only three of the participants in this sample stated that they had not done a presentation in their current year. The other 28 participants all stated that they had completed a presentation. Students were also asked whether they believed doing a presentation had benefited them. 5 of these participants stated that it had helped them to develop “presentation skills”, there were also 5 students who mentioned the development of “public speaking”. As well as this, as the majority of students’ presentations were in a group format, there was also a talk of “developing teamwork”. Furthermore, several students also mentioned that they felt it was needed for “real life” and “jobs” so it was important to be able to do. In addition to this, a number of students discussed how presenting had helped their “anxiety” and “confidence” as they had begun the year by not being able to “speak in front of others”. However, there was one student who stated that they are still “nervous as anything” about presenting and this has not changed therefore they don’t feel it has benefited them. This suggests that the majority of students find presenting very beneficial to their own personal development and in terms of going into their career.

“I do believe presenting has benefited me because I have gained a lot more confidence from them and realised I don't need any notes to prompt me if needs be. I feel like it has made me feel confident in a way that I am able to talk to people without being shy all the way through.” – Participant 17.

As previously discussed, a part of the group work that students were able to engage in was presentations. However, 6 participants stated that they had not completed any group work during the year. Several students discussed the group work helping with “team working skills” with others mentioning “communication” and “organisation”. In addition, participant 2 also stated that it is a good “opportunity for a lot of development” as someone often will have to take a “leader role”. Participant 18 said that group work “made coursework more fun” and that they “felt the group worked well”. Furthermore, like with the presentation skills, students also stated that it would be beneficial when “in a proper job” and that it gave an “insight” into “what it is like to try and work towards a goal with other people”. However, there were also three students who stated that they felt like it was not beneficial. They stated that they “ended up doing 90% of the work” and that it was “unfair” as other students would “benefit even though they never attended”. Participant 31 even stated that they found the situation “stressful”.

"I have gained communication skills and team working skills. However, there is a lot of group work and it can sometimes be annoying if people don't put their effort in. This has started to get tiring at the end of this year!" – Participant 24.

Students were also asked about their discussion of exams and assessments with other students. This was seen as part of personal development as students will develop communication skills. One participant did not respond to questions about discussing with other students, perhaps suggesting they don't engage in this behaviour. Four of the students stated that they don't discuss with other students, one stating it was not for "any particular reason". However, reasons that student gave for believing this to be useful was so that they were able to discuss their "concerns about assessments". Some students stated that they "help each other with coursework and revise together". Furthermore, participant 14 stated that they have discussed it more with other students as their "workload has increased", this was supported by participant 15 who stated that they "talk more when assignments are due". In addition, some students even stated that they used other students to help to "improve" as they can "compare" with other students if they have "received higher grades". Finally, participant 29 specifically stated that they used this as a way to know they weren't "alone struggling" which gave a sense of group dynamics that occur within a student population.

"I think it's normal to discuss these things with other students. We've been doing this for a long time, so nothing has changed in that sense. Having said this, I tend to focus on my grades and try not to compare too much to others. If I see someone with a higher grade than me, I might talk to that person a bit more in order to learn something from him or her." – Participant 27.

In summary, it appears that students have more opportunity for presentation skills than they do of group work. However, the students generally seem to see the benefit of engaging with both of these kinds of assessments in their personal development. Overall, students understood that these assessments will improve their skills for when it comes to their career. In addition, students have also seen the benefits in discussing their exams and assessments with other students. Student qualitative data has shown some potential reasoning as to why some students may be different from the vast majority.

7.4.2. Triangulating the Data

The final analysis to be conducted for this project is a final triangulation across the staff qualitative data, student qualitative data and quantitative outcomes. This will help to validate the outcomes of the quantitative data and the assessment tool itself as it is hoped that there will be some matching across these subjects. The summarisation of the triangulation with the staff qualitative data and quantitative outcomes, suggested some differences. Although a decrease in attendance could be seen, and was suggested, by members of staff; there was some misalignment with the reasons why this may occur. However, by talking to students themselves this may have helped to shed some light on the potential reasoning for this.

7.4.2.1. Engagement and Motivation

The first area that was noteworthy in the qualitative outcomes of students was that it appeared that students, who believed that their attendance was important in order to gain higher grades and to do well, would be more likely to attend. As students were asked questions on importance in the quantitative data, it can be seen whether this aligns together and therefore would be further validation for the assessment tool. In order to see if there is a relationship between these two variables, a correlational analysis will need to be conducted. Spearmans Rho Correlation was deemed to be the most appropriate as the data is that of ordinal level as well as a monotonic relationship being found between the two variables. Therefore, a Spearmans rank-order correlation was run to determine the relationship between the student's lecture attendance and their view of the importance of attending lectures in order to succeed in university. A weak positive correlation was found at time one which was statistically significant ($P>001$). Although the relationship is fairly weak, it does go some way to triangulating with that of the outcomes of the student qualitative data as it has shown that there is in fact a positive correlation between these two variables. A cross-tabulation table was created for these two variables in order to gain further insight into this relationship. It was found that 50.8% of students stated that they attended lectures “all of the time” AND selected the “Strongly Agree” option for attending lectures being important to succeed at university. This suggests that there is a triangulation with the qualitative outcomes as the largest percentage of students both attended all lectures and believed this to be important to succeed.

When looking at attendance, the qualitative student data appeared to indicate that the majority of students were highly engaged and would stay that way. This may support the idea that the students who were highly engaged were more likely to participate in these kinds of assessment tools and therefore the outcomes may not reflect the majority of the students on the course. However, 9 of the students did state that their attendance decreased over time as they were more likely to just use the online materials as they found theseto be just as effective. This aligns with the quantitative outcomes and the staff qualitative outcomes as the staff reported that attendance decreases as the year progresses and this was also found to be the case with the quantitative data. In addition, the two students who stated that their attendance had actually increased gave reasons that alluded to the fact that the levels which they are at now count towards their final grade. This was further supported in the qualitative staff outcomes as some of the staff members stated that level 6 students’ attendance would often increase as they knew this was the time to be “serious”. This is also reflected in the quantitative data as can be seen by table 7.18, level 6 attendance appears to start off higher than other levels except from foundation level (level 3) suggesting that students may begin the year with the notion that this final year is in fact more “serious” and therefore requires more attendance. This suggests that there is some validation between these three sets of data as they appear to be showing patterns which converge together. The fact that there are some differences is also beneficial as it explains some of the individual differences of students.

		Attendance Lectures		Attendance Workshops	
		Mean	SD	Mean	SD
Level 3	Time Point One	4.85	0.38	4.92	0.28
	Time Point Two	4.43	0.54	4.29	0.95
	Time Point Three	4.14	0.38	4.14	0.90
Level 4	Time Point One	4.56	0.53	4.76	0.51
	Time Point Two	4.25	0.64	4.42	0.74
	Time Point Three	4.02	0.81	4.12	0.86
Level 5	Time Point One	4.39	0.78	4.33	0.69
	Time Point Two	4.14	0.90	4.29	0.76
	Time Point Three	4.14	1.07	4.14	1.07
Level 6	Time Point One	4.57	0.50	4.57	0.74
	Time Point Two	4.18	0.60	4.09	0.83
	Time Point Three	4.15	0.69	4.08	0.76

Table 7.18: Attendance in Lectures and Workshops Split by Level and Time Points

Moreover, independent learning hours had yet to be discussed as this was something that staff would likely be unable to estimate. When students were asked about this, there was a variety of answers. The majority of students stated that their independent hours had increased as the year had progressed due to workload. However there were some students who stated that they had managed to find the number of hours which worked best for them and therefore, the number of hours stayed the same. In addition, there were also students who stated that there were circumstances which led their number of hours to decrease. This may go some way to explain the lack of logical pattern that was seen in the quantitative data. It appears that students have their own experiences with the number of independent hours and what it is that works best for them in order to achieve the grades which they want to achieve. This goes some way in explaining the fact that there will be individual differences across students; some students will not need to put in as many independent learning hours as others to achieve the same grade.

Furthermore, it was found that the majority of students (74%) stated that their engagement with answering and asking questions in lectures and workshops stayed the same. It was found that generally if a student was confident, they would start off answering questions and continue to do so as the year progressed and if students were shy, they would never ask questions and continue to do so. However, there were also found to be some individual differences where students would state that this has increased as their confidence has grown or decreased as they found it more difficult. In order to test whether this triangulates with data it would be beneficial to take a small sample of individual data to see whether this is in fact true, that their ability to answer and ask questions stays stable over time. As previously discussed, it was found that 52 cases had information at all three time points. A random sample of 5 students (10%) would therefore be taken by using a random number generator to select participants in order to look at individual journeys in their engagement with questions during sessions.

			Engagement with Questions (Max 10)		
Subject	Level	Gender	Time Point One	Time Point Two	Time Point Three
Psychology	6	F	5	5	6
Psychology	4	F	3	3	3
Wildlife Conservation	4	M	6	4	6
Biology	4	F	5	3	5
Biomedical Science	4	F	5	6	6

Table 7.19: Individual Results for Engagement with Questions.

It would appear that the majority of the students that were randomly selected tended to stay fairly stable over time, with no significant differences across the time points ($P>.05$). This suggests that this data triangulates with that of the student qualitative data as there is a clear pattern that students tend to stay the same with their engagement with answering questions. When students were asked about their use of Canvas, a high percentage (61.29%) stated that their use of Canvas stayed the same across the year. This is supported in the quantitative data which was utilised in the triangulation with the staff qualitative outcomes. The change in Canvas appeared to stay stable across time as the differences between these time points were not significant ($P>.05$). This suggests that students are not using Canvas more in order to compensate for a decrease in attendance, it appears that students may be finding what works best for them in order to achieve the grades which they want to achieve. Interestingly, this suggests that the assessment tool is representing the reality of what the students are stating.

In summary, it would appear that by looking at both staff and student qualitative data and triangulating this with that of quantitative outcomes, there appears to be some agreement when it comes to attendance. In addition to this, the qualitative data with students has found that there is alignment between the use of Canvas, independent hours and engagement with questions. This means that the use of both sets of this qualitative data has been useful in validating the outcomes of the assessment tool and suggests that perhaps these questions are being asked in a way which aligns with that of reality.

7.4.2.2. Critical Thinking

For critical thinking, this was not discussed as much in the qualitative survey with students. They were asked about whether they had learnt about critical thinking and if so, how this was done. A lot of students would not leave rich and detailed answers, making it difficult to understand and analyse this section of data. However, the data that was gathered is still able to be triangulated with the qualitative staff outcomes and quantitative outcomes and they both have rendered similar results. It was found that some students stated that they had not yet learnt about critical thinking, 70% of which were level 4 students. This aligns with the staff qualitative data in which many of the participants stated that students were not taught about critical thinking, or expected to use it, until the later years. These outcomes also align with those of the quantitative data, as in the triangulation with the staff qualitative data, it can be seen that generally level 6 students had higher levels of critical thinking than level 4 and 5 students. This would suggest that the assessment tool is able to distinguish between the different levels of students. However, across the time points it did appear that some of the levels would catch up with the other levels. At time point three there were no significant differences across the different levels ($P>.05$). This is likely due to retention rates rather than a true reflection of the data. Therefore, it would suggest that the assessment tool is valid in understanding the differences between the different students, further validating the measure.

7.4.2.3. Personal Development

The outcomes of the presentation and group work questions from the student qualitative survey suggested that students found these personal development opportunities to be advantageous towards their chosen career. In the personal development section of the quantitative assessment tool, students were asked about whether they believed that the development in their communication skills and the development in their teamwork skills would be useful towards their chosen career. Therefore, it would be beneficial to see whether the average student agreed with these statements, as it would align with the outcomes of the assessment tool. This was not discussed in the qualitative outcomes for the staff, this would suggest that it has been beneficial to discuss this information with the student population as more insights are coming from the surveys themselves than solely from interviews with staff.

	Communication		Team Work	
	Mean	SD	Mean	SD
Time Point One	4.42	0.66	4.32	0.74
Time Point Two	4.19	0.90	4.21	0.88
Time Point Three	4.21	0.83	4.14	0.82

Table 7.20: Personal Development Skills and Students Value to their Chosen Career.

As can be seen from Table 7.20, it appears that students highly agree with the idea of communication and team working skills being important for their chosen career. This also does not appear to waiver over time, as the differences across the time points were not significant (P>.05) and therefore were generally stable in their opinions on the importance of these skills. In addition to this, there were also several students who indicated that they did not like group work as they felt that this was “unfair”, this has also been confirmed by the staff qualitative outcomes in which some also stated that they would not “touch” group work due to the problems it arises. Within the personal development section of the quantitative assessment tool, there was a question to students as to whether they “avoid group work when possible” suggesting that they do not enjoy group work. It would be of interest to view the average opinion of students for this question and to see whether this triangulates with some students' dislike of group work.

	Avoidance of Group Work	
	Mean	SD
Time Point One	3.18	2.78
Time Point Two	2.78	1.23
Time Point Three	2.85	1.17

Table 7.21: Descriptive Statistics for Student Avoidance of Group Work.

From the descriptive statistics data, it appears that the average for each time point is around the midway point. However, the standard deviations for this are arguably quite large as the scale is a 5 point Likert scale. Therefore, it may suggest that the majority of students are indifferent to their avoidance of group work. However, it would also suggest that there are students whose individual differences can also be seen due to the spread of the scores that can be seen from the standard deviations. The differences across the time points were not significant ($P>.05$) suggesting that the students' opinions on this are fairly stable over time i.e. if a student avoids group work at the beginning of the year, they may also do the same as the year progresses. This further suggest that there is validation for the assessment tool as the differences across students can be seen, but the average appears to hold a fairly neutral stand point towards group work.

In summary, it is clear that the quantitative data appears to correspond with the qualitative data from students and explains some of the variance that was not explained by the experiences of members of staff. Students stated in their surveys that they were aware of the importance of developing communication and team work skills, this was also reflected in the quantitative data. Furthermore, there were individual differences towards attitudes of group work in the qualitative data from both staff and students, where one staff member even stated they would not "touch" group work. This can also be seen from the quantitative data as there is an average of the midway point however there is a fair spread of the scores that can be seen from the standard deviation.

7.5. Discussion

The results of this section were analysed in order to attempt to indicate whether the assessment tool was valid in its outcomes. This meant that there were qualitative interviews with members of staff in which they were asked to discuss their expectations and experiences with students. As well as a qualitative survey measure for students where they were also asked about their own experiences during their current academic year. This qualitative data was used in order to triangulate with the quantitative outcomes of the assessment tool. It was intended that the quantitative outcomes could be explained by qualitative outcomes and that similar behaviours would be reported.

When looking at staff qualitative data there were some issues with alignment with the quantitative data. However, it was found that these issues could be explained by the information that was analysed from the student qualitative data. Interestingly, with the staff qualitative data, the majority of these stated that attendance decreased over time, which was found to align with the quantitative outcomes, however, with the student survey the majority of students stated that they attended all the lectures and this stayed stable over time. Nevertheless, this can be explained by the likelihood that students who are highly engaged are more likely to voluntarily fill in a time-consuming assessment tool from which they seek no obvious benefit. In addition, the student qualitative data also showed an important area of discussion in which students' reasoning for attending was often that it was seen as important. This was also triangulated with the quantitative outcomes as there was a positive relationship between attendance and the student views that attendance is important to succeed at university. Furthermore, the staff data which looked into student questioning revealed that lecturers often had difficulty in getting students to engage by asking and answering questions. This was asked specifically in more detail for the student qualitative outcomes where students stated that they would stay the same in their engagement with this. This was found to also triangulate with the quantitative data as when looking at individual results it could be seen that this would stay stable over time. All of this information is part of the engagement subscale of the assessment tool. Therefore, this suggests that the assessment tool accurately and validly asks the questions in the correct manor so that the students are able to understand what is being asked and that this aligns with the reality.

Critical thinking was of high interest as this was a novel way in which to understand this skill from students by asking numerous questions which collectively showed critical thinking as a skill and which were specifically tailored to the cohort of students that were being targeted. This also appeared to align with that of the qualitative and quantitative outcomes. Staff data indicated that, for the most part, critical thinking would not be taught and expected in level 4 but may be slowly introduced. This was found to be the case also for the student qualitative data as 70% of those who stated they did not yet learn critical thinking were level 4 students. In terms of the quantitative data, it would appear that level 6 students were always higher than the lower levels. This is further validation for the measure in its subscales as it appears that the outcomes can be validated from the qualitative data.

Finally, when looking at personal development, there was also strong triangulation for the qualitative and quantitative data. It was found that there were no differences in the gains in presenting and group work skills across subjects, this was supported by the fact that the qualitative outcomes from members of staff stated that they would often do presenting and group work, usually in combination. As well as the student qualitative data in which some stated that they understood the importance of developing these skills towards their chosen career. This was also aligned with the quantitative outcomes as students were highly agreeable to the questions of communication and group work being important to their chosen career. In addition to this, group work avoidance was also looked at and there was found to be fairly high variance across students. This can also be seen from the student qualitative data as some students stated that they found this method of assessment to be “unfair”. This is showing further support and validation for the assessment tool and would suggest that each of the subscales is in fact measuring what it is that they intend to measure.

