THE INFLUENCE OF ASSESSMENT WASHBACK ON THE STUDENT EXPERIENCE OF THE INTERNATIONAL BACCALAUREATE MIDDLE YEARS PROGRAMME

JAMES SMITH

A thesis submitted in partial fulfillment of the requirements of Nottingham Trent University for the degree of Professional Doctorate in Education

March 2019

Copyright Statement

Abstract

The IB *Middle Years Programme* (MYP), an international curriculum for students between the ages of 11 and 16, is taught by 1,400 schools across the globe and aspires to embody a constructivist learning model in which students use inquiry to explore their world. Students are encouraged to develop conceptual understandings based on learning experiences that are driven by a statement of inquiry, and apply these understandings to novel situations. At the same time, students are subject to the need for measurement of achievement both by assessment practices within the MYP and, in most MYP schools, the need to prepare students effectively for the high-stakes assessment demands of the *IB Diploma Programme* (DP). Consequently, assessment practices within the MYP are a contested space in which competing understandings, and ideas about the purposes of assessment, are negotiated.

This research takes the form of an ethnographic study of a group of Year 5 (age 15-16) MYP English Language and Literature students at Bayside College in Hong Kong. It discusses the influence of assessment on the student experience in the MYP, identifying a range of mediating influences that are involved in the process of students attaching meaning to assessment. A series of metaphors is used to depict the complex multiplicity of roles that assessment plays in students' lives. Finally, several dimensions of washback are identified that speak to the influence of assessment washback on participants. Freirean and Foucauldian theoretical lenses are set against each other to explore the tensions between the emancipatory aspirations of the MYP and the insidious technologies of power in which students are frequently caught up during their schooling.

As a research endeavour in the context of a Professional Doctorate, the study also contributes new understands of the implications, particularly with regard to ethics, of insider/outsider positionality for school leaders conducting research in their own professional setting.

Acknowledgements

I would like to express my thanks to my supervisors, Dr Helen Boulton and Dr Andrew Clapham, for their tireless support as this work has developed. Their guidance, expertise and incisive questions have made all the difference to my development as a researcher; they have taught me to root my decision-making in a coherent ontological and epistemological stance, to employ theory to understand the wider social and political implications of the data, and ultimately to find my voice. For this, and for their encouragement and care, I am extremely grateful!

I am deeply grateful to my parents, Helen and Dennis, and my sister, Becky, for their love and support for my education. My aspiration to work towards a doctoral degree is a product of their belief in me, and their hard work to provide for me and my education growing up. They were my first teachers.

To my wife Elizabeth, and my children, Georgia and Luke, I would like to express my gratitude for their remarkable support over these past few years. Many a weekend and holiday has been utilised to do this work and their patience and understanding have gone well beyond the call of duty. I could not have succeeded without their love and their generosity towards me.

Finally, thanks are due to the participants in this research. The Bayside College community's willingness to support me in this endeavour has been vital, and I hope I have done justice to the impressive, thoughtful young people, and the committed teachers, whose voices are reflected in this thesis.

Contents

1   Context of the Research

This thesis represents the culmination of an ethnographic inquiry into the student experience of assessment at Bayside College, an international school in Hong Kong at which I am the Head of Secondary (the name of the school and of all participants are replaced within pseudonyms throughout). As such, it represents my explorations as a researcher within my own professional context, and into an aspect of the experience of my students which, based on anecdotal evidence from my 12 years of experience as an educator and school leader in Hong Kong, is profoundly impactful and significant. The research aims to illuminate the manner in which students experience, and respond to, assessment, and the role it plays in their learning, and their lives. This chapter will discuss three aspects of the context of the research, drawing from each the relevance of the inquiry as a response to its context:

- The International Baccalaureate (IB) programmes, and specifically the Middle Years Programme (MYP);
- Assessment in Hong Kong;
- Bayside College: its status, governance and the nature of the education it offers.

The starting point for these discussions will be the MYP as it is the programme of study in which the student participants of the research were engaged at the time of data generation, and therefore provides the curricular backdrop for the inquiry. As will be discussed in this section, current debates about the nature and purpose of assessment within the MYP, stemming from broader debates about

the philosophical underpinnings of the programme itself, play themselves out powerfully and influentially in the experiences of students and teachers in MYP classrooms.

1.1 The IB Middle Years Programme: emancipatory aspirations

The philosophical underpinnings of the *IB Middle Years Programme* (MYP) imply the importance of students acting as powerful agents of their own learning. The programme's student-centred approach is reflected in the Programme Model diagram, which features in one of the programme's most important guiding documents, *From Principles Into Practice* (International Baccalaureate Organisation 2014b):
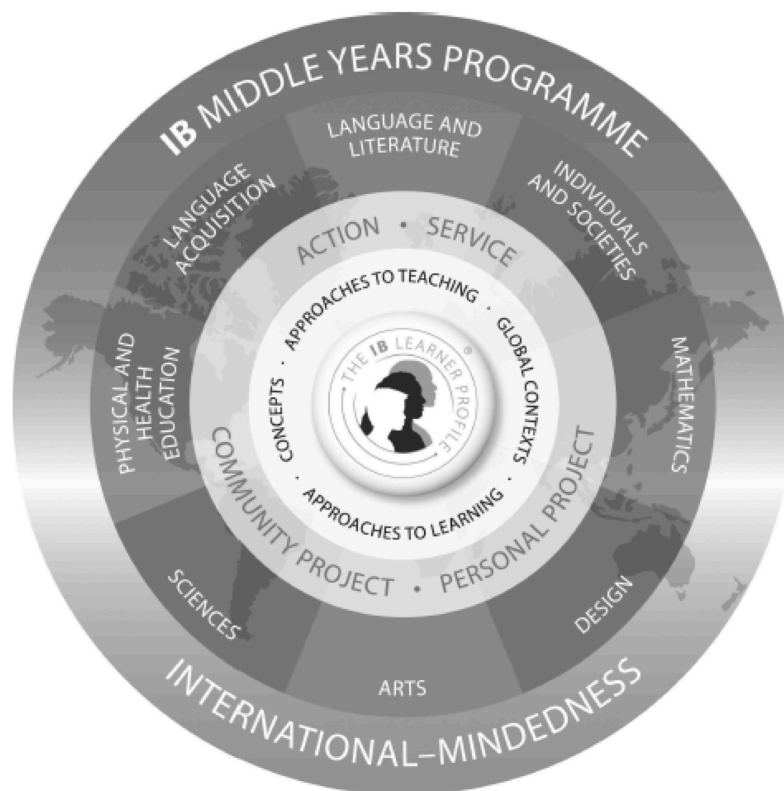
Figure 1 – MYP Programme Model

In this diagram, the student is shown at the heart of the programme, illustrated by the three human figures who represent different stages in a student's growth and development. The implication of this medial positioning is that the student's own role is afforded a primacy, and that all of the elements of the MYP learning experience revolve around her (male and female pronouns will be used interchangeably throughout this thesis when referring to generalised personas such as 'the student', 'the teacher', 'the researcher' etc.).

This aspiration towards an agentic student role is a result of the programme's history, theoretical underpinnings and the educational purpose it serves within the community of over 1,500 schools worldwide that deliver the programme. The MYP was developed during the 1970s and 1980s to fulfill a need for an IB programme in the middle years of secondary schooling that would prepare students effectively for the IB Diploma Programme (DP), a broad academic programme of study taken by students aged 16-18. Prior to the MYP's development, a significant number of DP schools (most of which were international schools and thus operated outside of the auspices of any one national education system) had adopted O-Levels (a British qualification in use between 1951 and 1987, generally taken by students at the age of 16 (Perry 2014)) as a means to provide a curriculum for students prior to the start of the DP (International Baccalaureate Organisation 2010). However, the community of DP schools perceived a "pedagogic disjunction" (ibid, p. 2) between O-Levels and the DP, and wanted to create a programme that placed emphasis on Inquiry as a pedagogy and the creation of an empowered student who had the skills, attitudes and understandings necessary to construct new knowledge from their

experiences. This constructivist learning philosophy is articulated by the

International Baccalaureate Organisation (IBO) in their curriculum materials:

> Teaching and learning in the IB celebrates the many ways people work together to construct meaning and make sense of the world. Through the interplay of asking, doing and thinking, this constructivist approach leads towards open, democratic classrooms. An IB education empowers young people for a lifetime of learning...

> (International Baccalaureate Organisation 2015, p. 4)

The allusion to democracy and the verb "empowers" both convey the importance

of power as a key theme in the IB programmes and the philosophical approach

they represent. Students are intended to be empowered, and active in

constructing meaning as they explore the world around them. In this way, the

IBO sought to address the "pedagogic disjunction" by developing a programme

that would facilitate effective transition to the DP, and indeed prepare students

for later life, by creating powerful learners. As the MYP was being developed

along these lines, the then-Director General of the IBO articulated the direction

the programme was taking in terms that made specific reference to students'

"powers":

> What matters not is the absorption and regurgitation either of fact or of pre-digested interpretations of facts, but the development of *powers* of the mind or ways of thinking which can be applied to new situations of facts as they arise.

> (Peterson 2003 - emphasis added)

1.2   Freire as a theoretical lens for understanding Emancipatory Education in the MYP

Peterson's words, and the approach to learning represented by the MYP, can be

understood in a theoretical context through the work of Freire (1970). Freire

offers a critique of what he terms the "banking model" of education, in which:

> ...knowledge is a gift bestowed by those who consider themselves
> knowledgeable upon those whom they consider to know nothing.
> Projecting an absolute ignorance onto others, a characteristic of the
> ideology of oppression, negates education and knowledge as processes of
> inquiry.
>
> <div align="right">(ibid., Location 986)</div>

In the banking model, students are seen as subject to a relationship of differential power because they lack knowledge, or the knowledge they do have is seen as lacking worth. Peterson's rejection of "pre-digested interpretations of facts" echoes Freire's concept of "cultural invasion", whereby the values of the powerful become a paradigm for those who lack power:

> For cultural invasion to succeed, it is essential that those invaded become
> convinced of their intrinsic inferiority. Since everything has its opposite, if
> those who are invaded consider themselves inferior, they must
> necessarily recognize the superiority of the invaders. The values of the
> latter thereby become the pattern for the former.
>
> <div align="right">(ibid., Location 2298)</div>

The resolution of the problem of cultural invasion therefore requires the resolution of the problem of the differential power relationship, whereby the student and teacher engage in a process of joint inquiry. The object of this joint inquiry is, in Freire's view, that the power that might otherwise be held *over* students is negated in favour of what Freire terms a "humanizing pedagogy", in which:

> ...the method ceases to be an instrument by which the teachers... can
> manipulate the students... because it expresses the consciousness of the
> students themselves.
>
> <div align="right">(ibid., Location 905)</div>

For this reason, inquiry as a pedagogy is a necessary foundation for the type of emancipatory education towards which Freire's work directs us. He sees the necessity of an altered teacher-student relationship in order to re-imagine learning as a process of "problem-posing" in which teacher and student examine

together the significance of their interactions with the world, and form

knowledge together based on an awareness of their own "incompletion" (ibid.,

Location 1174). Inquiry overturns the traditional power differential of the

teacher-student relationship because it recasts the teacher as working *with*,

rather than *on,* students:

> The teacher is no longer merely the-one-who-teaches, but one who is himself taught in dialogue with the students, who in turn while being taught also teach. They become jointly responsible for a process in which all grow.
>
> (ibid., Location 1110)

Seen through a Freirean lens, we might therefore see teaching in the MYP as

implying a different set of professional actions and relationships to those

required in other systems. Inquiry becomes a pedagogic tool of paramount

importance, respecting students' prior knowledge and placing them at the heart

of the learning process; and consequent to this is the shaping of the role of the

teacher to become that of a learner alongside students. This stance is reflected in

the MYP's guiding document, *From principles into practice*:

> Inquiry is a central idea in IB approaches to teaching… With inquiry there is a greater focus on students starting from a position of knowledge… and there is a reduced emphasis on the teacher being the keeper and transmitter of knowledge… [A] collaborative process of creating knowledge takes place in a learning community, as recognized in constructivist pedagogy.
>
> (International Baccalaureate Organisation 2014b, p. 73).

In realising Freire's aspiration for an "emancipatory" education, therefore, the

MYP seeks to cast students as powerful, active, and agentic, engaging with their

teachers in a process of joint inquiry in which knowledge is constructed from

their new understandings of their relations with the world.

1.3 Challenges to achieving emancipatory education through MYP assessment

There are, however, a number of factors that problematise the fulfillment of this aspiration in the student experience of the MYP at Bayside College, Hong Kong, particularly in the area of assessment. Assessment in the MYP is currently a source of debate and discussion as the community of MYP schools continues to navigate the challenges of implementing a broad range of programme reforms, collectively known as 'MYP Next Chapter' (International Baccalaureate Organisation 2013), that were implemented in 2014. Among other measures, one of the most significant changes made to the programme through 'MYP Next Chapter' was the introduction of a high-stakes examination termed 'eAssessment', and completion of this examination has now become a pre-requisite for students to receive the official MYP Certificate at the end of their final year of the programme (International Baccalaureate Organisation 2014a, p. 7). As a result, while schools are not required to take part in eAssessment, if they do not, they bear the consequence of their students being ineligible for the MYP Certificate. The certificate having previously been available to all MYP schools, the move to make it conditional on schools taking part in the new assessment regime is arguably a punitive measure designed to discourage non-participation. As such, it is reminiscent of Ball's observation of the "new set of incentives and disciplines" that accompanies the policy technologies of global education reform (Ball 2013b, p. 48). As the IBO moves to increase the prevalence and significance of high-stakes measurement tools within its programmes, it does so against a global policy backdrop in which similar moves are being made by governmental and international policy-making institutions (Sahlberg 2011). Such policy

technologies often bear noticeable similarities to each other as they utilise systems of incentive and punishment designed to exert pressure on schools to align with the intended reform, representing what Ball terms a policy "bricolage" in which ideas from different state and international actors are "cannibalis[ed]" and disseminated beyond their originally-intended scope (Ball 2013b, p. 34). The use of the MYP Certificate as a policy tool is arguably an example of this; one of the consequences of adopting eAssessment is that it delineates a list of subjects that 'count' towards the MYP Certificate (International Baccalaureate Organisation 2014a), effectively turning subject groups such as the Arts into 'optional' subjects in the final year of the programme (at Bayside College, Year 11). This bears a striking resemblance to the "Progress 8" measure implemented in the UK in relation to how school league tables are calculated based on the results of GCSE examinations (a qualification taken by students in Year 11), which has applied a similar subject delineation (Great Britain. Department for Education, 2018). The resulting pressure on schools to align their students' subject choices to the intentions of the reform has been attributed to a drop in participation in Arts subjects since its introduction in 2016 (Johnes 2017); while no research is yet available into the effects of eAssessment adoption on the curriculum offer in MYP schools, it would seem reasonable to be concerned that similar pressures might result in similar effects.

In 2015 Bayside College, amid such pressures, elected not to pursue eAssessment as it "does not align with the values of the school in providing a holistic, balanced education" (Bayside College 2015), but the debate that was held in reaching this view among the members of the College's governing body

reflects a wider debate across the community of MYP schools about what the

purpose and outcomes of assessment in the MYP are or should be, particularly in

the fifth and final year of the programme where eAssessment takes place. The

arguments in favour of eAssessment have varied over time. An early rationale

given by the Head of MYP Assessment, Gareth Hegarty, revolved around

expanding the programme's reach to include schools and school systems where

the presence of an accredited external examination for students aged 15-16 was

a systemic requirement (Hegarty 2015). However, since the first introduction of

eAssessment the IBO's language has been focused on what it perceives to be the

value of eAssessment in providing tools for measurement and evaluation. This

position is summarised by the IBO on its website:

> ...the examination process will be an engaging, positive experience for
> students... eAssessment will provide a rigorous and standardized
> summative assessment to motivate teaching and learning... it is perfect
> for... school districts to measure programme impact... With eAssessment,
> schools have the opportunity to consider students' learning against clear
> standards of achievement for this age group.

<div align="right">(International Baccalaureate Organisation 2017)</div>

It is arguable that these assertions constitute speculative claims rather than

evidenced findings, there being a dearth of research conducted into the student

experience at eAssessment schools (Hegarty 2017). The presumption that high-

stakes assessment possesses inherent value to MYP schools, students and

"districts" by virtue of its capacity to provide evaluation, comparison and

judgment locates the reform within a neo-liberal discourse of performativity; it

emphasises the role of assessment as measurement. Ball defines performativity

as:

> ...a technology, a culture and a mode of regulation that employs
> judgments, comparisons and displays as means of incentive, control,

attrition and change… The performances (of individual subjects or organizations) serve as measures of productivity or output, or displays of 'quality' or 'moments' of promotion or inspection. As such they stand for, encapsulate or represent the worth, quality or value of an individual or organization within a field of judgment.

(Ball 2003, p. 216)

The debate among the community of MYP schools about the adoption, or rejection, of eAssessment can therefore be seen as a debate over the role and purpose of assessment, setting its perceived functions of judgment and comparison against the wider approach to learning espoused by the programme; the latter emphasises the student as a powerful agent of their own learning, whereas the former sees students as subject to measurement and comparison, their "worth, quality or value" encapsulated by judgments made about them by others. As will be discussed below, this dichotomy represents a conflict between the Freirean emancipatory aspirations of the MYP and a Foucauldian conceptualisation of assessment that possesses the insidious characteristics of surveillance and normalising judgment (Foucault 1977). While the debate between these two visions of educational assessment in the MYP represents a tension of policy which is being played out at the level of programme requirements and structures, students are far from immune to the implications of such policy considerations, and policy on the 'macro' level can play a significant role in the student experience of assessment on the 'micro' level (Silfver, Sjöberg & Bagger 2016, p. 14).

1.4 Foucault as a theoretical lens for understanding tensions of power within educational assessment

Furthermore, this tension is not limited to the current debate about MYP eAssessment but may apply more broadly to conceptions and definitions of

assessment and its role within schooling. Wiliam links educational assessment to the concept of "decision-making":

> At its most general, an educational assessment is a procedure for eliciting evidence that can assist in educational decision-making.
>
> (1994, p. 5)

While this definition does not contain an extant discussion of power, the work of Foucault is instructive in helping to examine the latent power relationships that emerge as part of the act of "decision-making" based on "evidence" elicited by educational assessment. Foucault sees the two elements of educational assessment indicated by Wiliam above ("evidence" and "decision-making") as two "technologies of power" that are combined in order to "surveil":

> The examination combines the techniques of an observing hierarchy and those of a normalizing judgement. It is a normalizing gaze, a surveillance that makes it possible to qualify, to classify and to punish. It establishes over individuals a visibility through which one differentiates them and judges them.
>
> (1977, p. 184)

The basis of the power relationship between students and the examination as a form of "surveillance" is the students' own knowledge of their visibility, and the influence of this knowledge on what they do; for Foucault, the "chief function of the disciplinary power is to train" (ibid., p. 170), and the outcome of this "train[ing]" is a homogenising discipline that emerges from students' awareness that, through assessment, they are being compared, selected, classified, and in some cases, excluded (ibid., p. 182). Foucault utilises Bentham's *Panopticon* in order to explore the mechanisms of surveillance through which technologies of power such as the examination do this insidious work; the comparison he draws is that, as with the functioning of the Panopticon's central tower, the examination can be positioned so as to provide visibility on every student in a given

population. Further, it is a technology whose operation does not depend on a single operator, but rather, an examination can take the form of a standardised "procedure", to use Wiliam's term, which can be applied uniformly regardless of who might be administering it. Consequently, "visibility is a trap" (ibid., p. 199). The power exerted by assessment, as with that exerted within the Panopticon, exists independently of the human participants that might initiate it. Students merely have to know, or think they know, of its surveilling function in order for this normalising power to come into effect. The consequence of this is that:

> ...the surveillance is permanent in its effects, even if it is discontinuous in its action; that the perfection of power should tend to render its actual exercise unnecessary.
>
> (ibid., p. 200)

It is this permanence of effect that constitutes the most insidious implication of assessment as a technology of power; and it results from the combination of the two elements of Wiliam's definition above, the elicitation of "evidence" about students and the act of "decision-making" that emerges from this evidence (Wiliam 1994, p. 5). As Ball, applying Foucault's work to the context of educational assessment (Ball 2013a), finds, the result of the permanence of the surveilling effect of assessment is that in being constant, it takes on the quality of being hidden in plain sight:

> The learner is made visible, but power is rendered invisible, and the learner sees only the tasks and the tests which they must undertake.
>
> (ibid., Location 738).

It may be, therefore, that relations of differential power are inherent in the functioning of educational assessment. Consequently, a Foucauldian interpretation therefore leads us to question what the role of assessment can, or cannot, be within the aspiration to "emancipatory education" offered by the

work of Freire, and that is reflected in the ideological underpinnings of the MYP. This question helps to account for the tensions observed earlier in this chapter, between competing arguments for the purpose of assessment and the role it should, or should not, play in the experience of students. In response, this research wishes to examine the student experience of the MYP, to seek an understanding of how assessment influences this experience and posing the following overarching research question:

> **To what extent does educational assessment act as**
> **a means of student emancipation?**

1.5   Assessment in Hong Kong

Aspects of the professional context of the research suggest the importance of examining the student experience of assessment in the setting of an international school in Hong Kong. Green (2006) asserts that the influence of assessment on students and their experience of learning is likely to be greatest in contexts where students themselves attribute the greatest importance and significance to the potential implications of assessment. Hong Kong is one such context; Berry sees Hong Kong as possessing an "examination-oriented culture" (2011, p. 199), as a consequence of the combined legacies of British colonial government education policy and what Choi characterises as the "pervasive tenet of Chinese culture that academic credentials are superior to other qualifications" (Choi 1999, p. 405). Berry sees the influence of the former as having emphasised the role of assessment as a means of selection, with the colonial government

constructing a "pyramid education system" (Berry 2011, p. 201). The functioning of this system was to identify members of the local population who could be considered worthy by the colonial administration as potential employees. The consequence of such policies is to have created a culture of high-stakes assessment in which future success and prosperity is seen as being dependent on achieving success in examinations, and thereby being granted access, as part of a select few, to career opportunities.

