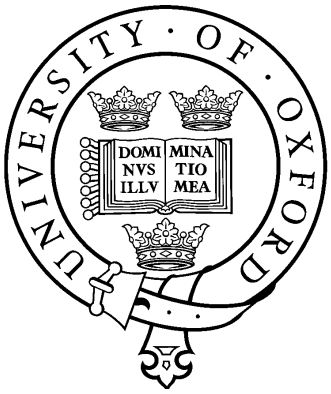
‘Backwash’ effects of high-stake degree level HE assessment on teaching practices.



Dr. Syed Ali Tarek
Candidate Number: 1056185
Supervisor: Prof. Jo-Anne Baird
St Stephen’s House
University of Oxford

A thesis submitted for the degree of
MSc Educational Assessment
Michaelmas Term 2022

DEPOSIT AND CONSULTATION OF THESIS

One copy of your dissertation will be deposited in the Department of Education Library via Oxford Research Archives where it is intended to be available for consultation by all Library users. In order to facilitate this, the following form should be inserted in the library copy of the dissertation.

Note that some graphs/tables may be removed in order to comply with copyright restrictions.

Surname	TAREK
First Name	SYED ALI
Faculty Board	DEPARTMENT OF EDUCATION
Title of Dissertation	'BACKWASH' EFFECTS OF HIGH-STAKE DEGREE LEVEL HE ASSESSMENT ON TEACHING PRACTICES.

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.

2. I understand that the Department requires that I shall deposit a copy of my dissertation in the Department of Education Library via Oxford Research Archives where it shall be available for consultation, and that reproductions of it may be made for other Libraries so that it can be available to those who to consult it elsewhere. I understand that the Library, before allowing my dissertation to be consulted either in the original or in reproduced form, will require each person wishing to consult it to sign a declaration that he or she recognises that the copyright of this thesis belongs to me. I permit limited copying of my dissertation by individuals (no more than 5% or one chapter) for personal research use. No quotation from it and no information derived from it may be published without my prior written consent and I undertake to supply a current address to the Library so this consent can be sought.

3. I agree that my dissertation shall be available for consultation in accordance with paragraph 2 above.

Abstract

The backwash phenomenon, or how assessment influences teaching and learning, is a firmly established field of inquiry within language testing research, particularly in the context of high-stake assessments. However, there remains a good deal of empirical research to shed light on similar phenomena in degree-level higher education (HE) contexts. This paper addressed that research lacuna by exploring summative tests in HE business studies and contributed novel insights into the impact of backwash influences in summative tests on academic practice and student behaviour.

Drawing on Alderson and Wall's (1993) washback hypotheses and Messick's (1989) conceptualisation of consequential validity, the research critically examined the nature and degree of assessment driven effects on learning and pedagogy. This research challenged claims by scholars such as White (2009) that argue higher education assessments are less susceptible to negative backwash effects due to their flexible nature in terms of design. Drawing on empirical evidence, the research provided insight into how summative high-stake assessments in HE can produce positive as well as negative consequences. This research learned that academic staff recognised the consequences of assessment design but often constrained by administrative and institutional demands, which suppressed their ability to innovate or improvise.

Teaching practices displayed a generally positive or neutral backwash effect, with lecturers being responsive and adaptable to the requirements of the curriculum rather than rigidly following preferred teaching methods. Learning was heavily influenced by summative assessment, as students focused on assessment tasks in ways that potentially undermine deep learning, indicating a clear negative backwash effect.

The research confirmed that any assessment with actual consequences, whether certification or academic progression, will inevitably lead to some sort of backwash. Findings validate that summative assessment in HE benefits teaching pedagogy with a positive backwash, hurts learning habits with a negative backwash, and has a neutral effect on teaching philosophy and techniques. This empirical research made a substantive contribution to the evolving academic dialogue on assessment design and highlight the need for more sophisticated policy and practice discussions within higher education assessment discourse.

Table of Contents

CHAPTER 1: INTRODUCTION

This dissertation reports a qualitative study of the 'backwash' effect of high-stake degree level higher education (HE) assessment on teaching practices. The study was conducted by interviewing academic and administrative staff members of a for-profit UK government-approved university in the UK.

1.1 Research problem

The backwash effect is the influence a test or assessment has on the way students are taught by the teachers (Alderson & Wall, 1993). In the assessment literature, at least from the 90s, test or examination impact on teaching and learning is referred to as washback (Alderson & Wall, 1993; Cheng & Curtis, 2003) or backwash (Biggs, 1995). In this dissertation, no differentiation is made between the etymology of both terminologies, and backwash is used throughout the thesis.

Heubert & Hauser (1999) argued that any assessment that has significant consequences for the test taker is a high-stake assessment. The assessment space within the HE differs from the commonly used high-stake tests. HE assessment has a wide range of assessment methods- portfolio submission, group assessment, time-constrained exams, essay, report writing and digital artefacts (presentation deck, poster, podcasts or recorded video). In HE, a high-stake summative assessment is the final stage of learning of a module where student-produced evidence is graded against marking rubrics, and criterion-referenced feedback is produced by the first marker to officially document the achievement of all the LOs or otherwise failures to meet the objectives (Dann, 2014). This research only considered summative assessment as the high-stake degree-level assessment because those assessment outcomes are likely to result in several crucial outcomes, including resits, failure to progress to the next level even failing to complete the programme of study. Even though the 'Backwash effect' is commonly discussed in language testing and other high-stake public examinations, there is a scarcity of research aimed at the HE assessment. White (2009) claimed HE assessment does not have to deal with negative backwash effects due to the absence of standardised tests. However, a growing body of literature suggests that assessment-driven learning practices within HE are likely to contribute to positive or negative backwash effects (Hussey & Smith, 2002; Elton & Brenda, 2002; Gibbs & Simpson, 2005; Bloxham & Boyd, 2012; Enriquez, 2020). The lack of empirical evidence on the above calls for primary research on the backwash effect of HE assessment.

1.2 Research Question

Based on the research problem, the following research question was confirmed.

RQ1: How does the 'backwash effect' impact teaching practices within higher education Business degree programmes?

1.3 Contextualising positive and negative backwash effects

Cheng et al. (2004) looked into the two possible categories of backwash: negative and positive. An assessment can have a negative backwash effect if it fails to reflect the learning outcomes of a course. On the contrary, the same assessment can have a positive backwash effect if it encourages covering the learning outcomes bringing harmony to what is taught and assessed.

Generally, all tests, especially language tests, are commonly criticised for negative 'Backwash effects' (Biggs & Tang, 2011; Cheng et al., 2004). In such a case, a test's content is a narrow subset of the curriculum, constraining teaching and learning possibilities. When a test has negative backwash, teachers tend to focus on covering the contents of the test, and students feel pressured to do well, resulting in incomplete delivery of the curriculum. Alderson & Wall (1993) proposed a set of 15 possible backwash hypotheses for language testing. The proposed hypothesises not only break down the backwash concept but also provides any researcher to use the hypothesises in other contexts beyond language testing. Since Alderson & Wall's (1993) initial ideation of the term, several studies have established negative impacts of test backwash in the form of narrowed curriculum, teaching to test, and a focus on surface learning. However, the number of studies that suggest no negative effects is low (Baird et al., 2017). Cheng et al. (2004) advocated further research on backwash to focus on the teaching and learning complexities associated with any assessment. The impact of HE assessment on learning continues to be less explored, and the claim made by White (2009) that HE assessment does not have to deal with negative backwash effects due to the absence of standardised tests requires the support of empirical research. Whether Alderson & Wall (1993) postulated hypothesises are applicable to the HE high stake assessment context is yet to be explored. Hughes's (1993) 'Trichotomy Backwash Model' can also be used in the context of the backwash effect in HE assessment. Hugh identified three key elements to backwash: Firstly, 'Participants'- students, classroom, teachers, administrators, course developers and publishers whose perception can be affected by a test. Secondly, 'Processes'- any actions taken by the participants affecting learning and thirdly, 'Products'- the quality of learning. This approach is likely to be helpful for the HE assessment as it covers a wide range of circumstantial elements that are common within a HE teaching and learning environment. Nevertheless, further categorisation of micro-macro contextual factors may become necessary in identifying the extent of positive or negative backwash.

The researcher also identified Watanabe's (2004) 'five dimensions' of backwash that academics could use in determining backwash effects. Specificity is the first dimension where the effect can be just a general assumption, such as all tests motivate test takers to study, or it can be specific such as the inclusion of a recorded assessment. Intensity is the second dimension that postulates if the exam is at high stake, then the exam can overpower the whole teaching and learning process. Length is the third dimension; Watanabe (2004) suggested that some tests have a short backwash effect, but some are prolonged. For instance, if the effect of the admission test continues after the admission, it can be termed a prolonged washback. Intentionality is the next dimension where Watanabe considered intended and unintended outcomes. The length of the backwash effect is a concept that requires further investigation using empirical data. Messick (1989a) implied that backwash could have both intended and unintended outcomes related to test interpretation and its use. The possibility of unintended consequences and side effects is essential, as is understanding the intended outcomes of an assessment. The final dimension is 'value'; to what extent the assessment impact is positive or negative can ultimately be determined by the evaluator of the assessment. Outcomes that teachers find positive may be judged otherwise by the administrators. To what extent does HE assessment backwash demonstrate these five dimensionalities have not been academically explored yet with empirical evidence.

The researcher had three potential routes to investigate the problem area: First, to take Alderson & Wall's work as a benchmark and assess whether backwash exists in the chosen sector. The second option would be to see whether the 'Trichotomy Backwash Model' is applicable within the HE environment and finally explore Watanabe's five dimensions of backwash in the HE context. Due to the scarcity of empirical research on the backwash effect of assessment in the Higher education context, it was ideal for the researcher to choose Alderson & Wall's (1993) work and establish the existence of the backwash effect in assessment, followed by identification of whether those effects are positive or negative.

This research aims to inform hypothesises 1, 2, 3, 4, 6, 11, 12, 13, 15 of Alderson & Wall (1993) and the details of the hypothesises are presented in Table 1.

Table 1: Alderson & Wall (1993) Hypothesises

Hypotheses	**Covered in this research**
H1: A test will influence teaching	Yes
H2: A test will influence learning	Yes
H3. A test will influence how teachers teach	Yes
H4: A test will influence what teachers teach	Yes
H5: A test will influence what leamers leam	No
H6: A test will influence how leamers leam	Yes
H7: A test will influence the rale and sequence of learning	No
H8: A test will influence the rate and sequence of teaching	No
H9: A test will influence the degree and depth of leaming	No
H10: A test will influence the degree and depth of teaching	No
H11: A test will influence the attitudes to the content, method, etc. of learning and teaching	Yes
H12: Tests that have important consequences will have washback	Yes
H13: Tests that do not have important consequences will have no washback	Yes
H14: Tests will have washback on all learners and teachers	No
H15: Tests will have washback effects for some learners and some teachers, but not for others	Yes

In terms of the broader education discipline, my research will contribute to the following Knowledge Gap:

KG1: Lack of empirical evidence on positive or negative backwash effect of HE assessment on learning practices

CHAPTER 2: LITERATURE REVIEW

Backwash effects of high-stake assessment directly impact teacher's teaching and student learning. Where teacher's instruction is concerned, it is necessary to explore learning theories and pedagogies. Further to this, two areas to explore with students' learning are learning motivation and the impact of the high-stake assessment outcome. There are also overlapping learning and teaching areas where the effect of assessment is felt in both learning and teaching. Furthermore, wider areas go beyond learning and teaching but profoundly impact both. Consequences for students and teachers, mental health and accountability are essential elements that require consideration in this research. In the following sections, each of these elements is further investigated.

2.1 Backwash effects of assessment on teaching and learning: Learning theories and pedagogies

Exploring the belief systems about learning is vital while considering the influence of assessment on teaching. Traditionally learning was mainly seen as acquiring knowledge and skills, but in contemporary literature and research, 'learning' has a wider scope, including emotional, social, and societal dimensions (Illeris, 2009). This research deals with degree-level programmes. The type of students who enrol in those programmes is adult widening participation students who return to education after a long gap in academic studies. Learning and assessment are both nuanced for these students as there can be a stark difference in the way they have experienced learning and assessment in the past. Their previous educational experiences also contribute to their motivation for learning and preparation for assessment. Hence, it becomes essential to consider learning theories in this research as these theories outline the central tenets of knowledge construction, retention, and the knowledge recall process. A teacher having a comprehensive understanding of how adult students learn can potentially assist in customising teaching materials for maximum impact and constructing assessment materials that can precisely assess learners' attainment.

In this research, the importance of learning theories and pedagogic approaches are considered for following three practicalities:

2.1.1 Learning theories that are embedded within HE practitioner's teaching philosophy

Three main types of learning theories are behaviourist, cognitivist, and socio-constructivist (Baird et al., 2017). Formal teacher training programmes largely focus on these main learning theories; as such, it likely impacts how a teacher interprets classroom engagement in the learning and teaching context. Main contributors to behaviourism, such as Thorndike (1920), Pavlov (1927), Watson (1959) and Skinner (1974), saw learners as passive, starting from a clean state (black box) and responding to environmental stimuli. Behaviourism does not consider consciousness or internal mental states, so learning relies upon positive and negative reinforcements to change the learner's behaviour. Albert Bandura realised direct reinforcement cannot account for all types of learning; he added a social element and argued that people learn by observing, modelling and imitating others (Bandura, 1965, 1977). Bandura's social learning theory bridged between behaviourist and cognitivist learning theories.

The next direction came from cognitivist theorists who argued for opening the 'black box' of the mind to understand the learning process better. Piaget's (1923) 'Schema theory' elaborated on how

cognitive structures created knowledge by a continuous process of accommodating[1] and assimilating[2] information. Atkinson–Shiffrin (1968) believed in similarities between human cognition and filling cabinets. Their proposed linear three-store mental model included: a sensory register, short-term memory and long-term memory. Information detected by sensory organs may not have any impact if they are not acknowledged. Repetition was seen as the key to reaching the farthest end of the chain – the 'long-term memory'. As such, the more something is rehearsed, the more chances are it will have a lasting impact from a memory recall perspective. Though memory and learning literature are not the same, memory has a role in how learning is perceived and thus requires some consideration in this research. This theory was critiqued for being too linear and not considering other internal factors that may interfere with a person's cognitive ability (Craik & Lockhart, 1972). Despite that, this memory model influenced other learning and memory theories. Academics who believe in such memory models will be seen repeating important tasks and content in the lessons with an aim to make those recallable at any time. When considering a degree-level programme, there is a general trend of having foundation year students transitioning to year one with limited subject knowledge and with the minimum academic skills to succeed in their first year (Roberts, 2002). It is important to see how both students and teachers see assessment at the early stage and whether the teachers take a repetitive approach to improve the chances of success in the summative assessment.

Another critical cognitivist theory is Bloom's 'taxonomy of educational objectives' (1956). It proposed a set of categories of learning stages to assist with designing learning and developing assessments. Bloom's taxonomy provided the needed practitioner's blueprint to assign appropriate progressive verbs for educational attainment or mastery in learning. An important tenet of the model is that each level needs to be mastered before moving to the next. Breaking down learning into specific, easily explainable stages made it favourable to a wide range of stakeholders in the education sector. Eventually, this taxonomy emerged as one of the practitioners' most influential models of learning. Even 65 years later, HE academics frequently use it to design learning and assess learners (Enriquez, 2020). Elton & Brenda (2002) raised concerns about situations where a poorly designed programme includes learning outcomes that either is a poor fit to the programme or are misfit. Academics who are handed a poorly designed programme to teach have little control to go beyond what is set in the programme and include content or summative assessment that may better represent the market demand. A situation like this can create a complicated situation where an academic does not believe in the programme/curriculum or the assessment's currency but is compelled to deliver those in the way the 'paperwork' requires him/her to deliver those. It can negatively affect learning as the academic may focus on the assessment and not cover the entire curriculum. Such scenarios need careful checking in the context of this research.

Socio-constructivists see learning as an environment-dependent, experiential, and dynamic process where knowledge is constructed by an individual's interactions with their peers, teachers and the society and culture at large (Vygotsky, 1978). Learning becomes more contextualised and personal as each learner brings their own unique experience. Active engagement, reflection, and motivation become critical in a meaningful learning experience. Socio-constructivist learning theories underpin student-centric teaching methods that often differ from the previously mentioned techniques where teachers passively transmit knowledge to students. Constructivist learning and assessment are primarily seen in classroom-based assessments within a higher education context. Summative assessments based on peer assessment, negotiated assignment, reflective records, and portfolios are rooted in socio-constructivist theories.

[1] Accommodation is creating a new schema that fits better with the new environment or adjusting the old one.

[2] Assimilation is when people use a current schema to understand the world around them.

One of the strengths of this socio-constructivism approach is that students have ownership of their learning and assessment. One of the biggest disadvantages of this approach is that it lacks structure, and some students require a highly structured learning environment to make the most of the learning opportunities. Traditional grading strategies are also changed in this approach as more importance is placed on students evaluating their progress. A lack of standardised grades means teachers may not identify the struggling students.

Baird et al. (2017) acknowledged cognitive schools of thought are prevalent in the choice of assessment systems, and socio-constructivist thoughts are more visible in classroom-based assessments. As it is an accepted view that high state assessments change teacher and student behaviours (Stobart & Eggen, 2012), it is important to see to investigate whether degree-level summative assessments may also change teacher and student behaviour (e.g. backwash effect) and in that to what extent learning theories and pedagogies have a role to play. Especially the researcher intends to explore the backwash effect considering the highly structured cognitivist approaches to delivery and assessment in contrast to student-focused delivery with non-traditional assessment methods.