This section was the final chapter in order to validate the assessment tool in its ability to measure such subscales in the form of a multiple choice Likert scale format. It appears, from the results, that it is clear that it is able to do so. However, as there have been some slight difficulties with retention rates, there may be some cause for concern in terms of whether the outcomes are in fact accurate as it may be due to the students who took part in each of the different time points. However, when looking at individual data of students who took part across all three time points, this also appeared to triangulate with the qualitative outcomes suggesting that this potentially may not be the case. Further discussion will continue in the final chapters as to the limitations and potential future research.

7.6. Chapter Summary

To summarise, this chapter hoped to validate the assessment tool by triangulating the outcomes of the quantitative data with those of the qualitative data from both staff and students. From this analysis it would appear that all of the subscales have shown some validation from one or both of these qualitative data sets. This would suggest that the assessment tool could potentially be valid and is measuring the subscales which it claims to measure. This also suggests that the format of the multiple choice Likert scale is still potentially useful, as the experience of staff was also found to triangulate with the quantitative outcomes, further suggesting validity.

8. Discussion and Conclusions

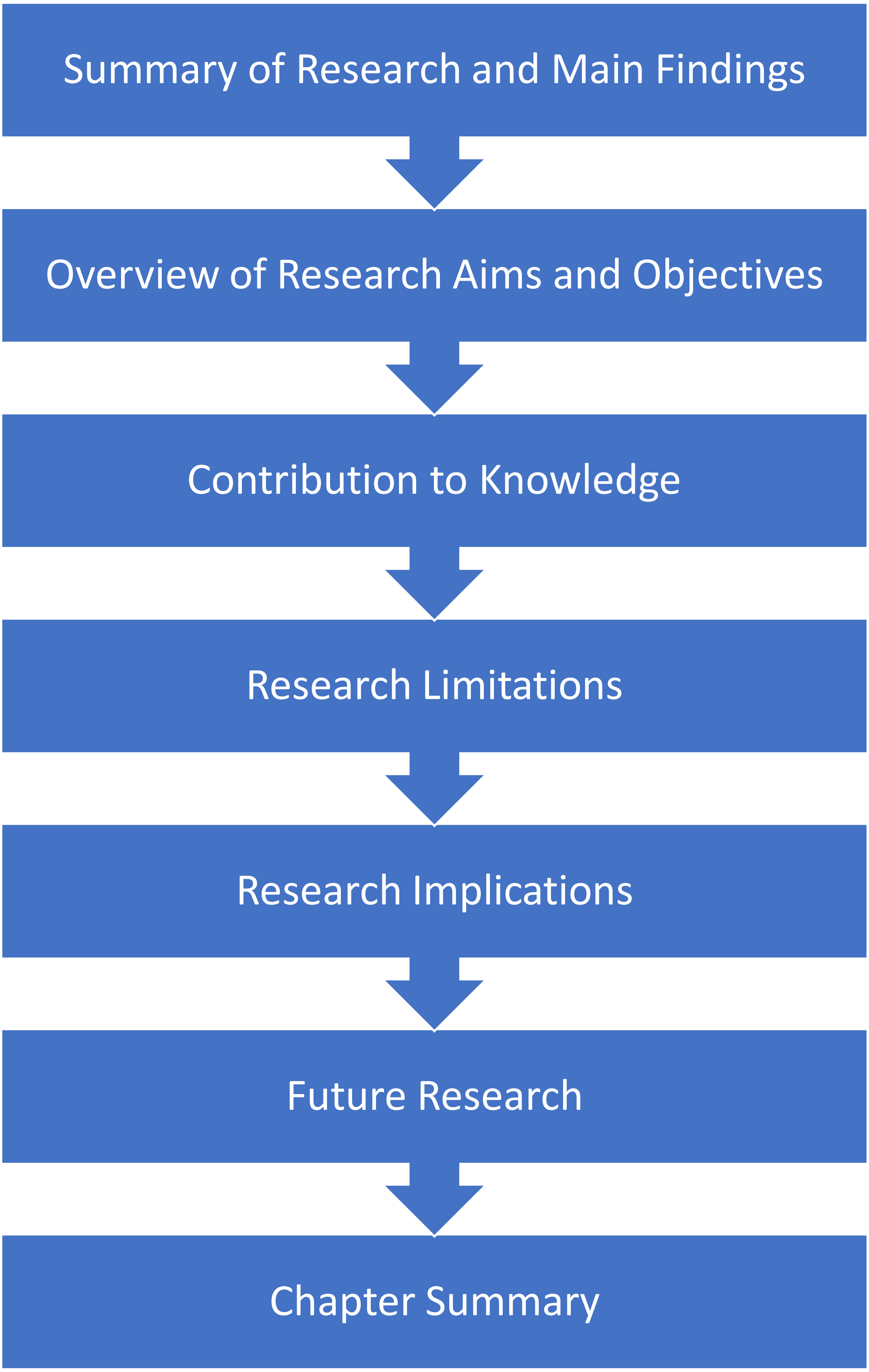

Figure 8.1: Overview of Chapter Progression for Discussion Chapter

8.1. Summary of Research and Main Findings

This research intended to create, implement and test the reliability and validity of a subject specific learning gain questionnaire. The initial stages of the research involved a preliminary inquiry with STEM students whereby interviews were conducted in order to understand the skills that they believed they have gained. Previous literature indicated that there were differences in the emphasis on certain aspects of “generic skills” with critical thinking aspects being different across subjects (Lai 2011; Bailin 2002; Ennis 1989). The literature also suggested a need to develop a measure of learning gain that utilised students as the consumer of the assessment tool (Howson 2017c; Kuh et al. 2015). The outcomes of the preliminary inquiry stage supported the previous literature in that there were different emphases across disciplines and led to a narrowing of focus down to the science-based cohort. This was due to these students focusing more on the evaluative side of critical thinking and suggesting less progress in some areas of personal development, as is aligned with the Wakeham Review of STEM subjects (BIS 2016b).

Furthermore, the skills that were to be utilised for the questionnaire were chosen based upon the outcomes of the qualitative interviews. These included; engagement, critical thinking and personal development. Engagement was referred to as important to measure by the students as it was seen as an extraneous variable which can affect the outcomes of the skills. This further aligned with the literature which suggested that engagement was an important aspect of student learning which was much more embedded into English HE as a proxy measure of student learning (Neves and Stoakes 2018; Buckley 2014; HEFCE 2018). However, it was highlighted that engagement should not be used alone and that current developments in learning gain within an English HE context have not addressed the need for a development of a subject-specific learning gain assessment tool (Howson 2019; Vermunt, Ilie, and Vignoles 2018). By exploring the skills of science-students, this paved the way to developing the assessment tool and narrowing the focus. The outcomes of this preliminary inquiry stage also indicated the skills needed to be focused on were that of; engagement, critical thinking and personal development. These also came from that of the literature on student learning.

Following this, the questionnaire was created. The items within this questionnaire came from mostly validated surveys that were already being utilised within an English HE context (HEFCE 2017c; HEA 2018; Center for Postsecondary Research 2018; ACER 2018). With the critical thinking subscale, the literature was also considered as to what is meant by critical thinking, this was also targeted to be an analytical and evaluative process rather than a problem solving process that was stated by the rest of STEM (Moore 2013; Ennis 1989; Halpern 1999). The creation of this questionnaire addressed the gaps in the knowledge in that a discipline-specific self-reporting tool was created. It was recognised in the literature that a self-report survey was the most efficient and effective method of measuring learning gain due to being able to be analysed longitudinally and increase student engagement with the measure (Howson 2019; Speight, Crawford, and Haddelsey 2018). However, this had yet to be explored until the development of this tool.

Following the creation of the questionnaire there was an initial pilot stage which involved checking the internal consistency through the Cronbach's Alpha as well as picking up any problems with the wording or understanding of the questions. As well as this, the face validity of the questionnaire was also tested as experienced members of staff were asked to check the questionnaire. It was found that there was a high reliability for the engagement subscale and critical thinking however, a low reliability was found for personal development suggesting more items were needed to gain a more reliable subscale. Furthermore, the validity of the questionnaire was tested by asking members of staff to comment on the wording and need for additional questions. The engagement section of the questionnaire was mainly found to have problems with the wording of the questions and potential need to split some questions into further items. The critical thinking subscale was the least modified as the only potential change suggested for this was the additional question of the linking evidence to theory. Finally, the personal development subscale suggested additional questions such as asking students whether they believed that they learnt more from other students than from lecturers. There were also additional questions which were added in order to help understand the behaviours of students as well as the actions such as “I avoid group work where possible”. This will be useful in understanding student actions. The final version of the questionnaire contained a total of 57 items, whereas the original version had only contained 31. This initial pilot stage of the questionnaire was an important phase within the research, it allowed for an initial testing of the assessment tool prior to using it in a longitudinal manner. This stage led to modifications which may have otherwise affected the outcomes of the final assessment tool.

The next stage of the research involved reliability testing. The questionnaire was sent out at three different times in order to get three different groups of science students from the same cohort, students were asked not to complete the questionnaire if they had done so previously. Having three independent groups of participants, allowed the researcher to look at the stability of the questionnaire. This stability was of importance as it would show one aspect of reliability and whether the averages of the students would stay the same when on different samples. This was found to be the case and suggested that the questions in the group were worded, asked and understood in a way in which the outcomes of the measure were stable and reliable. In addition to this, Cronbach’s Alpha was also run by grouping these three independent groups into one large sample of students. It was found that all three of the subscales were now highly reliable according to this statistic, suggesting good internal consistency with the questions and also showing an improvement on the pilot project where the personal development subscale was found to not be reliable. Furthermore, a confirmatory factor analysis was also run in order to check whether the items were found to appropriately fit to the subscale that they had been put into. It was found that this was the case. Therefore, this stage of the research suggested that the questionnaire was highly reliable and stable.

The final and perhaps most important stage of the research was to test the validity of the questionnaire. As discussed previously, there are several ways one could validate data. As the data from the questionnaire was of a quantitative nature it was decided to use qualitative data in order to triangulate the data together. The qualitative data that it was decided to use was that of two different perspectives; the staff and the students. The members of staff were found to be difficult to hold down into doing an interview, but were much easier to get to interview than the students. Therefore, it was decided that the student qualitative data would take the form of a qualitative survey in order to increase engagement. The staff interviews were taken first. Staff were invited to interview where they were told that they were going to discuss their experiences and expectations of their students and how this differs across the different levels. It was hoped that if a member of staff reported a certain level of engagement, this would be reflected in the quantitative data and would also be confirmed by the student survey data. It was found that there was some agreement with the staff data in areas such as engagement, particularly in ways of attendance, and critical thinking by varying levels, but there were also some slight discrepancies. When looking at student data, these discrepancies were able to be cleared as it was found that there was quite a wide range of student individual differences in many of these areas that could explain the variance that was being found which did not align with some of the staff discussions. In addition to this, it was also suggested that the cohort which took part in the questionnaire would have likely been a more highly engaged cohort as they were voluntarily taking part in the research. This may suggest that there are numbers of students who would not have taken part in this who were less engaged with the course and therefore may also be the reason for some of these discrepancies.

The areas which were found to be highly validated by the qualitative data were the attendance of students as this was found to decrease over time which was also reported by members of staff. As well as this, members of staff also reported that their students were most likely motivated by grades, which was also found to be the case within the data as the highest percentage of students stated that this was their main motivation for attending. Furthermore, members of staff also stated that they felt that workshops and smaller classes were more highly attended than larger classes such as lectures, this was also found to be the case within the quantitative data with the average indicating that students attend workshops more than lectures.

Having high reliability and validity in the questionnaire was of particular importance due to the method which was chosen, a self-report survey. The literature was contradictory in terms of whether it is a reliable method of measuring student learning or not with some stating that students are unable to accurately report learning (Bowman 2010b; Porter 2013) and others arguing that students are effectively able to report on their learning at the current point in time (Douglass, Thomson, and Zhao 2012; Berdie 1971; Pike 2011). The high reliability and validity of this measure would suggest the latter as the outcomes of the assessment tool aligned with that of the qualitative outcomes that were reported by both staff and students.

Overall, this research has helped to make a small step towards better and more robust learning gain measures within an English HE context. This research has found that the future trajectory for measurements which assess student skills has been confirmed to be that of ones which focus specifically on disciplines rather than that of generalisation due to the different nature in which each subject will be taught a similar skill. This research has supported the idea that generic measures cannot be utilised within an English HE context and the importance of allowing students to be involved in the creation of the measure. Although the questionnaire only targeted three subscales of engagement, critical thinking and personal development these three areas appear to be seen as highly important within HE, particularly in relation to future careers.

8.2. Overview of Research Aims and Objectives

The current research aimed to develop and validate a self-reporting skills assessment tool, that is able to demonstrate student learning gain, with a specific focus on English Science Undergraduate students. Through this the objectives of the research were explored;

1. To conduct a literature review to explore the current state of learning gain internationally.
2. To explore the developments of learning gain within an English Higher Educational context.
3. To identify the skills that are critical for students to achieve when studying a Science Undergraduate degree within an English Higher Education Institution.
4. To explore and review current measures and literature to define and operationalise the skills which have been identified to belong to Science Undergraduates in England
5. To develop and validate an assessment tool that is able to measure student learning gain with a focus on Undergraduate Science students.

For clarity, the aims and objectives will be presented within the discussion section, in turn each of these will be discussed as to how they were achieved. This section provides a summary of the outcomes of each of these objectives as well as a discussion for each.

8.2.1. Research Objectives

" 1. To conduct a literature review to explore the current state of learning gain internationally."

A literature review was conducted in order to meet the current objective. The outcomes of the literature review found that the US were currently the leading nation in learning gain measures. However, much of their developments were focused on generic standardised measures of student learning which were discussed to not be appropriate in an English HE setting where the degrees are much more specified and therefore, the learning outcomes of students will be discipline specific (Aloisi and Callaghan 2018b; Howson 2019). The outcomes of this objective in the research project highlighted the need for a learning gain assessment tool which has been developed with the English HE context in mind. The development of such a tool would be much more beneficial than an attempt to embed the existing measures from the US and other countries, where the educational system differs greatly from that within England (Bishop 2016; Rippner 2016; Steedle, Kugelmass, and Nemeth 2010). Understanding the international state of learning gain was an important step in this research project as it brought an understanding to the researcher in terms of where the gaps in the knowledge were as well as the weaknesses in the existing measures which are used in other nations educational systems. Not only were the measures criticised in terms of the idea of bringing into a different HE landscape, but they are also currently under scrutiny in their own countries in terms of the validity and reliability of the measures to be able to measure what they seek to (Possin 2013; Hardison and Vilamovska 2009; Aloisi and Callaghan 2018b). Therefore, the need for a measure to be developed within an English HE context became apparent.

"2. To explore the developments of learning gain within an English Higher Educational context"

With an understanding that there is a need for a learning gain assessment tool to be developed within an English HE context, there was also a need to understand what developments were currently occurring within England. Literature appeared to indicate that the most favourable measures of student learning within England were that of "proxy" measures such as engagement and satisfaction but that little instruments measured the learning gain that was occurring (McGrath et al. 2015). Grades were also found to be a favourable measure of student learning within England, despite the flaws in the reliability and validity of this method (Randall and Engelhard 2010; Allen 2005). Furthermore, current research into learning gain seemed to focus on the outcomes of the measures themselves with very little looking into the development of the most efficient measures of student learning gain (Scalise, Douskey, and Stacy 2018; Rogaten and Rienties 2018). Despite this, there appeared to be an awareness of a need for a discipline specific learning gain assessment tool as well as the literature pointing towards a self-report method to achieve this (Howson 2019; Speight, Crawford, and Haddelsey 2018). Therefore, the outcomes of this literature review suggested the need to not only develop a learning gain assessment measure within an English HE, but to also develop a measure which filled in the knowledge gap. The knowledge gap was identified to be that of a more discipline focused tool, of which was identified to be a self-survey method of measuring learning. The lack of development of these kinds of measures suggested a gap which was in need of filling. These measures would have implications from student to institution, which will be discussed later in this chapter.