While being in part a product of the history of colonial government policy, in the view of Zhao (2014) this sense of importance attributed to examinations coincides also with the legacy in Chinese culture of the Keju (科舉) system of examination. The Keju was a formal examination used in dynastic China from as early as AD605 in order to select civil servants to work in the Imperial Court, and consequently came to be seen, as with the three-tier education system in colonial Hong Kong, as a singular route through which prosperity and influence might be achieved by ordinary people. Zhao (2014) charts the influence of the Keju over the history of Chinese education policy, seeing it as the progenitor of the modern National Higher Education Entrance Examination, popularly known as the Gaokao (高考), a state-wide University Entrance examination system which is taken by over seven million school-age students in China each year and which is the most significant determining factor in gaining admission to a Higher Education institution (Jiang 2017). Zhao argues that the Keju and the Gaokao are historical manifestations of the significance placed by Chinese on examinations as a means of selection for future opportunity, and that this significance leads to what he describes as a "Prisoner's Dilemma" scenario, in which students, parents

and teachers, fearful of being outperformed in assessments, choose to allocate resources to schooling, textbooks, tutoring, test preparation activities and other means supposed to secure an advantage in assessments, often at the expense of wealth, health and happiness (Zhao 2014, Location 3331).

Consequent to these cultural factors is Berry's observation of the importance of assessment to the educational climate in Hong Kong. As noted above, this observation aligns with anecdotal evidence from my own professional experiences. As Head of Secondary at Bayside College, Hong Kong, and during my 12 years as an international educator and school leader in the city, I have seen the important influencing role that assessment seems to play in my students' lives. At times it seems to have shaped their decision-making, causing them to do things that they might not otherwise do in the cause of achieving success in assessment. At times I have observed students' emotional responses in reaction to assessment experiences, and at times I have felt it to colour their interpretations of their experiences of school. As noted above, one possible explanation for the powerful role that assessment appears to play in the student experience in Hong Kong is that given by Cheng's concept of "washback intensity" (Cheng 1997). Cheng asserts that the strength of influence of assessment is experienced differently by students in different contexts according, in part, to the stakes that they attribute to the assessment activities in which they participate. With Hong Kong being an environment in which, as observed above, assessment is often seen as possessing a high-stakes status, the intensity of its influence on the student experience implies that it is very worthy of study.

1.6　The Student Experience at Bayside College

Bayside College's relationship to the above-mentioned aspects of the context of Hong Kong's education culture is a complicated one, as it occupies a liminal positioning between the local and international school sectors. Bayside College is designated as a Private Independent School, a status which affords it a position outside of the local Hong Kong curriculum in that it is able to utilise international curricula such as those provided by the International Baccalaureate Organisation (2015). It utilises the IBO's Primary Years Programme (PYP), Middle Years Programme (MYP), Diploma Programme (DP) and the Career-related Programme (CP) to provide a curriculum to students between the ages of 5 and 18. It also employs a staff that is largely expatriate, and has a student population in which over 50 different nationalities are represented (Bayside College 2018), p. 29). All of the above respects lead to Bayside College regularly being seen as an "International School", for instance in the Hong Kong Education Bureau's online portal for international schools where it lists Bayside College as possessing this status (Hong Kong. Education Bureau 2016).

However, there is no universally agreed definition of an International School (Wickins 2014) and many schools occupying this status have some features that they hold in common with the international sector, and some held in common with their local context (Wickins & Edwards 2016). Wickins and Edwards apply the term "glocal" to this liminal positioning, seeing it as part of a growth area within the international sector fuelled by growing aspirations among local families for access to international schools:

> ...the recent rapid growth of international schools in Asia is not due to an expansion of schools committed... to serving the needs of 'global nomads'. The demand for international schools places now comes predominantly from non-globally mobile families who perceive comparative advantage in an international education.
>
> (ibid., p. 2)

In-keeping with this new "glocal" identity, Bayside College has some features that place it securely within the context of the Hong Kong education system. One of these features is a government requirement, as part of its status as a Private Independent School (PIS), that 70% or more of the students enrolled at the College have Permanent Resident status (Hong Kong. Education Bureau 2016). The implication of this requirement is that, where other schools might recruit from among a more transient population of families who are temporarily resident in Hong Kong for work purposes, Bayside College and other PIS organisations recruit primarily from a population of families with stronger ties to Hong Kong. Indeed, even where Bayside College's students might report their nationality according to the passport they hold, this might not be the only indicator of their geographic ties and it is regularly the case that a student might identify as British, American or Canadian but also possess Chinese cultural, ethnic and family ties.

The consequence of the College's liminal positioning is that it is not possible to speak of students being exposed to one single educational culture; rather, they find themselves at the meeting point of multiple, sometimes competing, cultures of educational thought and practice. As Ball (2013b) notes, these multiple cultural forces are also often bound up with each other, with national

policymaking being influenced by global comparisons and the spread of theory

and practice between different systems – what Ball terms:

> ...a process of Bricolage, a matter of borrowing and copying bits and
> pieces of ideas from elsewhere, drawing on and amending locally tried-
> and-tested approaches, cannibalising theories, research, trends and
> fashions...
>
> (ibid., p. 34)

Consequently, the student experience in schools is subject to the influence of

wider policy implications being played out not just on a national, but also on a

global, scale; this being the case, at Bayside College this policy influence should

be seen in the light of the different national and cultural forces that operate, and

interact, around students to shape their lived experience. This interplay of forces

echoes the work of Carless (2012) who, writing about the implementation of

formative assessment in Hong Kong schools, observes school-level practice as

being subject to the inter-relationship of three different macro-level discourses:

- Current (global) assessment policy and theory;

- Contextual and cultural influences presented by Hong Kong's Confucian
  Heritage setting;

- The city's post-colonial examination-oriented education system

(adapted from Carless 2012, p. 22)

Consequently, as will be discussed in Chapter 3, one of the challenges of research

design in the present study is to identify an approach that maintains a sensitivity

both to the local context of students' lived experiences of their school on a day-

to-day basis, and also to the wider policy structures and cultures that bear upon

what happens around and within their classrooms. By placing a focus on the

relations that exist between participants and Foucault's "technologies of power",

this research sees Ethnography as well-suited to the task of understanding how

students are situated at the nexus of multiple articulations of policy and

educational practice. It shares with Salzinger and Gowan a sense in which the

ethnographic endeavour should, by necessity, be engaged with questions of

power in order to build a trustworthy account of students' experiences:

> For us, to be sociologists at all is to recognize the relational, scalar, power-drenched world of the social and to locate the people and fields we study in those structuring practices, relations, and discourses. In this context, ethnography emerges as a privileged method, one that provides a ringside perch on the processes through which subjects act and are made in the asymmetrical interactions of daily life.

<div align="right">(2018, p. 61)</div>

2 Washback: A review of relevant literature

In studying the influence of assessment on the student experience, this research adopts the term "washback" as a significant underpinning concept. This term is derived from a body of literature on the influence of assessment, with its most pronounced usage occurring in the field of language testing, though the term is employed in a variety of different areas and disciplines. It is also, pertinently to this study, a term used in the research's professional context as it has been employed by the IBO in discussing assessment in the context of the final year of the MYP (Hegarty 2015). Consequently, the available literature on washback will be examined in order to establish a theoretical basis for inquiry into the role played by assessment in the experiences of students. At points in the discussion, links will also be made to the role of ethnographic research in illuminating students' experience of washback, in order to provide a grounding for the approaches adopted in this research.

Washback is defined by Cheng and Curtis as "the influence of testing on teaching and learning" (2004, p. 4). The concept is a broad one and can apply to a variety of forms of assessment influence, as outlined by Alderson and Wall in their seminal article on washback, who break the phenomenon down to specify influence on different elements of the learning environment:

- Teachers and Teaching

- Learners and Learning

- The rate and sequence of learning

- The degree of depth of learning

- Attitudes

- etc.

<div align="right">(Alderson & Wall 1993, pp. 8-9).</div>

Being concerned with the student experience, this study will focus on "washback to the learner", as distinct from "washback to the programme" (Bailey 1996, pp. 263-4), and the following review of literature will therefore discuss aspects of washback insofar as they relate to learner washback.

From the literature I have identified three characteristics of washback to the learner that are relevant to the present study:

- washback is a subjective phenomenon that represents the meeting point of the student's experiences of assessment and aspects of their own identity and perceptions;
- washback is a mediated phenomenon, where students' appreciation of the assessment experiences in which they participate is constructed from their interaction with different aspects of the learning environment;
- washback is multi-dimensional, and consequently can influence many different aspects of the student experience.

2.1   Washback as a subjective phenomenon

The influence of assessment can be said to be a subjective phenomenon insofar as it relies on how students think about their experiences. As Segers and Tillema (2011) point out:

> There is evidence to suggest that not so much the *design* of the
> assessment practice but how students conceive and experience
> assessment influences student learning.
>
> <div align="right">(p. 49 - emphasis original)</div>

This notion is reflected in Cheng's concept of "washback intensity", discussed above (Cheng 1997), though it has long been acknowledged that students' perceptions about the potential importance of assessment condition their responses to it; as early as 1969, Bloom asserted the perceptual nature of the influence of assessment on students:

> Perhaps the main point to be made about the effects of examinations is
> that it is largely a perceptual phenomenon. That is, if students, teachers,
> or administrators believe that the results of an examination are
> important, it matters very little whether this is really true or false – the
> effect is produced by what individuals perceive to be the case.
>
> <div align="right">(Bloom 1969, pp. 44-5)</div>

In the same vein, a number of studies find a link between the importance ascribed to a particular assessment event or activity by students and the apparent extent of its influence on them (see, for instance, Cheng 1997; Alderson 2004; Watanabe 2004; Green 2006; Green 2013; Pan 2014). Zhan and Andrews (2014) examine the possibility of influence due to importance being attributed to a particular assessment activity being connected with students "internalis[ing] the hopes of others" (p. 85). Their study was of three learners studying for a language examination at a university in China, where the examination system had recently been reformed to try to encourage different forms of examination preparation behaviours among students. Zhan and Andrews found that, of the three informants, one adopted the intended examination preparation behaviours much more readily than the other two because she had internalised the aspirations of the examination designers for the type of language learner she saw

herself to be. Consequently, they conclude that this informant had "experienced more of the washback intended by the [examination] designers than the other two informants" (ibid., p. 85). This finding echoes the argument advanced by Cross and Markus (1994), who see the possible selves that students imagine as they participate in assessment as influential in how they respond to them. For Cross and Markus, their study of undergraduates at a US university undertaking a logic test identified and distinguished students who were 'schematic' for the domain featured in the test (problem solving / logic), meaning that they had had past experiences of the domain that led them to believe that they could perform well in the domain and that it bore some importance towards them as a learner. Schematic students, as distinct from their aschematic counterparts, were able to imagine positive possible selves that they project onto their assessment experiences, leading to an ability to maintain higher concentration and performance on a given task over time (ibid., p. 434). The above studies would therefore seem to indicate that students' responses to their assessment experiences are not a product merely of the features of the assessment per se, but of how students *think* about the assessment, its importance, and the possible selves they conceive of in undertaking it. Students' subjectivities are actively engaged in the process of constructing meaning from the act of participating in assessment.

A number of researchers see assessment as possessing a communicative quality. The idea that assessment can constitute a statement of value or worth is a familiar one, as reflected by Ericksen who drew it as a key lesson from an entire career in teaching:

> An examination is a revealing statement by a teacher about what is important in the course. In fact, faculty standards concerning A-grade performance may be the most significant single means by which teachers set the academic values of a college.

> (Ericksen 1983, p. 135, quoted in Crooks 1988, p. 447)

The assumption that assessment can influence members of an educational community to respond to "academic values" underlies the utilization, on a large scale, of assessment as an instrument for educational reform. Indeed, testing is purposely and routinely utilised by politicians, system leaders and others in a variety of different national contexts to bring about change (Shih 2009). In adopting a series of reform measures as part of MYP Next Chapter, discussed above, the IBO's intentions in seeking to change assessment practices can be seen in a similar light.

The idea that assessment influences how students learn falls within a Messickian tradition of washback, which sees the phenomenon as linked to what Messick terms "consequential validity" (1996, p. 2), which evaluates a test's validity according to how it influences educational practice. Messick sees valid assessments as those which bring about "good educational practices", because of the avoidance of two threats to validity – "construct underrepresentation" (the construct which is the intended focus of the assessment is not represented comprehensively in the requirements of the task) and "construct-irrelevant variance" (other constructs, outside of the intended focus of the assessment, are present in the task and allow for high performance without the need to utilise fully the intended construct):

> ...negative washback per se should be associated with the introduction and use of less valid tests and positive washback with the introduction

> and use of more valid tests because construct underrepresentation and construct-irrelevant variance in the test could precipitate bad educational practices while minimizing these threads to validity should facilitate good educational practices.
>
> <div align="right">(ibid., p. 8)</div>

In other words, good assessment will "precipitate" good educational practices because these practices are exclusive and comprehensive requisites of assessment success. It is a similar idea, perhaps, which causes Ericksen to see assessment as an opportunity to communicate "values" (1983). In a similar vein, Sultana (2018) invokes a connection between washback and the concept of "curriculum alignment", appealing for emphasis on assessment practices that are well-matched to broader curriculum goals and that therefore act as "statements of what the system believes students should know and do" (Webb 1997, p. 10, quoted in Sultana 2018, p. 154).

However, the subjective nature of washback means that students do not necessarily respond in a uniform manner to such "statements". Crooks sees the possibility of assessment as issuing "cues" to students about how they might change their learning behaviours so as to maximise the effectiveness of their strategy in seeking success in assessment; however, he finds that some students are "cue seekers" and will readily look for such cues in order to incorporate them into their strategy, while other students are merely "cue conscious", being aware of such cues but not actively seeking them, while others still are "cue deaf" (Crooks 1988, p. 445). Where Crooks discusses "cues", Moss (1998) appeals to the concept of "messages" in discussing the interaction of students' own subjectivities with the communicative elements of their assessment experiences:

> ...the meaning of these messages in local contexts is not a fixed property of the message itself. Rather, it depends on how the individuals draw on the resources available to them in their particular sociohistorical circumstances to understand the messages they receive.
>
> (ibid., p, 7)

The role that subjectivity plays in students' construction of meaning from their assessment experiences challenges the Messickian arguments above, and the consequent logic that suggests that assessment reform could be used to achieve an intended change in educacional practice. At the very least, it seems unlikely that assessment reform could lead to *uniform* or *fully predictable* changes in practice given the complex interaction of students' own identities, histories and contexts with that which assessment might 'communicate' to them.

Added to this complex interaction is the role played by the student's own psychology in constructing meaning from their experiences of assessment. Watanabe notes the "very complex manner" in which "the psychology of test users will be involved in the process" (Watanabe 2008, p. 1), and Fransson likewise finds that students respond differently to tasks based on their levels of anxiety and the degree to which their motivation to perform well in the task is intrinsic (Fransson 1984). In their study of two Language Arts classes in an elementary school in the US, Brookhart and DeVoge identify a link between the students' perceived self-efficacy – that is, how effective they felt they would be on a particular task – and their effort, with a general finding that students who see themselves as more efficacious will generally put in higher levels of effort, regardless of their objective ability, unless the student's past experiences convince them that they are so capable that they do not need to put in much effort in order to be successful (Brookhart & DeVoge 1999).

The meaning that students form from their assessment experiences therefore represents a negotiation between aspects external to the student – the task, the learning environment etc. – and internal aspects such as psychology and past experiences. Washback is a highly complex phenomenon that, Sambell and McDowell argue, forms a "hidden curriculum" in which students are "active in the reconstruction of the messages and meanings of assessment" (Sambell & McDowell 1998, p. 391). The idea of assessment as a hidden curriculum reflects the assertion, above, that assessment can play a powerful role in the student experience but, as Sambell and McDowell argue, this experience will be different between different students because each individual will interpret their experiences differently, and "the outcomes of assessment 'as lived' by students are never entirely predictable" (ibid., p. 401). This claim reinforces the numerous calls that have been made for ethnographic approaches to researching the student experience of assessment (Reay & Wiliam 1999; Wall 2000; Gosa 2004; Watanabe 2004), as it constitutes an endeavour which does not rely on external or pre-determined categorisations or conceptualisations but instead seeks to build an understanding of students' "inner, unobservable views" (Gosa 2004, p. 109).

2.2   Washback as a mediated phenomenon

If washback is a subjective phenomenon, then following Booher-Jennings, it "does not exist *a priori*", but rather is delivered through "subtle features of schooling, such as the way that activities, interactions, and social relationships are structured" (Booher-Jennings 2008, p. 150). Washback is not a property of

the assessment itself (Watanabe 2008, p. 1), and so "does not flow in a straightforward manner either directly from the test or from washback to the teacher" (Green 2013, p. 43), but rather is created through the interaction of students with their environment. It is, as Alderson asserts, "brought about by people in classrooms, not by test developers" (2004, p. xi).

Consequently, it is apt to describe washback as a mediated phenomenon because this interaction of students with their environment is instrumental to their making of meaning from their experiences. Watanabe appeals to the concept of "mediating… factors" to describe aspects of the students' environment that contribute in some way to this meaning-making (2004, p. 22). He distinguishes between "micro-context" factors (as being within the school setting) and "macro-context" factors (as being wider societal factors that exist outside of the students' immediate school environment).

Other research helps to illuminate the different roles that can be played by specific factors; for instance, Tsagari (2009), studying the experiences of Greek EFL learners in preparing for a high-stakes English exam, points to a crucial role for specific exam preparation materials such as textbooks, which can narrow the content that students cover and focus them on certain exam preparation strategies. Zhan and Andrews (2014), in their study mentioned above, speak to the importance of students' own past experiences of assessment as a significant mediating factor in them forming interpretations of their assessment experiences in the present; in studying learners preparing for a University examination in China, they found that informants adopted assessment

preparation strategies they had used in the past for other examinations, even when these strategies were out of place in relation to the tests that the informants now faced, which were dissimilar to those in which the informants had experienced previously (ibid., pp. 80-1). The researchers' conclusion is that students' previous experiences "appeared to mediate [students'] learning methods… to a great extent" (ibid., p. 82).

A mediating factor in students' learning experiences that is among the most regularly observed in the literature is the role played by the teacher (Brookhart & DeVoge 1999; Gosa 2004; Booher-Jennings 2008; Shih 2009; Silfver, Sjöberg & Bagger 2016; Rasooli, Zandi & DeLuca 2018). Gosa, in her diary study of ten students preparing for the Romanian Baccalaureate exam, identifies teachers' teaching styles as a significant factor in influencing how students interpreted their assessment experiences. In particular, she describes an incident in which two non-Romanian teachers who were teaching in the students' programme seemed to have a negative influence over students' perceptions of assessment. She ascribes this negative influence to a cultural mismatch in which the teachers' expectations of what students wanted out of their learning were at odds with students' own perceptions; the students perceived the non-Romanian teachers to lack seriousness through having attempted to bring "fun and games" into learning (Gosa 2004, p. 233). Rasooli, Zandi and DeLuca point to other ways in which students' thoughts towards teachers can influence their response to assessment events, identifying negative students' perceptions of teacher fairness as being linked to hostility, cheating and even truancy (Rasooli, Zandi & DeLuca 2018, p. 164). Shih adds to this in her research on English-language testing at a

Taiwanese University, showing that the interpretations teachers form of the nature and purpose of assessment tasks can, in turn, influence how students see the same tasks. Teachers are, consequently, a vitally important mediating factor in washback.

Part of the teacher's influence stems from the fact that they are responsible to a large extent for the language that is used in the classroom. Silfver, Sjöberg and Bagger (2016) examined a Swedish Primary classroom in the wake of the introduction of new national standardised high-stakes tests for Primary-age students. They found that classroom teachers can influence students' experiences through the construction of linguistic repertoires that start with the words used by a teacher, and are disseminated to become used in due course by students. They identified two such repertoires ("take-your-time" and "do-it-fast"), that were powerful influencers on students' behaviour during test and test preparation activities but that were in fact contradictory, leading to tensions that the students found themselves needing to navigate. This discourse was seen to affect different students in different ways, with some students reacting negatively to the use of a particular repertoire, despite the best intentions of the teacher, as it might create pressure or anxiety (ibid., p. 11). In a different context, Booher-Jennings (2008) identified, in her study of a Texas Primary school, the formative role that teacher language played in classroom discourse connected with male / female student achievement. The differing language adopted by the teacher depending on whether the student being spoken with was male or female demonstrated, in Booher-Jennings', view:

> ...the ways in which the school and its educators unwittingly constructed
> gender by presenting and reinforcing the achievement ideology
> differently to girls and boys.
>
> <div align="right">(ibid., p. 158)</div>

Consequently, the language teachers use should be seen as part of the basis of their role as a significant mediating factor in assessment washback, something which is reflected in the findings of this study as detailed in Chapter 4.