2.1.2 Adult learning theories and the role of student motivation in higher education

Within the field of education, it is commonly found that theories that were proposed for children are applied to adult learning contexts without differentiation. Whether such differentiation is required or not is commonly debated, but there remains a continuous void in academia to establish clear guidance on to what extent the classic learning theories can be applied to adult learning without any change at all. Knowles (1973) was a prominent scholar who attempted to isolate 'andragogy'[3] from 'pedagogy' by distinguishing adults as growth-oriented autonomous learners. Around a similar time, 'Self Directed Learning (SDL)' strategies also started to emerge, bringing focus to adult learning (Rachal, 1983; Candy, 1991; Merriam, 2001). Alongside 'Self Directed Learning', similar concepts such as Self-Regulated Learning (SRL) (Zimmerman, 1989) and autonomous learning (Little, 1991) are also in existence, sharing similar meanings and contexts. The difference between these terminologies are narrow and often inconsistent; hence these are interchangeably used by many researchers (Loyens et al., 2008). In the late 90s, Lave and Wenger's 'Situated learning theory' and 'Community of practice' theories postulated how learners could learn through active participation with a common interest in constructing knowledge (Lave & Wenger, 1991, Lave, 2009, Wenger, 1998).

An adult learner entering higher education brings some form of institutionalised learning experience and their interpretation of the world around them. Being independent in the learning environment liberates adults to take the initiatives to identify learning needs, determine goals, scout sources, choose a suitable learning strategy and finally be able to evaluate their learning outcomes with or without external support (Knowles, 1975). Thus, adult learning can be seen as an interaction between three dimensions- sociological, pedagogical and psychological (Candy, 1991). SDL, SRL and similar other learning theories emphasise on intrinsic motivation[4] as a crucial competent and a lack of this is likely to result in failure in the learning process (Loyens et al., 2008). In this research, the researcher is interested in only the 'intrinsic motivation' aspect of learning and teaching as it is seen as crucial by an array of adult learning theories and the 'backwash' effect is likely to trigger intrinsic motivation for both students and teachers which requires further empirical studies.

[3] Art and science of helping adults learn.

[4] Intrinsic motivation is when one is motivated by personal satisfaction instead of external stimuli like reward or punishment.

Additionally, past experiences are a key to adult learning, providing learners with much-needed reflection. Dewey (1938), Kolb (1984) and Gibbs (1988) acknowledged learning through self-reflection. Self-reflective learning can make a HE learner succeed by continually improving through the process. A learner's aim is to reach a transformative stage where the learning makes a deeper impact. Mezirow (1997) defined 'transformative learning' as the process that transforms core assumptions through which we understand our experiences, expectations, perceptions, cognition and feelings. Such changes are bold and defining, which may be triggered with a single event or through a series of events lasting much longer (Mezirow, 1997; Cranton, 1994; Cranton, 2002; Kegan, 2008).

If we take the transformative stage as the ultimate stage of learning, then there must be intermediary transitional phases that lead to the end goal. 'Transitional learning' by Wildemeersch & Stroobants (2009) can fit the context quite well. Transitional learning defines those learning moments that challenge a learner to learn to anticipate, recognise and handle dynamic change in life and within a transforming context. Though these theoretical propositions exist in isolation, there is no empirical research done to combine these propositions to merge into a singular overarching learning proposition which the researcher believes sums up a HE student's learning journey.

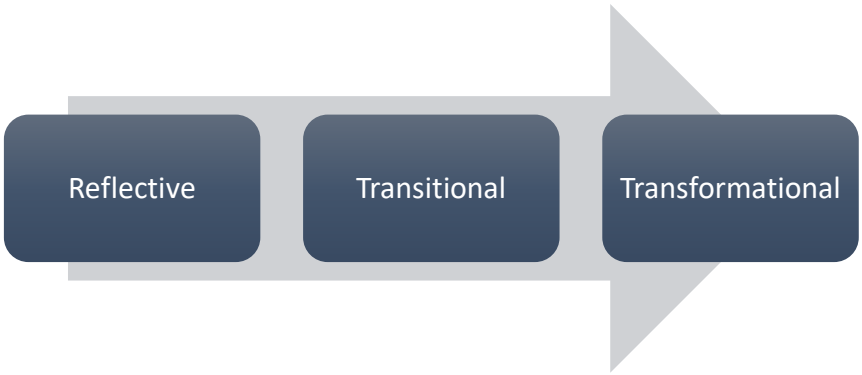


Figure 1: A HE Student's Learning Journey

As one of the areas of focus of this research will be on student motivation, whether there is a role of 'Transformational learning' in the HE context will be explored, covering three key theoretical strands of learning covering reflective learning (Kolb, 1984), transformational learning (Mezirow, 1997) and transitional learning (Wildemeersch & Stroobants, 2009) and transformational learning (Mezirow, 1997).

2.1.3 State of the direction on Teaching and Learning strategies from the UK HE Regulatory body, Professional Statutory and Regulatory Bodies (PSRBs)

Learning in HE is not just between the learner and teacher interacting within an instructional environment. There are other more powerful stakeholders in the system that very much dictates the way they expect learning to take place. In the UK, the regulatory body for higher education is Office for Students (OfS), and another significantly important independent body is Quality Assurance Agency (QAA). OfS legally holds universities to account for their teaching quality. QAA monitors and advises on the standards and quality of UK higher education to ensure students working towards a UK qualification receive their desired higher education experience.

As HE education is more outcome-based, learning outcomes are central to the delivery model. Regulatory bodies and PSRBs expect to put learners at the centre to ensure the learning is meaningful academically and provides value for money. It inevitably brings in two rather spaced-out teaching paradigms: 'active learning' for making learning meaningful and 'traditional- behaviourist/cognitivist

learning' to ensure learning is verifiable for economic prospects and accountability (Enriquez, 2020). Interestingly, in modern teaching, graduate attributes are deeply embedded in new programmes and academics are reminded to embed graduate attributes in their teaching. Johnston (2002) argued, that many of these attributes refer to 'soft skills', which are better produced within a constructive learning environment through collaborative learning, peer observation, team presentation and negotiated learning.

Within the HE, there is an increased focus on enhancing employability, embedding graduate attributes in the curricula, which is favoured by Professional Statutory and Regulatory Bodies (PSRBs). Many professional bodies validate academic programmes that embed a set of desired learning outcomes or deliver prescribed content/activity in a prescribed manner. Such validation can be led to an exemption from formal examination leading to membership in the professional body. These specific requirements have an evident influence on the programmes created for validation, modules written within that programme, and assessment strategies sought to evidence achievement within those modules/programmes. In this research, a desk review of PSRB (CMI, CIM and ILM) policies, indicated there is no prescribed teaching strategy for attainment of the learning outcomes even though there is clear guidance on what content needs to be covered. It suggests a clear gap in the PSRB expectations on assessments informing learning outcomes and the classroom teaching practices that deliver those learning outcomes.

2.2 Backwash effects of assessment on teaching and learning: A review of the complexities of assessment methods

It's not just teaching practices that are informed by learning theories; assessment design and assessment methods are also influenced by how practitioners and subject matter experts perceive how knowledge is constructed and the learner's role in that process. Cronbach (1988) and Messick (1989) advocated that learning theories should inform valid assessment practices. James (2008) categorised learning theories and their implications for assessment practices. In this mapping exercise, she proposed three generations of assessment: the first generation tied up with behaviourism, the second generation with cognitive constructionism and finally, the third generation of assessment with socio-cultural theories. James recognised the major theorists of the third generation focused more on learning than assessment, and there is a gap in knowledge in aligning learning with assessment within the third generation. However, to what extent such an equitable, like-by-like match can be made requires further support from empirical evidence.

Among the other leading approaches, Biggs & Tang (2011) proposed ‘Constructive Alignment' to create a synergy and coherence among intended learning outcomes (ILOs), teaching and learning experiences, and assessment tasks. The subsequent comprehensive research also encompasses the SOLO taxonomy (Biggs & Collis, 1982). While both Bloom's Taxonomy and SOLO taxonomy were not developed solely based on research on student learning, the former relied on judgements from educational administrators; thus, as per Biggs & Tang (2011), Bloom's Taxonomy and its other revised versions lacked clarity on student's level of understanding. Though SOLO is more teaching and learning-focused, it can also bring complexities in delivery where students may not learn at the same pace and fail to follow a 'one size fits all' approach to delivery. Placing students in predetermined slots based on their understanding level may discourage them from learning. Within SOLO, 'transitional' learning stages are well founded, but it doesn't include 'transformative' learning. Hence, some critiques want to see learning go beyond 'surface' and 'deep' learning and take the shape of 'lifelong' learning (Colding, 2020). Constructive alignment calls for teachers to have an active role in

determining the ILOs and assessment tasks that will be further linked with teaching activities. That is not always the case within the HE environment, as the lecturer delivering a module might have absolutely no involvement in the module or assessment design. Furthermore, a university student spends considerably less time with the lecturer, so 'qualitative assessment' of learning can be limited (Colding, 2020). Negotiated assessment though a possibility but can create issues with quality assurance unless the lecturer is well versed in such assessment design. Thus, constructive alignment becomes largely untenable in the HE environment.

One of the achievements of constructive alignment is that Biggs and Tang emphasised three key components that directly affect learning. The nature of higher education is dynamic and complex, making it seemingly difficult to bring issues such as what to teach, what will give students the subject knowledge and how to cultivate graduate skills. Expecting university students to achieve graduate attributes calls for learning that is situated in the society; learning that prepares students to tackle the 'unseen' problems of tomorrow, and acquire 'soft skills' that will prepare them to lead a diverse pool of people (Barrie, 2004; Bridgstock, 2009). Such learning calls for socio-cultural theories of learning. However, we cannot ensure total 'fairness', 'validity', 'reliability' unless the assessment methods are standardised. Within HE, assessments are mostly internally moderated, and the external examiner acts as a 'critical friend', so the risk is too high. Hence, an assessment paper that uses 'Bloom's taxonomy' can be seen favoured by the university administration on suitability and transparency. Such favouritism can lead to lecturers' departure from their belief system on learning and assessment (Bonner, 2016). Literature and research based on Argyris and Schön's (1974) 'espoused theory' and 'theory-in-use' can be relevant in this discussion as such departure can lead to a scenario where the theoretical beliefs on learning mismatches with the theories put into practice. In recent years, Annan et al. (2011) and Annan (2020) developed a tool- 'Learning Theories Profile (LTP)' for practitioners to help them identify the gap between theories used and believed. Though the authors anticipate academics can use it for university teaching, there is a lack of empirical data, and the choice of learning theories within the LTP is limiting.

In contrast, within the field of education, scholars continue to challenge the need for a universal or grand theory of learning that can amalgamate all the proponents of learning. Baird et al. (2014) acknowledged the challenges of mapping learning theories with assessment practices, especially where the theoretical underpinnings of commonly used learning theories are far apart. Sfard (1998) took a different approach in justifying the use of a specific learning theory. Sfard used two metaphors of learning – 'acquisition' and 'participation' to deal with the complexity. The acquisition is a gradual accumulation of concepts (basic units of knowledge) over time; on the contrary, participation is constant action. The acquisition metaphor sees the influence of cognitivist as well as constructivist ideas of knowledge construction. In the participation metaphor, learning occurs in the Community through constant discourse and communication. Sfard argued against the 'danger' of favouring one over the other and felt both could coexist similarly, like 'Euclidean' and 'non-Euclidean' geometry in mathematics. To Sfard, the complexities of learning cannot be seen by excluding one metaphor from another. It is worth exploring to what extent personal belief in learning theories and practices can impact the delivery or assessment in general, leading to positive or negative backwash.

2.3 Backwash and other types of assessment

Students encounter various forms of assessment throughout their higher education. From a curriculum instruction perspective, formative and summative assessments are the key assessments students must go through. It is often complicated to fully separate these two assessments making their separate existence blurry at times, raising debates and concerns among scholars (Bennett, 2011). Summative assessment is high stake, whereas formative assessment is a low stake assessment. Within

a HE learning environment, assessment of learning or 'summative assessment' is the ultimate indicator of students' learning.

Historically, Scriven (1967) elaborated on the distinctive nature of both types of assessment which was rather further contextualised by Bloom (1971) in his 'Mastery Learning' model'. Black & Wiliam (2003) argued the current understanding of both terms still resembled Bloom's interpretation. The most notable contribution to formative assessment in the last two decades came from the Assessment Reform Group (ARG) through their research in 'Assessment for Learning (AfL). Harlen & James (1997), Black & Wiliam (1998) of ARG, Brookhart (2001) and Hopfenbeck (2018) agreed that AfL is one of the leading interpretations of formative assessment. Within HE, commonly found AfL tasks are short essay submissions, progress tracking quizzes, take-home tasks, and peer assessment. The formative assessment comes in the form of a formal conversation between the first marker (in most cases, it is the lecturer) and the student on progress made towards achieving the Learning Outcomes (LOs) of the module. The lecturer identifies inaccuracies or inadequacies within the current submission and guides the student on how to achieve all the objectives in the summative assessment (Johnston, 2002). There is an ongoing debate among scholars to agree on the interrelation between formative (AfL) and summative assessment (Baird et al., 2014). Harlen (2007) is of the opinion that assessment created for summative assessment can be used for formative purposes. However, this may result in a tunnelled approach to coach to pass the summative assessment within the HE context. Whether academics practise such, needs to be investigated in the context of this research.

Moreover, Bennett (2011) found that interpretation of assessment feedback is less discussed in the formative assessment literature. This was also previously identified by Taras (2009), who claimed summative submission receives the most detailed feedback with indicative grades, whereas formative stage submissions do not always receive detailed feedback; furthermore, as they are non-graded, it is taken less seriously by the academics and students. Harlen (2005), White (2009) and Greenstein (2010) think that teachers who undervalue the role of AfL in the delivery miss out on opportunities to motivate and incentivise students to do well in the summative assessment. In addition, Tight (2004) found that HE academics are often reluctant to explicitly express their theoretical perspectives related to assessment practices which makes it important to consider whether the issue is in having profound knowledge of assessment methods and the applicability of those methods (i.e. 'assessment literacy') or it is because the delivery is skewed towards fulfilling the needs of high-stake summative assessments on learning.

2.4 Assessment literacy, politics of assessment and consequential basis of validity

Popham (2011) saw assessment literacy as an academic's understanding of the fundamental assessment concepts and procedural knowledge related to assessment. Stiggins (1995), Popham (2011) argued the concept of 'Assessment Literacy' is still in its early stages despite being around for 30 years (JISC,2015). Holroyd (2000) and Harlen (2007) argued that if academics see assessment as just a tool to measure and judge knowledge, then they are at risk of missing out on learning opportunities that these assessments provide. Assessment within HE should provide students with the opportunity and clear scope to self-regulate their learning, take ownership of learning, and further enhance the learning experience by achieving graduate attributes. It was reflected in Ball et al. (2012) reasoning of assessment literacy within the HE. They saw it as a cyclic process rooted in curriculum design and implementation, resulting in assessments that allow students to get 'unhurried' opportunities to develop complex skillsets and knowledge required to progress to the next level. Barrie (2004) and Bridgstock (2009) argued whether commonly used assessments like essays or examinations could make an informed judgement on students attaining soft skills such as negotiation,

team building or teamwork. Such concerns justify Johnston's (2002) call for innovating traditional methods to foster reliable assessment for assessment judging academic knowledge and graduate skills. In such cases, we also need to explore to what extent an academic has access to bring innovative assessment practices into the delivery.

Messick (1989b) proposed 'Consequential validity' as an integral part of test validity to investigate the actual and potential social consequences of testing. Test impact, washback/backwash, or social consequences are synonymously used to refer to consequential validity. In the early years, the terminology largely remained less discussed; it also lacked scholarly agreement on the exact meaning of the terminology and scope of the term (Mehrens, 1997). Messick (1995) proposed two dimensions of the consequential basis of validity: value implications and social consequences. Messick proposed test developers and takers assess a test's potential and actual consequences. Shepard (1997) later extended the social consequences component to include both intended and unintended consequences of a test. Construct validity is mainly at the centre of assessment validity-related discourse to ensure high stake summative assessment represents the content of the curriculum. Consequential validity is often ignored as it is seen as a 'great burden' (Hubley & Zumbo, 2011). In this research, there is a possibility to check whether academics see assessment validity discussions can go beyond 'instruments' used to assess to consider the 'inferences' of the assessment.

Ecclestone (2007) argued the importance of the context in which learning, and assessment take place. Likewise, Baird et al. (2017) suggested that assessment is also policy-driven, and educators are to follow the narratives of policymakers. Research exploring the backwash effect of assessment thus cannot ignore the scope of policymaking. Two lasting political agendas in the UK contribute to the assessment discussions; firstly, a push towards getting more under-represented BAME students in higher education using a widening participation strategy (David, 2012) and secondly, establishing value for money of academic programmes by improving the employability of students through soft skills embedded in the curriculum as graduate attributes (Johnston, 2002; Hill et al., 2016). The increasing pressure at all levels of the UK education system now equates achievement in learning with good grades. Students seen as 'consumers' in the UK universities resulted in more students expecting favourable, accessible assessments and good grades (Ecclestone, 2007). The UK education system now presents dropout and non-completion as failures and sees them as undesirable outcomes. Ecclestone's empirical research found that, despite assessments carrying a high socio-emotional stake, students demonstrated a lack of anxiety about assessments. She thought this was due to the opportunities students now have in the form of AfL and repeat submission opportunities for summative assessments. However, in recent studies, there is growing concern about the reasons for student dropouts. Trowler (2010) argues the primary reasons are lack of motivation and confidence to cope with the academic pressure.

Identifying the lack of data, Bradley (2017) explored the reasons by researching UK students to find that many entered university education without mental planning to tackle the educational challenges, resulting in unplanned stress. Bradley argued this stress than likely to pass on to the academics who are also tasked with 'guiding' students to succeed in their academic studies. Whether assessment has an impact on raising anxiety among academics was also questioned by Alderson & Wall (1993). Torrance et al. (2005) raised concerns that such pressure on attainment can lead to extensive coaching strategies to get students through the assessments resulting in reduced quality and validity of achieved outcomes. They were also concerned that there is a growing risk of the assessment process dominating the learning experience and assessment criteria compliance replacing learning. In this research, there is a scope to further investigate to what extent students and academics are mentally affected by assessment practices and to what extent that causes a negative or positive backwash.

The literature review indicated the following Knowledge Gap.

Knowledge Gap 1: Lack of empirical evidence on positive or negative backwash effect of HE assessment on learning practices

Considering the identified knowledge gap, the following research question was established for this research is:

RQ1: How does the 'backwash effect' impact teaching practices within higher education Business degree programmes?