"3.To identify the skills that are critical for students to achieve when studying a Science Undergraduate degree within an English Higher Education Institution"

To address this objective, STEM students were interviewed to discuss the skills which they believed to have gained during their undergraduate degree. Engagement, critical thinking and personal development were identified to be the skills to be measured within the assessment tool, with a narrowing of focus down to science-student cohort. This outcomes aligned with the literature in that engagement was seen as an important aspect of student learning (Neves and Stoakes 2018) with critical thinking being seen as the most important outcome of completing a degree (Arum and Roska 2011) and personal development being important to explore due to the decrease in employability skills, particularly in science students (BIS 2016b). The identification of such skills paid an important role in the research project as it identified the focus of the assessment tool. In addition, the understanding of the differences across the disciplines in that it was highlighted that science students tend to view critical thinking as a more evaluative process. This added to the development of learning gain assessment tools as this had previously been an unexplored area in understanding the differences across subjects to be able to develop such a tool.

"4. To explore and review current measures and literature to define and operationalise the skills which have been identified to belong to Science Undergraduates in England."

Current validated and utilised self-report measures were explored both nationally and internationally to view potential questions to include within the current assessment tool. These included two measures from the UK; NSS and UKES (HEFCE 2018; HEA 2018) as well as the American NSSE (Center for Postsecondary Research 2018) and the AUSSE (ACER 2018). These measures were utilised to bring out statements and questions to be used in the assessment tool, particularly being used for the engagement and personal development subscales. Literature on critical thinking skills and the aspects which it contains were also identified in order to help write questions to break down the critical thinking skill into individual parts of which would be able to distinguish between levels and ability (Ennis 1989; Moore 2013; Halpern 1999). The breakdown of critical thinking into its distinctive parts is an area of self-report survey method which had not previously been explored, the parts were identified from that of the science student cohort specifically focusing on evaluation and analysis of sources. This was the most novel aspect of the assessment tool. In addition, the combination of all of the validated measures made for a more comprehensive measure of engagement and personal development which addressed all areas of these subscales. Additional questions were also added following the pilot project which were suggested by the members of staff; these included questions related to motivation, importance and career-readiness. The combination of all of these subscales, and the specificity of the critical thinking skill, further made for a unique assessment tool which filled the gap in the need for a development of a tool which is able to be used longitudinally. This measure will make for easy comparability due to the quantifiably of the outcomes as well as the reliability and validity of the measure.

"5. To develop and validate an assessment tool that is able to measure student learning gain with a focus on Undergraduate Science students"

The final objective identified the need to both develop and validate the assessment tool. The development of the questionnaire occurred with the aid of existing validated questionnaires on engagement and satisfaction which were used in both the engagement and personal development subscale. The critical thinking subscale was addressed through the literature. The validation of the assessment was achieved through staff interviews and a qualitative survey sent out to students of the target population which allowed a triangulation and identification of where the questionnaire was aligning and where it was not. This phase also highlighted some potential barriers to using this method of measuring learning gain, whereby it appeared that the students who were most engaged were the ones most likely to respond to the assessment tool and to continue to respond across the three time periods.

8.2.2. Research Aim

The aim of this research was;

"*To develop and validate a self-reporting skills assessment tool, that is able to demonstrate student learning gain, with a specific focus on English Science Undergraduate students."*

Through a mixed method research approach, this research was able to meet the aim of research project. This research identified that there were differences in English Undergraduate students skillset and the emphasis that they are given on specific skills, particularly in relation to cognitive skills. In this research, the focus of the questionnaire was that of Engagement, critical thinking and personal development as identified by the preliminary inquiry stage. The critical thinking aspect of the research provided a focus in English Science Undergraduates in that the questions were designed to be evaluative and analytical in nature. These skills were measured via a self-reporting questionnaire whereby the target population responded to a statement on a five-point likert scale. This allowed for better engagement from students, quantifiability of the outcomes and an approach which had been identified within the literature to be the most efficient. This assessment tool was then validated with the experienced of staff and students at the institution and was found to have high triangulation. Therefore, the aim of this research project was achieved as a learning gain assessment tool was indeed developed and validated with a focus on English Science Undergraduate students.

8.3. Contribution to Knowledge

This research has developed and validated a learning gain assessment tool which is focused on a specific discipline in HE as well as being unique in the presentation of measuring such skills. The literature indicated that there was no "gold standard" to measuring learning gain and that many areas still are in need of exploration (Evans, Howson, and Forsythe 2018).

Current literature has indicated that the measuring of cognitive skills such as critical thinking have relied heavily on already validated and utilised methods that have been taken from the

American educational system (Liu 2011b; CAE 2017a; Voluntary System for Accountability 2018). It has been noted that these particular methods of measuring student skills have flaws in their validity, reliability and particularly in the generic nature of the measure which may be unable to be translated into the specified nature of English HE degrees (Aloisi and Callaghan 2018b; Possin 2013; Ewell 2005; Shermis 2008). The current research takes a unique perspective on the measuring of these cognitive skills in that the self-report method is utilised of which, the method had been identified to be an efficient and appropriate measure for student learning gain but had yet to be explored (Speight, Crawford, and Haddelsey 2018). In addition, the research also breaks down the cognitive skill of critical thinking into its individual components, of which was taken from the literature (Moore 2013; Ennis 1989). This way of measuring critical thinking was a unique approach which had yet to be explored in the measuring of student learning gain.

Furthermore, the literature also identified the need for a more subject-specific approach to measuring student learning, with an awareness of this is what is needed within the English HE landscape yet a lack of development within this area (Banta and Pike 2012; Shermis 2008). The current research looked into identifying the differences in the “generic” skills that are often measured for identifying student learning gain. This was achieved through the discussion with students themselves as well as the literature review into what skills may be “domain-specific”. This is closing the gap in the knowledge whereby much research in this area has been developed from practitioners such as HE providers, academics and educational boards/public bodies. With an untapped research main beneficiaries of the HE, which are those of students themselves. Therefore, this research intended to have a more focused approach in that students were involved in an exploratory phase in the research whereby an inductive approach was taken to give focus and direction in the research project. An understanding of the personal experiences of students appears to be a vital and missing aspect of developing measures of student learning gain. As already, from this research project, it is clear that there are domain-specific skills within HE that are in need of having measured which are more tailored towards specific subjects.

The research identified that critical thinking appeared to differ in the science students as they were more focused on the evaluative and analytical approach to critical thinking in that the focus was on the evaluation and critique of existing sources as well as the application of this into their own research. This research therefore, built upon the idea of a subject-specific assessment tool as the differences across subjects began to emerge and were addressed within this measure of student learning gain. Furthermore, the identification of critical thinking across all STEM subjects paid a contribution to the philosophical discussion as to what HE intends to instil into students. This provided support to the idea that critical thinking is an important and vital aspect to HE. As well as this, engagement and personal development in terms of employability skills, were also identified. This further adds to the conversation of what needs to be measured when looking at student learning gain and the potential influences that may be held on these skills in terms of areas such as student motivation. This research has highlighted how skills may potentially interact and influence each other, for example there was a clear pattern of students who believed attendance to be important, were more likely to attend. This is aligning student attitudes with their engagement and is adding to the literature on the influences over the student experience.

8.4. Beneficiaries

The importance of research can often be conveyed by considering the implications and impact that this particular research project may have. As was previously discussed, learning gain is a particularly important research area as it has recently been implemented into TEF which will be utilised to measure HE’s teaching. This research, therefore, is helping to build on this and the measures which are potentially possible. This questionnaire will be of particular use to students, educators and the institution itself.

At a student level, should this questionnaire be regularly and continuously used, this will be useful for both prospective and current students. As occurs with the NSS, a yearly report could be written using the data to provide an overall outcome of the results. The VLE could also be utilised to provide the outcomes of the questionnaire by embedding this within a curriculum and within the Canvas modules. The availability of this information would then inform prospective students who intend to study a science-based subject to be able to make an informed choice on the institution which they wish to attend. It will provide evidence to these students that the university is able to build on the skills that are needed towards a chosen career as the critical thinking and personal development aspect, as well as showing the students' engagement with the course and therefore showing the students' potential enjoyment of what they are learning, this would best be suited to an overall report aspect that could occur following the implementation of the questionnaire With current students, they can utilise this data for their own personal gain. This data could be taken in a way so that students can access their own individual results across the time periods, potentially through the VLE as previously mentioned, individual progress has the potential to be mapped and followed across the time points to be able to potentially create individual student profiles whereby they are able to view their progress on the assessment tool over time. The individual subscales could show how students learning gain has improved or declined in engagement, critical thinking and personal development. The potential with this assessment tool is that it can bring an overall learning gain score whereby students totals on all three subscales could be totaled. This could then be compared over time to view how students have changed in their learning, as was done in a study by Vermunt et al (2018).

At an educators' level, this information could be used for improvement in teaching. This directly relates to the TEF and will allow educators to see where their strengths and weaknesses lie. This would mean that there could be a system whereby the questionnaire was taken in each individual class and could be tracked in such a way that the questionnaire could give averages for that particular class or lecturer. Having this information would allow the educator to see where the students are struggling and where they are excelling. This would then allow the educator to put more focus on the areas where the students do not understand. For example, if for one of the questions in the critical thinking subscale such as “I am able to distinguish between sources that are useful and not useful” results indicated that the average of students strongly disagreed with that statement, then the educator could implement activities in an attempt to improve that skill. Furthermore, the questionnaire revealing the motivations of students can also help the educator to ensure that the sessions meet those motivations in order to get more students to engage.

Finally, at an institutional level, this will help the institution to compare within itself as to what is working and what isn’t working within the faculty. They may find that there are some differences across subjects within the faculty and therefore can use this to implement changes that will ensure progress in the following year. Ideally, should this questionnaire be found to be able to be utilised in other institutions, then it could also be used to compare across different institutions to help with league tables and to help prospective students to have a comparison in order to decide upon the HEI which they would like to attend.

8.5. Research Limitations

When conducting research it is important to be aware of the limitations that the research may suffer from. Firstly, the retention rates of the students across the time points. It was expected that there would be a loss of students across all three time points as the first influx of students included level 4 students from psychology who would receive points for completing the questionnaire. However, the research did attempt to increase this retention by sending out four reminder emails across the span of a month which helps to increase the number of students that completed the questionnaire at each reminder. This retention is often expected with longitudinal research, nevertheless, it is clear that the outcomes were still aligning with those of the qualitative data suggesting that the retention rates did not appear to have a significant impact on these outcomes. Furthermore, the individual student progress had yet to be explored with this assessment tool and therefore, there was a need to view how this can be utilised. Nevertheless, the assessment was made clear to be high in reliability and validty and therefore, future uses of this assessment tool could follow the individual progress of the students.

In addition to this, the students that were utilised were science students from one single institution. This, arguably could be a limitation as it is unclear whether the questionnaire would also be valid for science students in other institutions and therefore, would mean that the questionnaire could potentially not be comparable across HEI's. The qualitative outcomes which were used in order to help validate the questionnaire may not be the same experiences in other institutions. For example, other institutions may be harsher on attendance recording and monitoring meaning that the attendance of the institution will not decrease over time as there may be harsher consequences than in the institution that was used to create this questionnaire. However, this research has laid some groundwork into potentially testing the questionnaire on other institutions in the future. In addition to this, the questionnaire focus was intended to be led by the students of the institution and therefore, it may mean that this questionnaire may not work for other institutions and further research into science students from different institutions would be needed.

Finally, one may also suggest that there is a limitation due to the fact that the questionnaire is subject focused and therefore cannot be comparable across subjects. However, this was explained to be the novelty of the measure as previous research and piloting projects into learning gain has found that there are too many discipline specific skills to be able to have one measure which can be utilised across subjects. This may mean that instead of attempting to compare subjects within the same institution, the institution should focus on simply improving the individual subjects by comparing with previous cohorts of students and making sure that the discipline is giving the students the skills which it intends to.

8.6. Future Research

This research has found that it is possible to create, implement and validate a student led and subject specific questionnaire which aligns with the experiences and expectations of staff and students. In terms of improving and expanding on the research itself, it may be beneficial to test the questionnaire with a larger Likert scale in order to attempt to be able to notice further minute differences in the data. In addition to this, the questionnaire could have also explored the area of research skills and writing skills. These skills were two areas which came up in the preliminary interview data as well as in some of the qualitative data that was used to validate the questionnaire. However, these skills were not explored within this questionnaire due to the length of the questionnaire in its current state as well as the focus wanting to be on that of critical thinking as this was stated to be the most important skill to take from university level degree. As was touched upon in research implications, there could also be modifications to the questionnaire that could allow it to be tracked so that individuals can view their progress, educators can view their class averages and perhaps even by the individual subjects within the science faculty.

To expand on this research, this project has solely focused on science-based students from one particular institution. Future research could intend to expand on both of these areas which have potential to be explored. Firstly, questionnaires could be created for other subjects such as technology based subjects, art based subjects etc. This would mean going through the same process of ensuring that the questions and skills that were asked in the questionnaire were specific to the discipline that is being focused on. As well as this, there is also the potential to further specify the questionnaires down to individual subjects and having questionnaires for specific subjects such as “Psychology” rather than the entirety of science, though from this research it would suggest this may not be needed but may be of interest to explore. In addition to this, as this questionnaire only focused on one institution it would also be beneficial to view whether this questionnaire could be utilised across HEI’s and if not, to potentially edit and expand the questionnaire so that it is able to be used to compare institutions.

In addition, with the emergence of TEF, it is clear that further research is needed overall in the area of learning gain in order to be able to understand such a complex phenomenon. Learning gain has caused great difficulty in transferring the measures from the US into an English HE context due to the subject specificity of our English HEI's. Although this research has made significant steps towards subject specific questionnaires, future research needs to expand on this knowledge further into other institutions and subjects so that there is the ability to compare in order to have further implications for HE.

8.7. Chapter Summary

In summary, this research makes a novel contribution to knowledge as there has yet to be such a subject focused questionnaire which breaks down the skills into subcomponents in order to gain a more accurate representation of where the student learning is at using a self-survey method. Although there are some limitations to this research such as the retention rates over the time period and the involvement of just one institution, it would appear that it has laid the groundwork for future research to further expand upon this area of learning gain to bring discipline specific questionnaires into more subjects and across institutions. This questionnaire can be utilised to help the current institution with their existing and prospective students as well as with their teaching staff to understand where their improvements and strengths lie. Therefore, this research formed a novel area that is widely under-researched in the English HE sector and although this formed a small step in the area, it is still a contribution that is paving the way to better and more appropriate learning gain measures.

9. References

ACER. 2018. "Australasian Survey of Student Engagement." [online] Available at: https://www.acer.org/gb/ausse. [Accessed on 25th October 2018]

ACT. 2017. "ACT CAAP." [online] Available at: http://www.act.org/caap. [Accessed on 12th September 2018]

Adami, M.F., and A. Kiger. 2005. "The Use of Triangulation for Completeness Purposes." *Nurse Researcher*, 12 (4), pp. 19–29.

Adams, A., and A.L. Cox. 2008. "Open Research Online." In *Research Methods for Human Computer Interaction*. Cambridge, UK: Cambridge University Press., pp.163-178.

Alhojailan, M.I. 2012. "Thematic Analysis : A Critical Review of its Process and Evaluation." In *WEI International European Academic Conference Proceedings*, pp. 39–47.

Allen, J.D. 2005. "Grades as Valid Measures of Academic Achievement of Classroom Learning." *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78 (5), pp. 218-223.

Almalki, S. 2016. "Integrating Quantitative and Qualitative Data in Mixed Methods Research — Challenges and Benefits." *Journal of Education and Learning,* 5 (3), pp.288–296.

Almond, R.J. 2009. "Group Assessment : Comparing Group and Individual Undergraduate Module Marks." *Assessment and Evaluation in Higher Education,* 34 (2), pp. 141-148.

Aloisi, C., and A. Callaghan. 2018a. "Threats to the Validity of the Collegiate Learning Assessment (CLA+) as a Measure of Critical Thinking Skills and Implications for Learning Gain." *Higher Education Pedagogies,* 3 (1), pp. 57–82.

Aloisi, C., and A. Callaghan. 2018b. "Threats to the Validity of the Collegiate Learning Assessment (CLA +) as a Measure of Critical Thinking Skills and Implications for Learning Gain." *Higher Education Pedagogies,* 3 (1), pp. 137–162.

Alshenqeeti, H. 2014. “Interviewing as a Data Collection Method : A Critical Review.” *English Linguistics Research,* 3 (1), pp. 39–45.

Amaratunga, D., D. Baldry, M. Sarshar, and R. Newton. 2002. “Quantitative and Qualitative Research in the Built Environment : Application of ‘Mixed’ Research Approach.” *Work Study,* 51 (1), pp. 17-31.

Arico, F., H. Gillespie, S. Lancaster, N. Ward, and A. Ylonen. 2018. “Lessons in Learning Gain : Insights from a Pilot Project.” *Higher Education Pedagogies,* 3 (01), pp. 190–206.

Arthur, S., and J. Nazroo. 2003. “Designing Fieldwork Strategies and Materials.” In *Qualitative Research Practice*. pp.109-137.

Arum, R., and J. Roska. 2011. *Academically Adrift: Limited Learning on College Campuses*. Chicago: University of Chicago Press.