As with the observation above that washback is highly subjective, the fact that washback is also highly mediated points to the importance of an ethnographic approach to researching it (Watanabe 2004). Brookhart and DeVoge point out that the complexity of the assessment environment is not merely a product of the multiplicity of mediating factors that contribute to it, but also a reflection of the complex interaction of these factors:

> ...there are many... classroom assessment events in typical classrooms,
> interlocked and intertwined together. The overall sense of expectations
> that these build up, the meaning or sense that students make out of this
> aspect of their classroom, comprises the classroom assessment
> environment.
>
> <div align="right">(1999, p. 410)</div>

Consequently, to study such an environment is to study a culture – that is, not merely the structures and processes that contribute to how individuals experience a particular context, but the interaction of these structures and processes and the ways in which participants form meanings from this interaction. Ethnography is, by heritage and by nature, well-attuned to the task of studying cultures (Madden 2010) and by engaging with students' own views on their experiences, it is hoped that this study can make sense of how the various mediating factors in their environment interact to form the assessment cultures to which they belong:

If students' perceptions of the learning environment are such an important intervening variable in student learning, students' views may offer us a way forward for improving our educational practice.

(Struyven, Dochy & Janssens 2005, p. 336)

2.3   Washback as multi-dimensional

As discussed above, in their early work on washback Alderson and Wall (1993) sought to challenge the monolithic treatment of the phenomenon that they perceived had dominated research up to that point, by breaking it down into different "hypotheses" of influence. In so doing, they established that, rather than being a singular phenomenon exerting an influence on a single dimension of learning, washback is in fact multi-faceted. This multi-faceted nature is reflected in the diverse claims that are made about washback, and especially about the various aspects of the student experience upon which it is said to have an influence. This section will therefore discuss claims made in the literature related to four different aspects of students' lives:

- Emotions

- Motivation

- Behaviour

- Interpretations

Students' emotional responses to their experiences of assessment are a familiar theme in washback literature, with some studies finding positive emotions connected with assessment. For instance, Brown et al. (2009), in their study of assessment practices in New Zealand secondary schools, identify "Assessment is liked" as one of a series of conceptions of assessment constructed by

participations. However, the majority of studies on students' emotions find that negative emotions are associated with assessment experiences, with stress and anxiety among the most common negative emotions discussed (Fransson 1984; Reay & Wiliam 1999; Scott 2007; Howell 2017; Stenlund, Eklöf & Lyrén 2017). Scott's study of the student experience of high-stakes tests in a UK Primary School illuminated the sometimes-extreme anxiety emotions felt by very young children facing the prospect of a test:

> In the lead-up to the statutory tests, Pa described her anxiety as a 'funny feeling' in her throat, 'rumbles' in her stomach and her legs 'were really badly shaking', but her nerves calmed down once she was concentrating.

(2007, p. 43)

Other studies link stress and anxiety to other factors; for instance, a study of students taking a university selection examination in Sweden by Stenlund, Eklof and Lyrén (2017), which found that low-achieving students experienced markedly higher levels of test anxiety than their more highly-achieving counterparts in the lead-up to the examination. Howell (2017) used students' drawings as a basis for her analysis of the influence of a high-stakes test in a Primary school in Australia, finding that students were capable of inferring the significance of an examination in a way that was completely unintended by the test designers, and that this inference of importance generated significant levels of anxiety and worry among the children taking the test (p. 584). The high stakes of the test in question were ostensibly applicable only at the school level, as the test had been introduced as a method of generating school accountability for student progress and achievement – however, the sense that the students themselves were accountable for their individual results on the test, and that these results could therefore have significant consequences for them, had seeped

38

into their thinking via a variety of channels, including teacher language, parent beliefs and the media (Howell 2017), prompting a set of unintended negative emotions among the participants of the study.

The division between positive and negative outcomes of washback is echoed in much of the literature on motivation in assessment washback. Studies by Fransson (1984), Natriello (1987), Crooks (1988), and Johnson and Crisp (2009), all seek to understand assessment influence in terms of how motivating it is for students to complete assessment tasks, and underlying these studies is the assumption that if students are more motivated to work towards a given assessment outcome, then this constitutes a positive outcome. Johnson and Crisp (2009) focus on the washback effects of "Pre-release Examination Materials", source materials which are released shortly before the commencement of an examination, on student motivation, finding that the point of release of these materials seems to generate a "generally positive motivational affect [sic] on the learners" (p. 56). It is important to note that Johnson and Crisp are employees of the examination board by whom the research was commissioned. However, their findings reflect the general tenor of the studies in this area, that aspects of assessment can be motivating for students.

Some research indicates, however, that the motivational effects of assessment are not straightforward or linear. Fransson utilises the work of Yerkes and Dodson (1908) to explain why the participants in his study of English as a Foreign Language (EFL) university students experienced a fall-off in their motivation at a point where they perceive the level of challenge in an assessment

to become insurmountably high (Fransson 1984). Brookhart and DeVoge make similar findings in discovering that students in their study of a US Primary school who had had sufficient past experience of a particular type of assessment to convince them of their own high level of capability would experience a similar fall-off of motivation (Brookhart & DeVoge, 1999, p. 422). Finally, Harlen and Deakin-Crick (2003), in their review of studies on assessment motivation, identify a number of variables including "self-esteem", "interest", "goal orientation" and "locus of control" as influencing motivation beyond the nature of the assessment task itself (p. 183). The conclusion that can be drawn from the various studies on the relationship between washback and motivation is that where motivation, positive or negative, results from students' experiences of assessment this does not occur in a linear or uniform manner.

Discussion of a positive-negative dimension of washback is found additionally in studies relating to student behaviours and strategy. Studying a group of students preparing for a school-leaving exam in Iran, Damankesh and Babaii utilise a "think-aloud methodology" (2015, p. 64) in which they asked students to articulate what they were doing by way of classroom learning strategies, during the course of their studies. They found that students responded to the requirements of the examination they were sitting by adapting their strategies, including down to the level of the specific skills required by each question item in the exam (ibid., p. 68). The researchers see, in this behaviour, the potential for interpreting students' actions as either positive and negative:

> The findings of the study revealed that high school final examinations had both detrimental and beneficial effects on students' learning behavior. In fact, these examinations negatively influenced learners by directing them

> towards a measurement-driven (learn-to-the-test) approach of learning
> so that they could get well-prepared for the exams.
>
> <div align="right">(ibid., p. 66)</div>

This finding reflects the idea that students behave strategically in response to

assessment, aiming to maximise their chances of successful outcomes (Bailey

1996, pp. 264-5), and is similar to Klein's conclusions that resource decisions,

taken by individuals and at times by whole institutions or systems, can be

understood in terms of a strategic intent to maximise the outcomes of

assessment (Klein 2017). Other studies also seek to distinguish between

strategic behaviours that are seen to be "beneficial", and those that are perceived

to be "detrimental", in Damankesh and Babaii's terms. Struyven, Dochy and

Janssens (2005) note the distinction between "surface" and "deep" strategies:

> ...assessment procedures that are perceived to be 'inappropriate' ones
> tend to encourage surface approaches to learning. This finding suggests
> that a surface approach to learning is easily induced, whereas promoting
> the deep approach seems to be more problematic.
>
> <div align="right">(p. 336)</div>

while Sambell, McDowell and Brown (1997) conclude that some forms of

assessment experience lead to "shallow" learning:

> [The students] frequently believed that the quality of their learning was
> actually polluted or contaminated, because they set out, quite consciously,
> to achieve second-rate or "poor" learning for the purposes of a particular
> assessment point.
>
> <div align="right">(p. 357)</div>

Here, the connotations of the words "polluted" and "contaminated" reinforce the

assumption that different forms of learning behaviour that might result from the

influence of assessment are able to be categorised between positive and negative

poles. The absence of what is being termed "shallow" learning is the absence of

pollution or contamination, implying that the student experience would

otherwise be pristine and unsullied.

However, the idea of an unqualified positive / negative polarisation of the influence of assessment on behaviour is problematised by Elwood, who sees washback in terms of ethics and argues consequently that different ethical schools of thought result in different judgments about the positivity or negativity of the influence of assessment, and that such judgments are therefore contextually situated (Elwood 2013). This argument is echoed by Cheng and Curtis, who assert that:

> ...if the consequences of a particular test for teaching and learning are to be evaluated, the educational context in which the test takes place needs to be fully understood. Whether the washback effect is positive or negative will largely depend on where and how it exists and manifests itself within a particular educational context..."
>
> (2004, p. 11)

Consequently, rather than the influence of assessment on students' learning behaviours being categorised as straightforwardly positive or negative according to whether the behaviours in question are objectively worthy or desirable according to certain educational goals (Bailey 1996, p. 268), Cheng and Curtis see the evaluation of assessment consequences as being bound up with the circumstances of the assessment.

In the same vein, Silfver, Sjöberg and Bagger (2016) interpret the behaviours of students in response to assessment not in light of a positive / negative categorisation, but as a form of positioning in which the students seek to be seen in accordance with the characteristics of an "appropriate test-taker" (p. 12), as this is conceived by their teacher. Their research, discussed briefly above, takes the form of an ethnographic study of a class of 9-10 year old students in Sweden in the lead-up to a high-stakes test, and they find that students adapt their

behaviours to navigate the tensions of the different and competing verbal repertoires ("do-it-fast" and "take-your-time") that the teacher establishes through her classroom language. For example, one student finished a test as quickly as possible so as to be seen to be conforming to the "do-it-fast" repertoire, but then held on to the test paper while the remainder of the exam time elapsed so as not to fall foul of the "take-your-time" repertoire (ibid., p. 11). The researchers interpret his behaviour in light of a pressure to conform to an appearance of being the "appropriate test-taker" in the eyes of the teacher.

A similar analysis is carried out by Lunneblad and Carlsson in their study of testing in Swedish schools (2012). They observe teachers and students taking part in what they term "delusion or pretence", brought about by the official requirements of an assessment task which asks students to match up a poem and a picture and then explain their choice:

> There is no need for the students to say what they mean or to mean anything with what they say; 'All you have to do is just to explain why you chose this poem for this picture'. It is a matter of credibility more than an expression of true feelings about the poem or, rather, a plausible interpretation of the poem *as if* the choice of the poem and the picture built on true feelings, true personal preferences or a personal liking.

(p. 304 - emphasis original)

The behaviours of both the teachers and the students can be understood as a form of "game-playing", which Ball sees as concomitant with a culture of performativity:

> What is produced is a spectacle, or game-playing, or cynical compliance… which is there simply to be seen and judged – a fabrication…

(2003, p. 222)

The outcome of this game-playing is that assessment causes students to adapt their behaviour as a form of *performance*, a fulfillment of an expectation that their actions will conform to a determined set of characteristics that are, to use the term employed by Silfver, Sjöberg and Bagger, above, "appropriate" (2016). Such a conclusion presents a challenge to any aspiration towards emancipation, locating the influence of assessment on behaviour within a performative paradigm in which assessment performance is recognised as *artificial*:

> ...performativity works best when we come to want for ourselves what is wanted from us, when our moral sense of our desires and ourselves are aligned with its pleasures... [the] subject is malleable rather than committed, flexible rather than principled – essentially depthless. A consequence... is for many a growing sense of ontological insecurity; both a loss of a sense of meaning in what we do and of what is important in what we do.
>
> (Ball 2012, p. 31)

As will be discussed in Chapter 4, the idea of assessment as *artifice* finds relevance in the outcomes of this study.

In addition to bearing upon students' emotions, motivation and behaviours, evidence from the literature suggests that assessment experiences can influence students' interpretations of aspects of their learning environment, and of aspects of themselves. For example, in research by Reay and Wiliam (1999) on the influence of the introduction of SATs to UK Primary schools in the late 1990s, the researchers observed how, over the course of the period leading up to the exams, the relationship between one student and his peers was affected negatively because of the changing circumstances of his classroom environment:

> Before he had simply been recognized as clever; now he was increasingly labeled as 'a swot' by both girls and boys. There are frequent entries in the field notes which testify to a growing climate of hostility towards Stuart. For example:

> ...
> Alice: Stuart's such a clever clogs that's why no one likes him
> Diane: But you said you liked him.
> Alice: That's before he started showing off.
> But Stuart had not started to show off. Rather, the classroom practices in 6S over the spring term had dramatically increased processes of differentiation, which in turn had led to a growing polarisation among the peer group. In particular, the relationship between Stuart and the rest of 6S noticeably worsened.
>
> <div align="right">(p. 352)</div>

This finding that students can reinterpret their relationships with other people in their learning environment in light of their experiences of assessment echoes Gosa (2004), discussed above. In her research on washback in the Romanian Baccalaureate exam she observed that the relationship between participants in her study and their teachers also worsened during the course of an examination preparation period as students re-evaluated their interpretations of the teachers in light of changing circumstances associated with the assessment (p. 233).

Perhaps the most striking findings associated with assessment influencing students' interpretations concerns the thoughts that students have towards themselves. Assessment experiences have been linked to students' perceived self-efficacy beliefs (Brookhart & DeVoge 1999; Alkharusi et al. 2014), drawing on the work of Bandura (1977). Bandura's definition of a self-efficacy belief as "the conviction that one can successfully execute the behaviour required to produce [a desired] outcome" (p. 193) is linked reciprocally, in his work, to performance:

> Mastery expectations influence performance and are, in turn, altered by the culmulative effects of one's efforts.
>
> <div align="right">(p. 194)</div>

This reciprocity illustrates the point made by Brown and Hirschfeld that students' interpretations made in light of their assessment experiences are subject to a "chicken-and-egg problem" (2008, pp. 13-14); students form interpretations based on the experiences they have but, with washback being a subjective, highly-mediated phenomenon in which students are active in reconstructing the messages of assessment according to their own subjectivities (Sambell & McDowell 1998), in turn their interpretations play a formative role in the way they experience assessment. Consequently, students' interpretations are to be seen as being formed culmulatively over time, in alignment with Zhan and Andrews' point, discussed above, that students' experiences of assessment in the present are significantly influenced by the interpretations they have formed of their experiences in the past (Zhan & Andrews 2014). The consequence of this is, in the view of Huhta, Kalaja & Pitkänen-Huhta, the formation of "interpretative repertoires" that aid students in making sense of their encounters with assessment (2006). Their study, which focused on the experiences of students taking school-leaving examinations in Finland, identified four such repertoires that contributed, among other things, to the manner in which students attributed the causes of success or failure:

- Mr/Mrs Hard Work
- Mr/Mrs Skilled
- Mr/Mrs Lucky
- Mr/Mrs Cool

(adapted from Huhta, Kalaja & Pitkänen-Huhta 2006, p. 334)

Rather than being fixed aspects of students' interpretations of their own selves, the researchers found that students appeared to move fluidly between their various conceptions of assessment self:

> Test-takers appear to be multi-faced and their identities variable and situated in nature. Importantly, their identities are not only variable and changing during the test-taking process but they can also be complex and multi-layered within any given point in the process, which became evident in the way the repertoires changed from one moment to another within one single situation.
>
> <div align="right">(ibid., p. 456)</div>

At its height, the formation of interpretative repertoires is not restricted to aiding students in forming interpretations of assessment experiences, but seeps over into an interpretation of identity. This insidious implication is the subject of further findings by Reay and Wiliam (1999), who observe a "metonymic shift" in which the outcomes of assessment were internalised by the students to the point that they became statements about who the students felt they were (p. 346). One participant, fearful of performing poorly in a forthcoming test, feels that she may score an imagined "Level Zero" and, in anticipation of this, labels herself "a nothing":

> She is an accomplished writer, a gifted dancer and artist and good at problem-solving yet none of those skills make her somebody in her own eyes. Instead she constructs herself as a failure, an academic non-person, by a metonymic shift in which she comes to see herself entirely in terms of the level to which her performance in the SATs is ascribed.
>
> <div align="right">(ibid., p. 346)</div>

This process of internalising the outcomes of assessment as representative of the self is not restricted only to students, but is played out in the words of others within the learning environment as well. In her study, Booher-Jennings attends an event at which the school Principal is speaking to the assembled students about the forthcoming test:

> Principal: I want you to make me a promise. I want everyone who's taking the TAKS test tomorrow, just promise to do your very best. Because it's just going to *show everybody what kind of person you are.*
>
> <div align="right">(2008, p. 153 - emphasis added)</div>

The findings of both studies constitute the power of assessment to influence interpretations of the self. Here we see realised the potential of disciplinary power to work upon the self; in Foucault's words, to:

> ...operat[e] a differentiation that is not one of acts, but of individuals themselves, of their nature, their potentialities, their level or their value. By assessing acts with precision, discipline judges individuals 'in truth'.

(1977, p. 180)

A question posed by this research, therefore, is whether this potential is ineluctably realised in the assessment experiences of students in their final year of the MYP at Bayside College, or whether the power of assessment to seep into interpretations of the self can be resisted, reshaped, or reversed.

2.4   Research Questions

In Chapter 1 an overarching Research Question was introduced, which seeks to address questions about the emancipatory possibilities of assessment in the MYP:

> **To what extent does educational assessment act as a means of student emancipation?**

In light of the emergent theory of washback, articulated above, as a subjective, mediated and multi-dimensional phenomenon, this study seeks to address three further subsidiary questions, which will help to structure the research process around these theoretical considerations:

1. **What meanings and constructions do the students attach to the assessment in which they participate?**

2. **What aspects of the students' learning environment act as mediating influences on the formation of washback effects?**

3. **In what ways does washback influence the student experience?**

3   Research Design

3.1   Ontological and epistemological underpinnings

In order to reflect the complexity of washback as a subjective, mediated, and

multidimensional phenomenon, this research draws on social constructivism in

making its underpinning ontological assumptions. An aspiration to engage with

the complexity of the social world recognises that it is necessary to avoid

"atomiz[ing] phenomena into a restricted number of variables" such that one

may "miss the necessary dynamic interaction of several parts" (Cohen, Manion &

Morrison 2011, Location 2168). Instead, a social constructivist ontology

encourages the researcher to recognise that:

> …individuals seek understanding of the world in which they live and
> work. They develop subjective meanings of their experiences… These
> meanings are varied and multiple, leading the researcher to look for the
> complexity of views rather than narrowing meanings into a few
> categories or ideas… The questions become broad and general so that the
> participants can construct the meaning of a situation, a meaning typically
> forged in discussions or interactions with other persons.

(Cresswell 2002, p. 9)

The necessity to avoid "narrowing meanings into a few categories or ideas" is

demonstrated by the many different dimensions on which washback is said to

operate. Consequently, it is necessary to adopt an approach to research that

remains alive to the voices of people in the field rather than seeking solely to

generate *a priori* theory which is tested through narrow questions determined in

advance. It is partly for these reasons that a number of researchers call for the

adoption of ethnographic approaches to researching learner washback (see, for

instance, Gosa 2004; Watanabe 2004). This study adopts a definition of ethnography which:

> ...involves the researcher participating, overtly or covertly, in people's daily lives for an extended period of time, watching what happens, listening to what is said, and/or asking questions through informal or formal interviews, collecting documents and artefacts – in fact, gathering whatever data are available to throw light on the issues that are the emerging focus of inquiry.

<div align="right">(Hammersley & Atkinson 2007, Location 210)</div>

Being an immersive endeavour, the ethnographic approach seeks to utilise the researcher's own sensitivity to the field as the instrument of data generation in order to build a trustworthy understanding of a culture (Mills & Morton 2013), implying an interpretivist epistemology which, Bryman notes, creates an obligation to "gain access to people's 'common-sense thinking' and hence to interpret their actions and their social world from their point of view" (2012, p. 30).

However, this does not imply that the ethnographer's work is restricted to deductive ways of thinking; indeed, as will be discussed later my status as 'researching practitioner' implies a riding of the boundaries between practice and theory, with each helping to illuminate the former through an iterative process of inquiry and reflection. This iterative approach is familiar to the role of the ethnographer, and is spoken to by Yeo and Dopson in their reflections on a collaboration between two researchers, one of whom was an academic based at a university and the other was an educational practitioner based in a school. Their assertion is that:

...research creates a temporal space for organizational actors to transition between roles – as insiders and outsiders – and symbolically detach themselves from the expectations of their professional roles...

(Yeo & Dopson 2018, p. 24)

The benefit of this "detachment" within the "temporal space" of research is that theoretical perspectives, applied inductively from outside the research field, can create "theoretical rigor" (ibid., p. 4) in placing the interpretations formed by the research in context, while theory can in turn be "reproduced or contested by practice", revealing "hidden critiques of theory which could lead to a wider discovery of theoretical boundaries" (ibid., p. 5). Consequently, a researching practitioner conducting ethnographic research does so by combining deductive and inductive ways of thinking, leading Bryman to conclude of the interpretivist epistemology employed by such research that:

...[the researcher] is not simply laying bare how members of a social group interpret the world around them. The social scientist will almost certainly be aiming to place the interpretations that have been elicited into a social scientific frame. There is a double interpretation going on: the researcher is providing an interpretation of others' interpretations. Indeed, there is a third level of interpretation going on, because the researcher's interpretations have to be further interpreted in terms of the concepts, theories, and literature of a discipline.

(2012, p. 31)

Consequently, the researcher's own interpretations, and the theoretical basis she might employ to assist her in forming them, are vital parts of her role as research instrument. However, in being thus, they also require acknowledgement; it is necessary that the researcher recognises, and utilises, her own roles in the field and interprets her experiences in light of this recognition:

Acknowledging the fact that the ethnographer is the primary tool of research and an active participant in the ethnographic field also means that properly confronting the influence of the ethnographer on research

and representation is an unavoidable precondition of a reliable
ethnographic account.