CHAPTER 3: METHODOLOGY

In the literature review chapter, I established the lack of empirical studies dealing with the backwash effects of higher education summative assessment. Alderson & Wall (1993) proposed a set of possible backwash hypotheses for language testing, from the most general hypotheses to subject-specific hypotheses. In this research, I started with the generic hypotheses as a starting point and further explored the contextual areas. In this chapter, I will give a detailed account of the research philosophy, methods, data sources and reliability and validity of the research.

3.1 Research philosophy:

Creswell & Poth (2018) and Neuman (2014) emphasised that a researcher must have a defined research philosophy as it influences how the researcher undertakes the research from conceptualisation to completion. I believe in the subjective nature of reality, and to me, the reality is a social construct dependent upon complex factors such as culture, socio-economic conditions, politics, and reflection on life events. In a positivist approach, the reality is often seen as objective and measurable (Keat, 1981).

Interpretivism takes away the objective nature of reality and instils a socially constructed reality that is a continual process which varies with different circumstances (Berger & Luckman, 1967; Lincoln and Guba, 1985). In this research, I believe the research participants' views on the issue will not be objectively similar and will probably be more complicated than the current literature suggests. Hence, I have taken an interpretivist philosophical position in this research.

3.2 Research strategy:

Three common research strategies are Quantitative, Qualitative and Mixed Methods (Creswell & Poth, 2018). Each strategy has its strengths and weaknesses. Quantitative methodology is data-driven and focuses on objective reality, and quantitative methodology assumes human behaviours can be predicted and measured. On the contrary, qualitative research works with spoken words, pictures, videos etc., that carry subjective perspectives (Savin-Baden & Major, 2013). Qualitative research will suit well in research where the current knowledge of the phenomenon is less known, and further exploration is required to acknowledge the issues fully (Hammersley, 1992). In a mixed methods strategy, both qualitative and quantitative forms are combined to approach and answer the research questions (Lodico et al., 2006).

The research question is explorative in nature, and it does not intend to test any hypothesis but instead wants to explore and inform previously suggested hypothesises- 1, 2, 3, 4, 6, 11, 12,13, 15 of Alderson & Wall (1993) (see page 7, section 1.3). Sullivan & Sargeant (2011) argued that qualitative research is more suitable for understanding a research problem's 'how' and 'why'. Qualitative research is often used for 'hypothesis generating' rather than 'hypothesis testing'. The concept of 'backwash effect' within the higher education context is immature, as the literature review indicated a lack of empirical research on this matter. With such limitations in this research, the scope is to investigate the subjective nature of participant views to inform the reality of 'backwash' effect in the higher education context. Hence the narrowed-down segment of higher education covering only business degree programmes cannot give generalisability of the findings. It indicates the suitability of a qualitative strategy for this research.

3.3 Research participants:

The scope of this research was within a private, for-profit teaching university in the United Kingdom and Degree level business programmes only. The chosen university delivers programmes in blended learning and distance learning formats. All the undergraduate business programmes sit within one school. At the degree level, the university has a range of business programmes consisting of business management programmes as well as joint programmes with finance, HRM, data analytics, tourism, law, psychology, and computing. The university follows QAA and PSRB guidance on creating the programmes and assessments. Common assessment methods include a portfolio of evidence, report writing, case study, time-constrained examinations, presentations, and business simulations. Communications are top-down and follow a hierarchy; lecturers, including associates, are at the bottom in the rank, and PVC Academics is the highest rank at the university. Academic matters are generally communicated through the Head of School to the Programme Team Leaders, who further communicate the academic matters to the bottom layers of the school. Participants of this research belong to the school who are involved in delivering a range of business programmes, including BA (Hons) in Business, BA (Hons) in Business (Finance pathway), and BA (Hons) in Business (Tourism pathway). The participants also cover modules in the undergraduate joint programmes.

Purposive sampling was better suited for this research as I needed to have a representative sample based on the participant characteristics, guided by the scope of the research. For this research, it was needed to have participants who are delivering the programmes as well as those who are in the administrative decision-making positions both at the programme level and at the higher level. For fair representation, participants from all the layers were approached, and most of them were teaching staff members who would be able to give a detailed account of the impact of the assessment. I finalised a total of 12 participants for data collection, giving coverage for all the required levels of representation for this research. Inclusion and exclusion criteria were determined to assist the participation recruitment, which is further accounted for in the next section.

3.3.1 Participant recruitment

The researcher had access to the required participants through his professional network at the university. The researcher is the Deputy Head of the School, in charge of all the undergraduate and postgraduate business programmes within the chosen university so that the researcher can reach out to all target participants easily.

Inclusion criteria:

- All participants will be from the private university, academics/teaching staff, including associates who taught at this institution for at least one year.
- Research is also limited to degree-level business programmes only. Degree-level business programmes will include 1st, 2nd and 3rd year and 'Advanced entry to the Top up' stage.

Exclusion criteria:

- Teaching staff for foundation year programmes that are also part of a consolidated degree programme will be out of the defined scope.

Identification of the participants was via their professional appointments verified through the University website/university intranet and People services (where required). Upon meeting the inclusion criteria, all participants were personally invited via email. I forwarded all participants the welcome letter, information sheet and consent form (see Appendix -A, B and C). For those who agreed, I retained their signed consent forms. All the participant details are presented in Appendix E.

3.4 Research method:

For any research, a researcher must use methods suitable for the chosen research philosophy and methodology (Crotty, 1998). Guba and Lincoln (1994) argued interview is the most used qualitative data collection technique. An interview provides the researcher to have a conversation with a participant in an organised setting where participants are asked to express their feelings/beliefs/thoughts on the questions asked. Due to the exploratory nature of this research, it was essential to capture the issue of the 'backwash' effect of the high stake HE summative assessment from the participant's own perspective. I have used interviews as the only primary data collection method in this research, supporting my previously justified preferences toward research philosophy and strategy.

Kvale (1996) suggested possible structured, semi-structured and unstructured interview structures. In a structured interview, questions are predetermined from a set of research topics, and it stays that way throughout the interview without any scope of change. An unstructured interview offers flexibility as there are no predetermined questions; a researcher often starts with broad themes, but the subsequent questions result from the participant's response to the previous questions. Finally, a semi-structured interview provides the benefit of having a structured theme but with scope to branch out to other questions as seem necessary to the researcher.

Adams (2015) suggested a semi-structured interview for research that explores uncharted territories and in cases where participants are interested in the research outcome. In this research, due to the lack of pre-existing empirical research in the chosen area, I had to remain open to probing and expanding on answers to ensure the collected data provided a 'thick' depiction of the explored areas. Furthermore, all the participants were to be part of active practitioners in the field, so the outcome of this research was also of interest to them.

3.4.1 Designing and Conducting the Interviews:

Alderson & Wall's (1993) fifteen possible backwash hypothesises were at the centre of this research, followed by Baird et al.'s (2017). Based on those two key pieces of literature, I drafted thirteen questions for the semi-structured interview. All of the questions were mapped to inform Alderson & Wall's hypothesises (See page 20, Table 2). Furthermore, Wei (2017) categorised these fifteen hypothesises in four broader areas which was adapted in this research.

All the interview questions were categorised in one of these four broader areas:

- Teaching
- Learning
- Learning and Teaching
- Wider areas beyond learning and teaching

The fourth area was termed 'others' in Wei's original work, however the researcher felt 'Wider areas beyond learning and teaching' better represented that area and related hypothesises.

I anticipated in the earlier stage that some questions might require elaboration or further explanation. Rather than changing the questions altogether, I have opted to keep the questions as planned but rather explain the questions further when requested by the participants.

I followed general good practices in conducting the interviews (Cohen et al., 2018), such as empathy and creating rapport, active listening and careful probing throughout each interview session. During the interview, a set of questions was already structured, which were asked to all the candidates; further probing questions were brought forward when prompted by an answer of interest or concern. The following table summarises the mapping.

Table 2: Mapping of Interview Questions to Alderson & Wall (1993) Hypotheses

Area	Hypotheses	Interview question	Research question
Teaching	H1: A test will influence teaching	Q12. Based on the student outcome of an individual summative assessment, do you revise your teaching methods to ensure students are performing to their best of abilities in the summative assessment?	How does the 'backwash effect' impact teaching practices within higher education Business degree programmes?
	H3: A test will influence how teachers teach	Q4. Does the summative assessment affect HOW you teach, not just what you teach? -Can you give an example from your practice?	
	H4: A test will influence what teachers teach	Q2. While teaching a module, if you encounter a particular topic that is not relevant for the summative assessment, what do you do? -If you skip the content, follow-up question –Do you think skipping those seemingly less relevant topics for summative assessment impacts the student's overall learning and knowledge of the module?	
Learning	H2: A test will influence learning	Q5. Do you think summative assessment influences learning? If so, then can you elaborate with an example? Q3: Do you think summative assessment influences some students to show a lack of interest in some topics or lessons as they are less important or unimportant at all for the summative assessment?	
	H6: A test will influence how learners learn	Q1. Do you think student learning in your course should be bound within the areas covered in the summative assessment?	
Learning & Teaching	H11: A test will influence attitudes to the content and methods of teaching and learning.	Q6. The term 'Pedagogy' refers to the art and science of teaching. Pedagogy relates to the study and use of various teaching theories, including instructional models and how those influence students and their learning. To what extent do you believe there needs to be a synergy between the pedagogic model/teaching method and assessment methods (such as assignments, essays, presentations, time-constrained exams and portfolios etc.), especially those used for summative assessment?	
Wider areas beyond learning and teaching	H12: Tests that have important consequences will have washback	Q10. Within your practice area, to what extent is assessment design (type of assessment, weightage for each assessment segment etc.) influenced by the summative assessment results?	
	H13: Tests that do not have important consequences will have no washback	Q7. Apart from the summative assessment, what other assessment methods (such as formative, diagnostic etc.) do you actively use in delivering your module? Q8. Do you think the summative assessment is the only assessment that influences learning and teaching in contrast with formative assessments?	
	H15: Tests will have washback effects for some learners and some teachers, but not for others	Q9. To what extent do you find summative assessment to influence all students and teachers, or just some groups? -If some groups are indicated, follow-up question- Which groups in particular and are there any specific observations about those groups? Q11. Does your student's performance in the summative assessment impact your mental health in any possible way? Leading to an increased sense of accountability? Probe- Do you feel such pressure on student performance can lead to more enforced accountability? Q13. What is your current understanding of the positive or negative effects of summative assessment on learning and teaching within higher education?	

Interviews lasted between 30 minutes to 1 hour. The overall duration was within Adams's (2015) suggested maximum length for semi-structured interviews. The interview format was to start with permission for recording, an introduction, and a brief overview of the research followed by the interview questions. I asked questions in a sequence that covered the above categories, allowing me to probe the participants on topics of interest for this research. I used Microsoft Teams as a face-to-face interview was impossible due to the ongoing issues with post-COVID safety measures followed at an individual level.

3.4.2 Researcher's Positionality

Lack of rapport during an interview can lead to a 'less fluid' conversation between the researcher and the interviewee (Merton, 1972). As a member of the senior academic team, I line manage the line managers of most participants barring two participants. Prior professional ties with the participants helped me to create a great rapport with them at the onset of the interviews.

As I held a senior position among most participants, considerations of the power dynamics were needed. Thornborrow (2002) saw power as continually negotiated and constructed between the interviewer and the interviewee. Hoffman (2007) considered that the interviewer might deliberately take a less influential role. Despite that, the interviewer may appear to have greater power than the interviewee. In this research, I reduced the power balance by reminding participants of the focus of the research and how it is above and beyond my regular role at the university. By carefully managing the power dynamics, I ensured the collected data closely represented the participants' beliefs.

To ensure the quality of their participation was unbiased, I have clearly outlined the research agenda in the information sheet and in the pre-interview introduction. All the interview responses were going to be anonymised so their views would remain confidential. I was open to the possibility that participants would bring institutional issues that were not relevant to the research during the interview; however, I informed them before the interview that such queries need to be brought outside this recorded data collection meeting. Setting a clear ground rule enabled me to avoid any confrontational situation during the participation engagement. I faced no confrontational issues during the data collection process.

3.5 Ethics Approval

Ethical review is an integral part of any research as having ethical approval reaffirms the ethical standards for the researcher to follow (Hammersley, 2014). Ethical research protects participants from harm or distress and provides a clear account of how the data is handled and who has access to the collected interview data (Faden & Beauchamp, 1986). The University of Oxford requires all research to follow high standards in research and ethics. I sought a formal CUREC application and approval before commencing primary data collection to ensure this study followed all the ethical guidelines. I also followed BERA (2018) guidelines for educational research in the research execution. A full ethics approval was granted with favourable remarks on the research design. A copy of the CUREC approval is attached in Appendix D.

In this research, the main ethical considerations were to ensure all participants gave informed consent and voluntarily engaged in the interview process, additionally ensuring there was no undue direct influence from the researcher. All participants received an information sheet alongside the invitation email. Upon receiving the signed consent forms, the participants were officially recruited. Furthermore, establishing there was confidentiality and anonymity in the shared information. Any names brought by the interviewees were anonymised.

There was a potential risk of misinterpreting or interpreting data using a theory the participants were unfamiliar with. Such cases were averted by repeating questions or answers where required for clarity. Similarly, in the information sheet, I identified all the key literature for the participants to get familiar with. Given the nature of the data, I ensured all collected data was safely stored locally in encrypted storage and in the One Drive of the researcher's Oxford Nexus365 -two steps secured; password-protected account. The final consideration was to ensure potential conflict of interest was minimised or eliminated by taking a less powerful role in the interview process. Throughout this research, I respected participants' rights, dignity, and confidentiality of their engagement.

3.6 Qualitative data analysis:

Qualitative data analysis is a context-specific in-depth meaning-making of rich subjective data. Patton (2015) acknowledged that though qualitative data analysis converts data into findings, there is no simple formula for this process. Qualitative data analysis is often heavy on interpretation and often an ongoing process that can take place throughout the research. It generally involves coding, categorising and making sense of the studied phenomenon (Cohen et al., 2018; Creswell, 2014).

Creswell & Poth (2018) suggested a five stage 'Data Analysis Spiral' model which starts with managing and organising the data and completes with representing and visualising the data. Wellington (2015) proposed a similar seven-stage process which took a more broken-down approach. I have used the 'Data Analysis Spiral' model for this qualitative data analysis as it gave a comprehensive coverage of the data analysis stages envisioned for this research. All of these stages are discussed below:

Stage 1: Managing and organising the data

This research had only one primary data source: the interview transcripts. Microsoft Teams generated transcripts soon after I stopped the recording. Each transcript was further refined to resolve any issues with the quality of the auto transcription. Each interview took roughly two hours to transcribe. In this research, secondary data sources included books, peer-reviewed journals, reports, and other scholarly articles that dealt with the nuances of the backwash effect in assessment.
MS Teams generated auto transcripts. However, the quality of the transcripts was required to be further checked and validated to ensure it is fit for purpose. In cases where the transcription was not correct, further corrections were made by reviewing the original recording. The decision to record the interviews proved to be a useful judgement. Once cleaned and processed all transcripts were moved to nVivo12 for further processing. I have also linked all recordings with relevant interviews so that if required the original clips can be revisited.

Stage 2: Reading and memoing emergent ideas

All data sources in this research were textual. In this stage, reflecting on the research problem, research question and insights from the literature review provided the key prompts. In the interview stage, all questions could be categorised into several evolving areas as discussed in section 3.4.1. Additionally, during the initial interviews, some patterns started to emerge which were also noted and compared with the rest of the codes as they emerged throughout the other interviews. Both the deductive and inductive reasons were important to consider. It is discussed further in the next section.

Deductive and induction approach to coding in this research:

In a deductive analysis, a theory is applied to a data set to test the theory. In this 'top down' approach of analysis, codes are often predetermined and generated from the literature or the theories that the researcher is evaluating. An inductive analysis applies an emergent strategy where the researcher processes collected primary data and looks for possible codes. It is a 'bottom-up' approach where codes are generated through participants' words and by constant comparison of codes.

Thomas (2006) argued qualitative data analysis is often done through an inductive process. However, I believe, it is quite important to consider the importance of both approaches for qualitative analysis. Having a deductive approach gives structure to the data collection and an inductive approach allows to ensure no emerging codes/themes are missed out. I also agree with Dey (1993) that all themes are not possible to be foreseen unless data analysis has taken place. To sum up, the data analysis of this research is drawn upon both deductive and inductive analysis to provide comprehensive and rigorous qualitative research.

Stage 3: Describing and classifying codes into themes

Codes are in essence 'units of meaning'. It was imperative to identify codes that could be used to classify the collected data. Böhm (2004) suggested open, axial and selective coding in qualitative research where thematic analysis is applied.

Open Coding:
Where similar words were used including variants and synonyms those were required to be grouped to reduce the number of codes. Ryan & Bernard's (2003) suggested using 'Word Lists' and 'Key Words in Context (KWIC'' techniques to process qualitative data. These strategies were utilised by having a generic word list generated using Nvivo's 'Autocode' feature. This gave a list of the most recurring words. At this stage, similar words were merged to the closest match codes.

Axial Coding:
After that KWIC technique was applied where words that were most frequent were identified within the context in which they appeared. This is when axial codes started to merge into categories.

To ensure the software generated codes were meaningful, following three checks on the quality of the information were applied as suggested by Creswell & Poth (2018):

a. Expected information
b. Surprising information from the data
c. Unusual information/ potentially interesting information

Selective Coding:
I have also manually investigated words that raised my interest, this was the open coding process. Common words were combined under one code until no further codes can be generated from the data.

It was found that the analysis was much better suitable at a concept or idea level rather than at the level of words (Appendix I, K). The interview questions and their categorisation contributed to getting the codes suitably organised for further analysis.

The next stage was to create a codebook. In qualitative research, a codebook is a collection of codes to be used for the data analysis. Open, axial and selective coding strategies were used to create the codes.
At this stage of the research, codes were categorised to create themes. Themes in this research captured important pattern or relation in the collected interview data sets. Ryan & Bernard (2003) stressed the importance of discovering themes for research analysis that deal with social aspects. It was an iterative process involving continuous moving back and forth through the collected data. With emerging themes, inter relation between themes started to emerge. Themes generated from the data is presented in section 4.1.