Association of American Colleges and Universities. 2014. "Rising to the Challenge: Meaningful Assessment of Student Learning." [online]
Available at: https://files.eric.ed.gov/fulltext/ED519804.pdf [Accessed on 27th January 2019]

Australian Government. 2009. "Transforming Australia’s Higher Education System." [online]
Available at: https://www.voced.edu.au/content/ngv%3A14895 [Accessed on 11th February 2019]

Babcock, P.S., and M. Marks. 2010. “The Falling Time Cost of College”, *The Review of Economics and Statistics,* 93(2), pp. 468-478.

Bailin, S. 2002. “Critical Thinking and Science Education.” *Science and Education,* 11 (4), pp. 361–375.

Banta, T.W., and G.R. Pike. 2012. "Making the Case against -One More Time. The Seven Red Herrings About Standardized Assessments in Higher Education." [online]
Available at:

https://learningoutcomesassessment.org/documents/HerringPaperFINAL.pdf [Accessed on 13th October 2018]

Baume, D. 2018. “Towards a Measure of Learning Gain . A Journey . With Obstacles.” *Higher Education Pedagogies,* 3 (1), pp. 1–3..

Baxter, P., and S. Jack. 2008. “Qualitative Case Study Methodology : Study Design and Implementation for Novice Researchers.” *The Qualitative Report,* 13 (4), pp. 544–559.

Becker, H.S. 1956. “Interviewing Medical Students.” *American Journal of Sociology,* 62 (2), pp.199–201.

Berdie, R.F. 1971. “Self-Claimed and Tested Knowledge.” *Educational and Psychological Measurement,* 31 (3), pp. 629–636.

BIS. 2011. "Higher Education : Students at the Heart of the System." [online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/31384/11-944-higher-education-students-at-heart-of-system.pdf [Accessed on 12th March 2019]

BIS. 2015. "Green Paper - Fulfilling Our Potential: Teaching Excellence, Social Mobility and Student Choice." [online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/474227/BIS-15-623-fulfilling-our-potential-teaching-excellence-social-mobility-and-student-choice.pdf [Accessed on 15th March 2019]

BIS. 2016a. "Success as a Knowledge Economy: Teaching Excellence, Social Mobility and Student Choice." [online] Available at: https://www.timeshighereducation.com/sites/default/files/breaking_news_files/higher-education-white-paper-success-as-a-knowledge-economy.pdf [Accessed on 14th March 2019]

BIS. 2016b. "Wakeham Review of STEM Degree Provision and Graduate Employability."

[online] Available at:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/518582/ind-16-6-wakeham-review-stem-graduate-employability.pdf [Accessed on 4th March 2019]

Bishop, L. 2016. “Another Country: The Trans-Atlantic Alliance and the Benefits of Studying Abroad in the USA.” *Innovations in Practice,* 10 (1), pp. 11-14

Blaich, C.F., and K.S. Wise. 2011. "From Gathering to Using Assessment Results: Lessons from the Wabash National Study." [online] Available at: https://www.learningoutcomeassessment.org/documents/Wabash_001.pdf [Accessed on 4th March 2019]

Blandford, A. 2013. “Semi - Structured Qualitative Studies.” In *The Encyclopedia of Human-Computer Interaction*. 2nd edn. [online] Available at: https://discovery.ucl.ac.uk/id/eprint/1436174/2/semi-structured_qualitative_studies.pdf [Accessed on 8th April 2019]

Bloom, B. S., M. D. Englehard, E. J. Furst, and W. H. Hill. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals.* New York: David McKay Company.

Bourner, J., M. Hughes, and T. Bourner. 2001. “First-Year Undergraduate Experiences of Group Project Work.” *Assessment & Evaluation in Higher Education,* 26 (1), pp. 19-39.

Bowman, N.A. 2010a. “Assessing Learning and Development Among Diverse College Students.” *New Directions for Institutional Research*, 2010 (145), pp. 53-71.

Bowman, N.A. 2010b. “Can 1st-Year College Students Accurately Report Their Learning and Development?” *American Educational Research Journal*, 47 (2), pp. 466-496.

Brace, I. 2013. *Questionnaire Design: How to Plan, Structure and Write Survey Material for Effective Market Research*. 3rd edn. London: Kogan Page Ltd.

Braun, H., and C. Bily. 2013. “OECD Assessment of Higher Education Learning Outcomes

(AHELO): Further Insights - Volume 3" [online] Available at: https://www.oecd.org/education/skills-beyond-school/AHELOFSReportVolume3.pdf [Accessed on 19th March 2019]

Braun, V., and V. Clarke. 2008. "Using Thematic Analysis in Psychology.", *Qualitative Research in Psychology,* 3 (2), 77-101.

British Psychological Society. 2014. *Code of Human Research Ethics*. 2nd edn. The British Psychological Society: Leicester.

Brooks, P. 2012. "Outcomes, Testing, Learning: What's at Stake?" *Social Research* 79 (3), pp. 601–611.

Browne, J. 2010. "Securing a Sustainable Future for Higher Education." [online] Available at: http://www.educationengland.org.uk/documents/pdfs/2010-browne-report.pdf [Accessed on 18th March 2019]

Bryman, A. 1984. "The Debate about Quantitative and Qualitative Research: A Question of Method or Epistemology?" *The British Journal of Sociology,* 35 (1), pp. 75-92.

Bryman, A. 2012. *Social Research Methods*. 4th edn. London: Oxford University Press.

Buckley, A. 2014. *UK Engagement Survey 2014: The second pilot year*. York: Higher Education Academy.

Burrell, G., and G. Morgan. 1979. *Sociological Paradigms and Organisational Analysis*. New York: Heinemann Educational Books.

CAE. 2017a. "CLA + for Higher Education." Council for Aid to Education. [online]. Available at: http://cae.org/flagship-assessments-cla-cwra/cla/. [Accessed on 12th September 2018]

CAE. 2017b. "Comparing CLA to CLA+." [online] Available at: https://cae.org/images/uploads/pdf/cla_vs_cla_plus.pdf [Accessed on 14th January

2019]

Callender, C., P. Ramsden, and J. Griggs. 2014. “Review of the National Student Survey,” [online] Available at: http://www.hefce.ac.uk/pubs/rereports/year/2014/nssreview [Accessed on 25th October 2018]

Cameron, A., Y. Wharton, and J.B. Scally. 2018. “An Investigation into the Comparative Learning Gain and ‘ Value Added ’ for Students from Widening Participation and Non-Widening Participation Groups : A Case Study from Sports Degrees Degrees.” *Higher Education Pedagogies,* 3 (1), pp. 40–59.

Carnegie Results. 2007. “The Collegiate Learning Assessment: A Tool for Measuring the Value Added of a Liberal Arts Education.” [online] Available at: https://www.carnegie.org/media/filer_public/fd/14/fd1415a3-1045-479f-8039-09b611c8cb5c/ccny_cresults_2007_collegiate.pdf [Accessed on 7th March 2019]

CCC. 2012a. *CAAP 2012: "*Collegiate Assessment of Academic Proficiency." [online] Available at: https://calhoun.edu/overview/planning-research-grants/assessment-tools-surveys/ [Accessed on 7th March 2019]

CCC. 2012b. *CAAP 2014: "*Collegiate Assessment of Academic Proficiency." [online] Available at: https://calhoun.edu/overview/planning-research-grants/assessment-tools-surveys/ [Accessed on 7th March 2019]

CCC. 2016. *CAAP 2016: "*Collegiate Assessment of Academic Proficiency." [online] Available at: https://calhoun.edu/overview/planning-research-grants/assessment-tools-surveys/ [Accessed on 7th March 2019]

Center for Postsecondary Research. 2017. “Survey Findings on the Quality of Undergraduate Education Engagement.” [online] Available at: http://nsse.indiana.edu/NSSE_2017_Results/pdf/NSSE_2017_Annual_Results.pdf. [Accessed on 16th September 2018]

Center for Postsecondary Research. 2018. “National Survey Student Engagement Survey Instruments and Modules.” [online] Available at: http://nsse.indiana.edu/html/survey_instruments.cfm. [Accessed on 16th September 2018]

Center for Studies in Higher Education. 2018. “Uses of SERU Data.” Berkeley [online] Available at: https://cshe.berkeley.edu/seru/about-seru/uses-seru-data [Accessed on 22nd September 2018]

Chen, Yan, Marcus Henning, Jill Yielder, Rhys Jones, Andy Wearn, and Jennifer Weller. 2015. “Progress Testing in the Medical Curriculum: Students’ Approaches to Learning and Perceived Stress.” *BMC Medical Education,* 15 (1), pp. 1–8.

Cheng, J.H.S., and H.W. Marsh. 2010. “National Student Survey: Are Differences between Universities and Courses Reliable and Meaningful?” *Oxford Review of Education,* 36 (6), pp. 693–712.

Chirk, J. 2006. “Designing a Questionnaire.” *Malaysian Family Physician,* 1 (1), pp. 32–35.

Christie, H., P. Barron, and N. D’Annunzio-Green. 2013. “Direct Entrants in Transition: Becoming Independent Learners.” *Studies in Higher Education,* 38 (4), pp. 623–637.

CIWC. 2016. “Wabash National Study 2006–2012.” Center of Inquiry at Wabash College. [online]. Available at: http://www.liberalarts.wabash.edu/study-overview/. [Accessed on 24th September 2018]

Coates, H., and S. Richardson. 2012. “An International Assessment of Bachelor Degree Graduates’ Learning Outcomes.” *Higher Education Management and Policy,* 23 (3), pp. 1-19.

Cohen, D., and B. Crabtree. 2006. “Semi-Structured Interviews” [online] Available at: https://sswm.info/sites/default/files/reference_attachments/COHEN%202006%20Semistructured%20Interview.pdf [Accessed on 19th March 2019]

Cox, T.D., and M.A. Lemon. 2016. “A Curricular Intervention for Teaching and Learning: Measurement of Gains of First-Year College Student Learning.” *Journal of the Scholarship of Teaching & Learning,* 16 (3), pp. 1–11.

Creswell, J.W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th edn. California: Sage Publications.

Cronbach, L.J., and L. Furby. 1970. “How we Should Measure " Change "— or Should we ?” *Psychological Bulletin,* 74 (1), pp. 68–80.

Crotty, M. 1998. *The Foundations of Social Research: Meaning and Perspective in the Research Process.* London: Sage Publications.

Crowe, S., K. Cresswell, A. Robertson, G. Huby, A. Avery, and A. Sheikh. 2011. “The Case Study Approach.” *BMC Medical Research Methodology,* 11 (100), pp. 1-9.

DeBoer, Fredrik. 2016. “Standardized Assessments of College Learning: Past and Future." [online] Available at: https://static.newamerica.org/attachments/12843-standardized-assessments-of-college-learning/Assessment-White-Paper.72b9bbb2cf0447b4aacaabe6e6ea1824.pdf [Accessed on 29th October 2018]

Department for Education. 2018a. “Primary School Accountability in 2018,” [online] Available at: https://www.gov.uk/government/publications/primary-school-accountability [Accessed on 4th December 2018]

Department for Education. 2018b. "Teaching Excellence and Student Outcomes Framework : Subject- Level." [online] Available at: https://www.gov.uk/government/consultations/teaching-excellence-and-student-outcomes-framework-subject-level [Accessed on 4th December 2018]

Doran, H.C., and L.T. Izumi. 2004. *Putting Education to the Test: A Value-Added Model for California*. San Francisco: Pacific Research Institute.

Doran, H.C., and J.R. Lockwood. 2006. “Fitting Value-Added Models in R.” *Journal of*

Educational and Behavioral Statistics, 31 (2), pp. 205–230.

Douglass, J., G. Thomson, and C. Zhao. 2012. “The Learning Outcomes Race: The Value of Self-Reported Gains in Large Research Universities.” *Higher Education,* 64 (3), pp. 317-335.

Drost, E.A. 2004. “Validity and Reliability in Social Science Research.” *Education Research and Perspectives,* 38 (1), pp. 105–125.

Dubbs, M. 2011. “Collegiate Assessment of Academic Proficiency,” [online] Available at: https://www.unk.edu/academic_affairs/_files/assessment/CAAP10Freport.pdf [Accessed on 6th March 2019]

Edgerton, R., A. Astin, A.M. Cartter, D. Bennett, M. Broad, A. Gonzalez, and M. Nettles. 2003. “Converting Data into Action - Expanding the Boundaries of Institutional Improvement.” [online] Available at: https://scholarworks.iu.edu/dspace/handle/2022/23485 [Accessed on 3rd March 2019]

Educational Testing Service. 1993. *Annotated Bibliography of Tests: Achievement Tests, College Level*. New Jersey: Princeton.

Edwards, D., and J. Pearce. 2014. “Outcomes Assessment in Practice: Reflections on Two Australian Implementations.” In *Higher Education Learning Outcomes Assessment: International Perspectives,* pp. 89-112.

Ennis, R. H. 1989. “Critical Thinking and Subject Specificity: Clarification and Needed Research.” *Educational Researcher,* 18 (3), pp. 4–10.

Ennis, R.H. 1985. “A Logical Basis for Measuring Critical Thinking Skills.” *Educational Leadership*, 40 (10), pp. 44-48.

ETS. 2017. “ETS Proficiency Profile.” Educational Testing Service. [online] Available at: https://www.ets.org/proficiencyprofile/about. [Accessed on 12th October 2018]

Evans, C, C.K. Howson, and A. Forsythe. 2018. “Making Sense of Learning Gain in Higher Education.” *Higher Education Pedagogies,* 3 (1), pp. 1–45.

Ewell, P.T. 2002. “An Emerging Scholarship: A Brief History of Assessment.” In *Building a Scholarship of Assessment*, San Francisco: Jossey-Bass, pp. 3–25.

Ewell, P.T. 2005. “From the States.” *Assessment Update*, 3(3), pp.6-7.

Facione, P.A. 1990. *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations*. Newark: American Philosophical Association.

Field, A. 2017. *Discovering Statistics Using IBM SPSS Statistics*. 5th Edn. London: Sage Publications.

Fletcher, G.H. 2006a. “Spellings Commission on HEd.” *Higher Education,* 36 (7), pp. 14–16.

Fletcher, G.H. 2006b. “A Test of Leadership: Charting the Future of U.S. Higher Education: A Report of the Commission Appointed by Secretary of Education Margaret Spellings.” *Higher Education,* 36 (7), pp. 14–16.

Forsythe, A, and M Jellicoe. 2018. “Predicting Gainful Learning in Higher Education ; a Goal-Orientation Approach.” *Higher Education Pedagogies,* 3 (1),pp. 1–15.

Freeman, A., C. Van Der Vleuten, Z. Nouns, and C. Ricketts. 2010. “Progress Testing Internationally.” *Medical Teacher,* 32 (6), pp. 451–455.

Fulcher, K.H, and J.T Willse. 2007. “Value Added: Back to Basics in Measuring Change.” *Assessment Update*, 19 (5), pp. 1-16.

Gelder, T. V. 2005. “Teaching Critical Thinking : Some Lessons from Cognitive Science.” *College Teaching*, 53 (1), pp. 41 - 46.

Gibbs, G. 2010. "Dimensions of Quality." Higher Education Academy [online] Available at: https://s3.eu-west-2.amazonaws.com/assets.creode.advancehe-document-manager/documents/hea/private/dimensions_of_quality_1568036697.pdf [Accessed on 17th January 2019]

Gilbert, D. 2007. *Stumbling on Happiness*. New York: Harper Perennial.

Gillham, B. 2004. *Developing a Questionnaire*. London: Continuum.

Goe, L., C. Bell, and O. Little. 2008. “Approaches to Evaluating Teacher Effectiveness : A Research Synthesis.” [online] Available at: https://files.eric.ed.gov/fulltext/ED521228.pdf [Accessed on 9th March 2019]

Goldstein, H., P. Huiqi, T. Rath, and N. Hill. 2000. “The Use of Value Added Information in Judging School Performance.” [online] Available at: http://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/using-value-added-information.pdf [Accessed on 6th April 2019]

Gonyea, R.M. 2005. “Self-Reported Data in Institutional Research: Review and Recommendations.” *Survey Research: Emerging Issues. New Directions for Institutional Research*,2005 (127), pp. 73–89.

Greatbatch, D., and J. Holland. 2016. "Teaching Quality in Higher Education: Literature Review and Qualitative Research." [online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524495/he-teaching-quality-literature-review-qualitative-research.pdf [Accessed on 2nd April 2019]

Hagelskamp, C., D. Schleifer, and C. DiStasi. 2013. “Is College Worth It for Me? How Adults without Degrees Think about Going (Back) to School.” [online] Available at: https://kresge.org/sites/default/files/Is-College-Worth-It-For-Me-Public-Agenda-2013.pdf [Accessed on 2nd March 2019]

Halcomb, E.J., and L. Hickman. 2015. “Mixed Methods Research.” *University of Wollongong Research Online,* 29 (32), pp. 41–47.