(Madden 2010, Location 492)

3.2   Positionality and Power

In light of the above, it is important to recognise that insofar as this research is a
study of the assessment culture at Bayside College, Hong Kong, it is also 'about'
my own identity as a researcher, an educational practitioner, a school leader, and
a person. Gray points out that some element of "biographical work" is inherent in
ethnographic research:

> ...the ethnographer is simultaneously involved in biographical work of
> their own because they are part of, and interacting with, the field setting.

(2014, p. 455)

As Head of Secondary at Bayside College, I am simultaneously a researcher and a
practitioner. My responsibility for over 700 students who comprise the
secondary school, encompassing the students involved in this study who were all
in Year 11, the fifth and final year of the MYP, at the point of their participation,
means that I am in a position of power in relation to the study's participants. In
addition to the students involved, the two teachers connected to the study are in
a department for which I have line-management responsibility, meaning that my
role as Headteacher is relevant to my relationship with them as well. Alongside
my formal leadership role within the school, I am – in a more general sense – an
insider to the context; having worked at Bayside College since 2016, and prior to
that having worked at a sister organisation within the same group of schools for
8 years, I possess a detailed knowledge of the organisation, its workings, and its
people.

Hammersley and Atkinson illustrate that the status and privileged knowledge of the insider can be a productive and beneficial aspect of such research, allowing the researcher to "generate creative insight out of this marginal position of being simultaneously insider-outsider" (Hammersley & Atkinson 2007, Location 2528). Further to this, Reyes, in her work studying the experiences of Amerasian women in the Philippines and the US, points out that the aspects of a researcher's identity that are both similar to, and dissimilar from, the identities of other participants can both contribute productively to the research endeavour as they form an "ethnographic toolkit" that guides and assists the generation of knowledge (Reyes 2018).

The researcher's insider-outsider status can also present challenges for the research design and implementation, as pointed out by Cohen, Manion and Morrison who recognise the difficulties of working with the power differential that can result, for instance, from adult researchers working with child participants (2011, Location 9537). However, it is important not to treat power relationships as a monolithic element of the research endeavour; Mayeza, in his study involving school-age students in South Africa, asserts that the power dynamic between adult researcher and child participant can appear differently at different stages of the research, and in different contexts:

> ...the adult-child power relationship is not a monolithic enterprise but rather power needs to be understood as a social force that is fluid and constantly shifting: a process that sees both the adult researcher and the young research participant as capable of holding, exercising, resisting, negotiating, and challenging certain discourses of power during fieldwork.
>
> (Mayeza 2017, p. 4)

This echoes the work of Chavez, who observes the "shifting sands" of positionality that researchers must navigate as they move between the poles of "insider" and "outsider" status (2008, p. 478). Rather than conceiving of my relationship with the research participants in the form of a simplistic power differential, therefore, it may be more apt to regard the multiple roles that are ascribed to a researcher during the conduct of the research:

> In the course of fieldwork, then, people who meet, or hear about, the researcher will cast him or her into certain identities on the basis of 'ascribed characteristics'.

(Hammersley & Atkinson 2007, Location 2265)

Further to this, Brooks, Te Riele and Maguire note that these identity ascriptions will themselves evolve and become reworked over time; power and positionality are not static elements of the research enterprise, but rather are subject to constant flux in view of the complexity of the field as a social situation:

> Different aspects of identities may become foregrounded at different times, and may alter the dynamics of the research. Differences in roles, power and identity will be constantly reworked and renegotiated in the process of doing research.

(Brooks, Te Riele & Maguire 2014, p. 101)

Consequently, the responsibility of the researcher is to navigate the tensions and challenges of an evolving and uncertain positionality while seeking also to maximise the benefits of the identity ascription process to the research and the ethical obligations it conveys upon her (Chavez 2008).

The challenges of doing so are particularly acute in situations where the differing roles that might be ascribed to the researcher by different participants, or by the same participants in different contexts, come into conflict with each other or

create competing obligations. Such was the experience of Vernooij in conducting research on a medical project treating HIV sufferers in Swaziland, where she was simultaneously a doctoral researcher and also a clinician with involvement in the project. She found herself simultaneously adopting a critical stance towards aspects of the project in her research, and also being obliged to communicate positively about, and work positively on, the project in her medical role as she was accountable for its success (Vernooij 2017). Such experiences speak to the challenging decision-making processes that are often concomitant with practitioner research.

That Vernooij's dilemma was, in part, an ethical one speaks to the crucial role of ethics in considerations of positionality during practitioner research, and indeed to the pervasive role of ethics in ethnographic research more generally:

> At every phase of ethnographic research there is an ethical backdrop… The pervasiveness of ethical issues in ethnographic research means that at all stages ethnographers need to be aware of the range of possible consequences of their actions.
>
> (Madden 2010, Location 697)

Accordingly, as this chapter discusses each of the aspects of research design that contribute to the present study, it will also return regularly to issues of ethics to demonstrate how ethical considerations played themselves out at various stages of the research project, and in various different forms.

Indeed, any research endeavour in which the researcher moves between different roles and identities in his own mind, and those of the participants, is an endeavour in which ethical challenges will present themselves (Brooks, Te Riele & Maguire 2014, p. 5). Such challenges may result in part from the participants

possessing an imperfect knowledge (despite the best efforts of the researcher to inform them) of the purpose, nature, progress and implications of the research, meaning that they can be taken aback by how the information they have shared is used. In her ethnographic research in a fishing community in the US, Ellis (1995) describes the surprise of one participant after discovering that some of what she had shared with the researcher had featured in a publication:

> "I thought we was friends, you and me, just talkin'. I didn't think you would put it in no book."
>
> <div align="right">(p. 79)</div>

The closeness the participant and researcher had originally felt towards each other was a productive element of the research, in that it yielded information that contributed to the research outcomes. However, this closeness also invoked a dilemma. "The problem comes", as Ellis explains, "not from being friends with participants but from acting as a friend yet not living up to the obligations of friendship" (2007, p. 10). In the context of this research, Ellis' reflections convey a responsibility to be constantly aware of the ethical implications of the relationships I have with my participants, especially in light of the fact that I share with them membership of the College community alongside my role as a researcher.

A number of these ethical implications relate to how information is used. In particular, the "insider" researcher is so designated because she is already in the research setting in some sense, and her advantage is in bringing to bear her existing knowledge of the field to produce the "creative insights", as Hammersley and Atkinson (2007, Location 2528) put it, that can result. However, in discussing the concept of "Accidental Ethnography", an approach in which

ethnographic accounts can be written up after the event from data derived

through professional practice, Levitan, Carr-Chellman and Carr-Chellman (2017)

elucidate the ethical pitfalls of using what they term "post hoc practitioner data

and experiences". Among these, a significant dilemma is the question of whether

information gathered from the field at a point where consent was very unlikely

to have been sought can legitimately be used in published research (ibid., p. 14).

Similarly, this study has required me to confront the legitimacy of using my prior

knowledge of the field as a tool in the production and dissemination of the

research, and indeed this question is relevant to any research in which the

person conducting it is also in possession of prior knowledge of the field. In

addition to the question of legitimacy, prior knowledge can have implications for

relationships with participants; this is explored by Mansaray who, in studying

the role of Teaching Assistants (TAs) in a UK school setting, returned to an

organisation at which he had been formerly employed as a TA. However, in

returning as a researcher he experienced mistrustful attitudes from participants

who had formerly been his colleagues, in part because they were suspicious that

his new status might cause him to abandon old confidentialities that had been

established at the time of his previous employment (Mansaray 2012). Such

quagmires are further products of the multiple roles ascribed to the researcher

during the course of field work, and the differing and evolving power

relationships that may result.

Insofar as such role ascriptions may not be within the control of the researcher,

the consequent dilemmas that present themselves may become ineluctable in

conducting the research, causing the researcher to need to navigate them as best

he can. However, as discussed above it is not necessarily the case that formal roles or statuses will result in straightforward relationships of differential power; as Brooks, Te Riele and Maguire submit:

> ...positions [of power] cannot always be read off from roles or occupations. For instance, prospective participants such as school students can sometimes exercise power in ways that may be quite unpredictable. In consequence, we argue that what is needed is a reflexive stance towards power relations throughout the research work (and beyond), for these relations are not static.

(2014, p. 159)

The adoption of such a stance is the aspiration of this study; I seek to understand the shifting dynamics of power that exist within and around the field in order that, in the resultant account, I maintain a sensitivity to them and their implications for my research.

3.3   Access, Consent and Anonymity

Prior to the data generation period, I needed to secure access to the field through acquiring the consent of appropriate gatekeepers (Pole & Morrison 2003, p. 26). In the case of the present study, this began with an approach to my line-manager and, thereafter, to the class teachers responsible for teaching the Year 11 MYP English: Language and Literature course to the group of Year 11s who I had identified as possible participants in each of the two phases of the study. There are four classes of Year 11 students taking the course; Bayside College does not set students by ability and so all four classes are mixed ability, and all classes contain both boys and girls and represent a diversity of ethnicities, family and national backgrounds, and language profiles. Given this, for both phases of the

study I chose at random from among these four. The approach made to my line manager, to the two class teachers, and indeed subsequently to the student participants in the form of passive consent via a whole-class visit at the start of the data generation process, included an acknowledgement that ethical clearance had been sought from, and granted by, my university and that all aspects of the research were being conducted in accordance with the ethical guidelines laid out by the British Educational Research Association (2011). A sample participant consent form is included as Appendix A.

In relation to the class teacher in the first study phase, Emma (this name, and all other names used subsequently, are pseudonyms; a table giving a brief overview of all participants is included as Appendix D), I was struck by her response when asked if she would consent to me conducting research with her class:

> Emma: I'd be happy to. It's not the sort of thing that would bother me. I don't do anything differently when someone visits my classroom.

> (Field Note, 14th January 2017)

My initial interpretation was that this was the response of a confident classroom practitioner who does not have any concerns about her classroom being visited by a senior leader at her school who is also present in a researcher capacity. However, on further reflection I saw the possibility that Emma's response was part of a defensive reaction in which she felt obliged to point out, given my status in relation to her, that she had nothing to hide in her classroom practice that would warrant any fear of a visit. If this second interpretation were to be adhered to, it would suggest that Emma felt some degree of pressure to respond in a way she thought appropriate to the exchange between her and I, and by

extension, this might imply that her consent was offered in part because of the power relationship that existed between us. This calls into question the concept of "voluntary informed consent" where a participant feels obliged to give consent as part of a personal obligation towards the researcher, or one informed by a power differential (Brooks, Te Riele & Maguire 2014).

This concern about the veracity of Emma's seeming enthusiasm to participate remained with me throughout. It also caused me to think critically about other participants in the study, and whether their consent may likewise have been informed to some extent by the manner in which they related to me. In particular, I maintained a sense, following Brooks, Te Riele and Maguire, that school-age students may perceive a "pressure" created by a "desire to please", to "win favour" or to behave in a way that is perceived by their adult educators as courteous (ibid., p. 84). In light of this, and mindful of Wax's assertion that the acquisition of consent should be seen as a continuous, rather than static, process, where consent is revisited over time to ensure that participants remain informed and active in offering it (Wax 1982, p. 42), I made use of opportunities where possible to return to the question of consent – for instance, at the start of interviews where I outlined the details of the study and gave participants further opportunities to raise concerns or ask questions.

As part of gaining consent, I offered participants the option of withdrawing at any point prior to data analysis being conducted (British Educational Research Association 2011). I also sought to assure participants that their contributions to the research would be anonymised prior to dissemination; as Gray points out,

"the greatest risk [to research participants] arises at the time of publication" (2014, p. 453), and so, as noted above, all names of participants were replaced with pseudonyms in reporting on the research, as is "considered the norm for the conduct of research" (British Educational Research Association 2011). However, anonymisation is not always an effective means to protect the identities of research participants, particularly in ethnographic research which is capable of accessing very intimate details of participants' lives (Duclos 2019). Of particular concern are scenarios arising in connection with Tolich's concept of "internal confidentiality" (2004), where even if an external readership cannot identify a particular research participant from the details given about them, people from inside their own community can do so based on more detailed knowledge of the person in question. As Tolich contends:

> Internal confidentiality lies below the surface, going unacknowledged in ethical codes. Yet it too has the potential to scuttle both researchers and their informants.
>
> (ibid., p. 101)

In light of this, I maintained a cautious approach to reporting information about participations, disguising details where possible and omitting certain pieces of information that would otherwise have constituted a threat to "internal confidentiality" as described by Tolich, above.

3.4   Phases of the study

The research was divided into two phases, each phase involving one class of Year 11 students. The two phases encompassed two consecutive academic years (2016-17, and 2017-18), meaning that the students in phase 1 of the study are a

year older than those who participated in phase 2, though both groups were in Year 11 during the period of their participation and so all students were aged 15-16 at the time. The two classes had two respective teachers, Emma (Phase 1) and Tom (Phase 2), and were comprised as follows:

Phase	Females	Males	Total
1	12	10	22
2	16	12	26

Table 1 – Class membership by gender

The first phase of research sought to address Research Questions 1 (the meanings and constructions students attach to assessment) and 2 (mediating influences that contribute to assessment washback), as outlined in Chapter 2. The chosen duration of the phase was 12 weeks, to coincide with the teaching of a Unit of Inquiry focusing on a study text, *The Laramie Project* (Kaufman 2010). The Unit of Inquiry is the building block of MYP curriculum design; particularly in light of Research Question 2, on the mediating factors that contribute to washback, I decided that the length of one Unit would be a suitable duration as it would provide insight into a wide variety of different aspects of the learning environment. For instance, all Units of Inquiry feature specific formative assessment strategies designed to help students work towards at least one summative assessment task (International Baccalaureate Organisation 2014d), meaning that a study phase which took in a full Unit would encompass these various elements of learning design.

While phase 1 of the study was successful in this regard, I also felt, as I was concluding my data generation period at the end of the Unit in question, an inclination to see what might happen after the end of the unit. I had a sense from my experiences that students might continue to be influenced by the events of the Unit even after it had concluded and they had moved on to the next one. I saw resonance of this idea in the work of Scott (2007), who saw benefit in introducing a stage of research in her washback study that occurred *shortly after* the conduct of a summative assessment task, to examine how students responded to the task. Consequently, for the second phase of the study I chose to remain in the field for a full school term, encompassing one full Unit of Inquiry and a period at the start of the following Unit, and indeed this additional time proved valuable in terms of gathering data relevant to Research Question 3 (the influence of assessment on the student experience), which was the focus of the second phase.

3.5   Data Generation Tools

Ethnography challenges traditional conceptions of "research methods" as it is regarded more as a "way of seeing" the social world than as a specific method via which it can be accessed (Wolcott 2008). The ethnographer, as the research instrument, plays a vital role in being responsive to what is happening in the field and must deploy appropriate data generation tools so as to build a trustworthy understanding of the culture that is the focus of inquiry. As Fortun contends:

> For [ethnographic research design] to work, without becoming formulaic, students must engage with it as play… One moves through a research design process ready, quick on one's feet… Certitude about what one is doing should not be the goal.
>
> (2009, p. 180)

However, ethnography typically does make use of specific, planned data tools in order to aid the ethnographer in the research. This study uses two such tools to facilitate data generation and inscription from the complex and varied events and utterances experienced in the field:

- Field Notes

- Semi-Structured Interview

Field Notes were used as a means of generating data in connection with visits to the classroom, which occurred 4-5 times per week during the course of both phases of the study. They were also used to generate data from other moments, for instance from incidental conversations with teacher participations in the staffroom at break times. I chose an open format for the Field Note as this allowed for as much responsiveness as possible to the complex and varied events that took place in the field (Emerson, Fretz & Shaw 2011); consequently, the note format consisted of a small number of fields providing for the acquisition of basic data such as time, date, and location, followed by a large space into which jottings and prose reflections could be written. Following Hammersley and Atkinson (2007, Location 4103), I chose, wherever possible, to prioritise verbatim quotes over paraphrase so as to capture the thoughts of participants in their own words.

It is important to recognise, though, that even the recording of verbatim quotation is, in itself, an act of selection; indeed, the Field Note is necessarily a selective tool, in which the researcher privileges some observations over others in deciding what is to be written down, and how (Wolcott 2008, Location 1884). By choosing which of the participant's words were written down, where to begin the quotation, and where to end it, I found myself exercising my subjectivity in constructing the text of the Field Note. However, in an interpretivist epistemological framework this subjectivity has, in itself, a productive role to play in working towards the research outcomes:

> It does no good to treat subjectivity in note-taking as a 'private problem', rather it is better to engage with the fact that the perspective of the ethnographer, his or her own personal ethnographic gaze, will inevitably shape and form their notes. It is then that one can make more or less from the embedded subjectivity of the notes…
>
> (Madden 2010, Location 2193)

Consequently, in constructing field notes and exercising my "ethnographic gaze" in doing so, I sought to focus on those events, actions and utterances that connected with the research questions posed in order to achieve a greater focus of subject matter through my capacity to "self-consciously document a series of incidents and interactions of the 'same type' and look for regularities or patterns within them" (Emerson, Fretz & Shaw 2011, p. 28).

The productive role of subjectivity in the construction of the Field Note reflects the fact that analysis, rather than being a distinct stage of the ethnographic endeavour, pervades the data generation process and is therefore bound up with the act of inscription (Hammersley & Atkinson 2007, Location 4439). Indeed, as I generated data in each Field Note, as well as seeking to be descriptive in

accounting for what I was seeing and hearing, as an active participant in the field I naturally generated my own responsive thoughts about my experiences, that will have coloured my interpretations at the time. This aligns with the work of Gray, who anticipates such a mix of interpretative and descriptive elements as being expected in the production of Field Notes (2014, p. 439). In the spirit of Madden's call, above, to "engage" with my subjectivity, I felt it important to document these responsive thoughts; as Fine argues, in seeking to avoid a false separation between the researcher and the researched it is necessary to "probe how we are in relation with the contexts we study" (Fine 1994, p. 72). In so doing, I used red font to write asides and interpretative points and questions into the note, so that these were clearly demarcated, as interpretative text, from descriptive jottings and passages (Hammersley & Atkinson 2007, Location 4130; Emerson, Fretz & Shaw 2011, p. 72; Gray 2014, p. 455).

I chose, during the course of classroom visits, to make brief jottings using my laptop, and then to return to the note at the end of the day to write up a fuller account (Emerson, Fretz & Shaw 2011, p. 22). On the one hand, this strategy carries disadvantages in that notes taken at the time may be misremembered or misinterpreted later on, privileging the moment that is more distant in time to the original encounter as determining the final construction of the note; on the other hand, distance from the event may itself be advantageous as it allows for deeper reflection and processing of the events of the day. In balancing these perspectives, the view I took was that writing jottings, but not full prose notes, *in medias res* would allow for "an intense immersion in daily rhythms and ordinary concerns that increases openness to others' way of life (ibid., p. 22), while the

write-up at the end of the day was still close enough to the original point in time to allow for faithful accounts to be created.

The process of creating a jotting during an event or activity needs to be managed carefully as "the stance and act of writing are very visible to, and can influence the quality of… relationships with, those studied" (ibid., p. 23). However, in this regard I am fortunate that Bayside College is a 1:1 laptop environment in which it is normal for both staff and students to use computers incidentally during the course of the day; consequently, my use of a laptop was not likely to have been interpreted as an abnormal or unexpected behaviour (ibid., p. 39).

The use of Field Notes was associated with a number of circumstances in which ethical questions arose that required consideration. One such circumstance occurred in connection with the choice of study text for the class in the first phase of the study. I had heard from other colleagues in the English department, prior to the start of the research, that she intended to choose *The Laramie Project* as the text study for her Year 11 class as she felt the behaviour of some individuals within the yeargroup had exhibited intolerance and disrespect on previous occasions. *The Laramie Project* is a verbatim theatre text that depicts the murder of gay student Matthew Sheppard in Laramie, Wyoming, and discusses issues of homophobia, violence and hate crime.

Emma sought to use the text to generate discussion about these issues, in the hope that it would lead to more respectful behaviour and greater respect for diversity among members of the year group. As the data generation period

began, this reason of Emma's for why the text had been chosen – which she had not shared directly with the students – became relevant to the research. However, I faced a dilemma in that I only knew of Emma's reasons through information that had been gained second-hand, before the commencement of the research and therefore prior to the point at which I had gained Emma's consent to participate. I was therefore unsure about how she would feel if this information were included, particularly because she might have regarded it as sensitive information that could, if it became known, affect her relationships with the students.

This dilemma arose despite my best efforts to inform Emma about the research at the point of her giving consent, as it was not something I had predicted before the start of the research. It reinforces Gray's assertion that deception is "inherent in the ethnographic encounter" (2014, p. 456), as the researcher does not always know what information may be relevant to participants or how much they appreciate the ways in which information they have provided, or that others have provided about them, might be remembered, recorded and used. The dilemma finds resonance in Wiles' discussion of "off-the-record comments" (2012, p. 50), and indeed my choice of response echoes Wiles' suggestion that it is possible to "negotiate with study participants the ways in which they would be willing for… the information they have provided to be used" (ibid.). In light of this possibility, I chose to speak to Emma about why she had chosen *The Laramie Project* and this conversation yielded the same information as had been given to me second-hand several months before, meaning that I now felt I had a mandate to use the information as it had been given to me directly by Emma, in a context

where it was clear to her that I was receiving the information in my role as researcher as part of the study for which she had given her consent to participate. This occurrence acts as a reminder of Wax's comments that consent is a matter that must be revisited over time rather than being regarded as standing in perpetuity (Wax 1982).