Stage 4: Developing and assessing interpretations

Interpretation of qualitative data requires creative and careful considerations to ensure decisions made on themes and codes are meaningful to the research (Patton, 2015). Interpretation in this research meant going beyond the level of themes to larger abstraction that represented the data. Interpretations made here are social constructs and are subjective to findings from the data and the beliefs of the researcher. The latter is linked to the interpretation of data in relation to the wider literature reviewed in Chapter 2.

Stage 5: Representing and visualising the data

This is the final phase where findings from the data is presented to readers. Creswell & Poth (2018) discussed many forms of representation including visual images, matrix, hierarchical tree diagrams etc. as some common forms of data visualisation.

In the final phase of the 'Data Analysis Spiral' a researcher represents the data in a systematic manner using formats such as text (direct quotes from participants or paraphrased quotes), tabular, or figure. Miles et al. (2014) argued the importance of adding accompanying commentary to go with any of the forms in which data is represented.

3.7 Reliability and validity of the research

Lincoln & Guba (1985) argued that credibility, confirmability, dependability, and transferability are essential criteria for quality. These quality measures apply to this research.

Credibility contributes to the truth value of qualitative research as it is comparable to internal validity in quantitative research and measures whether the findings are accurate. Credibility is also rooted in the researcher and the selected methods (Bryman, 2008). In this research, a well-established research method- interview, was used for data collection.

Confirmability ensures the collected data in the research and the research findings represent the participants and not just the researcher's imagination (Shenton, 2004; Guba & Lincoln, 1994). In this research, I ensured all collected data were recorded and stored, and I checked interview transcripts against the recorded version to establish accuracy. The research findings are representative of the collected data.

Dependability is the reliability aspect of the research. All the stages of the study were meticulously detailed. As Shenton (2004) suggested, the research work followed a set systematic approach which is reported in a section detailing the research design and implementation, including how data was obtained and a reflection of the effectiveness of the research.

Transferability is the equivalent of external validity (Lincoln and Guba, 1985). 'Thick descriptions' of qualitative research provide a better opportunity for transferability of findings (Shenton, 2004). Given the nature of this research, generalisation of the findings may not be directly transferable to other settings, such as high-stake assessments for Post Graduate programmes, as those programmes have different entry criteria, a different set of assessments and assessment interventions. However, the generic findings should be able to inform similar research.

CHAPTER 4: RESEARCH FINDINGS

In this chapter findings through the data analysis are presented.

4.1 Research themes

From the data analysis several themes were identified, process of that is illustrated in Appendix F, H and I. The themes are:

1. Setting the boundary of Learning: Curriculum or Summative assessment
2. Covering topics that are not in the summative assessment
3. Whether students are influenced to skip curriculum because of summative assessment perimeter
4. Summative assessment influencing teaching behaviour
5. Synergy between pedagogic approaches and summative assessment methods
6. Other types of assessment and comparative importance between different types of assessment
7. Summative results impacting assessment design
8. Feedback and student motivation

Each of these themes are related to backwash effect and its various compositions. All of the themes are further elaborated in the following sections.

Setting the boundary of Learning: Curriculum or Summative assessment

One of the pre-determined themes was to explore how academic staff members set their boundaries of teaching and learning. Two clear emergent thoughts were presented by the interviewees in relation to curriculum and summative assessment. The curriculum was seen as the currency by some academics, and there were also views against it.

> *"When I was teaching strategic management, they (summative assessment) were still talking about a Ryanair case study published in 2015. While the impact of the pandemic was in there, in certain scenarios, I felt that what we were teaching wasn't reflected in the summative assessment. And, you know, students may struggle to relate to those current topics in their assignments that are based on old case studies." (Pa11)*

Pa11, for instance, gave an example that a case study on the profitability of Ryan Air was given as the summative assessment at a time when the whole aviation industry was affected by the pandemic. At the university, summative assessments are often set at the beginning of the academic year, and there is limited scope to change assessments midway. Similar concerns were raised by Pa9 as well, who felt contemporary issues might not be updated in the curriculum unless the programme goes through a periodic review. So, there is a danger of not going beyond the curriculum as students may miss out on essential academic or employability skills. Contrasting to this, Pa10 who teaches a finance module experience active role of PSRB- ACCA to ensure all of the required content is delivered otherwise they will withdraw their accreditation.

Pa5, felt too much attention is given to summative assessment, and learning should be more comprehensive than covering topics important for the assessment. Pa8 had a similar view but also expressed caution against the danger of overreliance on assessment to cover the curriculum.

Pa4 admitted that he did fall for the “trap of” being assessment-focused and losing sight of the curriculum. He argued that such happens when the term time is short, especially in six-week blocks.

Pa6 strongly felt the curriculum should be the boundary because it is approved by the gatekeeps of quality and higher education standards. Keeping an eye on the progress covering the curriculum and completing assessment tasks were of importance to Pa6.

Narrowing of the curriculum

In the interview, probes were used to check academics' intentions when they deal with topics that are not in the summative assessment. All participants shared their strategy and rationale behind covering those topics. As all participants were in favour of covering all the contents, the case for negative backwash effect was not found.

There were three emerging approaches:

1. Coverage depends on the timing of the content

Pa1 thought that based on when the less important content appears in the lesson may determine to what extent that can be covered. If less important content arrives at the beginning of a module, then there is a greater possibility that it will be covered at some length. The reason for that, as Pa1 elaborated, is that even though the content is not in the summative assessment, it can be essential to understand other content. Pa6 also had a similar line of thought and gave an example where ‘Porter’s Five Forces analysis’ may not be in the summative assessment but skipping this essential analytical tool will affect students' overall understanding of the ‘macro environmental’ analysis of a business.

2. Bridge building content

All participants thought all content within the curriculum must have some relevance; otherwise, it would not be in the curriculum. Pa7 put the case as some ‘bridge building’ topics may seem irrelevant, but those topics are essential for forthcoming content. Some participants were also interested in exploring how to connect those seemingly unconnected topics to the core content- classroom activity, and case studies were considered by Pa8, Pa7, Pa6 and Pa2.

3. Content relevance in comparison to student demographics

When courses are delivered to a group who are not domiciled in the UK, there are cases where a taught content may be of less interest to students, especially when they are not covered by the summative assessment. In that case, further contextualisation may become necessary. P10 gave examples of Berlin students who are likely to find lessons covering British taxation irrelevant, as German government tax policies differ from the UK policies. However, as students are on a UK university programme, the summative assessment will likely cover British taxation policies only, even though there may be scope to cover other approaches to taxation in the curriculum. An academic may want to bring some examples of the German taxation system to contrast with the British taxation system to make the teaching more relevant for the student. On the contrary, choosing to omit other taxation systems completely will result in a negative backwash as the curriculum is not covered to a fuller extent. In this case, the summative assessment had no apparent backwash effect on the teaching staff.

Whether students are influenced to skip curriculum because of summative assessment perimeter

This theme can be broken into phases. Within the context of the chosen university, all students receive their summative assessment simultaneously when they gain access to the module and its content online and offline.

Questions were asked to probe to what extent students' interest in learning is skewed by the summative assessment; all participants suggested a negative skew. Pa1 explained specific cases where he normally gets WP[5] students who may have been away from studies for a while; they need time to settle down in a new module and then move slowly to the summative assessment. There are also 'regular students' who understand the process very well and are likely to jump straight to the main points. Knowing what's in the final submission means the students are likely to look for the clues throughout the sessions and skip content if they identify the topic as unimportant.

Pa6 had a contrasting view and from his experience he found mature students often show an 'I already know attitude', so they start self-filtering the content.

> *"Yeah, I think it can be older (mature) students because they can sometimes exhibit the 'I know it already attitude. So, I don't need to listen'. But there again, older students are, in my experience, more likely to have already researched it, whatever the subject is and come across it and spotted the relevance or irrelevance if it's not in the final assessment." (Pa6)*

Many of these students end up creating a mental map that certain lessons are very important and others are not so. Pa6 also acknowledged the same group of students are likely to have spent more time in researching and pre-reading the topics in that filtering process.

Pa4 gave examples from his teaching experience that once students are aware of the summative assessment and they have worked out which topics are of greater importance, they may push the teacher to cover the most important topics first. With most important topics covered in the first few weeks, students are likely to show negligence in attending later lessons covering 'unimportant topics' for the summative assignment. Resequencing of content run order is a common case for Pa4 to give greater coverage to the summative assessment content. The participants felt, the risk of students missing out on learning fuller curriculum increases when summative assessment greatly influences the module's original delivery plan.

Pa3 gave insights into a wider issue where programmes are becoming industry driven and if content does not reflect the industry standards or likewise not found in the assessment then the need to learn those content becomes a secondary thought to the students. Pa5 found that for 'Time constrained exams' used for accounting modules students are at rush to prepare for important topics only. Pa8 also encountered instances where the student has provided a full draft of the final submission in the very beginning of the term. Some students are like that when it comes to summative assessment, expressed Pa8.

[5] Widening participation

Pa10 gave an interesting example where students were given a case study on Amazon and its environmental footprints as a summative assessment. One of the students of the cohort who performed well in that case study, eventually succeeded in gaining employment in Amazon. That student felt the case study prepared him for the interview and had a greater impact on his success in securing the role. Pa10 believed summative assessment certainly influences learning but also has an enormous potential to prepare students for employment.

Summative assessment influencing teaching behaviour

To assess how a summative assessment can influence teachers, probing questions were used in the interview and the findings indicates often teachers are influenced by the summative assessment. A range of examples were found such as – Pa3 mentioned in accounting and finance, students need to clear an exam which is also mapped with ACCA. ACCA as a professional body is very rigid to the way students are assessed. This improves reliability of the assessment but to ensure students are through, teachers end up turning their focus towards the final assessment. There is an element of coaching to pass the examination. Also, this was reiterated by Pa5, Pa7 and Pa10 who also teaches Finance modules. This indicates positive backwash effect as teachers are ensuring students are having good examination skills alongside knowledge of the curriculum. However, Pa10 argued for teachers who lack experience or newly qualified may struggle to cover all the content systematically and end up having a tunnelled vision to cover main topics only. In such a scenario, negative backwash effect can emerge as the whole curriculum is not covered as intended.

Similar issues were not cited by other modules that sits within CMI accredited programmes. To further explore the reasons, additional probing questions were used and in was clear where summative assessment is in an examination format, such coaching approach emerges and takes a pattern of negative backwash effect issues generally found with language tests.

There were also cases where the participants expressed the narrowed scope of summative assessment and the issue with following just the areas covered by that assessment. Pa9 had shared his views why he does not let summative assessment dictate his teaching.

> *“I'm not sticking only to the summative stuff (assessment) because usually sometimes they're very restrictive.” (Pa9)*

Pa9 asserted that teachers must understand the interdependency of topics. Some topics may not be in the summative assessment but is important for the students’ knowledge of the discipline. Similar was reiterated by Pa6. An example of this is, in a level 4 'Understanding Business Environment' module, where different types of business in the UK and key employment legislations are covered. It is unlikely that a summative assessment will comprehensively cover both topics in the summative assessment due to both topics' lengths. However, for a business management student, it is imperative to have a fuller understanding of both topics. Skipping this topic would mean the students will struggle when they move to the finance module, where various financing options for different businesses are covered. Similarly, the legislation part of the topic is essential for later HR-related modules. One of the critical issues with skipping topics will be that it will affect the progressive growth of self-regulated learning that is expected at a later level of the programme.

Pa8 gave examples from his teaching engagements where he needs to embed a lot of scaffolding in Level 4 assessments and then as the level of study goes up, the expectation is that students will become more autonomous learner and an active learning model takes over the scaffolded delivery. Eventually for assessments set at Level 4, it will be important to assist learners to gain the basics –

such as creating an engaging presentation, drafting a reflective essay. Such as- at Level 4, for a simple business topic like Belbin's team roles or Tuckman's Teamwork theory, if participants are asked to create a narrated group presentation, they might require assistance with creating such a presentation from scratch. They may not know how to record audio within the MS PowerPoint, or they may struggle to create formatted slides as they share the slides among the group members. However, the same type of assessments set at Level 6 will not require such assistance as the students are expected to be fully aware of various types of assessments including narrated group presentation.

Synergy between pedagogic approaches and summative assessment methods

In this research, it was fundamental to explore to what extent pedagogic approaches has role to play in the summative assessment. Influence of learning theories and pedagogies in the summative assessment context were explored. Pa11 felt the theoretical foundation is important in the early stage of teaching career, however, the practical classroom experience do not always resemble the theoretical presumptions. Pa11 asserted that topics like 'learning styles' are taught in the teacher training course but, learners cannot be just categorised in one learning style. Despite literature on learning styles proving such categorisation is a myth (Kirschner, 2017), there are books in circulation and on the current reading list of teacher training programmes that advocate learning styles. Pa11 brought that issue where an outdated perception is still taught to the trainee teachers. For a newly qualified lecturer to identify such learning styles does not exist is a task.

> *"You can read too much into the various pedagogic models etc., teaching, learning, assessment models, whatever you want to call them. I think it does depend on your personal psyche as a lecturer, and that's a good thing about lecturing … In particular, I think sometimes the pedagogies have got to suit. Because I've done my level 7 qualification in digital pedagogies and practices, quite honestly, many of the things I learned wouldn't suit my delivery style… I can put some models into practice and note where it's been successful. So, I think you need to be flexible and be ready that whatever you read, you may soon realise, it might not actually work on the ground level." (Pa6)*

Later, Pa6 further elaborated that, for him pedagogies must suit the level taught. From his recent teaching qualification, he discovered there are several new pedagogic approaches that may be useful for postgraduate teaching. He was also mindful that those approaches may not be suitable at a lower level where the preference will be to build up academic skills. He argued for 'flexibility' rather than having a procedural 'rigidity'.

Pa5 brought her experience of doing the same professional development programme PGCert in Digital Pedagogies and shared how she learn new techniques for assessment such as portfolio submission, podcasts but then in the practice she felt the assessment lacks imagination as most summative assessment she encounters are report writing or a PowerPoint presentation.

Pa12 found from his experience that as an academic certain level of fluidity is required when adapting a teaching strategy. Regardless of personal preference as a seasoned academic one must be able to change the teaching technique on the go so that it meets the needs of the classroom. Pa12 was further conscious of the differences online/hybrid delivery brought in comparison to campus-based delivery. He felt a sudden move to online delivery followed by hybrid delivery required some level of IT proficiency to adapt to digital learning pedagogies. There were cases where both the academic and students struggled in that learning environment. That had an impact on the summative assessment too as students were expected to almost replicate the assessment same way as they would do in

campus. An example of group presentation was given where the group presentation was replaced by narrated PowerPoint presentation and students struggled to understand and imitate all the steps. Pa12 felt summative assessment needs to adapt to the teaching environment and not just the teaching strategy.
The term pedagogy used over andragogy was questioned by some participants.

> *"…So is the pedagogy, the right word to say in terms of the adult learners, it should be Andragogy, should not it?" (Pa3)*

Not only Pa3 was curious why 'andragogy' was not used in the question but also further explored the prospects of digital pedagogies and how at present there is a gap in what teachers may have learnt 10-15 years back on digital learning and how that landscape has transformed over the last couple of years during the pandemic. The prospect of having a sudden digital transformation was happening on a large scale within the institute, despite being a digital first university, the changes meant teaching methods would evolve and assessments needed to be evolved too. Pa6 gave an example where in class group activities transformed into 'breakout rooms' and the summative assessment of 'group presentation' transformed into a 'narrated group presentation'. There was definitely a demand for synergy, but it was not always easy as any changes to assessment need to feed through the 'Quality and Registry' team and the process is time consuming. Pa6 also illustrated that, what impact those pandemic time changes in assessment would bring was unseen to the teachers and administrators. Hence, the backwash effect was mostly that teachers may have felt compelled to safeguard the students by teaching them how to attempt those modified assessments.

Pa10 thought the teaching approaches and assessments may again vary based on the discipline as certain discipline demands certain level of engagement with the course content. For instance, in health and social care, the assessment tends to be more patient/client focused, whereas in finance the assessment is more business focused and in business management assessment can be mixed bag. The same reflective log as an assessment tool will be much more comprehensive in health and social care discipline than the business discipline.

Other forms of assessment often fail to generate enthusiasm among learners.

In the interview, other recurring theme was that different types of formative assessments are used in the delivery but those are often less popular among the students and many struggles to submit a formative assessment on time and to act upon it. Various methods of formative assessment were found to be used and there is a sense of reflective practice around it. Pa5 elaborated on the tools she uses for formative assessment - Peer review and gamification (Kahoot!). Pa5 realised that Kahoot worked better when the classroom size was bigger, as that would make winning more competitive.

> *"…One thing that I did was to do Kahoot! Quizzes for formative, but I realised it works better in a bigger class size because in the Kahoot environment, winning matters. In one of my classes with 16 students, it was a success, but for another class for the same module with a low number of students, it did not work as expected." (Pa5)*

Pa3 raised concerns about the students' tunnelled vision to skip through all the other support activities and just focus on the summative assessment. She illustrated how students rarely engage in discussion forums. Those forums are meant to support learning but are often not well utilised by the students.

P3 further discussed how a 'patchwork assessment' can be used within her practice area to improve student engagement further.