Halpern, D.E. 1999. “Teaching for Critical Thinking: Helping College Students Develop the Skills and Dispositions of a Critical Thinker,” In *Teaching and Learning on the Edge of the Millenium,* New York: Jossey-Bass, pp. 69–75.

Hardison, C.M., and A.M. Vilamovska. 2009. *The CLA. Setting Standards for Performance at a College or University*. London: RAND Education.

Harrell, M.C., and M.A. Bradley. 2009. "Data Collection Methods: Semi-Structured Interviews and Focus Groups." [online] Available at: https://www.rand.org/pubs/technical_reports/TR718.html [Accessed on 4th March 2019]

HEA. 2017. “UK Engagement Survey.” [online] Available at: https://www.heacademy.ac.uk/institutions/surveys/uk-engagement-survey. [Accessed on 14th September 2018]

HEA. 2018. “UKES: UK Engagement Survey.” [online] Available at:https://www.heacademy.ac.uk/institutions/surveys/uk-engagement-survey. [Accessed on 14th September 2018]

HEFCE. 2016a. “National Mixed Methodology Learning Gain Project.” [online] Available at: http://www.hefce.ac.uk/lt/lg/mixed/.[Accessed on 26th October 2018]

HEFCE. 2016b. “Learning Gain.” [online] Available at: http://www.hefce.ac.uk/lt/lg/.[Accessed on 4th September 2018]

HEFCE. 2017a. “About the TEF.” [online] Available at: http://www.hefce.ac.uk/lt/tef/whatistef/. [Accessed on 16th January 2019]

HEFCE. 2017b. "Teaching Excellence and Student Outcomes Framework Specification." [online] Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/658490/Teaching_Excellence_and_Student_Outcomes_Framework_Specification.pdf. [Accessed on 4th March 2019]

HEFCE. 2017c. “National Student Survey.” [online] Available at: http://www.hefce.ac.uk/lt/nss/. [Accessed 2nd March 2019]

HEFCE. 2017d. “Learning Gain in Higher Education.” [online] Available at: http://www.hefce.ac.uk/pubs/rereports/year/2015/learninggain/. [Accessed on 12th September 2018]

HEFCE. 2017e. “National Student Survey 2017 Results Show High Levels of Student Satisfaction and Engagement.” [online] Available at: http://webarchive.nationalarchives.gov.uk/20180319124906/http://www.hefce.ac.uk/news/newsarchive/2017/Name,115244,en.html. [Accessed on 15th February 2019]

HEFCE. 2018. “National Student Survey.” [online] Available at: http://webarchive.nationalarchives.gov.uk/20180319121932/http://www.hefce.ac.uk/lt/nss/review/. [Accessed on 12th March 2019]

HESA. 2018a. “Who’s Studying in HE?” [online] Available at: https://www.hesa.ac.uk/data-and-analysis/students/whos-in-he. [Accessed on 14th March 2019]

HESA. 2018b. “What Do HE Students Study?” [online] Available at: https://www.hesa.ac.uk/data-and-analysis/students/what-study. [Accessed on 16th March 2019]

HESA. 2018c. “Definitions: Destinations of Leavers.” [online] Available at: https://www.hesa.ac.uk/support/definitions/destinations. [Accessed on 12th March 2019]

Higher Education Academy. 2013. “GPA Programme Outline,”. [online] Available at: https://www.heacademy.ac.uk/sites/default/files/downloads/gpa_programme_outline.pdf [Accessed on 6th April 2019]

Howson, C. B. K. 2019. "Final Evaluation of the Office for Students Learning Gain Pilot Projects." [online] Available at: https://www.officeforstudents.org.uk/publications/lg-pilot-projects-evaluation/ [Accessed on 4th August 2019]

Howson, C.B.K. 2017a. “Measuring Learning Gain Isn’t Easy, but It Has Become Necessary.” *WONKHE*. [online] Available at: https://wonkhe.com/blogs/analysis-measuring-learning-gain-isnt-easy-but-it-has-become-necessary/. [Accessed on 12th September 2018]

Howson, C.B.K. 2017b. "Evaluation of HEFCE’s Learning Gain Pilot Projects: Year 1 Report." *Higher Education Funding Council for England.* [online] Available at: http://www.hefce.ac.uk/pubs/rereports/year/2017/lgeval/. [Accessed on 17th September 2018]

Howson, C.B.K. 2017c. "Evaluation of HEFCE’s Learning Gain Pilot Projects." *Higher Education Funding Council for England*. [online] Available at: https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/learning-gain-pilot-projects/ [Accessed on 12th October 2018]

Howson, C.B.K. 2018. "Evaluation of HEFCE ’ s Learning Gain Pilot Projects Year 2." *Higher Education Funding Council for England.* [online] Available at: https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/learning-gain-pilot-projects/ [Accessed on 19th April 2019]

Hudson, R., and HEFCE. 2016. “National Mixed Methodology Learning Gain Project.” In *Learning Gain Conference Proceedings.*

Hyytinen, H., K. Nissinen, J. Ursin, A. Toom, and S. Lindblom-Ylänne. 2015. “Problematising the Equivalence of the Test Results of Performance-Based Critical Thinking Tests for Undergraduate Students.” *Studies in Educational Evaluation,* 44, pp. 1–8.

Immerwahr, J., and J. Johnson. 2012. “Squeeze Play 2009: The Public’s Views on College Costs Today--Public Agenda and the National Center for Public Policy and Higher Education.” [online] Available at: https://files.eric.ed.gov/fulltext/ED536382.pdf [Accessed on 17th March 2019]

Inayah, A.T., L.A. Anwer, M.A. Shareef, A. Nurhussen, H.M. Alkabbani, A.A. Alzahrani, A.S. Obad, M. Zafar, and N.A. Afsar. 2017. “Objectivity in Subjectivity: Do Students’ Self and Peer Assessments Correlate with Examiners’ Subjective and Objective Assessment in Clinical Skills? A Prospective Study.” *BMJ Open,* 7 (5), pp.1 -7.

International Education Specialist. 2016. “The US Education System.” [online] Available at: https://www.idp.com/usa/forstudents/whystudyintheus/education-system. [Accessed on 12th December 2018]

Jackson, R., D.K. Drummond, and S. Camara. 2007. “What Is Qualitative Research,”*Qualitative Research Reports in Communication,* 8 (1), pp. 21-28.

Javadi, M., and K. Zarea. 2016. “Understanding Thematic Analysis and Its Pitfall.” *Journal of Client Care*,1 (1), pp. 34-40.

Johnson, V. 2003. *Grade Inflation: A Crisis in College Education*. Washington: Springer.

Kerr, C. 2018. “Student Engagement and Learning Gains : Self- Reports , Direct Measures , and Instrument Specificity.” Masters Thesis, James Madison University, Harrisonburg.

Khan, S. 2014. “Qualitative Research Method - Phenomenology.” *Asian Social Science,* 10 (21) pp. 298–310.

Kim, H., and D. Lalancette. 2013. “Literature Review on the Value-Added Measurement in Higher Education.” *OECD*. [online] Available at: https://www.oecd.org/education/skills-beyond-school/Litterature%20Review%20VAM.pdf [Accessed on 2nd October 2018]

Kirkup, Catherine, Rebecca Wheater, Jo Morrison, Ben Durbin, and Marco Pomait. 2010. "Use of an Aptitude Test in University Entrance: A Validity Study." [online] Available at: https://www.nfer.ac.uk/publications/VAQ01/VAQ01.pdf [Accessed 3rd February 2019]

Klein, S. 2009. “The Lumina Longitudinal Study: Summary of Procedures and Findings.” Council for Aid to Education [online] Available at: internal-pdf://72.73.4.6/12_CLA_Lumina_Longitudinal_Study_Summary_Findi.pdf. [Accessed on 10th March 2019]

Klein, S., R. Benjamin, R. Shavelson, and R. Bolus. 2007. “The Collegiate Learning Assessment: Facts and Fantasies.” *Evaluation Review,* 31 (5), pp. 415–439.

Krause, K.L., and H. Coates. 2008. “Students ’ Engagement in First - Year University.” *Assessment and Evaluation in Higher Education,* 33 (5, pp.: 493–505.

Kuh, G., and S. Ikenberry. 2009. "More Than You Think, Less Than We Need: Learning Outcomes Assessment in American Higher Education." *National Institute for Learning Outcomes Assessment.* [online] Available at: internal-pdf://0978582587/niloafullreportfinal2.pdf. [Accessed on 12th January 2018]

Kuh, G.D. 1995. “The Other Curriculum.” *The Journal of Higher Education,* 66 (2), pp. 123–155.

Kuh, G.D. 2001. “The National Survey of Student Engagement: Conceptual Framework and Overview of Psychometric Properties.” [online] Available at: http://nsse.indiana.edu/pdf/psychometric_framework_2002.pdf [Accessed on 14th December 2018]

Kuh, G.D., T.M. Cruce, R. Shoup, J. Kinzie, and R.M. Gonyea. 2008. “Unmasking the Effects of Student Engagement on First-Year College Grades and Persistence.” *The Journal of Higher Education,* 79 (5),pp. 540-563.

Kuh, G.D., and P.T. Ewell. 2010. “The State of Learning Outcomes Assessment in the United States.” *Higher Education Management and Policy,* 22 (1), pp. 1-20.

Kuh, G.D., S.O. Ikenberry, N. Jankowski, T.R. Cain, P.T. Ewell, P. Hutchings, and J. Kinzie. 2015. *Using Evidence of Student Learning to Improve Higher Education*. San Francisco: Wiley.

Kuh, G.D., N. Jankowski, S.O. Ikenberry, and J. Kinzie. 2014. "Knowing What Students Know and Can Do: The Current State of Student Learning Outcomes Assessment in

U.S. Colleges and Universities." [online] Available at: https://www.learningoutcomeassessment.org/documents/2013%20Abridged%20Survey%20Report%20Final.pdf [Accessed on 23rd October 2018]

Kuhn, T. 1970. *The Structure of Scientific Revolutions*. Chicago: Chicago Press.

Lai, E.R. 2011. "Critical Thinking: A Literature Review." *Pearson Assessments.* [online] Available at: https://images.pearsonassessments.com/images/tmrs/CriticalThinkingReviewFINAL.pdf [Accessed on 15th April 2019]

Laird, T.F.N., T.A. Seifert, E.T. Pascarella, M.J. Mayhew, and C.F. Blaich. 2014. "Deeply Affecting First-Year Students' Thinking: Deep Approaches to Learning and Three Dimensions of Cognitive Development." *The Journal of Higher Education,* 85 (3), pp. 402–432.

Leech, B.L. 2002. "Asking Questions : Techniques for Semistructured Interviews," *Political Science and Politics,* 35 (4), pp. 665-668.

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), pp. 39–63. .

Lennon, M.C., and B. Frank. 2014. "Decentralised Learning Outcomes Assessments: Developments in Ontario, Canada." In *Higher Education Learning Outcomes Assessment: International Perspectives*.pp. 89-112.

Lester, S. 1999. *An Introduction to Phenomenological Research*. Taunton: Stan Lester Developments.

Lincoln, Y.S., and E.G. Guba. 1985. *Naturalistic Inquiry*. California: SAGE Publications.

Lindsay, T. 2013. "The Likelihood of Higher-Education Reform." *Society,* 50 (3), pp. 236–244.

Liu, O.L. 2011a. “Measuring Value-added in Higher Education: Conditions and Caveats – Results from Using the Measure of Academic Proficiency and Progress (MAPP™).” *Assessment & Evaluation in Higher Education,* 36 (1), pp. 81–94.

Liu, O.L. 2011b. “Value-Added Assessment in Higher Education: A Comparison of Two Methods.” *Higher Education,* 61 (4), pp. 445–461

Liu, O.L., H. Liu, K.C. Roohr, and D.F. McCaffrey. 2016. “Investigating College Learning Gain: Exploring a Propensity Score Weighting Approach.” *Journal of Educational Measurement,* 53 (3), pp. 352–367.

Liu, O.L., and K.C. Roohr. 2013. "Investigating 10-Year Trends of Learning Outcomes at Community Colleges." [online] Available at: https://www.ets.org/Media/Research/pdf/RR-13-34.pdf [Accessed 19th December 2018]

Lodge, J.M., and A. Bonsanquet. 2014. “Evaluating Quality Learning in Higher Education: Re-Examining the Evidence.” *Quality in Higher Education,* 20 (1), pp. 3-23.

Maguire, M., and B. Delahunt. 2017. “Doing a Thematic Analysis : A Practical , Step-by-Step Guide for Learning and Teaching Scholars .” *All Ireland Journal of Teaching and Learning in Higher Education,* 9 (3), pp. 1-14.

Marr, D. 1995. “An Investigation of the Construct Validity of the Long Form of the Academic Profile.” Educational Testing Service [online] Available at: https://origin-www.ets.org/s/proficiencyprofile/pdf/Validity_Academic_Profile.pdf [Accessed on 12th April 2019]

Maskell, E.C., and L. Collins. 2017. “Measuring Student Engagement in UK Higher Education: Do Surveys Deliver?” *Journal of Applied Research in Higher Education*, 9 (2), pp. 226-241.

Mason, J. 2002. *Qualitative Researching*. 2nd edn. London: SAGE Publications.

Mason, P. 2014. "*Research Philosophy and Research Design"*. In Researching Tourism, Leisure and Hospitality for your Dissertation. Oxford: Goodfellow Publishers, pp. 97 - 120.

Mathers, Nigel, Nick Fox, and Amanda Hunn. 2002. "Using Interviews in a Research Project." In *Research Approaches in Primary Care*. London: Taylor and Francis, pp. 6-28.

Mc Caffrey, D.F., J.R. Lockwood, D.M. Koretz, and L.S. Hamilton. 2003. *Evaluating Models for Teacher Accountability*.RAND Education. [online] Available at: https://www.rand.org/content/dam/rand/pubs/monographs/2004/RAND_MG158.pdf

McGrath, C.H., B. Guerin, E. Harte, M. Frearson, and C. Manville. 2015. *Learning Gain in Higher Education*. Cambridge: RAND Education.

Moore, T. 2013. “Critical Thinking: Seven Definitions in Search of a Concept.” *Studies in Higher Education,* 38 (4), pp. 506-522.

Moriarty, J. 2011. *Qualitative Methods Overview*. NIHR School for Social Care Research, London. [online] Available at: http://eprints.lse.ac.uk/41199/ [Accessed on 12th March 2019]

Morris, D. 2017. “TEF Results – How Do REF and TEF Results Compare?” *WONKHE*. [online] Available at: https://wonkhe.com/blogs/tef-results-how-do-ref-and-tef-results-compare/ [Accessed on 15th April 2019]

Muijs, D. 2004. *Doing Quantitative Research in Education*. London: Sage Publications.

National Audit Office. 2018. "Delivering STEM (Science , Technology , Engineering and Mathematics) Skills for the Economy." *Department for Business, Energy & Industrial Strategy and Department for Education.* [online] Available at: https://www.nao.org.uk/wp-content/uploads/2018/01/Delivering-STEM-Science-technology-engineering-and-mathematics-skills-for-the-economy.pdf [Accessed on 12th April 2019]

National Center for Educational Statistics. 2016. "Digest of Educational Statistics." [online] Available at: https://nces.ed.gov/programs/digest/

Neves, J. 2017. “The UK Engagement Survey 2017 Student Participation and Skills Gain.” Higher Education Academy. [online] Available at: https://s3.eu-west-2.amazonaws.com/assets.creode.advancehe-document-manager/documents/hea/private/hub/download/UKES%202017%20report%20FINAL%202_1568037563.pdf [Accessed on 21st April 2019]

Neves, J., and G. Stoakes. 2018. “UKES , Learning Gain and How Students Spent Their Time UKES , Learning Gain and How Students Spent Their Time.” *Higher Education Pedagogies,* 3 (1), pp. 219–221.

Nicol, D. 2007. “E-Assessment by Design: Using Multiple-Choice Tests to Good Effect.” *Journal of Further and Higher Education,* 31 (1), pp. 53 - 64.

Nijjar, A.K. 2018. “Learning Gain in English Higher Education: Progress Report,” *Higher Education Funding Council for England* [online] Available at: https://dera.ioe.ac.uk/31344/1/HEFCE2018_03.pdf [Accessed on 12th March 2019]

OECD. 2008. *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*. New York: OECD Publishing.