Other ethical issues, specifically to do with role conflicts, also arose during the course of the creation of field notes during classroom visits. At the start of one lesson, Tom, the teacher in the second phase of the study, set the class going with some work and then began inviting students to step outside of the classroom briefly so that he could offer them some individual, verbal feedback on a recent task. As he did so, members of the class pointed out that he had not yet distributed the grades from the task to students. Tom revealed that the grades were stored on a spreadsheet on his laptop, and he could not email out the spreadsheet in one go as he did not want students to be able to see each other's grades. Consequently, as he conducted the 1 to 1 conversations with students he asked me if he could email the spreadsheet to me, and have me distribute the relevant grades to each student via email. This request meant that I faced a dilemma; up to now I had been making a concerted effort to establish a relationship with the class that, in the context of my lesson visits, was distinct from the role that I held in relation to them as Headteacher. I felt that to accede to Tom's request would have asserted my previous role as professional educator in the school context, and that this would likely have been detrimental to my efforts in relationship-building. On the other hand, I felt a moral imperative to ensure that the students were provided with their grades in a timely fashion, and

had I not assisted Tom, it was clear that students would have been delayed in receiving them and that this would have been at odds with the school's assessment policy, to which I have the responsibility of ensuring compliance. Consequently, my practitioner and researcher roles came into conflict with each other at this point of decision-making. Such dilemmas are familiar in ethnographic research; for instance, Hill (2009), in her research in a New Zealand Primary School, felt compelled to clean up a mess in a classroom as the teacher was momentarily occupied with another task, even though she also wished to distinguish her role as researcher from that of a teacher in the school. Additionally, Puttick (2017) recounts how he felt ethically obligated to intervene with a school's lesson observation processes where he witnessed an observation outcome that he felt to be unfair. He concludes that failing to act would, in itself, be a form of action, realising that "…not saying anything… would actually have been to 'say' something" (ibid., p. 62). In the end, I felt similarly to Puttick that by refusing Tom's request I would have been actively working against the moral and policy imperatives that I felt to emanate from my professional role. Consequently, I acceded to the request, but sought to generate minimal attention to my function as conveyor of assessment information to the students by keeping the resultant emails curt and neutral and not discussing the grades directly with students. However, this incident, and its alignment with the experiences of Hill (2009) and Puttick (2017), above, serves to illustrate the necessity that the ethnographer may feel to navigate multiple, sometimes competing, roles in the course of research.

As was discussed earlier in the chapter, practitioner researchers may be particularly prone to such conflicts given their existing membership of the communities that are the focus of their work. Clapham and Vickers (2016) made similar observations in their study in which governors of Further Education colleges adopted roles as "governor ethnographers". During their participation in these roles, the governors felt a tension between the qualitative approaches to knowledge that were the basis of their research and the quantitative approaches which dominated college meetings and evaluation activities (ibid., p. 10). The conclusion that can be drawn from their reflections, and those observed above, is that it is "not always the choice of ethnographers themselves as to how they wish to be in the field" (Madden 2010, Location 1520). Facing the reality that aspects of fieldwork will occur outside of the ethnographer's immediate locus of control, Hammersley and Atkinson argue:

> All that can be required of ethnographers is that they take due note of the ethical aspects of their work and make the best judgements they can in the circumstances. They will have to live with the consequences of their actions; and, inevitably, so too will others.

> (2007, Location 6363)

Alongside field notes, semi-structured interviews were used to generate data in relation to the research questions. As noted by Hammersley and Atkinson:

> ...there are distinct advantages in combining participant observation with interviews; in particular, the data from each can be used to illuminate the other.
> (ibid., Location 2913)

The choice of a semi-structured interview format was designed to elicit such advantages, allowing greater freedom to the participants to lead the direction of the interview and providing space for questions and comments that were

responsive not merely to the interview schedule as determined in advance, but also to contemporaneous events that were occurring in the field; for instance, in one lesson from the first phase of the study the teacher had said that the summative assessment task students were completing was "mimicking" (Field Note, 17th March 2017) a task in the IB Diploma English Language and Literature course. One of the student participants, Ana, took part in an interview later the same day, and was able to set this word usage by the teacher in the context of her own perceptions about the purpose of assessment in the Year 11 English course:

> Ana: All the assessments we do in MYP are based on the ones we do in Diploma so it's preparing us, it's showing how the assessment works so you don't get there and you're like clueless and you don't know what to do.

The responsiveness of the interview to events in the field led to the generation of new knowledge between the interviewer and the interviewee; it reflects a conception of the interview within Kvale's metaphor of the "traveller":

> The interviewer-traveller wanders through the landscape and enters into conversations with the people he or she encounters. The traveller explores the many domains of the country... roaming freely around the territory. The interview traveller... walks along with the local inhabitants, asks questions and encourages them to tell their own stories of their lived world.

> (Kvale 2008, Location 619)

In this light, the function of the interview is seen, as is fitting the social constructivist ontological underpinnings of the research, as an opportunity to give primacy to the student experience itself; it aims to "become responsive to what others are concerned about in their own terms" (Emerson, Fretz & Shaw 2011, p. 16). The positionality of the researcher is relevant to the creation of

knowledge in such circumstances, as with the discussion earlier in the chapter,

as the interview:

> ...entails a hierarchical relationship with an asymmetrical power distribution of interviewer and interviewee. It is a one-way dialogue, an instrumental and indirect conversation, where the interviewer upholds a monopoly of interpretation.

(Kvale 2006, p. 484)

However, as pointed out by Vähäsantanen and Saarinen (2013), such power

asymmetries can be a productive and important part of the way that knowledge

is generated from the interview experience, sometimes affording the interviewer

a license to ask questions, by virtue of their dissimilarity of age, experiences etc.,

that would otherwise come across as unauthorized, unexpected, or naïve (p.

496).

 Kvale's "seven stages of an interview inquiry" (Kvale 2008, Location 937) were

used to construct and implement the interview schedule; using stage one,

"thematizing", I identified themes around which questions would be constructed

using the research questions relevant to each phase of the study (for phase one,

questions one and two; for phase two, question three). Themes arose from the

theoretical concepts that had emerged from the literature; for instance, in light of

Zhan and Andrews' assertion that students' responses to assessment can be

mediated by their past experiences (2014), I designed a question to elicit

responses around this theme:

> Tell me about a past experience of assessment that is memorable to you.
>
> - Why is this experience memorable to you?
> - How do you feel this assessment experience has influenced you?
> - How do you feel about this experience now?

The full interview schedules used in each of the respective study phases are included as Appendices B and C.

A total of 10 students were interviewed, 5 in each phase of the study (see Appendix D). Students were identified through "theoretical sampling", where:

> ...the analyst jointly collects, codes, and analyses his data and decides what data to collect next and where to find them, in order to develop his theory as it emerges. This process of data collection is *controlled* by the emerging theory...
>
> (Glaser & Strauss 1967, p. 45)

Accordingly, students were generally invited to interview where their speech or actions in, or outside of, class, indicated the potential applicability of their contribution to the research questions. Regard was also given to ensuring that different axes of identity were reflected in the sample, for instance by drawing on both males and females, with a variety of different ethnicities, family backgrounds and academic profiles. Ability was used as a particular point of emphasis given evidence that suggests the very different experiences of assessment that students of different levels of ability may have (Fransson 1984; Reay & Wiliam 1999; Harlen & Deakin Crick 2003; Struyven, Dochy & Janssens 2005; Booher-Jennings 2008). However, the spread of different ability profiles represented in the interview participants in both study phases, as with the other axes of identity listed above, was not an attempt to be *representative* per se, but

more an aim to capture contrasting experiences and profiles so that the sample did not become dominated by singular perspectives or backgrounds. Specific written consent, separate from the initial consent process at the start of the period of data generation, was sought from both the students and their parents ahead of the interview.

In both phases of the study, a stimulus was used as a basis for initiating the discussion. In the first stage, students were asked to bring along to the interview an image that, in their view, represented their experiences of assessment in the MYP. This technique drew on the work of Howell (2017), who made a similar research design choice in her analysis of the interpretations students made of their assessment experiences in undergoing a high-stakes test in an Australian school, and was conceived particularly with research question 1 (the meanings and constructs students attach to assessment) in mind. In the second phase of study, the approach was changed such that students were requested to write a short diary entry about their experiences of assessment as they were occurring at the time, seeking to address research question 3 in particular (the influence of assessment on the student experience). This choice of diary stimulus was guided by the research of Gosa (2004) and Tsagari (2009), both of whom used diaries in a similar fashion to:

> ...give access to the otherwise unobservable elements of a topic researched: that is, to what goes on in the respondents' minds.

(Gosa 2004, pp. 73-4)

The different forms taken by the respective stimuli in the two phases of the study reflected the different research questions that each phase sought to address, but

both conform to Hammersley and Atkinson's contention that there is value in exploring several different stimuli in the process of co-constructing meaning with the study's participants:

> ...there may be positive advantages to be gained from subjecting people to verbal stimuli different from those prevalent in the settings in which they normally operate. In other words, the 'artificiality' of the interview, when compared with 'normal' events in the setting, may allow us to understand how participants would behave in other circumstances...

> (Hammersley & Atkinson 2007, Location 3087)

The interviews were recorded using audio recording equipment and then transcribed using a laptop. In transcribing the interviews after they had taken place, I maintained an awareness that the act of transcription is also an act of transformation (Kvale 2008, Location 1957) – it results in a new articulation of the events and ideas of the interview, and so requires the researcher to balance an aspiration to capture, in some sense, the utterances of the interviewee with an aspiration to ensure the interview experience becomes sensible in its transcribed form. In order to aid this, I used a series of transcription conventions that made use of familiar typographic tools (these are included as Appendix E), and I made decisions about transcription as I listened, repeatedly, to the interview sound recordings so as to render the resultant transcription as semantically comprehensible as possible. For instance, pauses or stumbles were sometimes omitted or altered so as to preserve units of meaning, and where participants used grammatical structures that came across as nonsensical when rendered in writing, the grammar was altered to make the language intelligible when read while preserving, as much as possible, the prosodic qualities of the oral interview. In taking these decisions, I was exercising a prerogative "to invent

transcription conventions, format, and focus to fit the research purpose"
(Lapadat 2000).

I also opted to send a copy of each transcript to the respective interviewee by email after it was completed, with an invitation for them to make any further comment they wished. None of the interviewees took up this option, but had they done so, the new data would have been treated as an addition to, rather than a revision of, the original transcript. The intention of this step was to give greater power to the interviewee in the process of making meaning from the interview (Mero-Jaffe 2011), echoing the stance of the research in seeking to give primacy to the subjective experiences of the participants in seeking to understand their experiences.

3.6   Analysis

The unique role of the practitioner-researcher is as an explorer of the boundaries between the practices and experiences that make up the lived experiences of the research participants, and a theoretical landscape that can inform it and illuminate it. This role reflects the need for a recursive analytical process that moves backwards and forwards between inductive and deductive modes of knowledge creation (Clarke & Braun 2017). For this reason, thematic analysis was selected as the mode of analysis best placed to achieve this aim (Guest, MacQueen & Namey 2012, p. 10). Data from field notes, interview transcripts, diary entries (phase 2) and students' chosen images (phase 1) were coded according to emergent ideas. From these codes, themes were formed that

identified patterns in the data – Appendix F provides a map of codes and themes

to illustrate this. The formation of themes was guided by the emergent theory of

washback formed from the literature (Braun & Clarke 2006). However,

Hammersley and Atkinson comment on the ways in which theoretical literature

might, or might not, contribute productively to the analytical process:

> …the process of analysis cannot but rely on the existing ideas of the
> ethnographer and those that he or she can get access to in the literature…
> What is important is that these do not take the form of prejudgements,
> forcing interpretations of the data into their mould, but are instead used
> as resources to make sense of the data.

(Hammersley & Atkinson 2007, Location 4567)

Consequently, my approach to the relationship between the data and the

theoretical implications of the literature was to see the latter as a resource for

understanding the former, and additionally I was mindful of Yeo and Dopson's

argument that it is equally possible, and indeed useful, to "[use] practice to

theorize" (Yeo & Dopson 2018, p. 5). In light of this, I found myself moving

backwards and forwards between data and theory in forming the analytical

categories that have helped to shape this research, and this is reflected in some

of the structural choices made in reporting on it. For instance, the four

dimensions of washback that form an organising principle for the discussion of

the outcomes of research question three in the next chapter are also used as the

organising principle for the discussion of washback multi-dimensionality in

section 2.3. In working across both data and theory in this way, I aspired to make

best use of the possibilities of my status as practitioner-researcher, as someone

who is theoretically informed but who also brings a deep, immersive

understanding and experience of the people and places that constitute the field

to bear in creating new knowledge from the ethnographic encounter.

4 Discussion

This chapter presents findings in relation to the three subsidiary research questions posed during the study:

1. **What meanings and constructions do the students attach to the assessment in which they participate?**

2. **What aspects of the students' learning environment act as mediating influences on the formation of washback effects?**

3. **In what ways does washback influence the student experience?**

I have chosen to explore question 2 (on "mediating influences") first, followed by question 1 (on "meanings and constructions"), as what emerged during the course of study was that the meanings participants ascribe to their assessment experiences are intimately bound up with, and in many ways consequent to, the mediating influences that help to create their perceptions. The chapter then finishes with a discussion of question 3. The study's overarching research question, "To what extent does educational assessment act as a means of student emancipation?", is taken up in the final chapter by way of conclusion.

4.1 What aspects of the students' learning environment act as mediating influences on the formation of washback effects?

Echoing the findings of other washback studies (Tsagari 2009), my research identified the teacher as a significant mediating influence on students' experiences of assessment. One of the most powerful roles played by the teacher is in determining what is seen as important or significant within the assessment

culture of the classroom. For instance, both the main teacher participants in this study took actions reflecting an intention to de-emphasise the importance of numerical grades in favour of comments or qualitative feedback. This was reflected in a comment of Emma's to her class while discussing the feedback she had issued from a recent assignment:

Emma: As always, it's the comments we are really interested in…

(Field Note, 17th January 2017)

In a similar fashion, Tom made comments to his class about the respective importance that he wanted them to place in different types and forms of feedback:

Tom: The 'to improve' comments are the 'most important'…

(Field Note, 10th October 2017)

Both teachers' efforts to de-emphasise grades can be seen as part of a wider preoccupation among teachers at Bayside College that students can be too focused on grades. This was illustrated to me in a conversation in the staffroom (Field Note, 17th January 2017) in which a group of three teachers offered anecdotes of situations in which students had, in their view, placed too much emphasis on numerical feedback. Members of the group remarked that forces of culture and parental influence were at work in what they had experienced, seeing that parents' own educational experiences – especially in the Hong Kong context – were likely to have been ones in which grades were seen as highly important. Emma's use of the phrase, "As always…", above implies that her encouragement of students to focus more on comments is an habitual repertoire in her teaching practice, possibly a response to similar assumptions about how students at Bayside College might normally respond to feedback.

For Susan, one of the student interviewees from Emma's class, such repertoires seemed to contribute to an environment that she felt was positive and useful, and different from what she was experiencing in other classrooms:

> Susan: What you're given is the rubric and not your grade, so like we know that the grades are still there and that she just will still be putting it in Gateway and that your parents will still see it, but it's sort of like – the environment is different in terms of what your peers are looking at and what you're looking at.

Emma's actions therefore serve to mediate Susan's experiences of assessment by creating an "environment" that, in her view, reduced the pressure she might otherwise feel:

> Susan: We know she's not looking at the grade, she's looking more at – and like, just everything in English, it's less about the grade and that sort of mentality takes the pressure off.

However, it is notable that despite the effects Susan perceives Emma's actions to have, Susan still recognises, and positions Emma within, a network of different forces that also contribute to this environment. In referring to the school reporting system, Gateway, and the role her parents play in viewing her grades, she places Emma among a series of factors and depicts the interrelationship between these factors as the basis for the assessment "environment".

While Susan credited Emma for reducing "pressure", other students experienced the mediating role of the teacher differently as both Emma and Tom made attempts to de-emphasise grades. In one tense exchange, Tom explained to the class that he was not intending to provide numerical grades on a recent task as he was handing students their work back:

> Jane: Are we actually going to get a grade? Mr Bryant's class didn't, but the other teachers did.

Tom: (Sighs heavily) I haven't given you grades.

(Field Note, 10th October 2017)

At this point, the class responded with visible disappointment and a number of the more vocal students challenged Tom on this decision, arguing that it was unfair in relation to the other classes. Jane challenged the assumed view of students' behaviour in relation to grades:

Jane: Teachers think we only look at the grade, but we don't.

(ibid.)

Eventually, Tom gave way and promised to the class that they would be given a numerical grade to accompany the feedback comments he was issuing to them. However, he attached a caveat to this, that students who wished to receive this grade must email him before lunchtime on the same day to request this. The consequence of this caveat was that a number of students were clearly still resentful of Tom's stance by the end of the lesson, and in conversation with a small group afterwards the students remarked that his withholding of the grade had left them in suspense and feeling anxious about what it would turn out to be:

Nusrat: I want to know what I got wrong!

(ibid.)

Tom's approach to the situation, while sharing some of the qualities of Emma's approach, did in fact result in the opposite outcome to that described by Susan – in effect it generated additional pressure and anxiety. Part of the distinction between the two situations may have been the comparisons that students in Tom's class drew between his stance and that of other teachers on the course, a comparison that was not made at any point in my experiences of Emma's class. The students' response to what they saw as unfairness and unequal treatment

suggests that the teacher's role as a mediating influence should, in line with Susan's comments about Emma, be seen as nested within an institutional and systemic context in which the teacher's actions are contributive, *alongside a number of other significant forces*, to the assessment environment. The teacher has some influence on what aspects and purposes of assessment are seen as most important, but this influence is at times limited by the counteracting influence of other such forces.

This interpretation of the teacher's mediating influence helps to put into context why the student participants often did not draw a clear line between the role they saw assessment playing in their lives, and the role they saw for their teacher. At times, students saw their teacher possessing a pervasive, almost inexplicable gaze into their daily activities as the teacher made judgments about student performance. In his interview, Parth, a member of Emma's class, made the assumption that in the absence of a specific piece of graded work on which to base a Semester grade decision, Emma must have based her judgment on "notic[ing] everyone":

> Parth: …the grade has to be out on Gateway before we get the project actually in for the grade so sometimes I think she just analyses what we do in class and she has to notice everyone I guess.
>
> JS: How do you know that she might be doing that?
>
> Parth: Oh, because… actually I don't know why she might be doing that. I just think that's a possibility.
>
> JS: What has given you that indication?
>
> Parth: Erm… that she kind of checks some of the stuff. She always comes around to the different tables and she asks us, like, what we've done, asks us to read out what we've done and stuff. And what we think about

whatever she's talking about.

This 'all-seeing' interpretation of the teacher's assessment judgments goes some way to explaining the observed effect of the teacher's presence on activity and productivity during lessons – in relation to both Emma and Tom, there were several occasions (of which the below is typical) on which their movements around the room seemed to condition students to show themselves as engaged and responsive:

> As Tom moves around the room, students again appear to change how they engage with the material being studied – in this case, the poster analysis task. They become enlivened when he is near. He is like a plate-spinner – as one plate 'wobbles', he drifts over and it goes back on its stick. As he does so, another plate wobbles. And so on and so forth.

(Field Note, 9th November 2017)

Here, expectations of the task and the expectations of the teacher are elided, with students visibly keen to show themselves as adhering to the requirements of both. Similar to this was an occasion in a lesson of Emma's, on which she asked students to share their views on the different "types of violence" that they had experienced or witnessed:

> The group I am with circles around the question of whether non-physical violence can be considered violence. As they struggle with this question, Nick asks, "What does she (meaning Emma) mean?". Wendy says, "I don't think that's what she means". It seems like there is a hidden construct here – 'identify different types of violence *as defined by my teacher*'.

(Field Note, 6th March 2017 – emphasis original)

The irony of this episode is that even in a task where the students are asked to share their individual views based on experiences personal to them, they look to second-guess their teacher's intended meaning. As it turns out later in the lesson, Emma did very much intend for more abstract and conceptual understandings of "violence" to be considered. This episode, and the series of occasions noted

85

above relating to the influence of the teacher's physical presence on student engagement and positioning, reflect the conclusions of Silfver, Sjöberg and Bagger in their study of washback in a Swedish primary school that students desire to position themselves as "appropriate test-takers" in response to the combined expectations of the teacher and the assessment task in which they are engaged (2016), including by navigating and attempting to resolve any tensions between the two. Students in the present study seemed to draw very few dividing lines (if any) between their experiences of assessment and what they saw the expectations of their teachers to be in relation to assessment, often treating the two as being identical or overlapping.

Students also viewed the teacher as a mediating element of their relationship with what they saw as the "rules" of assessment, relating both to the school's systems of reporting and to the role they see the IB as playing in their learning. For some students, these "rules" are seen as unfair – Luca, who studied previously at another international school in Hong Kong, compares Bayside College's reporting structures unfavourably to his previous school:

> Luca: Here I feel like, sort of, the four criterions that we get assessed in don't really get influenced by what you do in class, or with the other ones, with like effort, organiz[ation]… social… I think those are sort of separated, where I was used to in my old school, it's sort of mixed together, so you can sort of help your grades, how good you do in class.