> *".....for example, in patchwork (assessment) on a professional development module will have the diagnostics, the self-auditing, the personal development (plan), looking at how you fit into practice and the sector in the future of that type of thing. So, it's kind of like (an ongoing) personal development planning. Students don't have to do it all at one go but can slowly build up the assessment evidence. I prefer this type of assessment" (Pa3)*

Pa3 felt such activities could help students understand the value of those informative assessment opportunities. Pa3 had also observed how motivation has a role to play; some students are enrolled in an academic programme to build their network. They are self-driven to engage in every opportunity but also many of the students are just purely driven by the summative assessment because the end grade will determine whether they progress further or not in their academic studies. This issue was found in another participant's response; Pa9 was concerned the way students interpret formative assessment. He found students are submitting the whole assignment well before the taught part of the module is completed. In his words:

> *"For formative assessment, I have one very big concern and I'm trying to change the mode of that (assessment). You know, the usual route is that they will send you the whole assignment by week four, which I find it very strange. I'm stressing to them what you are asking me effectively is to pre-mark the assignment. This is not (formative) feedback. What they need to do is send me parts of the assignment with specific questions attached to them and indicate to me what feedback they require to develop their work further." (Pa9)*

To understand the issues further when probed Pa9 believed students are so consumed by the summative assessment that they rush through the formative stages and try to get a confirmation that what they are producing will at least give them a pass grade. Pa8 gave examples of how learning can be expanded through formative assessment. Example given by him was that, in his tourism module, 'Butler's Tourism Area Life Cycle' model is very similar to the 'Product Life Cycle' model. In the formative stage students can make connections to those models and learn more about both whereas the later may not be in the summative assessment.

Pa5 elaborated on the tools she uses for formative assessment - Peer review and gamification (Kahoot!). Pa5 realised that Kahoot worked better when the classroom size was bigger, as that would make winning more competitive.

> *"…One thing that I did was to do Kahoot! Quizzes for formative, but I realised it works better in a bigger class size because in the Kahoot environment, winning matters. In one of my classes with 16 students, it was a success, but for another class for the same module with a low number of students, it did not work as expected." (Pa5)*

Pa5 also brought the issue of culture, as the participant grew up in a country where written examination was the only method of summative assessment so a lot of preparatory examinations would take place to prepare the students for the final examination. She felt the focus is less on such examinations in the UK and rather it is more towards essay and report writing. So, the tools like Kahoot

that she likes to use in the classroom will be of less use when the summative assessment is of a completely different type.

> *"…. personally, as a student I grew up in … (omitted for confidentiality) where all we had was written examinations…. I felt comfortable writing exams. I still feel comfortable writing a written exam in comparison to writing a report…. I think the cultural background of the student and what s/he has experienced up until this degree level study has a role to play." (Pa5)*

The common issues found in this theme is students are too eager to complete the summative assessment, so they forego the opportunities of truly engaging in formative assessment. There is also potential complexity around setting expectations from the formative assessment. Lack of such results in students interpreting summative assessment as the more important assessment in comparison to other forms of assessment that are used.

Summative assessment results impacting assessment design

Whether summative assessment results have an impact on the way these assessments are renewed or revised were also questioned. Intention behind this was to check to extent the outcome impacts delivery and to see what extent backwash effect of summative assessment impacts the assessment design.

Pa5 had experienced this in her various roles within the university. When she is just lecturing on a module, she has limited scope to bring any change but when she is module leader, she gets to have an official opportunity to suggest changes to assessment. She gave examples of a module where withdrawal rates were too high and eventually the academic team decided perhaps the students are not fully understand the summative assessment.

> *"I rarely have seen lecturers being able to radically make changes to assignment briefs and I think that's down to fear really, that if I make substantial changes and then if the initially results of student outcomes is bad, then you know I'll be told off." (Pa9)*

Pa9 was cautious that any radical change to the summative assessment may not work and if the student outcome is poor then the module leader will be held accountable for the negative impact.

Pa3 discussed a case where the programme team leader changed a module's assessment to include innovative approaches like podcasts and recorded narrated PowerPoint presentations compared to a traditional report writing assessment. However, the next cohort of students found the new assessment problematic and preferred a standard case study or report writing. Pa3 felt students should also be part of any assessment change consultation as students are essential stakeholders in inclusive assessment practices. Commenting on the recent drives to make assessments more approachable, innovative and inclusive, she reflected it is just hard to achieve. A narrated group presentation can be daunting for people with social anxiety. During the pandemic, there were several such cases where students struggled to follow the modified summative assessment. Pa3 felt that in challenging times, a general lack of adaptability to summative assessments could cause anxiety among students, resulting in tunnel vision to study the bare minimum content to get a pass.

Feedback and student motivation

Pa12 discussed how students are increasingly becoming demanding to understand the assessment outcome and breakdown of the grading. He felt students can be benefitted if they are guided sufficiently by using summative assessment feedback as a tool. There are cases, where feedback lacks a detailed account of achievements, including shortcomings. Pa3 shared examples where students officially complained because of lack of details in the summative feedback. Students use the feedback in variety of ways and one of those is for self-improvement. From an academic perspective Pa3 felt, learning referencing skills, writing skills, ICT skills are often gradual. At the undergraduate level, students may need several attempts before acing that skill. From a teacher's end it becomes imperative to give those additional improvement feedback.

Another interesting finding was that students are continually becoming aware the final summative submission is not the very final opportunity to pass the module. Students get multiple resit opportunities up to 4 attempts in the chosen university with also a scope for an exceptional 5th attempt. It works well for some students as they can then prepare for the examination. However, it impacts on the 'high stake' nature of the summative assessment. With multiple resubmission options, students become less stressed with a negative outcome. Intrinsic motivation does strive the students to succeed but expected level of progression varies from student to student. Pa3 spoke about it and gave examples where the expectation from feedback is to give indication of how the student could have achieved a higher score, but then some students discard that feedback as they were not attempting a higher score anyway. Pa3 suggested individualised feedback that concentrates and celebrates student's achievement in the high-stake summative assessment rather than creating a report of things that were not done or a list of things that could be done in a better way.

CHAPTER 5: DISCUSSION

This chapter relates the research findings to the previous literature as discussed in Chapter 2. This chapter also considers the limitations, future research, and contribution to knowledge.

5.1 Significance of the findings in relation to the existing research - Alderson & Wall (1993) Hypotheses

The research conducted aimed to inform hypothesises 1, 2, 3, 4, 6, 11, 12, 13, 15 of Alderson & Wall (1993). All of these hypothesises were categorised in four broader areas (See Appendix G). Such categorisation was trialled by Wei (2017) and was reported to be adequate.

Area	Hypotheses	Covered in this research
Teaching	H1: A test will influence teaching	Yes
	H3. A test will influence how teachers teach	Yes
	H4: A test will influence what teachers teach	Yes
Learning	H2: A test will influence learning	Yes
	H6: A test will influence how leamers leam	Yes
Learning & Teaching	H11: A test will influence the attitudes to the content, method, etc. of learning and teaching	Yes
Wider areas beyond learning and teaching	H12: Tests that have important consequences will have washback	Yes
	H13: Tests that do not have important consequences will have no washback	Yes
	H15: Tests will have washback effects for some learners and some teachers, but not for others	Yes

The collected data was sufficient to inform hypotheses 1, 2, 3, 4, 6, 12 and 13. Hypotheses 11 and 15 require additional data; thus, this research could not fully inform those two hypotheses.

5.1.1 Summative assessment in the context of the wider areas beyond learning and teaching

Alderson & Wall's (1993) hypotheses H12, H13 and H15 were considered. The H12 hypothesis stated that tests that have important consequences would have washback. The collected data indicates students and teachers are both aware of the importance, and additional effort or support is given to ensure the student's summative assessment outcomes are favourable. Messick (1989b) investigated the element of 'consequences', and the researcher found that the academics also acknowledged the social aspects. A lot of efforts are placed in the assessment designing stage to ensure the summative assessment has 'construct validity'; however, academics agree that the social consequences are often undermined at the designed stage. Teachers brought issues like social anxiety that may directly affect a student's performance in a group presentation type summative assessment. Still, in the paperwork, the task may appear to be a good assessment, assessing subject knowledge and soft skills such as collaborative work. Teachers argued that such consequences are not possible to judge unless encountered and high-stake summative assessments at times lack 'reasonable adjustment' arrangements to ensure the reliability of the assessments. As discussed in section 2.4, academics acknowledge the 'inferences' of assessment and knowledge of such inferences can lead to positive reinforcements to ensure better student performance in summative assessment. This research

suggests that summative assessments with high-stake results will have backwash and are likely to be positive within the higher education undergraduate setting.

The next hypothesis, H13, stated that tests that do not have important consequences will have no washback. A range of other types of assessment, such as diagnostic assessment, initial assessment, and assessment for learning (AfL), are considered low-stake assessments. It was found that students clearly showed a reduced level of engagement in the other forms of assessment. Since other assessments like AfL are seen as less important by students and teachers, teachers were not too keen on pushing students for formative submissions. Where students were intrinsically motivated to do well in the summative assessment, the opposite level of engagement was widely claimed by academics. There were convincing reasons for that; the university where this research is situated run six weeks of study blocks so formative assessment come at an early stage on the third week. The participants agreed that such a short time does not allow students to fully grasp the contents to write a formative piece of work. There were also concerns where students write up a whole piece of work aimed for summative submission- such cases would generate alarms of potential ghost-writing. In principle, teachers showed less concern for other forms of low stake assessment, and largely they use it to guide students to improve writing for the summative stage. There was no push to get a good-quality submission at the formative stage. Hence, intrinsic motivation, as discussed in section 2.1.2 was seen as more prominent for summative assessment; there was a lack of motivation for low-stake assessments indicating no backwash effect.

The final Alderson & Wall's (1993) hypothesis that this research intended to contribute was H15 stated that tests will have washback effects for some learners and some teachers but not for others. The scope of this hypothesis was found to be too broad, and the data collected in this research was not fully able to inform the hypothesis. However, the collected data could indicate several considerations. Some of the participants found the assessment to be too restrictive and lacking innovative approaches. Lack of imagination, administrative support and resources were the main reasons cited for the lack of innovative assessment approaches. Johnston's (2002) call for 'innovative assessment' (see section 2.4) was reiterated by most of the participants who felt there exists a disconnect between placing students at the centre of delivery, allowing them to harness graduate attributes and then the assessment lacking in allowing students to showcase those soft-skills or graduate attributes. Innovative assessments may reduce the chances of having a negative backwash effect by effectively covering all the curriculum materials and the program's intended outcomes. This research indicates, to achieve that 'assessment literacy' needs to be further improved among the academics who are designing and delivering the programmes (see section 2.3). It will also enable a highly structured summative assessment to move to more student-centric flexible options like student-teacher negotiated assessment.

Though H15 did not directly mention consequences, however, high stake summative assessment also has potential consequences in the form of anxiety, mental health concerns (and an increased sense of accountability. It is partially covered in section 5.1.4. In addition, findings indicated that seasoned academics are prepared to embrace the outcome of summative assessments as they are likely to predict the outcome at an early stage. Participants indicated that there is potentially some impact, especially for those who are new in the teaching profession delivering a higher education programme, as they will have to demonstrate their credibility and suitability for teaching the module. In students' cases, the consequences will also include lack of motivation and dropping out of the programme, which were out of scope for this research. From the data, there were no concerns raised by academics on mental health issues because of summative assessment outcomes. Based on the available data, there was insufficient evidence to convincingly suggest whether the backwash effect exists for some students and some teachers.

5.1.2 Summative assessment in the context of learning and teaching

Alderson & Wall's (1993) hypothesis 11 was that 'A test will influence attitudes to the content and methods of teaching and learning.' The researcher explored personal beliefs on learning theories and pedagogic aspects related to content delivery in this research. It was anticipated that participants might prefer specific learning theories and methods of instruction and there will be a possibility of having a mismatch of delivering the curriculum in the expected way. However, the data suggests the choices were more fluid (i.e. having no fixed positioning and being able to take new forms based on external stimuli), and teachers almost intuitively switch between various modes of delivery to ensure all student needs are 'satisfactorily' met. 'Satisfactorily' here does not have an objective definition. Teachers who mentioned such goal setting were found to make their case of what they think will be 'satisfactory' for their students. The findings on the fluidity are in line with Sfard's (1998) suggestions to see learning in the middle ground of 'acquisition' and 'participation'. What it means for backwash is whether a set of learning theories guides teachers, or the teachers have a tunnelled vision of seeing learning from one angle only. It is important to consider that assessment requirements within a degree programme may involve both acquisition and participation. It could potentially contribute to backwash by missing curriculum elements because of a particular teaching technique such as missing out on building collaborative skills because of a teacher-centric delivery model. As not much thought is given to whether the teaching technique is in line with their teaching philosophy or any specific teaching or learning theory, the academics did not show possible summative assessment-related backwash on the methods and techniques they use to teach.

In this research, the researcher expected to be able to understand the way these three learning theories – 'Reflective learning' (Kolb, 1984), 'Transitional learning' (Wildemeersch & Stroobants, 2009) and 'Transformational learning' (Mezirow, 1997) could play a role in students learning journey (section 2.1.2). While some perspectives on this were collected in the participant interviews, as no data was collected from student groups, a conclusive remark could not be made on whether those three learning theories can be merged to make a better sense of the student's learning journey in a HE undergraduate programme. So, whether a test can influence attitude toward the content and methods of teaching and learning can be split into two halves- the teaching aspect shows no backwash or neutral state, and the learning aspect is out of scope as not enough data was available.

5.1.3 Summative assessment and teaching

The first area of concentration was 'teaching', only to explore whether the summative assessment can influence teaching. Collected data indicated a positive backwash effect in this aspect as teachers become mindful of the elements of assessment upon knowing the methods of assessment and guide students through curriculum alignment. Key determinant of whether this is positive or a negative backwash- i.e. narrowing of curriculum is dependent upon whether less important topics are ignored in delivery. The researcher found that HE teachers did not omit any content of the curriculum even though they are aware some of the content is of less significance in the summative assessment. Participants felt those content might become necessary later in their study, and some participants illustrated such with examples from their practice. Additionally, there were cases where participants felt some curricula could 'age fast', thus lacking innovation in the assessment required from that programme. It was in line with the concerns raised by Enriquez (2020) and Elton & Brenda (2002) who thought at times academics are forced to follow programmes that lacked currency (see section 2.1.1). However, it was encouraging to find from the data that, if need be, teachers do tend to cover additional topics by aligning emerging topics to existing curricula as supplementary knowledge to bring currency to the taught modules. It was a common phenomenon for all the participants.

There were discussions around teaching approaches, and the participants indicated they amended their choice of teaching methods as per the demand of the module. The teachers claimed they frequently switch between different approaches to accommodate that need. The participants recognised the nature of highly structured summative assessment, but it did not occur to them as a significant barrier to delivering student-centric teaching. Some participants also felt this is an attribute of 'seasoned academics' and being able to switch from own philosophical beliefs about teaching to another suitable one, is a key competency for a modern classroom. Some participants acknowledged the slightly contrasting non-innovative assessment methods, but those were isolated cases and primarily due to PSRB, such as ACCA requirements. Academics cannot change the assessment methods of those accounting and finance modules as otherwise, the PSRB will invalidate the assessment outcome with further consequences of losing PSRB accreditation. The findings correlate to the literature review (see section 2.1.3) that student-centric teaching is not always harmonious with the assessment methods chosen by the PSRBs. In this case, the teacher had to almost follow a behaviourist style of delivery for the accounting module and create mock MCQ tests as formative assessment.

The findings indicate summative assessment within a degree-level business programme influences teaching in general, including what the teachers teach and how the teachers teach. Participating academics acknowledged the need for more scaffolded delivery in the early undergraduate stage, especially in the first year (see section 2.1.3). They felt a good number of students remain weak to embrace the challenges of a full-time degree-level programme. Though they learn the most fundamental aspects of academic studies in the foundation year, they need frequent prompts and signposting of tasks and guidance to tackle business and management-related questions. Participants agreed that the university policy of fostering more self-guiding learning as they progress from level 4 onwards is more appropriate as the students have a general sense of understanding of the necessary academic skills and subject knowledge to succeed in their studies. Teachers at the university identify 'at risk' students based on the performance of previous modules and the current module progress to determine whether the student may require additional support to succeed in the summative assessment. These support mechanisms increase the opportunity to learn the curriculum content and do well in the examination. Such an additional support mechanism helps complete the entire curriculum while supporting the weak students.

Elements such as gamification came into a discussion which can transform a classroom experience. Then the final summative work can be a report with a reflective account where students can use the classroom experience. Similarly, if a summative assessment requires students to create a digital poster, then those digital skill sets are taught apart from the subject knowledge. The skills needed to be taught if the summative assessment asked to produce a paper poster would be slightly different if not digitally created. Here, the drive for certain skills attainment (in this case the digital poster creation skill) is driven by the summative assessment. The type of backwash effect here is positive, as it fosters learning and fulfilment of the curriculum.

5.1.4 Summative assessment and learning

The second area of interest was students' learning. Though no student data was collected, and the narratives were based on the teachers' perspectives, there was convincing evidence to reach some conclusive remarks.

Participants discussed the demographics of students and how that has an impact on the way learning is perceived. As most of the students are from widening participation, the way students perceive their study is rather complex as they juggle between work, family, and studies. Participants expressed how the level of motivation varies depending on the taught topic's importance. Students within a degree-

level programme go through several summative assessments in their journey and become accustomed to systemic assessment procedures. Students were found to be very keen for summative submission and it had a direct impact on other low stake assessments which includes AfL. At the university, formative assessment is generally either submission of first 1000 words or a one-page document outlining the structure of the planned summative submission.

Participants shared that; many students treat this formative submission point to submit a full draft instead with an aim to get feedback on the likelihood of getting at least a pass grade for the summative submission. There are also a minority group of students who either do not bother submitting anything for formative assessment or submit some work in progress draft so that they are not identified as 'at risk' students who may or may not submit the summative assignment. Within the university's degree-level undergraduate programmes, summative assessments are used for formative purposes, which aligns with Harlen's (2007) suggestion. However, participants did not feel they were coaching students to pass the summative assessment.

Key enquiries were whether assessment influences learning and how learners learn. The evidence suggested a learning pattern where students show a keen interest in the most important topics, almost always focusing on those areas as a priority. If a less experienced academic lead the classroom, there are chances that the students will steer towards leading conversations on the topics important for the summative assessment. Students were also seen as less interested or motivated to engage with unimportant topics for the final submission. It has further connections in the assessment and its interpretation, as students may ignore the formative feedback received as they see the assessment as unimportant which is in line with Dann (2014)'s work. There is an overarching tendency of students to skip to the most important topic; hence whether summative assessment within a degree-level business programme does influence learning, the collected evidence indicates yes, it does. The backwash effect is negative as it hinders the learning of the curriculum.