OECD. 2011a. "Building a High-Quality Teaching Profession: Lessons from around the World" [online] Available at: https://files.eric.ed.gov/fulltext/ED518775.pdf [Accessed on 12th February 2019]

OECD. 2011b. “A Tuning-AHELO Conceptual Framework of Expected/Desired Learning Outcomes in Engineering.” [online] Available at: http://www.oecd.org/edu/skills-beyond-school/43160507.pdf%0Ahttp://www.oecd.org/dataoecd/46/34/43160507.pdf. [Accessed on 14th February 2019]

OECD. 2015. “AHELO Main Study.” *OECD Supports the Assessment of Learning Outcomes* *in Higher Education*. [online] Available at: http://www.oecd.org/education/skills-beyond-school/ahelo-main-study.htm. [Accessed on 13th February 2019]

OECD. 2018. "The Future of Education and Skills: Education 2030." [online] Available at: https://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf [Accessed on 11th February 2019]

Office for Students. 2018a. "The Regulatory Framework for Higher Education in England." [online] Available at: https://www.officeforstudents.org.uk/advice-and-guidance/regulation/the-regulatory-framework-for-higher-education-in-england/. [Accessed on 4th March 2019]

Ontario Government. 2005. "Reaching Higher: The McGuinty Government Plan For Postsecondary Education." [online] Available at: https://news.ontario.ca/opo/en/2005/05/reaching-higher-the-mcguinty-government-plan-for-postsecondary-education.html [Accessed on 2nd February 2019]

Oppenheim, A.N. 1972. *Questionnaire Design and Attitude Measurement*. London: Heinemann Educational Books.

Oppenheimer, D., F. Zaromb, J.R. Pomerantz, J.C. Williams, and Y.S. Park. 2017. "Improvement of Writing Skills during College: A Multi-Year Cross-Sectional and Longitudinal Study of Undergraduate Writing Performance." *Assessing Writing,* 32 (2015), pp. 12–27.

Ospina, S. 2004. "Qualitative Research." In *Encyclopedia of Leadership*, New Delhi: SAGE Publications, pp. 1-14.

Padilla-diaz, M. 2015. "Phenomenology in Educational Qualitative Research : Philosophy as Science or Philosophical Science ?" *Journal of International Excellence,* 1 (2), pp. 101–110.

Palmer, E.J., and P.G. Devitt. 2007. "Assessment of Higher Order Cognitive Skills in

Undergraduate Education: Modified Essay or Multiple Choice Questions? Research Paper." *BMC Medical Education,* 49 (2007), pp. 1–7.

Panke, D. 2018. *Research Design and Method Selection*. London: Sage.

Parker, P., E. Cleaver, and Y. El-Hakim. 2016. "Reflections on the Teaching Excellence Framework." *Educational Developments* 16 (4), pp. 1-6.

Pascarella, E.T., T.A. Seifert, and C. Blaich. 2010. "How Effective Are the NSSE Benchmarks in Predicting Important Educational Outcomes?" *Change: The Magazine of Higher Learning,* 42 (1), pp. 16–22.

Patton, M.Q. 2001. *Qualitative Evaluation and Research Methods*. California: SAGE Publications

Patton, M.Q. 2002. *Qualitative Methods in Social Work Research*. 2nd edn. London: Sage Publications.

Pike, G.R. 1995. "The Relationship between Self Reports of College Experiences and Achievement Test Scores *Research in Higher Education* , 36 (1), pp. 1-21.

Pike, G.R. 2011. "Using College Students' Self-Reported Learning Outcomes in Scholarly Research." *New Directions for Institutional Research,* 2011 (150), pp. 41-58.

Polit, D, and C.T. Beck. 2008. *Nursing Research: Generating and Assessing Evidence for Nursing Practice.* 8th edn. Philadelphia: Lippincott Williams and Wilkins.

Porter, S.R. 2013. "Self-Reported Learning Gains: A Theory and Test of College Student Survey Response." *Research in Higher Education,* 54 (2), pp. 201-226.

Possin, K. 2013. "A Fatal Flaw in the Collegiate Learning Assessment Test." *Assessment Update*, 33 (3), pp. 8–12.

Raban, C., and D. Cairns. 2015. "Where Do We Go from Here?" *Perspectives: Policy and Practice in Higher Education* 19 (4), pp. 107–115.

Raddon, A. 2010. “Early Stage Research Training:Epistemology and Ontology in Social Science Research.” [online] Available at: https://www2.le.ac.uk/colleges/ssah/documents/research-training-presentations/EpistFeb10.pdf [Accessed on 19th March 2019]

Rae, B. 2005. “Ontario: A Leader in Learning.” *Postsecondary Review*. [online] Available at: http://www.ontla.on.ca/library/repository/mon/10000/250004.pdf [Accessed on 18th April 2019]

Randall, J., and G. Engelhard. 2010. “Examining the Grading Practices of Teachers.” *Teaching and Teacher Education,* 26 (7), pp. 1372 - 1380.

Randles, R., and A. Cotgrave. 2017. “Measuring Student Learning Gain: A Review of Transatlantic Measurements of Assessments in Higher Education.” *Innovations in Practice* 11 (1), pp. 50–59.

Rea, L.M., and R.A. Parker. 2014. *Designing and Conducting Survey Research: A Comprehensive Guide*. 4th edn. San Francisco: Jossey-Bass.

Rebolj, B. 2017. “The Case Study as a Type of Qualitative Research.” *Journal of Contemporary Educational Studies*, 1 (2013), pp. 28-43.

Reichertz, J.O. 1995. “Abduction, Deduction and Induction in Qualitative Research,” In *a Companion to Qualitative Research*, London: SAGE Publications. pp. 299–310.

Research Methodology. 2018. “Epistemology.” [online] Available at: https://research-methodology.net/research-philosophy/epistomology/. [Accessed on 22nd March 2019]

Rippner, J.A. 2016. *The American Education Policy Landscape*. New York: Routledge.

Robbins, J.K. 2011. “Problem Solving, Reasoning, and Analytical Thinking in a Classroom Environment.” *The Behavior Analyst Today,* 12 (1), pp. 40-47.

Rodgers, T. 2007. “Measuring Value Added in Higher Education: A Proposed Methodology for Developing a Performance Indicator Based on the Economic Value Added to Graduates.” *Education Economics,* 15 (1), pp. 55–74.

Rogaten, J., B. Rienties, R. Sharpe, S. Cross, S. Lygo-baker, and A. Littlejohn. 2019. “Assessment & Evaluation in Higher Education Reviewing Affective , Behavioural and Cognitive Learning Gains in Higher Education.” *Assessment & Evaluation in Higher Education,* 44 (3), pp. 321–337.

Rogaten, J., and B.C. Rienties. 2018. “Which First-Year Students Are Making Most Learning Gains in STEM Subjects ? Which First-Year Students Are Making Most Learning Gains In.” *Higher Education Pedagogies,* 3 (1), pp. 161–172.

Roohr, K.C., H. Liu, and O.L. Liu. 2016. “Investigating Student Learning Gains in College: A Longitudinal Study.” *Studies in Higher Education*, 42 (12), pp. 2284 - 2300.

Rosenthal, R., and R.L. Rosnow. 2008. *Essentials of Behavioral Research: Methods and Data Analysis*. New York: The McGraw-Hill Companies.

Rowley, J. 2011. “Designing and Using Research Questionnaires.” *Management Research Review,* 37 (3), pp. 308-330.

Ryan, G.W., and H.R. Bernard. 2003. “Techniques to Identify Themes.” *Field Methods,* 15 (1), pp. 85–109.

Saunders, M., P. Lewis, and A. Thornhill. 2016. *Research Methods for Business Students*. 7th ed. Essex: Pearson Education.

Saunders, M. 2016. "Understanding Research Philosophies and Approaches." In *Research Methods for business Students*, 7th edn, Essex: Pearson Education. pp. 122-161.

Scalise, K., M. Douskey, and A. Stacy. 2018. “Measuring Learning Gains and Examining Implications for Student Success in STEM.” *Higher Education Pedagogies,* 3 (1), pp.10–22.

Schmidt, Henk G., Stephanie L. Wagener, Guus A.C.M. Smeets, Lianne M. Keemink, and Henk T. van der Molen. 2015. "On the Use and Misuse of Lectures in Higher Education." *Health Professions Education,* 1 (1), pp. 12–18.

Schuwirth, L., G. Bosman, R.H. Henning, R. Rinkel, and A.C.G. Wenink. 2010. "Collaboration on Progress Testing in Medical Schools in the Netherlands." *Medical Teacher,* 32 (6), pp. 476–479.

Scouller, K. 1998. "The Influence of Assessment Method on Students' Learning Approaches: Multiple Choice Question Examination versus Assignment Essay." *Higher Education,* 35 (4), pp. 453–472.

Segal, D.L., F.L. Coolidge, A. O'Riley, and B.A. Heinz. 2006. "Structured and Semistructured Interviews." In *Adult psychopathology and diagnosis.* 7th edn. New York: Wiley, pp 103 - 129.

Seifert, T.A., and A.M. Asel. 2011. "The Tie That Binds: The Role of Self- Reported High School Gains in Self- Reported College Gains."*New Directions for Institutional Research,* 2011 (150), pp. 59-72.

Shank, G. 2002. *Qualitative Research. A Personal Skills Aproach*. Merril Prentice Hall.

Shermis, M.D. 2008. "The Collegiate Learning Assessment: A Critical Perspective." *Assessment Update,* 19 (6), pp. 9–11.

Siekmann, G., and P. Korbel. 2016. "Defining 'STEM' Skills: Review and Synthesis of the Literature," [online] Available at: https://files.eric.ed.gov/fulltext/ED570655.pdf [Accessed on 14th February 2019]

Simpson, O. 2006. "Predicting Student Success in Open and Distance Learning." *Open Learning,* 21 (2), pp. 125–138.

Speight, L., K. Crawford, and S. Haddelsey. 2018. "Towards Measures of Longitudinal Learning Gain in UK Higher Education : The Challenge of Meaningful Engagement

Education.” *Higher Education Pedagogies,* 3 (1), pp. 196–218.

Spradley, J.P. 1979. *The Ethnographic Interview*. Belmont: WADSWORTH CENGAGE Learning.

Stake, R.E. 1995. *The Art of Case Study Research*. California:Sage Publications.

Steedle, J., H. Kugelmass, and A. Nemeth. 2010. “What Do They Measure? Comparing Three Learning Outcomes Assessments.” *Change: The Magazine of Higher Learning,* 42 (4), pp. 33–37.

Steinbock, K. 2014. "Eastern Wyoming College Outcomes Assessment Summary Report." *Eastern Wyoming College.* [online] Available at: https://ewc.wy.edu/documents/2018/03/2016-2017-outcomes-assessment-report-final.pdf/ [Accessed on 12th October 2018]

Sternberg, R.J. 1986. "Critical Thinking: Its Nature, Measurement, and Improvement." [online] Available at: https://files.eric.ed.gov/fulltext/ED272882.pdf [Accessed on 13th February 2019]

Suddaby, R. 2006. “What Grounded Theory Is Not.” *Academy of Management,* 49 (4), pp. 633–642.

Surridge, P. 2008. "The National Student Survey 2005-2007: Findings and Trends." [online] Available at: https://www.hefcw.ac.uk/documents/council_and_committees/council_papers_and_minutes/2008/TQI%2008%2079%20Annex%20A.pdf [Accessed on 18th November 2018]

Tam, M. 2004. “Using Students’ Self-reported Gains as a Measure of Value-added.” *Quality in Higher Education,* 10 (3), pp. 253–260.

Tekwe, C.D., R.L. Carter, J. Algina, M.E. Lucas, J. Roth, M. Ariet, T. Fisher, and M.B. Resnick. 2004. “An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance.” *Spring,* 29 (1), pp. 11–36.

The Council of Independent Colleges. 2008. “Evidence of Learning: Applying the College Learning Assessment to Improve Teaching and Learning in the Liberal Arts College Experience.” Learning Outcomes Assessment [online] Available at: https://learningoutcomesassessment.org/documents/CLAreport.pdf [Accessed on 20th March 2019]

Thomas, D.R. 2006. “A General Inductive Approach for Analyzing Qualitative Evaluation Data.” *American Journal of Evaluation,* 27 (2), pp. 237–246.

Tremblay, K, D Lalancette, and D Roseveare. 2013. “Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study.” [online] Available at: http://www.oecd.org/education/skills-beyond-school/AHELOFSReportVolume1.pdf [Accessed on 7th January 2019]

Tucker, M., E. Mulliner, and H. Wilson. 2017. “Research Preparedness in the Undergraduate Property and Construction Curriculum : The Student Experience Research Preparedness in the Undergraduate Property and Construction Curriculum : The Student Experience.” *International Journal of Construction Education and Research,* 13 (3), pp. 203–224.

Turner, P., and S. Turner. 2009. “Triangulation In Practice.” *Virtual Reality,* 13 (3), pp. 171-181.

Tymms, P. 1999. “Baseline Assessment, Value-Added and the Prediction of Reading.” *Journal of Research in Reading,* 22 (1), pp. 27–36.

University of Minnesota. 2018. “Student Experience in the Research University (SERU).” [online] Available at: https://www.oir.umn.edu/surveys/seru. [Accessed on 13th January 2019]

University of Rhode Island. 2012. “Wabash National Study: Additional Information.” [online] Available at: https://web.uri.edu/assessment/wabash_study_info/ [Accessed on 17th

February 2019]

Vermunt, J.D., S. Ilie, and A. Vignoles. 2018. “Building the Foundations for Measuring Learning Gain in Higher Education : A Conceptual Framework and Measurement Instrument Measurement Instrument.” *Higher Education Pedagogies,* 3(1), pp. 266–301.

Voluntary System for Accountability. 2018. “Excellence in Assessment.” [online] Available at: http://www.voluntarysystem.org/eia. [Accessed on 9th December 2018]

Wiers-jenssen, J., B.J. Stensaker, and J.B. Grogaard. 2010. “Student Satisfaction : Towards an Empirical Deconstruction of the Concept.” *Quality in Higher Education,* 8 (2), pp. 183–195.

Williams, J., G.C. Ansfield, J. Williams, and G. Cappuccini-ansfield. 2007. “Fitness for Purpose ? National and Institutional Approaches to Publicising the Student Voice.” *Quality in Higher Education,* 13 (2), pp. 159–172.

Wilson, H.K., and A. Cotgrave. 2016. “Factors That Influence Students’ Satisfaction with Their Physical Learning Environments.” *Structural Survey,* 34 (3), pp. 256–275.

Wright, I., P. Blomfield, R. Fuller, P. Kyle, A. Milling, J. Reynolds, A. Solloway, et al. 2016. "The Teaching Excellence Framework: Assessing Quality in Higher Education." *House of Commons Business, Innovation and Skills Committee.* [online] Available at: https://publications.parliament.uk/pa/cm201516/cmselect/cmbis/572/572.pdf [Accessed on 8*th* March 2019]

Xia, H., and B. Zhong. 2014. “Student Learning Outcomes Assessment in China: Primitive Foundations.” In *Higher Education Learning Outcomes Assessment: International Perspectives*, pp. 113-134.

Yazan, B. 2015. “Three Approaches to Case Study Methods in Education.” *The Qualitative Report,* 20 (2), pp.: 134–152.

Yin, R.K. 2002. *Case Study Research: Design and Methods.* 5th edn. California: Sage Publications.

Young, J.W. 2007. “Validity of the Measure of Academic Proficiency and Progress (MAPPTM) Test.” *Educational Testing Service*. [online] Available at: https://www.ets.org/s/mapp/pdf/5018.pdf [Accessed 25th November 2018]

Zilvinskis, J., A.A. Masseria, and G.R. Pike. 2017. “Student Engagement and Student Learning: Examining the Convergent and Discriminant Validity of the Revised National Survey of Student Engagement.” *Research in Higher Education,* 58 (8), pp. 880–903.

Zlatkin-Troitschanskaia, O., C. Kuhn, and M. Toepper. 2014. “Modelling and Assessing Higher Education Learning Outcomes in Germany.” In *Higher Education Learning Outcomes Assessment: International Perspectives*, pp. 213 - 236.

10. Appendices

Appendix 1: Participant Information sheet, Consent and Debrief from STEM Interviews

LIVERPOOL JOHN MOORES UNIVERSITY

Title of Project*:* Learning Gain in STEM Subjects in U.K. Higher Education Institutions.

Researcher: Rebecca Randles

You are being invited to an interview. Before you decide it is important that you understand why the interview is being conducted and what it involves. Please take time to read the following information. Ask the researcher if there is anything that is not clear or if you would like more information. Take time to decide if you want to take part or not.

1. What is the purpose of the interview?

The purpose of the interview is to get information about what students hope to gain from doing their undergraduate degree. This includes discussing what students think makes a "good" degree, the skills they gain at university and how this translates onto employment.

2. Who can take part?

Students from STEM subjects who are over the age of 18 years old. This is due to the research being focused on STEM subjects and therefore it is important that the information comes from students who are on STEM subject courses.

3. Do I have to take part?

No. It is up to you to decide whether or not to take part. If you do you will be contacted by the researcher to ask about your availability to arrange a time. At the interview you will be given this information sheet and asked to sign a consent form. You are still free to withdraw at any time and without giving a reason. A decision to withdraw will not affect your rights/any future treatment/service you receive.