Luca laments the fact that his class performance seems to count for nothing, in his view, compared with the marking of the summative assessment. He links general class performance with grades given for generic criteria labeled 'Effort", "Organisation" and "Social Skills", which appear on the school report for all subjects, but he and other students (for instance, Chloe, who thinks that "because

it's not a summative assessment some people aren't taking it seriously, or they'll put less effort into it") see these generic criteria as less important than the main subject grade – for Luca this is influenced by the physical layout of the report itself:

> Luca: It is also in a separate table, let's say, on our report card so that just gives me the impression that it's separate and not really included in... the other one [the main subject grade].

Reporting is mediating the student experience of assessment by suggesting to students what is and is not included in the assessment constructs being used to report, and also by communicating particular messages about how reporting decisions are taken – some of which are viewed negatively by students, with the teacher consigned to take part, as suggested by Ana:

> Ana: I feel like the teacher is following the rules more than the emotional part, because some teachers might have a better relationship with some students. Then that will affect the grade. And I think that it's not fair. So the teacher's following some rules instead of going to the emotional part.

> JS: Who makes the rules?

> Ana: Erm, that's a great question. I have no idea! But I think, like, the IB? Yeah. Because – I'm not sure [laughter]. Like, they already have the rubrics or I think [they] make the rubrics.

Ana is seeing the teacher's role as necessarily ignoring aspects of the teacher's own relationship with students in making assessment decisions, relying instead on what Ana sees as the "rules" made by the IB. Her sense of the IB's involvement in the assessment decisions made about her comes from her experience of "the rubrics", by which she refers to the MYP Assessment Criteria for each of the MYP Subject Areas. Luca echoes this sense of teachers following a pre-determined remit in making assessment decisions:

> JS: So are the teachers deciding how students should be assessed?

Luca: No, I think that is, now, I think comes from – that's given to them as well. In how they need to assess the students, and then they just need to follow that and see what the student has written, and then either – like – and then see what strand it basically fits into the best.

So the IB as a mediating influence is seen as a setter of "rules" that teachers "follow". One student, Chloe, even sees the IB as a key agent in the process of assessing students, even where she is describing assessment practices that are being carried out in the school context:

Chloe: The IB is assessing you [laughter]. I mean, yeah, because you're being assessed against the rubric, and the rubric is from the IB, and also the teachers, so it's an interconnected process of assessing things. Yeah! [laughter]

Chloe's comments reinforce the sense in which the different mediating influences that shape students' relationships with their experiences of assessment cannot be delineated neatly from one another but should rather be seen as deeply interconnected and interwoven.

A further way in which the IB is seen by students as influential to their experiences of assessment is through the Diploma Programme (DP). As above, the respective roles played by Emma and Tom as teachers were also bound up with the role played by the DP; for instance, as a matter of routine both teachers made connections, in the way that they spoke to students, between the design of the Year 11 course and what the DP had in store for students in Years 12 and 13. When describing the analytical skills students were expected to develop in relation to the poetry they were studying, Emma positioned these skills as vital both to the writing of "your own poem in a few weeks" and to that which students need in order to be successful "next year for your Paper Ones" (Field

Note, 17th February 2017). Teacher and student participants used a range of

different words or phrases to describe this process, which occurred regularly in

both Tom's and Emma's classes and was also reported by interviewees in

relation to other teachers at Bayside College. Tom used the word "mirroring":

> Tom: Like everything we do in Year 11, this is *mirroring* something you are going to be doing next year.

> (Field Note, 28th November 2017 – emphasis added)

while Emma described the same phenomenon as "mimicking" the features of the

DP (Field Note, 17th March 2017), and Luca described it during an interview as

"bring[ing] stuff from the next two years down to Year 11". Adopting Tom's term,

this "mirroring" – drawing students' attention to the features of their current

experience that connect in some way with their likely or possible experiences in

the DP – was seen in general to be valuable and productive; when asked about it

during an interview, Elise remarked, "I feel like we need this kind of stuff".

Likewise Ana saw it as important in order that students could feel prepared:

> Ana: …all the assessments we do in MYP are based on the ones we do in Diploma so it's preparing us, it's showing how the assessment works so you don't get there and you're like clueless and you don't know what to do.

It appears, then, that one role of the DP as a mediating influence is therefore to

indicate to students a particular importance or significance to their experiences

of assessment in the present. However, for some students this role is of even

greater gravity – the DP puts into perspective for students the question of their

"readiness" to progress to further and higher forms of study:

> JS: What's the point of being assessed in English this year?

> Parth: Erm… to test the skills she's taught us, our analytical skills, our writing skills, and see if we're ready for IB or if we have to go in a different direction.

The prospect of University study also plays a similar mediating role, adding a layer of perspective to students' current experiences of assessment and, in particular, of grading. Celine professes a consciousness, when asked about the significance of assessment in Year 11, that Universities would be able to "see" all her grades, and makes it clear that some of her decisions in relation to how she conducts assessment work are made in light of this. Her appeal to the concept of "seeing" is notable here; it is clear that her assumptions about the visibility of her grades to the as-yet unidentified admissions officer of an as-yet unselected university are conditioning her decisions in the present.

In this respect, her experiences are reminiscent of the Foucauldian concept of surveillance where the effect of the Panopticon is to induce "a state of conscious and permanent visibility that assures the automatic functioning of power" (Foucault 1977, p. 200). Corresponding to the unknown identity of the admissions officer who Celine imagines will "see" her grades, the Panopticon does not depend on any specific individual for its operation but rather is a "machine for creating and sustaining a power relation independent of the person who exercises it" (ibid.). The effect of this on the student experience is to induce a hyper-awareness of the significance of grades in relation to the perceived, though not specifically known or understood, future consideration of a university application process.

Equally significant, and equally unknowable and obscure to student participants in Year 11, is the sense in which any piece of assessment information derived in the present could make a difference to some future threshold or requirement for

university admission. Luca articulates this in relation to a pressure he feels to

"keep his options open" through his performance in assessment tasks:

> Luca: Let's say, if the university requires a certain number of points you
> get, if you don't achieve that you sometimes aren't able to go join that
> university. Sometimes you can still join it but you first need to do an
> exam. And so that, it just makes it maybe a bit harder or limits your
> options. And it's sort of, let's say, the better you do sometimes, it's sort of,
> you have your options, more options stay open.

This sense in which everything is important because anything could potentially

make a difference to a future point of selection or gate-keeping – contributes

powerfully to the 'Arbiter' role played by assessment in the student experience,

discussed in section 4.2 below. It is also one way of accounting for the

perceptions of teachers at Bayside College, as discussed above, that students are

highly focused on grades. The mediating role played by the prospect of future

University study appears to be to create links in students' minds between their

present performance and the extent to which they will have access to

opportunities and "options" in the future.

For some students, this process of creating links between future and present

extends to more general aspects of their experience of school. Susan indicated

during an interview that when she thinks about assessment in general, she is

caused to think about the 'Diploma Centre', the school's exam hall, because this is

where she will sit her final DP Exams in two years time. However, though the DP

is a future prospect for Susan her knowledge of her future experiences in the

Diploma Centre is assumed rather than known:

JS: How do you know what the exam will be like?

Susan: I guess I don't really *know*, but there is this idea of what it *could* be

like.

Susan went on to explain that her sense of the importance of the Diploma Centre as a space symbolizing assessment came in part from past experiences of hers, specifically in sitting other tests in the same space a few years earlier. Likewise, Luca expressed a confident knowledge of what assessment will be like in the Diploma Programme generally because of his older sibling's past experiences:

> Luca: My sister just graduated last year. I know what she went through. And so it's – a lot of me I already know what is gonna happen or should happen and so within like a month you have, I don't know, like, over ten assessments, and sometimes back to back or something, and just being able to sort of start practicing, let's say, to learn if you have assessments close to each other that you practice that for then in the future when it really counts.

So for both Luca and Susan, past experiences intertwine with suppositions about future experiences so as jointly to mediate assessment experiences in the present.

Finally, students' interaction with their peers acts to mediate assessment experiences in a way that is much closer to home. Peers play a role in conditioning students' emotional responses to their assessment experiences, with Parth discussing the influence of his friends in the wake of receiving assessment feedback:

> Parth: Most of the time I'm satisfied, I think, 'cos I feel like I've done the best I could most of the time. Then I talk to my friends and they're like, 'oh we did this answer this way', and if I did it wrong then I'm like, 'oh my god that was such an easy thing I could have got', maybe a sense of regret sometimes. That I could have studied a little more harder.

In one lesson (Field Note, 16th January 2017), the effects of interaction between members of a group of which Parth was a member appeared to be to counteract the assessment environment, discussed above, that Susan saw as "taking the

pressure off" and encouraging students to look beyond the grade. Parth and his friends had just received some feedback from Emma and Parth asked his friend Robert what grade he had got. There was no grade marked on the feedback sheet, as is Emma's customary approach, but different parts of the rubric were highlighted meaning that Robert was being asked to guess at his grade based on where the highlights were. He handed Parth the sheet and said, "5, 6 I think". Parth then looked at his own sheet with a grin and showed Robert that he had got some highlights in the 'top' box, which equated to the mark range '7 – 8'. He asked Robert to say what he thought Parth's grade would be, and when Robert replied rather ungenerously to say that he thought Parth would get a '6', Parth and a third member of the group, Leonard, sniggered. There is clearly a strong sense of competition among the group and by the end of the lesson they were all still discussing their competing theories about what grades each of them would be awarded. Emma's strategy to reduce emphasis on the grade by not revealing it until after work had been done on the non-graded feedback sheets she had produced appeared not to have been effective in relation to this group, suggesting the prominent role that peers play in the way students respond to their assessment experiences.

For other students, the role of peers is also connected with emotions but in a more positive sense. Susan related that she routinely works together with a group of her friends in which they proofread each other's work prior to submission. She saw this as a helpful act of collaboration:

> Susan: It kind of takes the pressure off because you know, oh, so, you're reading other people's work as well so you know... the sort of standard where the assessment lies, and you can base yours off of that standard as

> well. You can say, 'oh so this person actually wrote a lot, so I should maybe write more for mine'... it sort of sets a standard and you kind of – you get more of an idea of how you can do this assessment well.

Alongside the assessment rubric and any information given by the teacher, Susan's peer group was playing an active role in generating her perception of the assessment construct; the group serves to flesh out her sense of what success looks like. This desire to seek the judgment of peers reflected Susan engaging, like Parth, in acts of comparison between her and her friends – when asked why she places store in the "standard" she observes in her friends' work she took the view that they are her academic superiors:

> Susan: Some of them are sort of overachievers, they always aim for, like, really really high grades so – like a 52 out of a 56 would be average for them [laughter]"

This awed view of her friends helps to explain why Susan's act of comparing herself with her peers manifested itself very differently from Parth's. Susan saw herself as academically less capable than the people she was comparing herself with and so, other than in certain areas like "Art" for which she would "trust more in myself to maybe make judgments", she habitually deferred to them. Consequently, the way in which a peer group acts to mediate a student's assessment experiences is highly dependent on the context of the individual student and how they see themselves within the group.

In summary, this study has identified a number of influences that mediate students' experiences of assessment, including:

- the teacher

- school reporting

- the IB

- Anticipated future study contexts, including the DP and University

- Past experiences

- Peers

4.2   What meanings and constructions do the students attach to the assessment in which they participate?

The above mediating influences help to foreground the different meanings that students ascribe to assessment. The manner in which I have come to understand these meanings during the course of carrying out the research is as a series of five metaphors. The use of metaphor in this way is drawn from a body of assessment literature in which metaphors are used to describe different aspects of assessment. For instance, Biggs (1998) uses metaphors of "tree trunks" and "the backside of an elephant" to characterize the relationship between formative and summative assessment (p. 108); Wiliam and Black (1996) see the functions of assessment in terms of a "cycle" (p. 537) and a "spectrum" (p. 540); and Taras (2007) seeks to represent assessment as a "cake" (p. 65). In drawing on metaphor in a similar way, I am aspiring to make use of the power of metaphor in "creating a bridge from old concepts and knowledge to new ones" (Taras 2007, p. 56), relating the meanings ascribed to assessment by the participants of this study to existing concepts in order to "bring [them] alive for the readers in other places and times" (Mills & Morton 2013, Location 277).

The five metaphors identified in my research as representing the meanings and constructions students attach to assessment are:

The Arbiter

The Boss

The Good Angel

The Mirror

The Mask

<u>The Arbiter</u>

With Year 11 marking the end of the MYP and a point of transition into the DP,

with University on the horizon, participants see assessment as performing the

role of arbiter in deciding the future directions they will follow in their learning.

Parth sees assessment as determining what his future study options will look

like:

> JS: How are your grades from this year going to affect you going forward?

> Parth: It could affect, like, the class I go to next year, and it could affect, like, if she told me I'm ready or not ready.

The question of his future study direction is one that Parth knows he himself will

have to answer as part of the process of subject choice, but his thoughts on his

own role versus that of his teachers reveal a tension between a perception of

choice and what he sees as the objective question of whether he is "ready or not":

> Parth: The teachers don't really – when they say you're not ready, they don't really decide what you do because it's still your choice. But they maybe make you rethink… Can you do better? Or should you just drop, because you're not at the level or the standard that you need to be for next year.

Parth's language positions the student not as decider, but rather as interpreter of

the information provided by assessment. When assessment acts as an arbiter, its

judgment is fixed and unchallengeable – if you "drop", it's because "you're not at the level" needed. The student's role is to "rethink" sufficiently to understand this reality.

Luca therefore feels that assessment will be key to him either being permitted to "continue" after the end of Year 11, or not as the case may be:

> Luca: I think for the next few months… I just know if I wanna continue I need to at least get the – what I need so I'm allowed to continue.

When pressed to say what this might mean in practice, Luca is unable to say what it is he feels he "needs" in order to be "allowed to continue", or relate this to any school policy or structure. In fact, there is no such policy that would realistically cause him to leave the school based on his grades in the coming months – but the authority of assessment to determine his future is more ephemeral and obscure than a specific policy or process. It is a hidden power that manifests itself in his own anxieties. Elise is also subject to the pervasive, ephemeral power of anxiety over her future when she articulates the role of assessment as arbiter of her future DP subject choice, and conflates this with the much broader considerations of university and career:

> Elise: Right now you need to be thinking about what subjects you're going to take next year. And if you don't know what you're going to be doing, you don't know what you're going to do in university, you don't know what job you want to do, like me, you're going to feel really lost, you're going to, you know, be very stressed.

Elise's rapid crescendo from DP subject choice all the way to employment in later life is startling in revealing the connections she makes between assessment as arbiter of her imminent study directions and the subsequent re-enactment of this same arbitration process on a grander scale in the future. It echoes

observations made by Reay and Wiliam in their research on the student

experience of assessment in a UK primary school that students, even at a very

young age, connect the role of assessment in the present with the determination

of their future life chances (Reay & Wiliam 1999). For Elise, assessment is not

merely an arbiter for Year 11, but for the foreseeable future; its function as such,

and the assessment information that is derived about Elise, will stay with her.

The Arbiter metaphor depicts assessment as a Foucauldian "technology of

power" that sees the individual "as he may be described, judged, measured,

compared with others… [the description of the individual] is no longer a

monument for future memory, but a document for possible use" (Foucault 1977,

p. 191).

The Boss

This power relationship is regularly seen by participants as going beyond the

singular relationship between the student and the apparatus of assessment.

Others are brought into the power relationship when assessment is considered

in light of the Boss metaphor. For instance, when assessment becomes *necessary*,

the teacher becomes its stooge. In introducing a summative assessment task

during a lesson, Emma undergoes an uncharacteristic linguistic shift:

> The task requires you to…
> The Written Task's objective is to…
> What we want to see is…
> The success criteria explain in a little more detail what the rubric means…
> What is Criterion C asking for?
>
> (Field Note, 17th March 2017)

Throughout her introduction to the task, Emma repeatedly made the task itself, or elements of the task such as the rubric, the subject of her sentences. Where she featured herself as the subject, she expressed this using the plural pronoun, "we", and otherwise she was absent as a subject. The task is seen to "require", to "ask", to possess an "objective". This anthropomorphizing of an assessment task is indicative of the role Emma plays – she delivers news of an entity that has actions, intentions and requirements of its own, and that conditions what everyone in the class – the students and Emma herself – will work on for the next few lessons. The assessment is the Boss.

Students pick up on this power relationship in describing the MYP rubric as having been "given" to the teacher (Luca) whereby the teacher becomes a follower of "rules" (Ana). The students might only have a vague sense of the identity of the Boss who is conditioning the teacher's actions, as in this example where Luca mistakenly believes that the assessment criteria used for Year 11 Reporting were designed by the federation of schools to which Bayside College belongs:

> JS: So when you say they are 'given' [the rubric], who gives it to them?
>
> Luca: I think it's made by [the federation], now in Bayside College. But I don't – that is – I don't know.

But in the Boss metaphor this identity does not matter. When, as Chloe puts it, "it's the assessment that's assessing you", it assumes an anthropomorphic power that reduces others, such as the teacher, to being its agents.

The Good Angel

Despite its role in the power relationships of their learning environment,

students also see assessment as helpful. Its ability to help them is connected with

motivation, as will be discussed in greater depth in section 4.3, below:

> JS: Could you see a situation where you could learn that body of
> knowledge without assessment?

> Chloe: Um, yeah. But then I think it would take a lot more encouraging.
> Like, inner, like self-talk, to achieve the same standard.

Here, Chloe sees assessment as fulfilling a role that what she terms "self-talk"

would otherwise need to fulfill. She alludes here to a dichotomy that the

washback literature recognizes as "intrinsic and extrinsic motivation" (Fransson

1984; Benmansour 1999), seeing assessment as beneficial because it acts as an

extrinsic motivator in scenarios where it would be harder for her to become

intrinsically motivated. In order to do so, she would need to engage in "self-talk",

talk by herself, to herself.

The paradox of the idea of "self-talk" is rooted in the splitting of the self – here a

part of Chloe, which seeks, or recognizes the importance of, motivation towards

a particular learning goal would need to talk to a part of Chloe who may need

motivating. Assessment, in the stead of the former part, performs this role of

"talking" to the recalcitrant or reluctant side of the student in order that they

become motivated to do something.

Hence, Chloe is seeing assessment within the Good Angel metaphor. Assessment

stands in place of students' better selves, and offers an easier, more expedient

route to action by "talking" to students. Parth also articulates the importance of assessment as a mechanism to "force" learning:

> Parth: I think the assessments are useful to help people learn because then they're forced to, to put it all into, put everything they've learned into one project or test.
>
> JS: So it forces you to learn it?
>
> Parth: Yeah. Even if you don't want to. So I think that's useful.

When questioned about his seemingly paradoxical stance that being forced to act is something that could be sought or desired, Parth echoes Chloe's idea of self-talk with his own "inside / outside" dichotomy:

> Parth: You don't want to because sometimes it might just be too much work, but *inside* you know that this *long term* is going to help you. Like, you – nobody really wants to do tests. But they know that it's going to help them. So that's why they will do it.
>
> <div align="right">(emphasis added)</div>

So within the Good Angel metaphor, assessment is representative of the "inside" / "long-term" part of students' own thinking. It embodies the part of them they see as their best selves, regulating the behaviour of parts of their identity that they feel need to be "talked" to.

The Mirror

As detailed in relation to the Arbiter metaphor, above, participants routinely conceive of assessment as objective, valid and not subject to challenge or question. The basis of this conception is a mistrust by students of their own ability to form judgments:

> JS: Do you think we should give you more charge over your learning and your assessment?

> Amy: I don't think, like, the teachers should, because – because, like, doing things ourselves, I don't think we have that good of an indication, an understanding, of what is actually happening to do that. And like, because we always think us ourselves are right, right? So we don't really know what is good or not.

Assessment is therefore positioned as a source of objectivity, in contrast to students' sense of their own subjectivity. Despite the fact that he knows himself to be highly capable in English, Gladwin sees the necessity of assessment as a form of "proof" that he could not otherwise derive by himself:

> Gladwin: I get to prove to myself and have definite proof that I know the subject enough to get a good grade or, whatever grade I get.

The consequence of students' subordination of their own judgments to the authority of assessment is to position assessment as a Mirror, in which that which *seems* or *appears* becomes that which *is*. In the study by Reay and Wiliam discussed above (1999), the researchers observe what they term a "metonymic shift" occurring for one student as a result of her experiences of assessment:

> She is an accomplished writer, a gifted dancer and artist and good at problem-solving yet none of those skills make her somebody in her own eyes. Instead she constructs herself as a failure, an academic non-person, by a metonymic shift in which she comes to see herself entirely in terms of the level to which her performance in the SATs is ascribed.

(p. 346)

The metonymic shift occurs when the student in question expresses her fear that she will "be a nothing" – that is, get a notional Level Zero on the test. The language is important because the student in question has internalised the results of the test as revealing, in her view, a truth about herself which is objective, valid and not subject to challenge or question.

Students in this research use similar language when ascribing meaning to assessment as a mirror of themselves. For instance, in speaking about having had Emma as a teacher for several years and therefore being able to make valid assessment decisions about him, Parth says of Emma that she "knows how I am". In a similar way, Susan, in giving her view of the role of assessment feedback in the learning process, says:

> Susan: I think [feedback on assessment] is really important as well, because you kind of need to *know yourself*.
>
> (emphasis added)

When assessment is seen as a means to gain self-knowledge, it becomes a Mirror for students as they internalise the outcomes of assessment as revealing things about who they are.