5.2 Limitations of the research

This research was conducted as planned with the available resources. The key limitations of the research are listed below to assist future research work and transferability of this research:

1. There was a scarcity of empirical research on the backwash effect of high stake undergraduate higher education summative assessment. Existing empirical research would have further assisted the researcher in narrowing the research scope to contribute to the major knowledge gaps first.
2. The researcher researched on a very constricted schedule, so the researcher could only reach participants of just one university, which narrowed the scope of reaching out to a larger population covering more than one university. Coverage of more universities with more participants would have contributed to a broader generalisability of the research.
3. Participants often referred to shared knowledge, which would sometimes make it complicated to find quotations to support conclusions. Any prompts for participants to make answers more explicit would have risked disrupting the flow of the dialogue.
4. As argued by James (2008), aligning learning theories with third-generation assessment also requires further empirical evidence. After the data analysis, the evidence collected strongly suggested having student opinions, too, so that this research could conclusively inform Alderson & Wall's (1993) hypotheses H11 and H15.
5. The singular data source of this research was 'self-reported data'. In self-reported data, it can be prone to selective memory where the participant may not remember experiences as they occurred. It is also prone to 'attribution' bias, where positive events and outcomes are

attributed to own engagement, and the negative consequences are shifted to external courses (Rosenman et al., 2011). In the data collection stage, checks were made in the way of prompts to ensure the claims were valid. Also, the researcher being part of the senior academic team of the same school within the university, helped identify whether the participants were exaggerating their claims.

6. The researcher could have used additional data sources to triangulate (Patton, 1999), resulting in a comprehensive understanding of the backwash phenomenon. Denzin (1978) and Patton (1999) saw triangulation as a way to test qualitative research's validity through information from multiple sources. In this research, additional data sources could have been students and observation. Triangulation would limit 'selective memory' and 'attribution' biases of 'self-reported data'.

5.3 Contribution to knowledge

The significant contribution of this research was to bring the academic conversation on the 'backwash effect' for summative assessments found within degree-level higher education programmes. The researcher found the research data to be effective in contributing to the hypothesises. The introduction and literature review chapter identified that most of the scholarly work contributing to 'backwash' or 'washback' literature is concerned with language testing and other high-stake examinations related to academic achievements at a lower level than a university level study. There were also claims that, HE assessment does not have to deal with negative backwash effects due to the absence of standardised tests (White, 2009). Given the nature of the high-stake summative assessments, it was important to consider the rising concerns among HE academics (Bloxham & Boyd, 2012; Enriquez, 2020) that the assessment practices are increasingly becoming more of 'paperwork' that does not fully take into consideration the 'consequences'.

This research further explored Messick's (1989b) 'Consequential validity' to see whether HE assessment has consequences and if so, then to what extent they are – positive, negative, or neutral. In Appendix F, a full account of the findings is provided, which indicates there are different levels of backwash effect that we can find in the high-stake summative assessment within Higher Education degree level business programmes. It is the first of its kind of empirical work that investigated the backwash effect of HE summative assessment. It was found academics are aware of the consequential elements of assessment; however, there is, at times, an administrative burden that does not allow them to bring the necessary changes to the assessment. The fluidity of the teaching techniques came to the forefront as the researcher found the academics to have a less rigid view of their teaching philosophy, and there was no evidence of preferring their own teaching techniques over the need to deliver the curriculum. Alderson & Wall (1993) called for exploring the concept of backwash and the account or extent of that. Findings of this research categorically can report that the summative assessment exhibits a positive backwash effect in teaching.

Conversely, in learning, assessment has a significant negative backwash effect as students tend to prioritise summative assessment as that determines their progress and progression. In terms of methods of teaching, there was no backwash or a neutral state. Most importantly, the research had convincing evidence that any tests that have consequences will have backwash, and in the context of this research, it was found to be positive.

5.3 Future research

This research, though it contributes significantly to the current understanding of the backwash effect within the higher education context, it did not consider all the hypothesises of Alderson & Wall (1993).

Hypothesises that were not considered are:

Area	**Hypotheses**
Teaching	H8: A test will influence the rate and sequence of teaching
	H10: A test will influence the degree and depth of teaching
Learning	H5: A test will influence what leamers leam
	H7: A test will influence the rale and sequence of learning
	H9: A test will influence the degree and depth of leaming
Wider areas beyond learning and teaching	H14: Tests will have washback on all learners and teachers

This research could not fully inform the following hypothesises:

Learning & Teaching	H11: A test will influence the attitudes to the content, method, etc. of learning and teaching
Wider areas beyond learning and teaching	H15: Tests will have washback effects for some learners and some teachers, but not for others

In the subsequent research, the remaining hypothesises (H5, H7, H8, H9, H10 and H14) and also H11 and H15 covering four areas – teaching, learning, learning & teaching, and wider areas can be considered by adding additional data sources – student interviews and classroom observations as discussed in section 5.2.

CHAPTER 6: CONCLUSION

Through this dissertation, the researcher investigated a critical aspect of high-stake summative assessment within higher education. The nature of high-stake assessment is not acknowledged within the higher education assessment discussions though the issues of 'drop-outs' are discussed widely. This research is one of the first attempts to contextualise to what extent scholars should initiate the conversation on backwash for higher education. There is a need to acknowledge the types of backwash effects of assessments within higher education and take appropriate actions to ensure the summative assessment is not driving the curriculum delivery. We should do so because current and future graduates are expected to be able to harness soft skills and graduate attributes that make them bankable in the job market and prepare them for future challenges. Summative assessment within higher education reliably considers 'construct validity,' i.e., they are accessing the curriculum; however, the consequences are not always well recognised at all levels within the institute. It is expected that future research in this topic area will go closer to the delivery of the curriculum, engage with students and triangulate the findings for a comprehensive evaluation of all the hypotheses of Alderson & Wall (1993).

REFERENCES:

Adams, W. (2015). Chapter Sixteen: Conducting Semi-structured Interviews. In K. Newcomer, H. Hatry, & J. Wholey (Eds.), *Handbook of Practical Program Evaluation* (4th ed., pp. 492–505). Jossey-Bass, A Wiley Imprint.

Alderson, J. C., & Wall, D. (1993). Does Washback Exist? *Applied Linguistics*, *14*(2), 115–129. https://doi.org/10.1093/applin/14.2.115

Annan, J. (2020). *The Learning Theories Profile : A metacognitive tool for reflecting on professional practice*. *21*(2), 1–19.

Annan, J., Bowler, J., Mentis, M., & Somerville, M. P. (2011). Between Theory and Practice Falls the Shadow: The Learning Theories Profile. *Journal of Cognitive Education and Psychology*, *10*(3), 238–252. https://doi.org/10.1891/1945-8959.10.3.238

Argyris, C., & Schön, D. A. (1974). *Theory in practice : increasing professional effectiveness*. Jossey-Bass.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*, 89–195. http://doi.org/10.1016/S0079-7421(08)60422-3

Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: fields apart? *Assessment in Education: Principles, Policy and Practice*, *24*(3), 317–350. https://doi.org/10.1080/0969594X.2017.1319337

Baird, J.-A., Hopfenbeck, T., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *State of the Field Review Assessment and Learning*. Oxford.

Ball, S., Bew, C., Bloxham, S., Brown, S., Kleiman, P., May, H., Mcdowell, L., Morris, E., Orr, S., Payne, E., Price, M., Rust, C., Smith, B., & Waterfield, J. (2012). *A marked improvement : transforming assessment in higher education*. Higher Education Academy.

Bandura, A. (1965). Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology*, *1*(6), 589–595. https://doi.org/10.1037/h0022070

Bandura, A. (1977). *Social Learning theory*. Prentice-Hall.

Barrie, S. C. (2004). A research-based approach to generic graduate attributes policy. *Higher Education Research and Development*, *23*(3), 261–275. https://doi.org/10.1080/0729436042000235391

Barrie, S. C. (2004). A research-based approach to generic graduate attributes policy. *Higher Education Research and Development*, *23*(3), 261–275. https://doi.org/10.1080/0729436042000235391

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

BERA. (2018). *Ethical guidelines for educational research, Fourth edition (2018)*. BERA. Retrieved June 14, 2022, from https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018

Berger, P., & Luckman, T. (1967). *The social construction of reality*. London.

Biggs, J. B. (1995). Assumptions underlying new approaches to educational assessment. Curriculum Forum, 4(2), 1–22.

Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning : the SOLO taxonomy (structure of the observed learning outcome)*. Academic Press.

Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University : What the Student Does* (4th ed.). Mcgraw-Hill, Society For Research Into Higher Education & Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. In *International Journal of Phytoremediation* (Vol. 21, Issue 1). https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. In *International Journal of Phytoremediation* (Vol. 21, Issue 1). https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2003). "In Praise of Educational Research": Formative assessment. *British Educational Research Journal*, *29*(5), 623–637. https://doi.org/10.1080/0141192032000133721

Bloom, B. S. (1956). *Taxonomy of Educational Objectives : The Classification of Educational Goals. Handbook 1: Cognitive Domain*. Longman.

Bloom, B. S. (1971). Mastery learning. In J. H. Block & P. W. Airasian (Eds.), *Mastery learning: theory and practice.* (pp. 47–63). New York: Holt, Rinehart And Winston.

Bloxham, S., & Boyd, P. (2012). Accountability in grading student work: securing academic standards in a twenty-first century quality assurance context. *British Educational Research Journal*, *38*(4), 615–634. https://doi.org/10.1080/01411926.2011.569007

Böhm, A. (2004). Theoretical Coding: Text Analysis in Grounded Theory. In U. Flick, E. von Kardoff & I. Steinke (Eds.), *A Companion to Qualitative Research* (1st ed., pp. 270-275). London: SAGE Publications Ltd.

Bonner, S. M. (2016). Teacher's Perceptions About Assessment: Competing Narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 21–39). Routledge.

Bradley, H. (2017). "Should I stay or should I go?": Dilemmas and decisions among UK undergraduates. *European Educational Research Journal*, *16*(1), 30–44. https://doi.org/10.1177/1474904116669363

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. http://doi.org/10.1191/1478088706qp063oa

Bridgstock, R. (2009). The graduate attributes we've overlooked: Enhancing graduate employability through career management skills. *Higher Education Research and Development*, *28*(1), 31–44. https://doi.org/10.1080/07294360802444347

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *International Journal of Phytoremediation*, *21*(1), 153–169. https://doi.org/10.1080/09695940123775

Bryman, A. (2008). *Social research methods* (3rd ed.). Oxford: Oxford University Press.

Candy, P. C. (1991). *Self-direction for lifelong learning*. Jossey-Bass.

Cheng, L., & Curtis, A. (2003). Washback or backwash: A review of the impact of testing on teaching and learning. *Washback in Language Testing: Research Contexts and Methods*, 3–17. https://doi.org/10.4324/9781410609731

Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in language testing: research contexts and methods*. Lawrence Erlbaum.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Colding, J. (2020). *A critical reflection on constructive alignment in theory and practice*.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal behavior*, 11, 671-684.

Cranton, P. (1994). Self-Directed and Transformative Instructional Development. *The Journal of Higher Education,65*(6), 726. doi:10.2307/2943826

Cranton, P. (2002). Teaching for Transformation. *New Directions for Adult and Continuing Education*, *2002*(93), 63–72. https://doi.org/10.1002/ace.50

Creswell, J. W., & Poth, C. N. (2018). *Qualitative Inquiry & Research Design: Choosing among Five approaches* (4th ed.). SAGE.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE.

Cronbach, L. J. (1988). Five Perspectives on Validity Argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 5–18). Routledge.

Crotty, M. (1998). *The foundations of social research meaning and perspective in the research process*. London: Sage Publications.

Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy and Practice*, *21*(2), 149–166. https://doi.org/10.1080/0969594X.2014.898128

David, M. E. (2012). *Widening Participation in Higher Education : Casting the Net Wide?* Palgrave Macmillan.

Denzin, N. K. (1978). *Sociological methods: A sourcebook*. McGraw-Hill Book.

Dewey, J. (1938). *Experience and education*. New York: Macmillan.

Dey, I. (1998). *Qualitative data analysis a user-friendly guide for social scientists.* London: Routledge.

Ecclestone, K. (2007). Learning assessment: students' experiences in post-school qualification. In N. Falchikov & D. Boud (Eds.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 41–54). Chapter, Routledge.

Elton, L., & Brenda, J. (2002). *Assessment in universities: a critical review of research*. http://eprints.soton.ac.uk/59244/

Elton, L., & Brenda, J. (2002). *Assessment in universities: a critical review of research*. http://eprints.soton.ac.uk/59244/

Enriquez, J. (2020). *Paper-work : what documents have to say about assessment practices*.

Faden, R., & Beauchamp, T. L. (1986). *A history and theory of informed consent*. Oxford Univ. Press.

Gibbs G (1988). Learning by Doing: A guide to teaching and learning methods. Further Education Unit. Oxford Polytechnic: Oxford.

Gibbs, G., & Simpson, C. (2005). Conditions Under Which Assessment Supports Students' Learning. *Learning and Teaching in Higher Education*, *1*, 3–31.

Greenstein, L. (2010). *What Teachers Really Need to Know About Formative Assessment*.

Guba, E. G., & Lincoln, Y. S. (1994). Competing Paradigms in Qualitative Research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage Publications.

Guba, E. G., & Lincoln, Y. S. (1994). Competing Paradigms in Qualitative Research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage Publications.

Guskey, T. R. (2010). Formative Assessment: The Contributions of Benjamin S. Bloom. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 106–125). Routledge.

Guskey, T. R. (2010). Formative Assessment: The Contributions of Benjamin S. Bloom. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 106–125). Routledge.

Hammersley, M. (1992). Reconstructing the qualitative-quantitative divide. In M. Hammersley (Author), *What is wrong with ethnography? Methodological explorations* (pp. 159-173). London: Routledge.

Hammersley, M. (2014). On the ethics of interviewing for discourse analysis. Qualitative Research, 14(5), 529–541. https://doi.org/10.1177/1468794113495039

Harlen, W. (2005). Teachers' summative practices and assessment for learning – Tensions and synergies. *Curriculum Journal*, *16*(2), 207–223. https://doi.org/10.1080/09585170500136093

Harlen, W. (2007). Formative classroom assessment in science and education. In J. H. Mcmillan (Ed.), *Formative classroom assessment : theory into practice*. Teachers College, Columbia University.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *International Journal of Phytoremediation*, *21*(1), 365–379. https://doi.org/10.1080/0969594970040304

Heubert, J. P., & Hauser, R. M. (1999). *High stakes testing for tracking, promotion, and graduation*. National Academy Press.

Hill, J., Walkington, H., & France, D. (2016). Graduate attributes: implications for higher education practice and policy: Introduction. *Journal of Geography in Higher Education*, *40*(2), 155–163. https://doi.org/10.1080/03098265.2016.1154932

Hoffman, E. (2007). Open-ended interviews, power, and emotional labour. Journal of Contemporary Ethnography, 36, 318-346. doi : 10.1177/0891241606293134.

Holroyd, C. (2000). Are assessors professional? *Active Learning in Higher Education*, *1*(1), 28–44. https://doi.org/10.1177/1469787400001001003

Hopfenbeck, T. (2018). Classroom assessment, pedagogy and learning–twenty years after Black and Wiliam 1998. *Assessment in Education: Principles, Policy and Practice*, *25*(6), 545–550. https://doi.org/10.1080/0969594X.2018.1553695

Hubley, A. M., & Zumbo, B. D. (2011). Validity and the Consequences of Test Interpretation and Use. *Social Indicators Research*, *103*(2), 219–230. https://doi.org/10.1007/s11205-011-9843-4

Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript, University of Reading, England.

Hussey, T., & Smith, P. (2002). The Trouble with Learning Outcomes. *Active Learning in Higher Education*, *3*(3), 220–233. https://doi.org/10.1177/1469787402003003003

Illeris, K. (2009). *Contemporary Theories of Learning* (K. Illeris (ed.); 1st ed.). Routledge.

James, M. (2008). Assessment and learning. In S. Swaffield (Ed.), *Unlocking Assessment: Understanding for reflection and application* (pp. 20–36). Routledge (David Fulton). https://doi.org/10.4135/9781446250808

JISC. (2015). *Assessment literacy*. Jisc. https://www.jisc.ac.uk/guides/transforming-assessment-and-feedback/assessment-literacies

Johnston, B. (2002). Some Basic Assessment Dilemmas with Particular Reference to Portfolios. In *assessment in universities: a critical review of research* (pp. 35–49). Learning and Teaching Support Network (LTSN) Generic Centre. http://eprints.soton.ac.uk/id/eprint/59244

Johnston, B. (2002). Some Basic Assessment Dilemmas with Particular Reference to Portfolios. In *assessment in universities: a critical review of research* (pp. 35–49). Learning and Teaching Support Network (LTSN) Generic Centre. http://eprints.soton.ac.uk/id/eprint/59244

Keat, R. (1981). *The Politics of social theory: Habermas, Freud and the critique of positivism*. Oxford: Basil Blackwell.

Kegan, R. (2008). What "form" transforms? A constructive-developmental approach to transformative learning. In K. Illeris, *Contemporary Theories of Learning* (1st ed., pp. 42-43). Oxon: Routledge.

Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education*, *106*, 166–171. https://doi.org/10.1016/j.compedu.2016.12.006

Knowles, M. S. (1973). *The Adult Learner: A Neglected Species*. Gulf Pub. Co.

Knowles, M. S. (1975). *Self-Directed Learning: A Guide for Learners and Teachers*. Cambridge: Englewood Cliffs.

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. *Journal of Organizational Behavior*, *8*, 359–360. http://doi.org/10.1002/job.4030080408

Kvale, S. (1996). The interview situation. In *Interviews. An Introduction to Qualitative Research Interviewing* (pp. 124–135). London: Sage Publications Ltd.