4. What will happen to me if I take part?

If you decide to take part in this interview, you will be invited to a set location on Liverpool John Moores University premises. The interview will be recorded for transcribing purposes but no real names will be used in the report. You will be asked a series of questions surrounding university and your experience on your course. The session will be around 30 minutes long and you will be given a debriefing sheet at the end of the session.

5. Are there any risks / benefits involved?

There are no intended benefits associated with taking part however the findings will help the researcher in developing a questionnaire on learning gain for STEM subjects in university.

6. Will my taking part in the study be kept confidential?

Any information you provide will be kept strictly confidential. You do not have to provide a name on any questionnaires or information sheets and therefore the information you provide cannot be identified with your name. The interview will be recorded and transcribed but this will only be heard/seen by the researcher. You are requested to provide the researcher with a signed or initialled consent form. This will be kept by the researcher separate from the any other information you provide. This will be stored securely and destroyed within 24 months of completion of the study. Pseudonyms will be used in all transcripts and reports relating to the study to protect the identity of individuals and organisations.

7. Has this study been approved by an ethics committee?

LJMU Research Ethics Panel has assessed the research study and approved it. (REC: 16/BUE/010, approved on the 15th of December 2016).

8. Who to contact with enquires about this study?

Rebecca Randles; R.Randles@2013.ljmu.ac.uk

If you have any concerns regarding your involvement in this research, please discuss these with the researcher in the first instance. If you wish to make a complaint, please contact researchethics@ljmu.ac.uk and your communication will be re-directed to an independent person as appropriate.

LIVERPOOL JOHN MOORES UNIVERSITY

Project title: Learning Gain in STEM Subjects in U.K. Higher Education Institutions.

Researcher: Rebecca Randles

1. I confirm that I have read and understand the information provided for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason and that this will not affect my legal rights. ☐
3. I understand that any personal information collected during the study will be anonymised and remain confidential. ☐
4. I agree to take part in the above study and I agree to this session being recorded by the researcher. ☐

Name of Participant **Date** **Signature**

Name of Researcher **Date** **Signature**

LIVERPOOL JOHN MOORES UNIVERSITY

DEBRIEFING SHEET

Thank you for participating in this interview.

The aim of the interview was to get as much information as possible, from a students perspective, on the skills they hope to gain from their university degree. This information will be used in order to create a questionnaire to measure students Learning Gain at University.

It is not possible to provide you with your individual data or to withdraw your data after debriefing, as it has all been kept completely anonymous.

If you wish to know the overall findings of the study or would like more information or to comment on your experience please feel free to contact the Researcher: Rebecca Randles (R.Randles@2013.ljmu.ac.uk)

Thank you again for your participation

Appendix 2: Example of Interview Transcript from STEM Subjects

Participant: 5

Age: 21

Gender: Female

Subject: Geography

Level: 6

Interviewer: Can you just start by telling me how old you are, what you are studying and what level.

Participant 5: I’m 21, I’m studying Geography and I’m Level 6, third year.

Interviewer: Okay. So, can you tell me about the skills that you’ve developed whilst you’ve been on your course.

Participant 5: I’ve definitely developed like, people skills, I know that’s not one that you would think to go with geography but, like the ability to speak to people of different ages, especially people older than me and relay information to them – that I’ve obviously got skills in from doing my degree and I can give the education to other people. Erm, more related to the course, I’ve learnt practical skills, like field based skills which I think Geography at John Moores is really good at doing; learning like field based skills and measuring stuff. I’ve learnt computer skills, which I think would be really good if you wanted to go into GIS, personally, I don’t want to do that, but if you did want to go into doing GIS or erm, like mapping skills on the computer, it would be good to do that.

Interviewer: What about critical thinking?

Participant 5: Yeah. Like, erm, you’ve got to look at information that’s been given to you and then say whether you think that’s right or not. Information that you know, erm, then you look at papers and analyse whether you think what information they are saying compares to information B and what you think compared to that, which I probably didn’t do before I came to uni, I just read it and was like “yeah that’s right”.

Interviewer: And essay writing?

Participant 5: Erm, I wouldn’t say we write essays as such, erm, I’d say more reports. And what we tend to do is we tend to do some practical work and then write a report on it. So you would say how you did it, your method and then your results, you discuss your results. Then you’d compare it to other papers and evaluate it. So, it’s not like a – we do a sort of essay when we do lit reviews but I’d say that’s the closest we come to doing an essay really.

Interviewer: and do you think you have developed your writing since being on the course?

Participant 5: Oh yeah. Definitely, erm, I did English as an AS and I did dance as an A level and they were the only things I would write in because I did quite science based A levels. Erm, so my essay writing and ability to write long pieces of work was shocking. Absolutely terrible. Erm, and I've definitely, since coming to uni, like I've learnt how to use referencing tools and stuff which I had never been able to do before uni. And like formatting, which I think is such a key thing especially as we have just done a erm, stakeholder analysis, which you definitely do in the working world which you would never have learnt different types of styles of writing. When I was at school it was just 4 mark answer, 5 mark answer, 10 mark answer; you just was prepping for the exams, where as at uni you are prepping for actual real life writing.

Interviewer: Erm, do you do like, problem solving?

Participant 5: What do you mean?

Interviewer: Like erm, do you have things that go wrong and you have to change what you're doing

Participant 5: I don't really, erm, because I've picked quite – subjects that – most of the modules and subjects that I choose are not really problem based, but some of my friends do like major lab work which goes wrong and then they have to change the whole method. Where as I don't really do that.

Interviewer: Are there any skills that you thought you might develop that you haven't?

Participant 5: Erm, I thought I would do more lab based work, I thought I would do a lot more of the chemistry side of geography. But, then again, when I've come to uni I have kind of realised – well, I hated A level chemistry – so I kind of was like 'no more of that!' and I stuck to more human geography. I think it's by my own choice really.

Interviewer: do you think that there are some skills that are more important than others? And why?

Participant 5: erm, yeah, I think being able to speak to people is a huge thing. And I think every course, whether it's geography or English or any course, I think that needs to be developed because all you do through your whole life is meet people, talk to people, give them information. And I think a lot of people – like I'm quite confident and will speak to pretty much anybody and I don't mind presenting in front of people but I know a lot of people in my class hate it and the thought of doing it is petrifying and I think that should be developed more. I'm not sure how, because we do do it a lot, hmm, I think that's something that needs to be developed through-out all people but I think a lot of it is not necessarily what you do in class like I do a lot of extra curricular, always have done, and I think that works hand in hand with school.

Interviewer: in terms of STEM subjects, do you think that there are skills which run across all of them?

Participant 5: I definitely think there's skills in all subjects that run across all subjects – from the arts to the sciences. I think one thing that comes from science subject is the ability to write reports and erm, take real life problems that are going on now - analyse them and relay information to people. I think it's – a lot of the subjects it's kind of, what you find and then explaining it to people. That differs to other subjects in that, in other subjects such as dance or English, you are using somebody else's work and writing about it but really not relaying the information to anybody else just kind of what you think about it. I don't know, that might be a generalisation.

Interviewer: Do you think you have gotten more independent in your learning since you've been on your degree?

Participant 5: Yeah, I was always very independent anyway. Erm, I always have been, but I think when you come to uni there is a lot of time where you have to actually go and do work, like you have to get up and go to the library. Whereas, in school, you might not have done that. Theres a lot more erm – self encouragement? I don't know what the word is. Like, forcing yourself to do things. Time management? I don't know.

Interviewer: if you were going to be measured on these skills, how would you prefer to be tested?

Participant 5: I hate things online so definitely offline. I hate it, I don't think you can portray enough information online. Do you know when you do like a survey monkey or something?

Interviewer: Yeah

Participant 5: I just don't – if you have to answer a multiple choice and your answers not there? And I know for someone who is doing a research project or are trying to get information, they need like answer, answer, answer. But then, I've also seen it from the other side of being the researcher and how you can manipulate that information. Whereas if you get someone to write it or speak to them, you get every single side of it. And if I was to get assessed on my reading and writing online, it would be – it's not a fair comparison because online has spellcheck, copy and paste, when you're actually writing, that's there. I know you'll do better online because you can cheat, so to say, but it's not your actual reading and writing. And as well, reading online is totally different to reading actual physical stuff. I always print stuff off and read it on a piece of paper.

Interviewer: in terms of your personal development, do you think you've had a lot of opportunity to develop that?

Participant 5: Yeah but again, I think that's a personal thing because I choose to do the open days, I choose to do – I choose to turn up for group activities where as some people don't. So, again, I think yeah there is the opportunity to do it but whether you take those opportunities. Which I think, a lot of people don't and I also think it is a lot of the same people who do the same thing. I think if you put yourself forward for something then you will benefit from it but if you don't then you won't. if you turn up to a group piece of work and sit in silence then wont work.

Interviewer: So, you think that the engagement with the course is very important?

Participant 5: Yeah massively. I think engagement, not even just in the course but I think in the community and societies and sports, I think it's a total all rounded thing – like if you engage in something then you're obviously going to do better in it because you've got a whole round opinion on something.

Interviewer: So why did you choose geography?

Participant 5: Well, I was originally going to do pharmacy at john moores and then I decided in about March of A levels that I didn't want to do chemistry again because it was just killing me, not literally. So, I rang john moores and I said I'm meant to be doing pharmacy but I want to do geography what can I do and they said there is not a lot you can do now you'll just have to go through clearing, so I went through clearing to do geography – but I love geography! I always have. I think it was quite competitive for me though because my sisters all did geography and my family and friends so I wanted to do better than them. I'm super competitive.

Interviewer: Do you think you've gained a lot of knowledge in Geography?

Participant 5: Yeah, I wouldn't say it was a general knowledge, I've got a very in-depth knowledge on certain subjects. Whereas, I think at school it's a lot more - lots of little bits, I think at uni one of those little bits, you go into in depth. So, yeah I think I've gained a lot of knowledge at uni but very in depth on certain subjects.

Interviewer: So, do you have options in your final year?

Participant 5: Yeah we had options in second year as well. There was some core modules in the second year, and in third year it was all optional apart from research. What I think is really good is in geography there is like the option but then in that there is options as well. Usually we get – say for example, we just did a coasts piece of work and you think like seaside, pollution and stuff like that. From that you've got to pick something else so I picked wind energy, whereas one of my friends picked marine protection areas. So, in the choices there is further choices.

Interviewer: So, there is lots of opportunity to do what you are interested in?

Participant 5: Yeah and then there's opportunity to do what you think you are interested in and then realise that you're not.

Interviewer: So, thinking about your grades, do you think they accurately reflect the knowledge you have?

Participant 5: Sometimes. Sometimes absolutely not because sometimes you put a lot of work in something and get a rubbish mark and it's like "oh" but other times you think "that wasn't great, I didn't really know a lot about that and I winged it" and you get a better mark. So, yes and no, I think – I think its harder to blag something with a science subject because you're right or wrong. So, maybe, but then I don't know, I think yeah usually it reflects it but as I say with science it is harder to blag something.

Interviewer: What about exams compared to coursework? Are they different in portraying knowledge?

Participant 5: Yeah, I hate exams, I'm terrible at them, it makes me clammy thinking about it. And as well, this is a weird example, but I got mumps last year during my exams so I was really poorly during my exams. And I feel like that wasn't fair to reflect my whole year on those few days when I have my exams when I was really really poorly because in one of my exams I got a mark that didn't reflect my marks for the full year and I think with coursework you've got – if you have a bad week then you've got time to pick up from it. Whereas exams, if it's a bad day, it is a bad day, that's it, final, done. I do totally get why they do exams, it's just different ways of marking people. For example, if you're in the working world and you had a bad day in a presentation then that's it you've had a bad day, I get it, but then I also see it from my point of view where I was really poorly and it reflected badly on the whole year because of that. I think it happens a lot with people as well.

Interviewer: Do you know what you want to do for your career?

Participant 5: Nope.

Interviewer: No idea at all?

Participant 5: Nope.

Interviewer: Do you know it's going to be in Geography though?

Participant 5: Nope. Erm, I'd quite like to be a mum. I'm really sorry! I was asked in year 7 what I wanted to be and I said a mum. Its terrible. I really – I honestly don't know. I'm not that bothered either. I want to have lots of money and lots of children.

Interviewer: this kind of makes this next question redundant but do you think your course has prepared you for employment?

Participant 4: Yeah I do. Do you know what, like as bad as it sounds that I don't know what I want to do, I think my course really has because it's given me a wide range of skill sets, it's took me to countries where I have had to speak to foreign people, find information from them. I've been away twice with my course, overseas field trips, I've met people from all different walks of life which you will probably do in the working world. Erm, I've presented in front of people, I've wrote my CV and had it checked by god knows how many people. I know I don't know what I want to do, but I think it has. To me, I always say like I love my subject, I absolutely love geography, I think it's a bigger picture than that, the whole uni thing, is not just your subject.

Interviewer: So, that's everything, thank you.

Appendix 3: Coding Report for STEM Interviews

(2 Pages were taken to give an example rather than the entire 18)

Aggregate	Classification	Coverage	Number Of Coding	Reference Number	Coded By Initials	Modified On

Node

Nodes\\Critical thinking

Document

Internals\\Participant 1

No		0.0283	2			
				1	RR	25/01/2018 14:01

we're doing like a lot of coding and stuff and that takes like… like you learn a method but you can't just apply it like exactly,

				2	RR	25/01/2018 14:03

Definitely critical thinking I suppose because, I'm just thinking in terms of jobs really, its not going to be black and white.

Internals\\Participant 2

No		0.0149	1			
				1	RR	25/01/2018 14:08

Interviewer: What about critical thinking? Do you use that much?

Internals\\Participant 3

No		0.0144	2			
				1	RR	25/01/2018 14:11

critical thinking

				2	RR	25/01/2018 14:12

Participant 3: Critical analysis is key. Erm, I'd say the ability to problem solve in STEM fields and is arguably the most

Aggregate	Classification	Coverage	Number Of Coding	Reference Number	Coded By Initials	Modified On

Internals\\Participant 4

No		0.0153	1			
				1	RR	25/01/2018 14:15

Critical thinking. In some ways, I have, in some ways I think I'm being more inward. As in, I think in a way, I keep it to myself,

Internals\\Participant 5

No		0.0325	2			
				1	RR	26/01/2018 11:10

you've got to look at information that's been given to you and then say whether you think that's right or not. Information that you know, erm, then you look at papers and analyse whether you think what information they are saying compares

				2	RR	26/01/2018 11:10

Then you'd compare it to other papers and evaluate it.

Internals\\Participant 6

No		0.0437	4			
				1	RR	26/01/2018 11:46

there's a lot of critical thinking actually, this year, I guess that has been quite a big focus of level 6

				2	RR	26/01/2018 11:47

The forensic module. That's the only module where I have had to like, problem solve and that was more with an applied

				3	RR	26/01/2018 11:48

definitely critically evaluating

				4	RR	29/01/2018 13:14

critically evaluating what I'm given; research and papers.

Appendix 4: Questionnaire in Bristol Online Surveys

(2 pages were taken to give an example rather than all 27)

Page 4: Engagement

Below, you are about to see items which ask you about your independent learning and engagement. Please carefully read the questions and answer to the best of your ability.

10. I attend the lectures on my course.

- All of the time
- Most of the time
- Some of the time
- Rarely
- Never

11. I attend the workshops/seminars on my course.

- All of the time
- Most of the time
- Some of the time
- Rarely
- Never

12. My main motivation to attend lectures and workshops is

- My interest in the subject
- To obtain a better grade
- To gain skills for employment
- To satisfy course requirements

7 / 27

- Other

12.a. If you selected Other, please specify:

13. I prepare for my lectures.

- All of the time
- Most of the time
- Some of the time
- Rarely
- Never

14. I prepare for my workshops/seminars.

- All of the time
- Most of the time
- Some of the time
- Rarely
- Never

15. I ask questions during lectures.

- All of the time

Appendix 5: Example Comments From Student Pilot

18 Are there any comments you have in regards to the wording/structure of the questions?. If not, please leave blank.

Showing all 5 responses	
- You can't attend "all" of the lectures on a course "some of the time". You can attend some of the total number of lectures on a course. Same applies for seminars/workshops.	281502-281494-24219739
It seems odd that you can select more than one answer for each question- may skew your results.	281502-281494-24222209
Q13 uses a different scale, perhaps for continuity of questions it should follow the same as the others also for SPSS this will be one less variable	281502-281494-24225111
You should have the lists either all below each other or all across the page. The 2 by 2 lists came across a little confusing.	281502-281494-24225967
Q16 read a bit funny	281502-281494-24227822

Appendix 6: Example Comments from Staff Validation Pilot

16 Are there any items which you believe should be reworded? If so, please leave the number below and what you think should be changed. If not, please leave blank.