The Mask

The first four metaphors discussed above place emphasis on assessment as a technology of power, in which some of the insidious aspects of a Foucauldian interpretation of assessment are realised (Foucault 1977). However, some of the meanings students ascribe to assessment constitute sites for resistance and challenge. The Mask metaphor possesses an antagonistic relationship with The Mirror in that it represents a conscious recognition, on the part of the student, of assessment as *artifice*.

Ana ascribes meaning to assessment that adheres to the Mask metaphor as she describes a past experience of assessment at her school in Brazil, where she grew up before moving to Hong Kong at the age of 15:

> Ana: We were supposed to draw a part of the book we were reading. And I was never good at drawing… so I had to work extra hard on it and I remember I asked help from my parents and friends and family members, and I worked very hard and when I turned it in the teacher looked at me and she was, like, 'that was not what I asked you to do'. And I was like, 'what??'. I worked so hard and the teacher just came and said it was wrong, but then I went home and then I thought about it and then when I looked at my work I was like, 'Damn! That's so good, that's like a masterpiece for me', so after analyzing I was able to see that the things the teacher said was just so superficial compared to the things I was feeling and how proud I was of my own work.

Despite being a painful memory in some ways, this past experience mediates Ana's assessment experiences in the present by opposing the Mirror metaphor:

> Ana: I believe assessments, they give you a grade but it's not, like, what represents you, because you might have learned more than you showed in the assessment.

Consequently, Ana comes to see assessment as a Mask that stands in front of, but in some way obscures, the reality of the individual being assessed. This causes her to doubt the validity of assessment judgments made about her:

> Ana: I feel like the teacher only saw, only, she only saw what she expected and not the things I learned… so she couldn't see, like, all the things behind it.

Ana's doubt about the validity of the teacher's judgment reflects the findings of a study by Booher-Jennings (2008), in which a group of lower-ability boys who had expended a lot of effort in relation to an assessment task but were told by their teachers that they had not performed well responded by doubting the fairness of the policies the teachers had used to assess them (ibid., p. 158). This conception of assessment contributes to the negation of its status as a technology of power because the assessment is seen as inferior to, less valid than, that which sits "behind it". This new conception of assessment, and particularly of the role of the teacher within it, is not a comfortable positionality for Ana and she

struggles to reconcile what she sees in retrospect, recalling her previous

experiences in Brazil, to be two competing views of her work:

> JS: So you look back at that experience and you see failure?

> Ana: Yeah, I see failure because of the things the teacher told me… I still
> have the, the poster I did, and when I look at it I see how – I see two
> things. I see how hard I worked… and then I remember what the teacher
> said, so it's like two perspectives of the same thing.

Here the power of the teacher as a mediating influence is seen in the way that,

despite her convictions about the nature and quality of her work and the

unfairness of the teacher's judgment, Ana still associates the memory with

failure. The power of teachers either to sustain or to damage the confidence of

their students with the words they choose is evident from this vignette.

However, despite the tension of her choice to see assessment as artifice, the

Mask metaphor gives Ana a power to see her own importance within the process

of assessment:

> Ana: It's like there's this photo of a ballerina where you only see the
> performance… and all of the gracious things she does, but you never see
> how hard and how, how much time she spends on practicing… So that's
> how assessments work, so the teacher might be able to see the good
> things but she, like, the teacher won't be able to see how hard you work.

> JS:…So who is it that's assessing you?

> Ana: Erm… I think the teacher and yourself. 'Cos it's a thing about you
> deciding, either it's good or bad. It's more – it's – I think it's more about
> you assessing yourself than the teacher. Yeah.

Within the Mask metaphor lies the potential for students to engage with

assessment on their own terms. By recognizing assessment as artifice, Ana is

able to see her own role as newly empowered and respect her own judgments

alongside those of her teachers. Seen in the context of the "emancipatory"

education of the Freirean imagining, the Mask metaphor is therefore one in which "the teacher's thinking is authenticated only by the authenticity of the students' thinking" (Freire 1970, Location 1063). The recognition of assessment as artifice may therefore offer possibilities for a more humanizing, student-centred approach to learning:

> Education must begin with the solution of the teacher-student contradiction, by reconciling the poles of the contradiction so that both are simultaneously teachers *and* students.

(ibid., Location 986)

THIRD PARTY COPYRIGHT MATERIAL REMOVED

Figure 2 – "photo of a ballerina"

4.3   In what ways does washback influence the student experience?

As noted in the previous chapter, a recursive process of analysis, which drew on both the data from the study and the theoretical literature underpinning it, resulted in the identification of four dimensions of washback – "Interpretations", "Motivation", "Emotions" and "Behaviours" – and these dimensions form the organizing principle for this chapter.

<u>Interpretations of assessment as giving authority to other aspects of the learning environment</u>

There is a reciprocal relationship between the manner in which students form meaning from their experiences of assessment (as discussed in the two sections above) and the way in which these meanings cause students to (re)interpret other aspects of their learning environment accordingly. One example of this emerges in relation to students' thoughts about the MYP subject Criteria for English Language and Literature, which is referred to generally by students and teachers as "the rubric". Significant store is placed in the role of the rubric within the process of assessing student work. Echoing the occasion discussed above on which Emma's classroom language served to anthropomorphise assessment, Chloe takes a similar position in relation to the rubric:

JS: So who is assessing you?

Chloe: The teacher. And the rubric. It's like a little bit of both I think.

Gladwin expresses his positivity towards the use of criterion-based assessment in the MYP, contrasting it with experiences at his previous school in which he studied iGCSEs:

> Gladwin: In Bayside College it's been more criterion-based. And it makes it better to grade – I think it makes the grading more reliable. As the criteria are given to us beforehand so we always know what we need to do in order to achieve a grade. So it's not – it's largely black and white so we know what is expected of us. Less subjective. And in iGCSE it's not as clear of a criterion on what the grading is based on.

Gladwin's comments above reflect the generally positive attitudes of the participants towards the idea of a rubric. However, when pushed to articulate what the stated criteria for MYP Language and Literature actually contain, each of the participants asked struggled to recall or explain it:

> Luca: I think it asks mostly within each criteria, I think in English, let's say, It's how you write it, like what you write. I don't quite know what all four are but it's more like just, it just looks at the assessment paper itself, let's say, and what you did on that.

The criteria in question, according to the relevant MYP subject guide (International Baccalaureate Organisation 2014c), are labeled:

- A: Analyzing
- B: Organizing
- C: Producing Text
- D: Using language

with an example descriptor from strand C being as follows:

> Produce texts that demonstrate insight, imagination and sensitivity while exploring and reflecting critically on new perspectives and ideas arising from personal engagement with the creative process.

> (ibid., pp. 9-11)

Part of the basis of students' difficulties in articulating the criteria from memory may be, as Luca puts it, that they are "written pretty fancy" – he explains that many teachers tend to rewrite the criteria (in IB parlance, a "task-specific

clarification" (International Baccalaureate Organisation 2014b, p. 27)) for students to be able to access more easily. However, despite the absence of any detailed recall of the criteria themselves, it is clear that students *actively reinterpret the criteria* in light of other stimuli. In general participants tended, when asked what assessment constructs they thought the criteria promoted, to lean towards issues of language accuracy and written style:

> Parth: Usually it's like, the language features you use, the vocabulary you use, and the thesaurus and the language features help raise the grade and help improve.

Luca felt similarly, when pressed to comment on what he believed the rubric emphasised, that the crux of the criteria were to do with style and accuracy:

> Luca: I think it's mostly just being able to like show, like, good language… like, how you're able to communicate something that, is it clear… yeah.

While such observations may aptly correspond to the stated descriptors within Criterion D: Using language, the remaining three criteria are largely neglected in participants' thinking. This even extended to one occasion on which students were going through teacher feedback on an assessment task, with a physical copy of the rubric in front of them. When handed her task back, Elise quoted loudly to the group of students on her table from a highlighted section of the assessment rubric pertaining to her work:

> Elise: 'Few errors in spelling'. Yes! [Clenches fists in celebration] That's literally all I care about!

(Field Note, 10th October 2017)

Elise's responses in interview elucidate the store she puts in this particular aspect of her performance – in describing people in her class who she feels are the most capable in English, she specifies certain aspects of language construction as being indicative of high capability:

> Elise: If you've got good spelling, you've got good grammar, if you know what you're talking about then it's obviously going to be a lot easier.

Elise's belief that highly capable students tend to possess high levels of language accuracy and written sophistication may help to shed some light on the basis for participants' tendency to gravitate towards accuracy and style as they reinterpret the rubric. Despite the breadth of the rubric as it is stated in official documentation, her interpretation privileges constructs that she believes a priori to be significant to high academic performance in a broader sense.

Amy's comments are likewise informative on this point, suggesting that she processes the rubric through the lends of what she believes "society needs us to learn". She regards the MYP's learning values, as represented in the rubric, as being similar to what she knows of other systems and programmes:

> Amy: It's like, most school curriculums, like, even they follow different programmes but the knowledge we gain is sort of similar, so I believe it revolves around, the society needs us children to learn to be, to join, like, and be part of society.
>
> JS: What do you mean by 'joining society' in this context?
>
> Amy: I think that society sort of urges every person to be a good citizen and to get a job and to go to university and so on. And so, like, I think the moral views, what the society wants, makes what we learn.

This profound reflection is confirmatory of Ball's comment that, in a culture of performativity,

> …pedagogical and scholarly activities [are reoriented] towards those which are likely to have a positive impact on measurable performance outcomes for the group, for the institution and increasingly for the nation…
>
> (Ball 2012, p. 31)

Herein lies the resolution of the apparent disjuncture between students' sense of the importance of the rubric and their limited knowledge of what it actually

contains. Its actual wording is less important than the *idea* of criteria, leaving space for the rubric itself to be re-interpreted. Wyatt-Smith and Klenowski (2013) reflect this process in their research on the grading practices of Australian teachers; from their findings, the reality of assessment judgment practice is that, even where rubrics are used as a resource by teachers and others making assessment decisions, other forms of "criteria" – "latent criteria" and "meta-criteria" – which are not extant in written rubrics but are more implicit aspects of the psychology of assessment decisions – are highly influential in the way that such judgments are made (Wyatt-Smith & Klenowski 2013). In light of their experiences of assessment, students actively reorient their understand of the formal written assessment criteria around much broader conceptualisations of high performance and in response to the implications of a climate of performativity.

Emotional responses and relationship with the teacher

As well as prompting students to reinterpret aspects of their learning environment such as the MYP criteria, students' experience of assessment can also colour their relationships with other people involved in the learning process. One example of this came in relation to one of the teacher participants, Tom, whose relationship with several students in his class was adversely affected at various moments during the study. One such moment occurred shortly before the deadline for a piece of summative assessment work (the "Further Oral Activity") that the class was asked to complete. As this deadline neared there was an appreciable change in the mood of the classroom:

The atmosphere seems quite altered from the start of the year. Where at the time the teacher seemed to be initiating a lot of the activity, now the students seem to be initiating a lot more. This comes through with questions about the FOA task and the exemplar:
>   Celine: Is there a structure we should follow?
>   Nusrat: Where do we find films for the FOA?
>   Anson: So did this example actually get an 8 in the end?

(Field Note, 12th October 2017)

The deadline in question represented the first significant assessment deadline of the year, and where the atmosphere in earlier lessons had been more jovial and relaxed, students began to express disquiet about how they were being prepared for assessment:

> I ask two students sitting nearby whether they felt that the lessons were different now than at the start of the year, or whether things had stayed roughly the same.
>   Elise: Yes, different. I think he (the teacher) wanted to deliberately hold off on doing things related to assessment. We didn't do any practices or anything, and then all of a sudden we were asked to do an assessment. The other classes have done practice essays, FOAs, loads of things.
>   Jane: He thinks it will stress us out less if we hold off doing assessment stuff, but it stresses us out more. We watch what other classes are doing and then we get really stressed and anxious.

(Field Note, 12th October 2017)

These feelings are similar to the findings of Gosa, discussed in Chapter Two, that assessment experiences influenced the relationship between students in her study and two of their teachers as the teachers' approaches were seen to be poorly aligned to the changing circumstances of students' assessment preparation (Gosa 2004). Here, Elise and Jane reflect a general sense of anxiety and unease in this preparatory lesson, which Elise expanded on when interviewed:

> Elise: I feel like he does help when it's almost time that it's, something's due… but it's very stressful at the same time. Like, you feel really pressured to ask all these questions about you're not – they don't really

help because you're under so much stress already. And if he's not working and helping throughout the whole span of an assessment it's hard.

Elise is one of a number of students who feel that Tom is not being as proactive as the teachers of other Year 11 classes in preparing them for assessment tasks. However, none of this unease had been evident prior to the introduction of the assessment task. As students enter into cognizance of the requirements of their first formal assessment task of the year, their relationship with their teacher is negatively affected and this in turns contributes to a changed classroom atmosphere and in the emotional responses that students demonstrate towards their experiences.

Motivation

The influence of assessment was linked to motivation by each of the interviewees, in the sense that they all saw it is something that made them work harder. As Gladwin says:

> Gladwin: It drives me to work harder... And instead of being, sort of, spaced out I try to finish the work and like try to research more in depth about it. So I can make a – I can create a better piece of work.

This motivational effect was frequently attributed to the sense in which assessment was linked to the future; however, this link was at times only able to be described in vague terms, with the DP being the most-cited example of how assessment was connected with future considerations. Elise expressed this link in terms of "confidence":

> Elise: I guess if you're getting good grades now you're going to feel more confident going into the IB, but obviously if you're getting average or bad grades you're not going to feel confident, you're not going to feel prepared, and you're going to feel like you're just going to fail straight

away, so what's the point of even trying?

However, Elise also reflected in the same interview on the fact that in reality a poor performance in specific assessments in Year 11 would not have any major negative consequences for progression into Year 12 or for other future considerations – leaving a question as to why, if she accepted that the potential negative consequences of poor performance were relatively minimal, she continued to feel pressured by a sense in which assessment was contributing to her future. Celine addressed this point in her interview, in comments that are congruent with the connections made by Elise (discussed above in relation to the Arbiter metaphor) between her Year 11 grades and her future prospects in general:

> Celine: I guess getting a bad grade, it wouldn't take away my house, it wouldn't take away my family, it wouldn't make me lose an arm or something, I wouldn't be affected by it critically but I guess it would affect my overall grade which goes into the system which, you know – and in order to be accepted into a certain university they expect a certain mark, and if you don't meet that mark you don't get accepted, and in order to, I guess, the people tell people in my year that if you want to do well in life you have to get into a good university, and to get into a good university you have to get good grades.

Celine's comments are telling in relation to the tension between perceived stakes and actual stakes – the string of potential consequences, each building successively upon the other, that Celine connects with her experiences of assessment serve to amplify the importance of assessment in the present based on the possibility that, though a current assessment event such as a grade is only one of many contributing factors to someone's life chances in the future, insofar as it is a contributing factor it may at some stage make an incremental difference – and, in Celine's logic, one thing may then lead to another. This amplification effect helps to explain why the connection students feel between their present

experiences of assessment and their future prospects only needs to be a vague

connection in order to exert a power over them.

However, as well as being subject to this power, students also exert a power of

their own in relation to their motivation, evoking the same sense of duality as

was observed above between students as being subject to power and

simultaneously being wielders of it. Ronnie was, when asked during an interview

about what motivated him to work towards assessment outcomes, able to

rationalise the dual purpose of his assessment experiences as being about

evaluating his progress and at the same time being, as he puts it, "for life". Here

he comments on a task the class is working on at the time, which is about

exploring ethical dilemmas:

> Ronnie: This is for life. And for instance if you are found in that situation
> (a moral dilemma of the same type) you have something to relate to, and
> you'll act accordingly. So it doesn't really matter if I am being assessed
> right now or not.

Here Ronnie places himself in a position of power by prioritising a purpose of his

current activities, and therefore a form of motivation to pursue those activities,

which runs to his benefit, and that is not related by necessity to summative

assessment but merely coincides with the assessment activities intended to be

undertaken in the unit.

<u>Hiding one's true self as an assessment-influenced behaviour</u>

Ball asserts that one of the "costs" of performativity is:

> ...a kind of *values schizophrenia*... where commitment, judgement and
> authenticity within practice are sacrificed for impression and
> performance.
>
> <div align="right">(Ball 2003, p. 221)</div>

One participant whose experiences seemed to affirm this concept of *values schizophrenia* in the choices she felt compelled to make in light of the requirements of assessment was Celine. Celine is an extremely bright student who came across as comparatively more outspoken than her classmates and who holds strong views on a variety of the social issues touched upon in lessons during the study. In the final few weeks of the study, the class turned their attention to preparing for an assessment in response to a playtext, *The Laramie Project*, which discusses a variety of issues connected with the theme of sexuality. When asked about her feelings on this forthcoming assessment, Celine expressed a sense of frustration about what she perceived were the restrictions on what she was permitted to write:

> Celine: If the assessment were to ask my opinion or my thoughts, then I
> would give them my opinion or thoughts... but it really depends on the
> assessment I guess, because if the criteria doesn't allow it, then, you
> know, because the whole point of an assessment is to grade the individual
> who's being assessed, right? And if you were to do something that the
> criteria for the assessment didn't agree with, then they would lower the
> possible grade that you could get...

It becomes clear from her responses that Celine's preference would be to be able to write an opinion-based response to *The Laramie Project* as she feels strongly about the issues that the text raises. However, her impression is that the criteria are mandating her to take a more analytical stance and that this will restrict her response. She considers whether she might be able to find a way round this restriction:

> Celine: ...if I could potentially phrase it in a way that still answered what
> the assessment were asking me to do whilst still portraying what I

thought about it, then I could go about that, but I wouldn't – I don't think I will have the possibility to just present my – my thoughts and opinions in the open, I guess.

The contrast Celine draws between her thoughts and opinions "in the open" and the necessary path she will take in response to the current assessment is a point of tension for her. She considers the assessment to be a barrier to her fully expressing herself.

Her experiences are echoed in the work of Lunneblad and Carlsson (2012), discussed above, who see "a hint of delusion or pretence" in the classroom experiences of the students in their research. The teacher in their study has, in response to the requirements of a high-stakes test, to request of students that they offer certain types of "acceptable" response to an analysis task based on poems and pictures. The resultant responses are, according to the researchers,

> …a matter of credibility more than an expression of true feelings about the poem or, rather, a plausible interpretation of the poem *as if* the choice of poem and the picture built on true feelings, true personal preferences or a personal liking.

> (Lunneblad & Carlsson 2012, p. 304)

In much the same way, Celine is being called upon to respond to *The Laramie Project* as if she felt differently than she does about the text or the prompts to which she is responding. She realises that, in some sense, she still has the prerogative to write the piece as she wishes but that this prerogative is set against a system of incentives whereby she would have to suffer consequences if she chose what she views as a transgressive course of action:

> JS: Let's pretend that you decided that for whatever reason you were just going to fully express yourself in response to this task. What would happen?

117

Celine: …I would possibly be given the opportunity to re-do it in a way that would be acceptable.

JS: And do you think you'd take that opportunity?

Celine: It depends how passionate I would feel about what I did and how much I stand by it. But then again how much I would care about the grade that I got, because despite how much I felt about it in the end of the day it's still a grade which is what universities would see, so – I would have to care about it at some point.

Consequently, the influence of assessment on her behaviour is to require of her that she hide her true self rather than expressing her feelings "in the open"; assessment demands of her that she "[has] to care about it at some point", and the power it wields means that any capacities Celine feels she has for transgressive behaviour are ultimately subordinated.

5   Conclusion

One of the contributions of this research has been to offer new insights into the experience of students in the context of the IB Middle Years Programme, and in the specific setting of Bayside College, Hong Kong. By examining the influence of assessment in this specific context, this research illuminates the complex interplay of forces that act variously to facilitate, or to deny, the emancipatory aspirations of the programme as explored in Chapter One. A key finding, in examining both the mediating influences on assessment washback in this context, and the dimensions in which assessment washback is seen to its influence, is that students' experiences of assessment are linked to factors in the assessment environment that go well beyond the campus walls. Students respond to assessment while situated at the meeting point of factors including the IBO, universities, education policy, and assessment discourses pertaining to different geographies and cultures.

The view of assessment washback that emerged in the second chapter of this thesis was as a highly subjective, mediated and multi-dimensional phenomenon. Correspondingly, my study has sought to respond to this complexity in the range of ways in which the concept has been explored. It has identified a range of mediating influences that play a part in how students form meaning from their experiences. It has also offered insights into what the resultant meanings are for the participants of this research, and finally it has also identified a number of different dimensions in which the influence of assessment is seen to operate.

Some of the most powerful mediating influences observed are those that are physically present in the learning environment. The teacher is a central figure in conditioning how students experience assessment, and what the teacher says and does plays an impactful role in what is seen as important or significant. A similar role is played by students' peers, and by formal school processes such as reporting. However, mediating influences that are more distant (both in terms of time and geography) also play a weighty part in how students form meaning of their experiences; in particular, students process their experiences in the present in light of a knowledge (often highly speculative or uncertain) of what lies in wait for them in the future. With students in Year 11 situated at the boundary between the end of one programme, and the beginning of another, there is a vivid sense of future implication running through the thoughts of the participants as they go about their learning.