Lave, J. (2009). The practice of learning. In K. Illeris, *Contemporary Theories of Learning: Learning Theorists ... In Their Own Words* (1st ed., pp. 200-209). Oxon: Routledge.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. *Learning in Doing*, *95*, 138. http://doi.org/10.2307/2804509

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, Calif.: Sage Publications.

Little, D. G. (1991). *Learner autonomy. definitions, issues and problems*. Authentik Language Learning Resources.

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2006). *Methods in Educational Research*.

Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-directed learning in problem-based learning and its relationships with self-regulated learning. *Educational Psychology Review*, *20*(4), 411–427. https://doi.org/10.1007/s10648-008-9082-7

Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issues and Practice*, 16, 16–18.

Merriam, S. (2001). Andragogy and self-directed learning: Pillars of adult learning theory. *New Directions for Adult and Continuing Education*, *89*, 3–14. https://doi.org/10.1002/ace.3

Merton, R. K. (1972). Insiders and Outsiders: A Chapter in the Sociology of Knowledge. In *American Journal of Sociology* (Vol. 78, Issue 1, pp. 9–47). https://doi.org/10.1086/225294

Messick, S. (1989a). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). New York, NY: Macmillan.

Messick, S. (1989b). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher*, *18*(2), 5–11. https://doi.org/10.3102/0013189x018002005

Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.

Mezirow, J. (1997). Transformative Learning: Theory to Practice. *New Directions for Adult and Continuing Education*, *1997*(74), 5–12. https://doi.org/10.1002/ace.7401

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A sourcebook of new methods* (3rd ed.). Thousand Oaks, CA: Sage.

Neuman, W. L. (2014). *Social research methods: qualitative and quantitative approaches*. Boston: Pearson.

Patton, M. Q. (2015). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice* (4th ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Patton, M.Q. (1999). Enhancing the quality and credibility of qualitative analysis. Health Sciences Research, 34, 1189–1208.

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*(G. V. Anrep, Ed.). Mansfield Centre, CT: Martino Publishing.

Piaget, J. (1957). *Construction of reality in the child*. London: Routledge.

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *Teacher Educator*, *46*(4), 265–273. https://doi.org/10.1080/08878730.2011.605048

Rachal, J. (1983). The Andragogy-Pedagogy Debate: Another Voice in the Fray. *Lifelong Learning: The Adult Years*, *6*(9), 14-15.

Roberts, G. (2002). *SET for success: The supply of people with science, technology, engineering, and mathematics skills*.

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International journal of behavioural & healthcare research*, *2*(4), 320–332. https://doi.org/10.1504/IJBHR.2011.043414

Ryan, G. W., & Bernard, H. R. (2003). Techniques to Identify Themes. *Field Methods*, *15*(1), 85–109. http://doi.org/10.1177/1525822X02239569

Savin-Baden, M., & Major, C. (2013). *Qualitative Research: The Essential Guide to Theory and Practice*. Routledge.

Scott, I. (2011). The Learning Outcome in Higher Education: Time to Think Again? *Worcester Journal of Learning and Teaching*, (5), 1–8.

Scriven, M. (1966). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of Curriculum Evaluation Chicago* (pp. 39–83). Chapter, Univ. of Colorado.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, *27*(2), 4–13. https://doi.org/10.3102/0013189X027002004

Shenton, A. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*, 63–75. http://doi.org/10.1111/j.1744-618X.2000.tb00391.x

Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16, 5–8,13, 24.

Skinner, B. F. (1974). *About behaviorism*. New York: Knopf.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*(3), 238. https://www.proquest.com/scholarly-journals/assessment-literacy-21st-century/docview/218532914/se-2?accountid=13042

Stobart, G., & Eggen, T. (2012). High-stakes testing – Value, fairness and consequences. *Assessment in Education: Principles, Policy & Practice*, 19(1), 1–6. doi:10.1080/0969594X.2012.639191

Sullivan, G. M., & Sargeant, J. (2011). Qualities of qualitative research: part I. *Journal of graduate medical education*, *3*(4), 449–452. https://doi.org/10.4300/JGME-D-11-00221.1

Taras, M. (2009). Summative assessment: The missing link for formative assessment. *Journal of Further and Higher Education*, *33*(1), 57–69. https://doi.org/10.1080/03098770802638671

Thomas, D. R. (2006). A general inductive approach for analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, *27*(2), 237–246. https://doi.org/10.1177/1098214005283748

Thornborrow, J. (2002). Power talk: Language and interaction in institutional discourse. Harlow: Longman.

Thorndike, E. L. (1920). *Educational psychology*. New York: Teachers college, Columbia University.

Tight, M. (2004). Research into higher education: an a-theoretical community of practice? *Higher Education Research & Development*, *23*(4), 395–411. https://doi.org/10.1080/0729436042000276431

Torrance, H., Colley, H., Garratt, D., Jarvis, J., Piper, H., Ecclestone, K. and James, D. (2005) The Impact of Different Modes of Assessment on Achievement and Progress in the Learning and Skills Sector, Learning and Skills Development Agency (available at: https://www.lsda.org.uk/cims/order.aspx?code=052284&src=XOWEB).

Trowler, V. (2010). *Student engagement literature review*. [online] The Higher Education Academy. Available at: https://www.heacademy.ac.uk/system/files/studentengagementliteraturereview_1.pdf

Vygotsky, L. (1978). *Mind in Society: Development of Higher Psychological Processes* (p. 86). Cambridge: Harvard University Press. (Original manuscripts [ca. 1930-1934])

Watanabe, Y. (2004). Methodology in Washback Studies. In *Washback in language testing : research contexts and methods* (pp. 19–36). Lawrence Erlbaum.

Watson, J. B. (1959). *Behaviorism*. Chicago: The University of Chicago Press.

Wei, W. (2017). *A Critical Review of Washback Studies: Hypothesis and Evidence*. *January 2017*, 49–67. https://doi.org/10.1007/978-3-319-32601-6_4

Wellington, J. (2015). *Educational research: Contemporary issues and practical approaches* (2nd ed.). Bloomsbury Academic.

Wenger, E. (1998). *Communities of practice*. Cambridge, U.K.: Cambridge University Press.

White, E. (2009). Are you assessment literate? Some fundamental questions regarding effective classroom-based assessment. *OnCUE Journal*, *3*(1), 3–25. http://jaltcue-sig.org/files/OnCUE/OCJ3-1articles/OCJ3-1-White-pp3-25.pdf

Wildemeeersch, D., & Stroobants, V. (2008). Transitional learning and reflexive facilitation: the case of learning for work. In K. Illeris, *Contemporary Theories of Learning* (1st ed., pp. 219–232). Oxon: Routledge.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, *81*(3), 329–339. https://doi.org/10.1037/0022-0663.81.3.329

Appendix A: Written consent form

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Principal Investigator: Prof. Jo-Anne Baird

PI Telephone number: +44 1865 274002
PI Email address: jo-anne.baird@education.ox.ac.uk

Primary researcher: Dr. Syed Ali Tarek, MSc Educational Assessment student
Telephone number: +44(0)1865 274024
Email: syed.tarek@education.ox.ac.uk

Consent to take part in 'Backwash' effects of high-stake degree level HE assessment on teaching practices.

Central University Research Ethics Committee (CUREC) approval reference: **CIA-22HT-082**

Purpose of Study: Exploring the relationship between assessment and beliefs about learning particularly at a degree level.

	Please initial each box if you agree with the statement
I confirm that I have read and understand the information sheet for the above research. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
I understand that my participation is voluntary and that I am free to withdraw at any point without giving any reason.	
I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.	
I understand that I will not be identifiable from any publications or presentations.	
I consent to being video recorded.	
I understand how videos will be used in research outputs.	
I give permission for you to contact me again to clarify information.	
I understand how to raise a concern or make a complaint.	

Template written consent form, version 4.1, January 2022

I agree to take part.[1] ☐

_____	dd / mm / yyyy	_____
Name of participant	Date	Signature

_____	_____	_____
Name of person taking consent	Date[2]	Signature

[2] *To be signed and dated in the presence of the participant. Once this has been signed by both parties the participant should receive a copy of the signed and dated participant consent form. The original signed and dated consent form should be kept with the project's main documents, which must be kept in a secure location.

Appendix B: Invitation letter

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Principal Investigator: Prof. Jo-Anne Baird
PI Telephone number: +44 1865 274002
PI Email address: jo-anne.baird@education.ox.ac.uk

Primary researcher: Dr. Syed Ali Tarek, MSc Educational Assessment student
Telephone number: +44(0)1865 274024
Email: syed.tarek@education.ox.ac.uk

Tuesday, 19 April 2022

Letter of Invitation

Dear Colleague,

Good morning, I hope you had a good Easter. I am currently recruiting participants for my research exploring the relationship between assessment and beliefs about learning. My research area is less academically discussed within Higher Education, however, from practice, I believe it is something that is worth having a deeper investigation.

I will be grateful if you manage to share your valuable insights in a one-one interview (between 45 min to 1 hour) via Teams. Your participation in this research will help me to move the current academic literature to consider higher education practices.

You can confirm your participation via email as soon as you receive this email but please try to let me know by 22/04/2022. I am attaching the participant information document with this email so you can take an informed decision.

Thank you very much for assisting me in this important research.

Best wishes,
Syed Tarek.

Appendix C: Participant information sheet

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Principal Investigator: Prof. Jo-Anne Baird
PI Telephone number: +44 1865 274002
PI Email address: jo-anne.baird@education.ox.ac.uk

Primary researcher: Dr. Syed Ali Tarek, MSc Educational Assessment student
Telephone number: +44(0)1865 274024
Email: syed.tarek@education.ox.ac.uk

'Backwash' effects of high-stake degree level HE assessment on teaching practices.

PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee Approval Reference: [**CIA-22HT-082**]

1. Introductory paragraph

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part.

2. Why is this research being conducted?

The impact of a test or an examination on teaching and learning is referred to as 'Backwash'. 'Backwash' effect is commonly discussed in language testing and other high stake public examinations. However, there is a scarcity of research aimed at the HE assessment. I will be narrowing the scope to Degree level business programmes and build up from the current tensions among the practitioners on the gaps in learning theories used in the classroom delivery and the theories that dominated assessment practices. Within a deeper concentration on the impact of 'backwash effects', the study aims to explore the relationship between assessment and beliefs about learning.

3. Why have I been invited to take part?

You have been invited because you are currently teaching on a degree level business programme or you are in an administrative role making decisions on the teaching and learning aspects of a degree level business programme. In this research, the scope of the degree level business programmes is inclusive of the 1st, 2nd and 3rd year and 'Advanced entry to Top up' stage and exclusive of foundation year degrees.

4. **Do I have to take part?**

No. It is up to you to decide whether or not to take part. You can withdraw from the study, without giving a reason, and without negative consequences, by advising me of this decision by email to syed.tarek@education.ox.ac.uk by 29/07/2022, 1700. If you decide to withdraw after the data was collected, the data will be securely erased without any chance of data retrieval.

5. **What will happen to me if I take part in the research?**

If you agree to take part in the research the following will take place:

- Consent form and other relevant documents will be emailed prior scheduling the interview.
- Interviews will be scheduled remotely via MS Teams calendar invite.
- The interview duration is expected to be between 45 min to 1 hour.
- With your consent, I would like to video record you because video recording is MS Teams default recording option and video recording will help me to have an accurate record of our conversation.
- You can ask to pause or stop the interview at any time.

6. **What are the possible disadvantages and risks in taking part?**

There is no foreseeable disadvantage for participation.

7. **Are there any benefits in taking part?**

While there are no immediate benefits for those people participating in the project, it is hoped that this research will lead to academics better understanding the impact of HE summative assessment.

8. **What information will be collected and why is the collection of this information relevant for achieving the research objectives?**

Data generated from this research will be:

- Your signed research consent form.
- Video recording of the interview.
- Interview transcript.

All documents will be stored in the One Drive of the researcher's Oxford Nexus365 -two steps secured; password protected account. Research data will be strictly restricted for sharing. Only the researcher will have primary access to the raw data. Research supervisor Prof. Jo-Anne Baird will be also able to access the data if required.

At the end of the research project the data will remain in the Nexus 365 One Drive until the researcher's date of completion of course comes to an end which is 30th Sept 2022. Once the terminal date is reached, the researcher will encrypt all data using a suitable 256 bit encryption software such as Encrypto (https://apps.apple.com/us/app/encrypto-secure-your-files/id935235287?mt=12) and store all the files in the OSX security protected SSD for up to 3 years. No forms of personal data will be stored beyond this point and will be destroyed using an appropriate digital tool such as Blancco Drive Eraser (https://www.ncsc.gov.uk/products/blancco-drive-eraser-6) to comply with UK GDPR Act.

9. **Will the research be published? Could I be identified from any publications or other research outputs?**

The findings from the research will be written up *in a dissertation, academic publications and conference presentations.*] The participant data will be anonymised and not identifiable. Pseudonyms will be used for participants and their institutions.

I would like your permission to use direct quotations *but without identifying you* in any research outputs.

A copy of my dissertation will be deposited both in print and online in the Oxford University Research Archive where it will be publicly available to facilitate its use in future research.

10. **Data Protection**

The researcher will not be holding any personal data. In general terms, the University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the study. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available at https://compliance.admin.ox.ac.uk/individual-rights.

11. **Who has reviewed this study?**

This study has received ethics approval from a subcommittee of the University of Oxford Central University Research Ethics Committee. (Ethics reference: **CIA-22HT-082**).

12. **Who do I contact if I have a concern about the research or I wish to complain?**

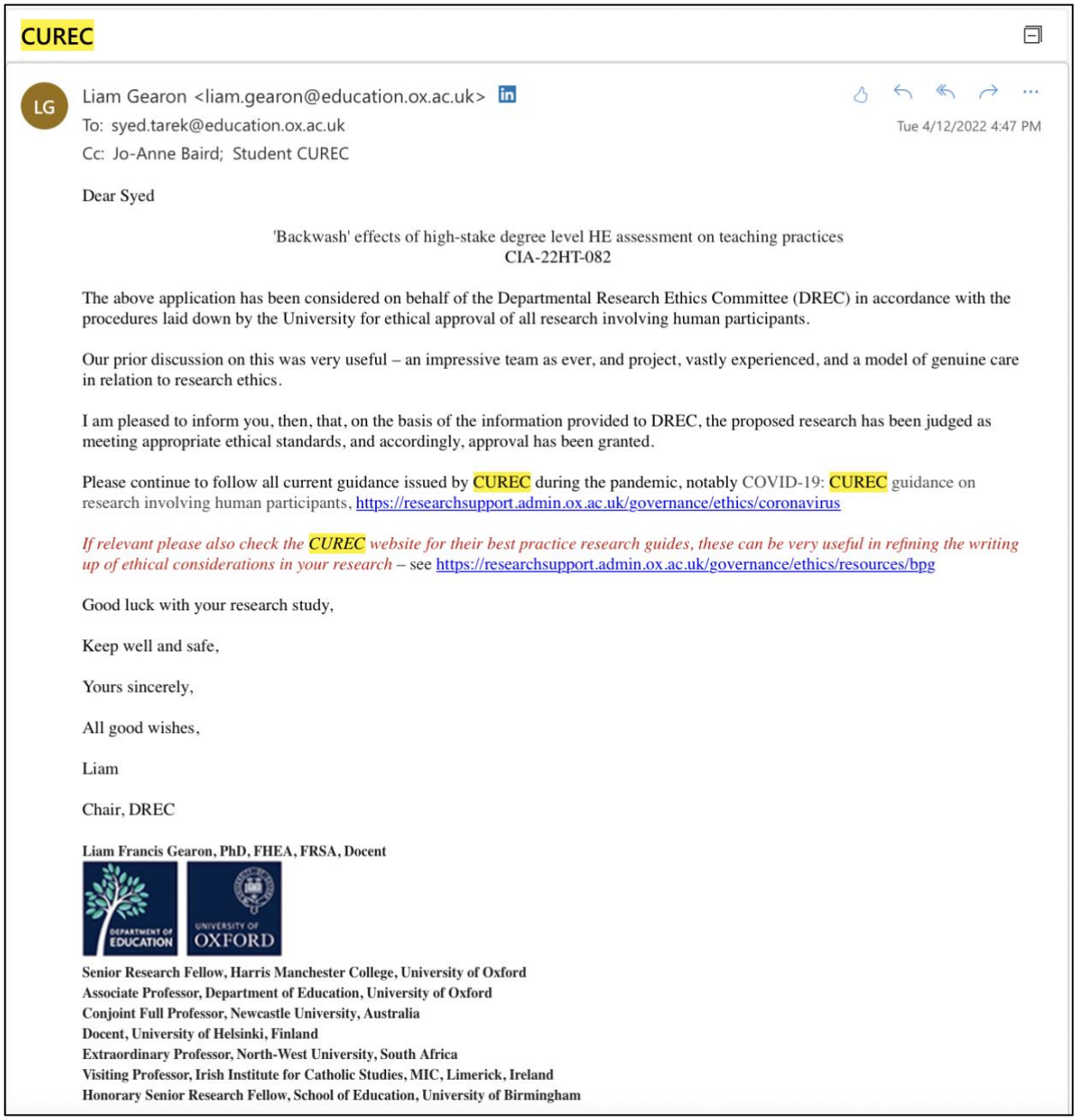
If you have a concern about any aspect of this study, please contact my supervisor, Professor Jo-Anne Baird (jo-anne.baird@education.ox.ac.uk).

13. **Further Information and Contact Details**

If you would like to discuss the research with someone beforehand or if you have questions afterwards, please contact:

Dr. Syed Ali Tarek
Email: syed.tarek@education.ox.ac.uk

Appendix D: CUREC approval



CUREC

LG

Liam Gearon <liam.gearon@education.ox.ac.uk>

To: syed.tarek@education.ox.ac.uk

Cc: Jo-Anne Baird; Student CUREC

Tue 4/12/2022 4:47 PM

Dear Syed

'Backwash' effects of high-stake degree level HE assessment on teaching practices
CIA-22HT-082

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

Our prior discussion on this was very useful – an impressive team as ever, and project, vastly experienced, and a model of genuine care in relation to research ethics.