Showing all 5 responses	
Q14 I think needs to include the word academic work otherwise some will interpret it as paid or voluntary work	281233-281225-24326469
Question 8 might be answered very differently depending on whether 'class' means lectures or workshops. you might want to specify/distinguish (or even add as separate questions perhaps?) I'd invite reflection on how to word this one.	281233-281225-24327177
Q10 asks about grades or assignments. these are different things so one question should be posed about each. Similarly, Q14 asks about assignments and exam revision. These should also be separated. It may be worth pointing out in relation to Q14 that it does not involve paid work.	281233-281225-24330999
13 - you could perhaps separate this into two questions, one on exam revision and one on coursework.	281233-281225-24332621
Q10 "discussed grades or assignments" - maybe this could be specified or developed further; wouldn't it be relevant to see whether feedback was discussed. "grades" sugests that it was after an assignment, whereas "assignments" could be before or after, i.e. not quite clear. Q14 "spend doing work outside of class" - the context is kind of clear, but in principle this could be related to any kind of work, not necessarily academic/course related Q14 response format: although unlikely to be more than 30 hours, by limiting the responses at the upper end the resulting data cannot accurately be treated as interval scale => limiting the statistical tests that can be used. Maybe consider to just let the participants enter the hours as free text?	281233-281225-24339671

Appendix 7: Example of Raw Data

(Descriptive Statistics and Correlation between Student Attendance of lectures and students importance)

Frequencies

Statistics

		Time1: I attend the lectures on my course.	Time1: Attending lectures on my course is important to succeed at university..
N	Valid	185	185
	Missing	15	15
Mean		4.58	4.70
Std. Deviation		.546	.525

Frequency Table

Time1: I attend the lectures on my course.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Some of the time	5	2.5	2.7	2.7
	Most of the time	67	33.5	36.2	38.9
	All of the time	113	56.5	61.1	100.0
	Total	185	92.5	100.0	
Missing	System	15	7.5		
Total		200	100.0		

Time1: Attending lectures on my course is important to succeed at university..

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	1	.5	.5	.5

	Neither Agree Nor Disagree	3	1.5	1.6	2.2
	Agree	46	23.0	24.9	27.0
	Strongly Agree	135	67.5	73.0	100.0
	Total	185	92.5	100.0	
Missing	System	15	7.5		
Total		200	100.0		

Nonparametric Correlations

Correlations

			Time1: I attend the lectures on my course.	Time1: Attending lectures on my course is important to succeed at university..
Spearman's rho	Time1: I attend the lectures on my course.	Correlation Coefficient	1.000	.329**
		Sig. (2-tailed)	.	.000
		N	185	185
	Time1: Attending lectures on my course is important to succeed at university..	Correlation Coefficient	.329**	1.000
		Sig. (2-tailed)	.000	.
		N	185	185

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix 8: Interview Schedule for Staff Interviews

Can you please start by explaining your role at LJMU and the teaching that you do.

General Skill Discussion

How do you tend to assess student knowledge and skills? What sort of assignments have they had this year?

- Exams
- Group work
- Presentations
- Coursework; Reports, Essays, Literature Reviews

What skills would you say students have developed this year?

- Level 4 vs Level 5 vs Level 6
- Academic vs Personal

Critical Thinking

Have students conducted any research so far?

Are there a lot of models/theory to the degree?

What types of sources are often used in assessments?

- Books
- Journal articles

Have you done any work on critical thinking?

- What has been done? How has it been taught?
- Does this increase? (4 vs 5 vs 6)

What kind of coursework pieces are written?

- Arguments
- Literature reviews
- Critical reviews
- Viewpoints

Independent Learning/Engagement

How would you describe student attendance throughout the year?

- Different from lectures to workshops

What would you say potentially motivates students to attend?

Do students often ask questions in lectures/workshops and does this vary depending on level or time of year?

Have you ever had students discuss anything academic with you?

- Grades
- Feedback
- Assignments

Do you think students utilise Canvas a lot? Do staff?

Working with others/Communication

Have students presented this year?

Have students done any group work?

What do you believe students think about group work?

Do you know of students discussing their exams/assessments with other students?

Final Questions

What would you say is the most important skill for your subject area?

Appendix 9: Example of Interview Transcript from Staff

Duration of Interview: 22 minutes

School/Programme: Natural Sciences and Psychology/ School of Law

Gender: Female

Names mentioned have been removed from the transcript to preserve anonymity.

Interviewer: Can you start by explaining your role at LJMU?

RS: I'm programme leader for forensic psychology and criminal justice. Which is a strange role because im based in the school of law rather than in psychology, but my background is in psychology. As well as managing the programme as a whole, I also teach at each level. I'm quite lucky because I teach each level all the way through the semesters. It's really nice because it gives me a good overview of when they come in as fresh little faces right through to when they are graduating.

Interviewer: How do you tend to assess students?

RS: It's a mix of everything, we do have an assessment strategy. The main part of the strategy is that they don't suddenly get to level 6 and have a new type of assessment that they've never done before. If it's lets say a presentation, we wont have any students taking that module who have never done a presentation before. The other strategy is to try and give everybody a chance to be good across the board. So, some students like exams, some like coursework, some like reports. So, we sit down and look at what assessments we have got across all the core modules at each level. At level 4, we make sure they have got a range of unseen exams, seen exams, reports, essays, presentations and a few other little assessments on the way. So by the time they get to level 5, they have done everything so they are not terrified when they see that there is an exam. So that's the two parts to it really. We want to make sure that we are rounding them off.

Interviewer: Do they do group work?

RS: They do do group assessments yeah, in psychology they do small groups and things like seminar work. Which isn't assessed as such but is important for their ability. We are quite focused on employment as well, so knowing that a lot of them will go into the criminal justice sector and knowing what they will need to do as part of those roles. They will need to write assessment reports and that kind of thing. So we try and weedle our assessments into similar formats, so that they're used to writing in that way when they go off and become a probation officer or whatever.

Interviewer: What about research?

RS: They do research skills at every level. At level 4 they do all the research skills over in psychology. They do a bit of qualitative and a bit of quantitative, really basic but it's about learning the key skills and getting the fear away from them, which doesn't always work. At level 5, it gets a lot more advanced. So, I teach them their qualitative and I teach the workshops for the quantitative skills but they have their lectures over in psychology. One of the reasons I've done that, despite the fact I hate statistics, is that because our students are forensic psychology students they were finding a disconnect with the main psychology cohort, in those research methods lectures. All the examples they were using were about rats in a maze and that kind of thing, and they were wondering how any of it was relevant. So I have been able to say, okay the example is using rats in a maze but lets look at drug treatments. So, I hope that its starting to embed their thinking about research in what they are actually interested in.

Interviewer: Are there a lot of models and theory?

RS: Oh god yeah. To the content or to the way we deliver?

Interviewer: Content.

RS: It's BPS accredited. So it has to fulfil all the theoretical requirements of the BPS in order to get that. Our students are either lucky or unlucky, depending on how you look at it, because not only do they have to fulfil all the core elements for psychology but they have to learn all the criminological theory as well on the criminal justice side. A lot of it will overlap in parts, but a lot of it is radically different. I think that's really good for the way they think about things because they don't get stuck in this "this is how you look at things, this is our approach", just as they are learning about biology and brain and behaviour, they are looking at the social injustices. I do think it opens their mind a little bit. At the end of level 4 their minds are a little bit blown because they have had to be both really wide and really focused. But I think that's good for them because it does change the way they think about things.

Interviewer: What types of sources do they use? Is it mainly journal articles?

RS: We hope!

laughter

RS: We hope that they mainly use journal articles, particularly for psychology as the research is more up to date but they are the standard theoretical texts that they will use. Criminal justice; there's the journal articles and there's the standard texts. But there is also a lot of contemporary news stuff because obviously there is stories in the news all the time. They're updated by the time they come in for the lecture, so we have to keep on top of it, we have to use the latest example or we look out of date. Even if the latest example only happened yesterday, so it's very fast moving.

Interviewer: Do you use case studies as well?

RS: Oh yeah, you name it, we use it. One of the great things they do in forensic psychology at level 6 is problem based learning. All the examples are based on a scenario that a real life forensic psychologist may come across, and they have to go away and research what may have happened and why.

Interviewer: What about critical thinking, how do you embed that?

RS: From my point of view, I think we are ahead of the game with critical thinking, because like I said before we force them to think about things from two different topics to start off with so straight away they are realising they cant just accept everything they're being told because someone else is telling them something from a really different point of view, that's being analysed in a different way. But, certainly my level 4 module, anything I tell them I then say "so whys it not true" and they're a bit like "well what do you mean you just told us it's a seminal study" and I'm like "yeah it is, but what might not be true about it" and start trying to critique it; whether that's the ethics, the methodology, the assumptions they've made based on the data, what doesn't fit with other things that they know. I want them to know from an early stage that they're allowed to think for themselves. Because they very much have the attitude that if its in a book it must be perfect and right, and that's not the case. We critique every theory as we go through so we will talk about attachment theory but say there are huge disadvantages of this despite the fact that it's still used today. Its limited in numbers, it was align with the political climate at the time of women not going to work. But despite all that we still use it with sex offenders in prisons now. So what does that mean; have we not got any good alternatives? Or is it because there's actually something in it which helps us understand people.

Interviewer: Do you actually have a lecture onto critical thinking or do you just embed it?

RS: Some of the modules do have a session on critical thinking. We do try and help them with study skills. Some of the criminal justice modules do have a session where they think about what is critical thinking. We also have quite a lot of formative assessments early on, so doesn't count towards their final mark but they do get a lot of feedback and that feedback will focus on things like critical thinking. So that if they don't really get it, hopefully straight away we will be explaining and giving examples of where they could be critical.

Interviewer: What kind of coursework pieces is it?

RS: One of the very first ones they do is what we call a positional statement where they have to take a position on an issue, and argue it. We will present a title such as "the prison service aims to whatever is it fit for purpose" and they have to come out for or against. They can come out either way but what they have to do is justify their approach. That is marked in a particular way because it is one of the first assessments they do at level 4, so they have to have at least

8 academic references. And that causes issues because they don't understand what an academic reference is. They have to get a balance argument, it has to be well structured and it has to have an element of critical discussion. And they are the marking scheme. So, its quite tight and they can actually get 100% because as long as they have hit all of those, the content is sort've irrelevant, but it's not because they are getting that critique in. But they're getting that feedback straight away about this is where you have been critical and that is where you could've elevated that to the next mark range by doing this, that and the other. It's practical examples. We move on from that to, book reports as critiqueing a book. The usual critical analysis, evaluative type essay. There's research reports as well where they have to critique their own methodology.

Interviewer: How would you say the attendance changes?

RS: In every programme it does drop off. That is affected by what time of day the lecture is. We get a lot of feedback saying they want more contact time which is great. But then they say but not on a Monday or a Friday or before 11 or after 3. The attendance on the FPCJ is actually pretty high. It's not something I struggle with, what tends to be very clearly at level 4 is you tend to have 90% of the cohort who is there 90% of the time and then theres that 10% who clearly aren't gonna stick it. They'll start the first few weeks but then tail off and you might seem them once or twice throughout the year. But by the end of the year, its clearly not right for them. But we don't have many of those inbetween ones that sort of come in when they feel like it and then have weeks when they don't. We are pretty good at chasing them up as well. If they miss two weeks, they get an email off me. Not a horrible email just saying I have not noticed you in the lecture theatre are you okay. And for the ones who have just been partying a bit it's a boot up the bum. And then for the ones that are struggling it prompts them to come in and talk.

Interviewer: Do you have options at any point?

RS: At level 6, yeah they have options. They have to do their research project and forensic psychology for obvious reasons. But then they have 60 credits worth of options to take. They have loads of options, some from psychology and some from criminal justice.

Interviewer: What would you say motivates students?

RS: I think the content of the lectures. If it's interesting, they'll come. Some of it is about practicality, about trying to get up in the morning. But I keep telling them, when you get a job you're going to have to get up in the morning. The feedback we've had is that they really like the lecturers, they feel like the lecturers are not only engaged with the subject but are out there doing research. Our team is quite lucky as a lot of them are ex-practitioners. So if I see everyone is nodding off I can tell them a little story and it wakes them up a little bit. So I think that's a motivator for them. The other thing we really like is guest lecturers. When we pull people in from local employers to talk about the day in the life, and we do that in level 4 and straight away they can think that it sounds interesting or not what they thought it would be like.

Interviewer: You say that students are quite engaged, so do you get students asking a lot of questions?

RS: It depends on the level. They're pretty shy at level 4, they do ask questions in the big lectures but as soon as you break them down into seminars they don't open up as much. But then you find you have a queue of people at the end wanting to talk to you one on one, which I'm perfectly happy to do. It's a shame sometimes because sometimes a question infront of everyone, answers everyones question. One of the things we are putting in place is discussion boards on Canvas. Someone can ask a question, it gets answered, but then everyone can read the answer. That seemed to be quite popular in psychology and its something we are starting to do over here.

Interviewer: Do you ever have students come and discuss assignments and feedback?

RS: What we do is, obviously we don't want to comment on another module leaders marking. But, what we will do is if someone comes with their feedback and says I don't understand what this means. I will sit down with them as programme leader or they can go to their personal tutor. And then we can say okay what they mean by you haven't been critical here is that you haven't discussed XYZ. These are things you could've put in and just explain what that feedback means in language that they get.

Interviewer: Do you utilise Canvas a lot then?

RS: Yeah, Canvas is fabulous. I think it's a lot better than blackboard. So, just as a basic overview of whats in Canvas for them. Theres the lecture slides, a reading list, accompanying information, volunteer links relating to each subject, documentaries, even movies, contacts of local experts, all information on assessments, discussion board and even quizzes they can fill in to check whether they have got all the key points.

Interviewer: What sort of presentations do you do?

RS: Normally what we'll do is we will set people either a title or a problem. And they need to go away and do a short overview. We always start with group presentations which is sometimes more trouble than they're worth, but just so they're not stood up on their own. So we normally say okay do a presentation on one of these from the list, and what you find is everybody picks the same thing from the list. And then they'll have to do maybe ten minutes between four of them. They tend to do really well in the presentation, what they think is they're going to do terribly because they're stuttering and shaking and sweating. But we aren't looking at that, we are looking at whether they have done the work. As they move through they end up doing their own individual presentations.

Interviewer: Do you think students discuss assessments when its not group work?

RS: I think they do. Sometimes that's really useful, sometimes its not. Because it only takes one person to have the wrong information and it spreads like wildfire. We have had particular issues with people using facebook groups that only students have got access to and whats happening is someone is happening a perfectly sensible question, somebody else gets the wrong information, other people start panicking and it whips them up into a bit of an anxiety bomb. And I come in to 42 emails all asking the same question. But because we cant access that group to calm it all down, you've got people hysterical. Its become quite serious in places because people have ended up arguing on what the due date is but its in the guide. So we are hoping the discussion boards are more of an open forum for that kind of thing as well.

Interviewer: What would you say is the most important skill?

RS: I don't think its an academic skill. All of those things are important and they are going to need them in the future. I think the most important thing they can take is that "I can do this" and I think that is the most important thing someone can take. If you get to the end of the course, you are a different person than who you were when you started; you're more confident, you're more able, you're more independent and you're ready. And when we see them at graduation, that's when you feel proud and that is what I want people to come out with at the end. Is just that ability to say c'mon world what have you got for me.

Interviewer: That's all, thank you.

Appendix 10: Example of Survey Data from Students

(These were some examples from asking students about whether they have learnt critical thinking)

You get taught throughout the years about critical thinking. You just keep being expanded on the knowledge and get introduced to websites that may help throughout the year / topic and learn in different ways	357122-357113-33971852
No.	357122-357113-33979966
I have been expected to use this when searching for literature and reviewing source material. I have been taught about it in tutorial sessions. With each new assignment I'm improving my skills, so my use of critical thinking has increased	357122-357113-34005504
yes, it is important to my degree as we our required to evaluate a topic fully and not take the 'face value'	357122-357113-34052720
No	357122-357113-34096712
Yes. I remember learning about it in research methods. So, we have to read a number of journal articles to give an overall point of view and not from just 1 journal article. I believe my critical thinking has changed since the beginning of Level 5.	357122-357113-34107711
No as a part time student I only attend 1 day a week and we haven't had anything on research methods or critical thinking yet, but it is implied in the lectures. I think at level 4 we are just learning facts and not done much critical thinking at all apart from in our poster presentation.	357122-357113-34269006
Our teachers often challenge us to critically think through asking us questions based on our initial answers. We are often taught how to give an informed opinion. I feel my ability to critically think has progressed from this course.	357122-357113-34270633
I can't think of a time where we have been taught this.	357122-357113-34270750
Ive never been thought what critical thinking is.	357122-357113-34272414
Had to use it but were not taught it very well, we had an exam last semester on critical analysis of a paper	357122-357113-34277559