I am pleased to inform you, then, that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

Please continue to follow all current guidance issued by CUREC during the pandemic, notably COVID-19: CUREC guidance on research involving human participants, https://researchsupport.admin.ox.ac.uk/governance/ethics/coronavirus

If relevant please also check the CUREC website for their best practice research guides, these can be very useful in refining the writing up of ethical considerations in your research – see https://researchsupport.admin.ox.ac.uk/governance/ethics/resources/bpg

Good luck with your research study,

Keep well and safe,

Yours sincerely,

All good wishes,

Liam

Chair, DREC

Liam Francis Gearon, PhD, FHEA, FRSA, Docent

DEPARTMENT OF EDUCATION | UNIVERSITY OF OXFORD

Senior Research Fellow, Harris Manchester College, University of Oxford
Associate Professor, Department of Education, University of Oxford
Conjoint Full Professor, Newcastle University, Australia
Docent, University of Helsinki, Finland
Extraordinary Professor, North-West University, South Africa
Visiting Professor, Irish Institute for Catholic Studies, MIC, Limerick, Ireland
Honorary Senior Research Fellow, School of Education, University of Birmingham

APPENDIX E: Participant details

Participant identifier	Subjects Taught/ Managed	Typical summative assessment methods in the taught subjects	PSRB approval for the programme	PSRB Directed Assessment	Typical formative assessment methods in the taught subjects
Pa1	Management, business strategy and marketing	Report, Individual presentation	CMI	No	Draft plan
Pa2	Business, Strategy, Skills development	Report, Individual presentation	CMI	No	First 1000 words of the summative assessment task
Pa3	Management, Leadership, HRM, Culture	Report, Simulation, Reflective essay	CMI	No	Draft plan
Pa4	Business, HRM, Operations, Supply chain	Report, Reflective essay	CMI	No	Draft plan
Pa5	Accounting, Finance	Time Constraint Exam/ Report	ACCA CMI	Yes, ACCA directs the assessment standard	Mock test /Draft of one summative assessment task
Pa6	Tourism, Management	Report, Essay, Case study	CMI	No	First 1000 words of the summative assessment task
Pa7	Finance, Leadership	Time Constraint Exam, Report, Simulation	ACCA CMI	Yes, ACCA directs the assessment standard	Mock test /Draft of one summative assessment task
Pa8	Tourism, Management	Report, Essay	CMI	No	First 1000 words of the summative assessment task
Pa9	Tourism, Management	Report, Essay	CMI	No	First 1000 words of the summative assessment task
Pa10	Accounting, Finance	Time Constraint Exam/ Report/ Case study	ACCA CMI	Yes, ACCA directs the assessment standard	Mock test /Draft of one summative assessment task
Pa11	Business, HRM	Report, Reflective essay	CMI	No	Draft plan
Pa12	Business strategy and marketing	Report, Individual presentation	CMI	No	Draft plan

APPENDIX F: Summery of the findings

Area	Alderson & Wall (1993) Hypotheses	Research question	Interview question	Research findings indication	Backwash effect indicator
Teaching	H1: A test will influence teaching	How does the ' backwash effect' impact teaching practices within higher education Business degree programmes?	Q12. Based on the student outcome of an individual summative assessment, do you revise your teaching methods to ensure students are performing to their best of abilities in the summative assessment?	Evidence suggested teachers are mindful of the assessment methods. If students have to create a poster for the summative assessment, teaching will be moulded in a way that covers how to create posters. Similarly for presentation or report writing teaching is adapted to that need. In terms of covering what is important for the assessment. Though issues like 'coaching' or 'tunnelled coverage of the curriculum' were brought in the discussion, however, none of the participants said they skip any of the curriculum's indicative content.	Positive backwash
	H3: A test will influence how teachers teach		Q4. Does the summative assessment affect HOW you teach, not just what you teach? -Can you give an example from your practice?	Teachers are pragmatic in their choice of teaching style. They are mindful of the type of teaching would likely deliver a positive outcome for the students.	Positive backwash
	H4: A test will influence what teachers teach		Q2. While teaching a module, if you encounter a particular topic that is not relevant for the summative assessment, what do you do? -If you skip the content, follow-up question –Do you think skipping those seemingly less relevant topics for summative assessment impacts the student's overall learning and knowledge of the module?	Data suggests teachers teach the curriculum and also ensures all the assessment items are also covered with a greater coverage of those key topics. Areas that are less prominent are covered in a way that corresponds to the low importance (for the assessment) nature of that area. There was also evidence that teachers go beyond the curriculum to ensure students are getting the concurrent knowledge and skills required for their life beyond the programme.	Positive backwash
Learning	H2: A test will influence learning		Q5. Do you think summative assessment influences learning? If so, then can you elaborate with an example? Q3: Do you think summative assessment influences some students to show a lack of interest in some topics or lessons as they are less important or unimportant at all for the summative assessment?	Yes, summative assessment influence learning. Students may tend to skip curriculum content that are not impotent for the summative assessment.	Negative backwash
	H6: A test will influence how learners learn		Q1. Do you think student learning in your course should be bound within the areas covered in the summative assessment?	Yes, other assessments like AfL are not seen as very important by learners in general as that assessment does not contribute to final grading.	Negative backwash
Learning & Teaching	H11: A test will influence attitudes to the content and methods of		Q6. The term 'Pedagogy' refers to the art and science of teaching. Pedagogy relates to the study and use of various teaching theories, including	From the data indications are that academics determine the needed teaching technique 'on the fly'. Not much a thought is given on whether the teaching	No backwash/Neutral (Teaching aspect)

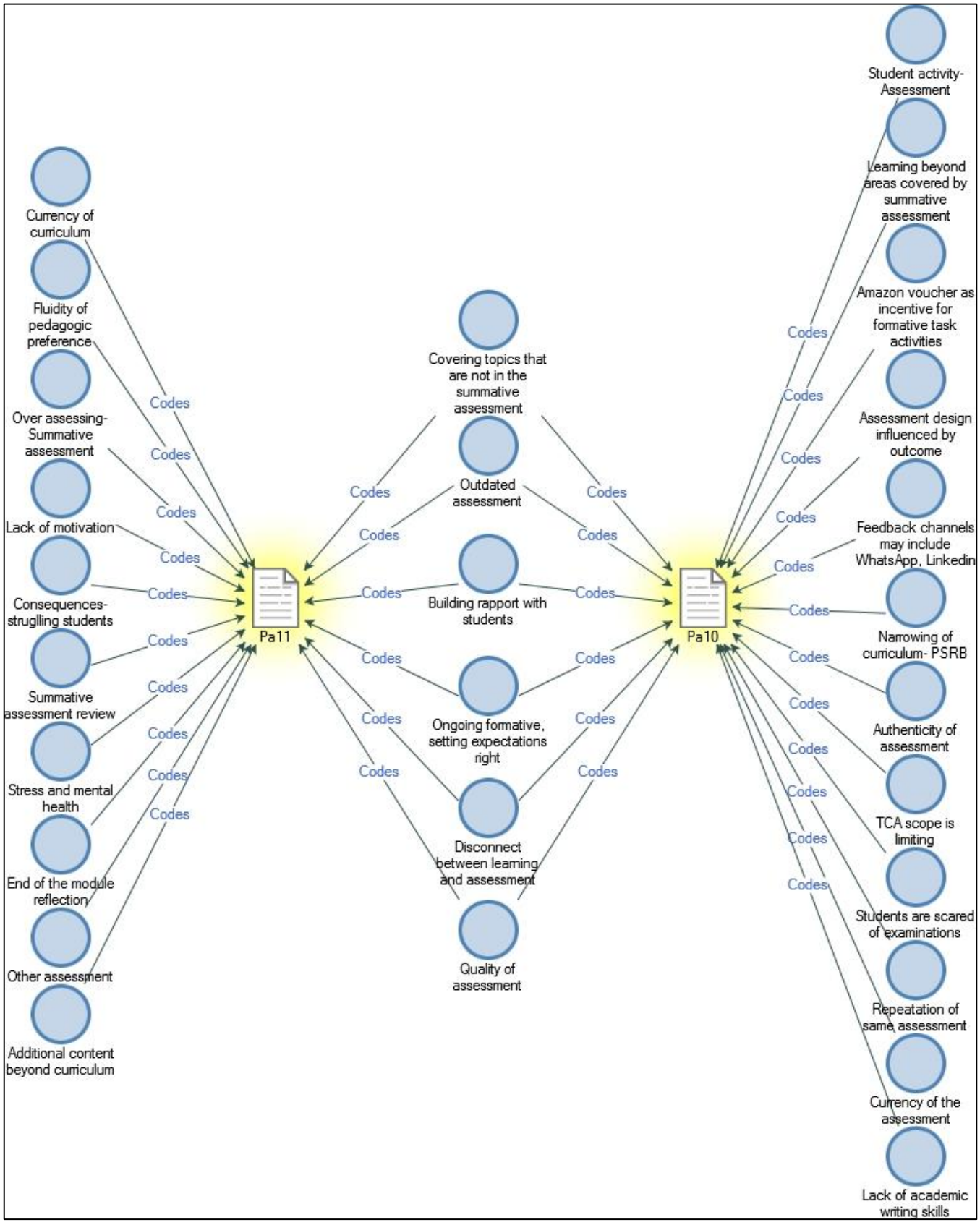
	teaching and learning.		instructional models and how those influence students and their learning. To what extent do you believe there needs to be a synergy between the pedagogic model/teaching method and assessment methods (such as assignments, essays, presentations, time-constrained exams and portfolios etc.), especially those used for summative assessment?	technique is in line with their own personal teaching philosophy or any specific teaching or learning theory. Teachers can move between various methods of teaching and summative assessment does not greatly influence that choice. Evidence suggests summative assessment may influence attitude to the content but whether methods of learning is influenced was beyond the scope of this research.	Not enough data collected/ Out of scope (Learning aspect)
Wider areas beyond learning and teaching	H12: Tests that have important consequences will have washback		Q10. Within your practice area, to what extent is assessment design (type of assessment, weightage for each assessment segment etc.) influenced by the summative assessment results?	Yes, students and teachers are both aware of the importance and a level of additional effort or support is given to ensure the summative assessment outcomes are favourable for the student and the teacher.	Positive backwash
	H13: Tests that do not have important consequences will have no washback		Q7. Apart from the summative assessment, what other assessment methods (such as formative, diagnostic etc.) do you actively use in delivering your module? Q8. Do you think the summative assessment is the only assessment that influences learning and teaching in contrast with formative assessments?	Yes, other assessments like AfL are seen as less important by students and academics are not too keen on pushing students for formative submissions.	No backwash/Neutral
	H15: Tests will have washback effects for some learners and some teachers, but not for others		Q9. To what extent do you find summative assessment to influence all students and teachers, or just some groups? -If some groups are indicated, follow-up question- Which groups in particular and are there any specific observations about those groups?	Whether assessment has washback effect on just some learner is too wide and complex to measure. From the evidence of teacher, it indicates there may be a possibility of isolated washback effect, but it needs further empirical study.	Overall, not enough data collected/ Out of scope
			Q13. What is your current understanding of the positive or negative effects of summative assessment on learning and teaching within higher education?	Some academics found assessment too be too restrictive and lacking innovative approaches. Academics all suggested a disconnect between placing students at the centre of delivery, allowing them to harness graduate attributes and then the assessment lacking in giving students an opportunity to showcase those soft-skills or graduate attributes.	No backwash/Neutral
			Q11. Does your student's performance in the summative assessment impact your mental health in any possible way? Leading to an increased sense of accountability? Probe- Do you feel such pressure on student performance can lead to more enforced accountability?	Those who are seasoned academics, see no issue with the outcome of the summative assessment. However, they did indicate, there is potentially some impact especially for those who are in new in the teaching profession delivering a higher education programme.	No backwash/Neutral

APPENDIX G: Alderson & Wall (1993) all hypothesises and categories

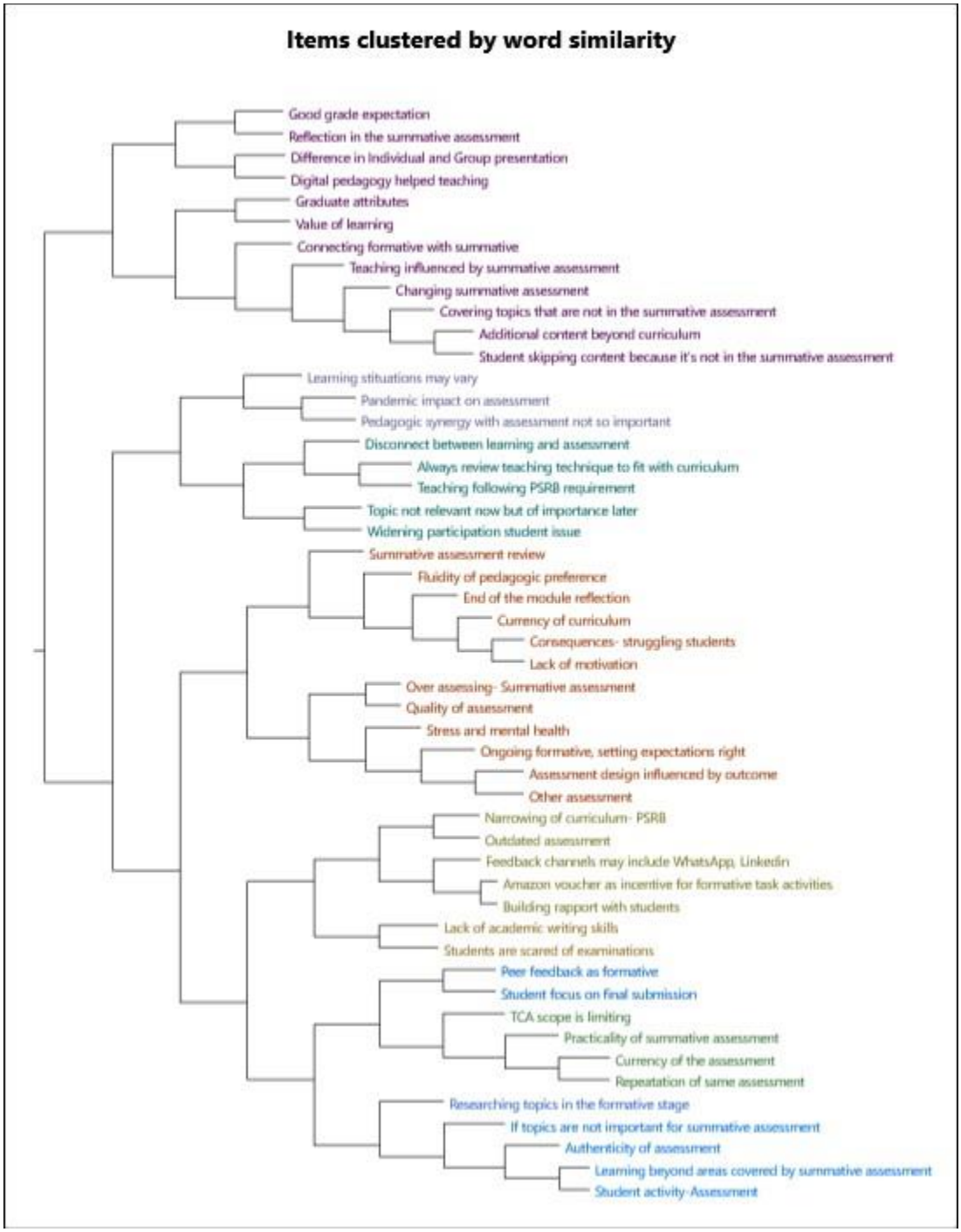
Area	Hypotheses	Covered in this research
Teaching	H1: A test will influence teaching	Yes
	H3. A test will influence how teachers teach	Yes
	H4: A test will influence what teachers teach	Yes
	H8: A test will influence the rate and sequence of teaching	No
	H10: A test will influence the degree and depth of teaching	No
Learning	H2: A test will influence learning	Yes
	H5: A test will influence what leamers leam	No
	H6: A test will influence how leamers leam	Yes
	H7: A test will influence the rale and sequence of learning	No
	H9: A test will influence the degree and depth of leaming	No
Learning & Teaching	H11: A test will influence the attitudes to the content, method, etc. of learning and teaching	Yes
Wider areas beyond learning and teaching	H12: Tests that have important consequences will have washback	Yes
	H13: Tests that do not have important consequences will have no washback	Yes
	H14: Tests will have washback on all learners and teachers	No
	H15: Tests will have washback effects for some learners and some teachers, but not for others	Yes

APPENDIX H: Nodes at a participant level

APPENDIX I: Comparison of Nodes



APPENDIX J: Sample of Items clustered by word similarity



APPENDIX K: Sample codes

- Additional content beyond curriculum
- Always review teaching technique to fit with curriculum
- Amazon voucher as incentive for formative task activities
- Assessment design influenced by outcome
- Authenticity of assessment
- Building rapport with students
- Changing summative assessment
- Connecting formative with summative
- Consequences- struggling students
- Covering topics that are not in the summative assessment
- Currency of curriculum
- Currency of the assessment
- Difference in Individual and Group presentation
- Digital pedagogy helped teaching
- Disconnect between learning and assessment
- End of the module reflection
- Feedback channels may include WhatsApp, Linkedin
- Fluidity of pedagogic preference
- Good grade expectation
- Graduate attributes
- If topics are not important for summative assessment
- Lack of academic writing skills
- Lack of motivation
- Learning beyond areas covered by summative assessment
- Learning situations may vary
- Narrowing of curriculum- PSRB
- Ongoing formative, setting expectations right
- Other assessment
- Outdated assessment
- Over assessing- Summative assessment
- Pandemic impact on assessment
- Pedagogic synergy with assessment not so important

- Peer feedback as formative
- Practicality of summative assessment
- Quality of assessment
- Reflection in the summative assessment
- Repetition of same assessment
- Researching topics in the formative stage
- Stress and mental health
- Student activity-Assessment
- Student focus on final submission
- Student skipping content because it's not in the summative assessment
- Students are scared of examinations
- Summative assessment review
- TCA scope is limiting
- Teaching following PSRB requirement
- Teaching influenced by summative assessment
- Topic not relevant now but of importance later
- Value of learning
- Widening participation student issue