The resulting meanings and constructions that students attach to assessment help to provide insight in relation to the overarching research question posed by this study:

> **To what extent does educational assessment act as
> a means of student emancipation?**

By offering a series of five metaphors to represent the different meanings students attach to assessment, this research contributes new understanding of how the various functions and roles ascribed to assessment are engaged in the process of making meaning out of assessment experiences. The metaphors provide a reference point for exploring assessment as a complex, multi-faceted

part of the experience of schooling for young people. Of the five metaphors

presented, four present a prima facie challenge to the emancipatory aspirations

of the MYP and the IB programmes in general. Where assessment acts as the

Boss, or holds a Mirror up to students who suppose it "judges individuals 'in

truth'" (Foucault 1977, p. 180), it is being conceived of in such a way that

students are caught up in a "technology of power", combining "the techniques of

an observing hierarchy and those of a normalizing judgment" (ibid., p. 184). In

contrast, one of the five metaphors – The Mask – provides, in the recognition of

assessment as *artifice*, a route via which students hold greater agency in the

process of assessment. However, the findings of this study also problematise the

notion that assessment experiences can be straightforwardly either

emancipatory or non-emancipatory. As each metaphor is invoked, students are

seen as both subject to technologies of power, and as agents in exercising their

own power. For instance, in the *Good Angel* metaphor students recognise

assessment as providing a version of their best selves; some participants actively

seek out this function of assessment as a means of deriving benefit, but

simultaneously they recognise a "splitting" of the self, the seeming paradox that

students might be being forced to act, and might also wish this force upon

themselves. The emancipatory and oppressive influences of assessment are held

in suspension as students navigate the various meanings and constructions they

form of the role of assessment in their lives.

A question left unanswered by this research is the extent to which students draw

on something like The Mask to form meaning from their assessment experiences,

versus other, more insidious conceptualisations. Where it occurred in this study

it demanded of the student an ability to recognise the subjectivity, and therefore the fallibility, of the judgments made by others about them and their learning. In the context of participants' deeply-held beliefs about the respective roles of student, teacher, school and system, this might in fact be seen to border on *disobedience*. One student, Ilya, illustrates the point as he is asked whether he would still work hard on a task if he knew it did not 'count' as part of the formal summative assessment for the unit he was studying:

> Ilya: Yes, because it is preparing me for the assessment.
>
> JS: How do you know?
>
> Ilya: I don't know.
>
> JS: What if the teacher told you in 2 minutes' time that this task was nothing to do with the skills involved in the formal assessment for this unit?
>
> Ilya: I would still do it, because that's what I'm used to. I'm used to traditional school where the style of learning is that the students do what the teachers ask them to do.

(Field Note, 9th November 2017)

If The Mask requires a form of disobedience that is at odds with views such as those expressed by Ilya, it behoves those of us within the community of MYP schools to consider how productive disobedience might best be fostered. Recognising and legitimising a role for productive disobedience within the IB programmes may be a mechanism for the development of student agency, a concept of central importance to the programmes (International Baccalaureate Organisation 2018) that belongs securely to the Frereian tradition in which the programmes are seen. This legitimised role would align with the finding of Wischmann and Riepe (2019), who see in their study of student agency in a

122

German primary school the necessity of what they term "resistant agency" in order for agency to occur at all (p. 140).

A further unanswered question that might be addressed through additional research is the relationship between the role of student agency in schools and Hong Kong as a Confucian-heritage educational context. Lu (2019), in her autoethnographic research on the role of Confucian philosophy in learning, notes that issuing challenge to authority in a Confucian school setting creates a difficult tension between the goal of student agency and the historic norms of Confucian teaching and learning. In the context of MYP assessment at a school like Bayside College, this tension would need to be understood in more detail in order for the "solution of the teacher-student contradiction" (Freire 1970, Location 986) to occur in a way that is responsive to cultural and educational context.

Additionally, this research has identified a number of dimensions of influence that assessment washback is seen to have on the lived experience of students. Students are promoted to (re)interpret significant aspects of their learning environment in light of this influence. Their relationships with others are bound up in their experiences of assessment, and on occasion it influences the choices they make in declaring, or hiding, aspects of their own self to others around them. Students seem also to tread a line between being 'motivated by' assessment, and 'using' assessment to motivate themselves – a line which is representative of the complexity of the overarching research question. Where participants reported being disempowered by assessment or caught up in what Ball describes as a culture of "performativity" (Ball 2003), it was also frequently

the case that they simultaneously retained autonomy and agency of their own, and often saw themselves as being placed in a position of power even as they were subject to authorities outside of themselves. Consequently, the question of whether educational assessment offers emancipatory possibilities cannot be answered straightforwardly and is not a binary consideration; the findings of this study echo Huhta, Kalaja and Pitkänen-Huhta's (2006) conclusions that the various roles students take on in their responses to experiences of assessment are not matters of fixed identity, but that at different times and in different contexts the extent to which students enact, or are denied, emancipatory positions in relation to assessment is in flux and subject to constant movement and change:

> Test-takers appear to be multi-faced and their identities variable and situated in nature… their identities are not only variable and changing during the test-taking process but they can also be complex and multilayered within any given point in the process.

(p. 345)

Finally, this research has contributed new understandings of the opportunities and challenges afforded to the researching practitioner in conducting research as part of a professional doctorate in his own school. The complexity of such research is understood particularly through the multiple and competing identities ascribed to the researcher in such circumstances, and a number of ethical dilemmas arising from this complexity are explored and examined. It is hoped that by shedding light on the actions and decisions taken as part of this research, and the often-intractable challenges these have presented along the way, this study may be of use to other researchers in similar circumstances.

At a time in which the role and nature of assessment in the MYP is debated and contested, the findings of this research speak to the importance of understanding assessment in terms of its influence on the student experience at what Silfver, Sjöberg and Bagger term the "micro-level" of the classroom (2016). The contribution of this study is to illuminate these experiences in the final year of the MYP, contributing to an understanding of what it means to navigate this liminal aspect of the programme and how students at Bayside College ride the tensions between the emancipatory possibilities of the programme and the performative assessment cultures that intersect it.

References

Alderson, J. C., 2004. Foreword. In: Cheng, L., Watanabe, Y., and Curtis, A., eds. *Washback in Language Testing: Research Contexts and Methods.* Mahwah, NJ: Lawrence Erlbaum Associates, 2004, pp. iv - vii.

Alderson, J.C. & Wall, D., 1993. Does washback exist? *Applied Linguistics.* 14 (2). 115-129.

Alkharusi, H., Aldhafri, S., Alnabhani, H. & Alkalbani, M., 2014. Classroom Assessment: Teacher Practices, Student Perceptions, and Academic Self-efficacy Beliefs. *Social Behavior and Personality.* 42 (5). 835-855.

Bailey, K.M., 1996. Working for washback: A review of the washback concept in language testing. *Language Testing.* 13 (3). 257-279.

Ball, S.J., 2003. The teacher's soul and the terrors of performativity. *Journal of Education Policy.* 18 (2). 215-228.

Ball, S.J., 2012. *Global Education Inc.: New policy networks and the neo-liberal imaginary.* [Kindle version]. (s.l.): Routledge. Available at: http://www.amazon.co.uk/ . [Accessed 30 August 2017].

Ball, S.J., 2013a. *Foucault, Power, and Education.* [Kindle version]. (s.l.): Routledge. Available at: http://www.amazon.co.uk/ . [Accessed 17 June 2018].

Ball, S.J., 2013b. *The Education Debate.* 2nd ed. [Kindle version]. (s.l.): The Policy Press. Available at: http://www.amazon.co.uk/ . [Accessed 22 July 2017].

Bandura, A., 1977. Self-efficacy: Towards a Unifying Theory of Behavioural Change. *Psychological Review.* 84 (2). 191-215.

Bayside College, 2015. *Minutes of School Council Meeting, Thursday 23 April 2015* [online]. Hong Kong: Bayside College. Available at: [URL REMOVED FOR ANONYMITY] [Accessed 11 April 2017].

Bayside College, 2018. *Annual Report* [online]. Hong Kong: Bayside College. Available at: [URL REMOVED FOR ANONYMITY] [Accessed 29 September 2018]

Benmansour, N., 1999. Motivational orientations, self-efficacy, anxiety and strategy use in learning high school mathematics in Morocco. *Mediterranean Journal of Educational Studies.* 4. 1-15.

Berry, R., 2011. Assessment trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Principles, Policy & Practice.* 18 (2). 199-211.

Biggs, J., 1998. Assessment and classroom learning: a role for summative assessment? *Assessment in Education: Principles, Policy & Practice.* 5 (1). 103-110.

Bloom, B.S., 1969. Some theoretical issues relating to educational evaluation. In: Taylor, R. W., ed. *Educational Evaluation: New roles, new means: The 68th yearbook of the National Society for the Study of Education, Part II.* Chicago: University of Chicago Press, 1969, pp. 26-50.

Booher-Jennings, J., 2008. Learning to label: Socialisation, gender, and the hidden curriculum of high-stakes testing. *British Journal of Sociology of Education*. 29 (2). 149-160.

Braun, V. & Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology.* 3 (2). 77-101.

British Educational Research Association, 2011. *Ethical Guidelines for Educational Research* [online]. London: British Educational Research Association. Available at: https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf [Accessed 10 October 2016].

Brookhart, S.M. & DeVoge, J.G., 1999. Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education.* 12 (4). 409-425.

Brooks, R., Te Riele, K. & Maguire, M., 2014. *Ethics and Education Research* [Kindle version]. (s.l.): Sage Publications. Available at: http://www.amazon.co.uk/ . [Accessed 10 January 2017].

Brown, G.T. & Hirschfeld, G.H., 2008. Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice.* 15 (1). 3-17.

Brown, G.T., Irving, S.E., Peterson, E.R. & Hirschfeld, G.H., 2009. Use of interactive–informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction.* 19 (2). 97-111.

Bryman, A., 2012. *Social Research Methods.* 4th ed. [Kindle version]. (s.l.): Oxford University Press. Available at: http://www.amazon.co.uk/ . [Accessed 18 February 2016].

Carless, D., 2012. *From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-Heritage Settings.* [Kindle version]. (s.l.): Routledge. Available at: http://www.amazon.co.uk/ . [Accessed 10 March 2018].

Chavez, C., 2008. Conceptualizing from the inside: Advantages, complications, and demands on insider positionality. *The Qualitative Report.* 13 (3). 474-494.

Choi, C., 1999. Public examinations in Hong Kong. *Assessment in Education.* 6 (3). 405-417.

Cheng, L., 1997. How does washback influence teaching? Implications for Hong Kong. *Language and education.* 11 (1). 38-54.

Cheng, L. & Curtis, A., 2004. Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning. In: Cheng, L., Watanabe, Y., and Curtis, A., eds. *Washback in Language Testing: Research Contexts and Methods.* Mahwah, NJ: Lawrence Erlbaum Associates, 2004, pp. 3-18.

Cheng, L., Watanabe, Y. & Curtis, A., 2004. *Washback in Language Testing: Research Contexts and Methods.* Mahwah, NJ: Lawrence Erlbaum Associates.

Clapham, A. & Vickers, R., 2018. Further education sector governors as ethnographers: five case studies. *Ethnography and Education.* 13 (1). 34-51.

Clarke, V. & Braun, V., 2017. Thematic Analysis. *Journal of Positive Psychology.* 12 (3). 297-298.

Cohen, L., Manion, L., & Morrison, K., 2011. *Research Methods in Education*. 7th ed. [Kindle version]. (s.l.): Routledge. Available at: http://www.amazon.co.uk/ . [Accessed 4 April 2015].

Cresswell, J. W., 2002. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. 2nd ed. Thousand Oaks, CA: Sage Publications.

Crooks, T.J., 1988. The impact of classroom evaluation practices on students. *Review of educational research*. 58 (4). 438-481.

Cross, S.E. & Markus, H.R., 1994. Self-schemas, possible selves, and competent performance. *Journal of Educational Psychology*. 86 (3). 423-438.

Damankesh, M. & Babaii, E., 2015. The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in Educational Evaluation*. 45. 62-69.

Duclos, D., 2019. When ethnography does not rhyme with anonymity: Reflections on name disclosure, self-censorship and storytelling. *Ethnography*. 20 (2)*.* 175-183.

Ellis, C., 1995. Emotional and ethical quagmires in returning to the field. *Journal of Contemporary Ethnography*. 24 (1). 68-98.

Ellis, C., 2007. Telling Secrets, Revealing Lives: Relational Ethics in Research with Intimate Others. *Qualitative Inquiry.* 13 (1). 3-29.

Elwood, J., 2013. Educational assessment policy and practice: A matter of ethics. *Assessment in Education: Principles, Policy & Practice.* 20 (2). 205-220.

Emerson, R.M., Fretz, R.I. & Shaw, L.L., 2011. *Writing ethnographic fieldnotes.* 2nd ed. [Kindle version]. (s.l.): University of Chicago Press. Available at: http://www.amazon.co.uk/ . [Accessed 15 November 2016].

Ericksen, S.C., 1983. Private Measures of Good Teaching. *Teaching of Psychology.* 10 (3). 133-136.

Fine, M., 1994. Working the hyphens: Reinventing the self and other in qualitative research. In: Denzin, N. & Lincoln, Y. S., eds. *Handbook of Qualitative Research.* Thousand Oaks, CA: Sage Publications, 1994, pp. 70-82.

Fortun, K., 2009. Figuring out ethnography. In: Marcus, G. & Faubion, J., eds. *Fieldwork Is Not What It Used to Be: Learning Anthropology's Method in a Time of Transition*. Ithaca, NY: Cornell University Press, 2009, pp. 167-183.

Foucault, M., 1977. *Discipline and Punish: The Birth of the Prison.* [Kindle version]. (s.l.): Vintage. Available at: http://www.amazon.co.uk/ . [Accessed 20 April 2016].

Fransson, A., 1984. Cramming or Understanding? Effects of Intrinsic and Extrinsic Motivation on Approach to Learning and Test Performance. In: Alderson, J. C., & Urquhart, A. H., eds. *Reading in a Foreign Language.* London: Longman, 1984, pp. 86-115.

Freire, P., 1970. *Pedagogy of the Oppressed*. [Kindle version]. (s.l.): Herder & Herder. Available at: http://www.amazon.co.uk/ . [Accessed 20 May 2016].

Glaser, B.G. & Strauss, A.L., 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine.

Gosa, G., 2004. *Investigating Washback: A Case Study Using Student Diaries*. Ph.D. thesis, Lancaster University.

Gray, D.E., 2014. *Doing Research in the Real World.* 3rd ed. London: Sage Publications.

Great Britain. Department for Education, 2018. *Secondary Accountability Measures: Guide for maintained secondary schools, academies and free schools* [online]. (s.l.): Department for Education. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/748497/Secondary_accountability_measures_guidance__-_October_2018.pdf [Accessed 19 December 2018].

Green, A., 2006. Watching for washback: Observing the influence of the International English Language Testing System academic writing test in the classroom. *Language Assessment Quarterly.* 3 (4). 333-368.

Green, A., 2013. Washback in language assessment. *International Journal of English Studies.* 13 (2). 39-51.

Guest, G., MacQueen, K.M. & Namey, E.E., 2012. *Applied Thematic Analysis.* Thousand Oaks, CA: Sage Publications.

Hammersley, M. & Atkinson, P., 2007. *Ethnography: Principles in Practice.* 3rd ed. New York, NY: Routledge.

Harlen, W. & Deakin Crick, R., 2003. Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice.* 10 (2). 169-207.

Hegarty, G., 2015. *MYP Programme Update (Assessment)*. Presentation given at the "*IB Global Conference 2015*" hosted by the International Baccalaureate Organisation on 21 March 2015 in Macau, China.

Hegarty, G., 2017. *MYP Programme Update (Assessment)*. Presentation given at the "*IB Global Conference 2017*" hosted by the International Baccalaureate Organisation on 30 March 2017 in Yokohama, Japan.

Hill, M., 2009. Ways of Seeing: Using Ethnography and Foucault's Toolkit to View Assessment Practices Differently. *Qualitative Research.* 9 (3). 309-330.

Hong Kong. Education Bureau, 2016. *Private Independent Schools* [online]. Hong Kong: Hong Kong Education Bureau. Available at: http://edb.hkedcity.net/internationalschools/private_independent_schools.php?lang=en [Accessed 11 October 2016].

Howell, A., 2017. 'Because then you could never ever get a job!': children's constructions of NAPLAN as high-stakes. *Journal of Education Policy.* 32 (5). 564-587.

Huhta, A., Kalaja, P. & Pitkänen-Huhta, A., 2006. Discursive construction of a high-stakes test: the many faces of a test-taker. *Language Testing.* 23 (3). 326-350.

International Baccalaureate Organisation, 2010. *History of the Middle Years Programme*. Cardiff: International Baccalaureate Organisation.

International Baccalaureate Organisation, 2013. *Transition Document*. Cardiff: International Baccalaureate Organisation.

International Baccalaureate Organisation, 2014a. *General Regulations: Middle Years Programme*. Cardiff: International Baccalaureate Organisation.

International Baccalaureate Organisation, 2014b. *MYP: From principles into practice*. Cardiff: International Baccalaureate Organisation.

International Baccalaureate Organisation, 2014c. *MYP Language and Literature Guide* [online]. Cardiff: International Baccalaureate Organisation. Available at: http://occ.ibo.org/ibis/occ/Utils/getFile2.cfm?source=/ibis/occ/home/subjectHomeMYP.cfm&filename=dSpace%2Fen%2Fm_1_langa_guu-nc_1409_3_e%2Epdf [Accessed 11 October 2016].

International Baccalaureate Organisation, 2014d. *MYP Unit Planner* [online]. Cardiff: International Baccalaureate Organisation. Available at: http://occ.ibo.org/ibis/occ/Utils/getFile2.cfm?source=/ibis/occ/home/subjectHomeMYP.cfm&filename=dSpace%2Fen%2Fm_g_mypxx_mon-nc_1312_2_e%2Epdf [Accessed 28 December 2016].

International Baccalaureate Organisation, 2015. *What is an IB Education?* [online]. Cardiff: International Baccalaureate Organisation. Available at: http://www.ibo.org/globalassets/digital-tookit/brochures/what-is-an-ib-education-en.pdf [Accessed 10 October 2016].

International Baccalaureate Organisation, 2017. *Benefits of eAssessment* [online]. Cardiff: International Baccalaureate Organisation. Available at: http://www.ibo.org/programmes/middle-years-programme/assessment-and-exams/assessment-from-2016/myp-eassessment/ [Accessed 15 January 2017].

International Baccalaureate Organisation, 2018. *The learner* [online]. Cardiff: International Baccalaureate Organisation. Available at: http://resources.ibo.org/data/the-learner_fc17a71a-2344-4b26-84cb-caca3a02750d/PRC-p-0-pypxx-gui-lrnr-1-e_d32875a1-8611-4de3-9f7d-14a22127adc2.pdf [Accessed 27 January 2020].

Jerolmack, C., and Khan, S., eds., 2018. *Approaches to Ethnography: Analysis and Representation in Participant Observation.* New York: Oxford University Press.

Jiang, J., 2017. Higher Education in China. *Oxford Research Encyclopedia of Education*. Oxford: Oxford University Press.

Johnes, R., 2017. *Entries to arts subjects at Key Stage 4* [online]. (s.l.): Education Policy Institute. Available at: https://epi.org.uk/wp-content/uploads/2018/01/EPI-Entries-to-arts-KS4-1.pdf [Accessed 19 December 2018].

Johnson, M. & Crisp, V., 2009. An exploration of the effect of pre-release examination materials on classroom practice in the UK. *Research in Education.* 82 (1). 47-59.

Kaufman, M., 2010. *The Laramie Project.* New York: Vintage.

Klein, J., 2017. How schools cope with the double challenge of excellence in high-stakes risk tests and investment in education. *Assessment in Education: Principles, Policy & Practice.* 24 (4). 474-488.

Kvale, S., 2006. Dominance through interviews and dialogues. *Qualitative inquiry.* 12 (3). 480-500.

Kvale, S., 2008. *Doing Interviews.* [Kindle version]. (s.l.): Sage Publications. Available at: http://www.amazon.co.uk/ . [Accessed 11 August 2016].

Lapadat, J.C., 2000. Problematizing transcription: Purpose, paradigm and quality. *International Journal of Social Research Methodology.* 3 (3). 203-219.

Levitan, J., Carr-Chellman, D. & Carr-Chellman, A., 2017. Accidental ethnography: A method for practitioner-based education research. *Action Research.* Manuscript submitted for publication.

Lu, J., 2019. Confucius stands on the London eye: an auto-ethnographic study. *Ethnography and Education.* 14 (1). 51-64.

Lunneblad, J. & Carlsson, M.A., 2012. Performativity as pretence: A study of testing practices in a compulsory school in Sweden. *Ethnography and Education.* 7 (3). 297-309.

Madden, R., 2010. *Being Ethnographic: A Guide to the Theory and Practice of Ethnography.* [Kindle version]. (s.l.): Sage Publications. Available at: http://www.amazon.co.uk/ . [Accessed 13 July 2016].

Mansaray, A.A., 2012. *The roles and positions of teaching assistants in two urban primary schools: An ethnographic study of educational work and urban social change.* Ph.D. thesis, Institute of Education, University of London.

Mayeza, E., 2017. Doing Child-Centered Ethnography: Unravelling the Complexities of Reducing the Perceptions of Adult Male Power During Fieldwork. *International Journal of Qualitative Methods.* 16 (1). 1-10.

Mero-Jaffe, I., 2011. 'Is that what I said?' Interview transcript approval by participants: An aspect of ethics in qualitative research. *International Journal of Qualitative Methods.* 10 (3). 231-247.

Messick, S., 1996. Validity and washback in language testing. *Language Testing,* 13(3). 241-256.

Mills, D., & Morton, M., 2013. *Ethnography in Education* [Kindle version]. (s.l.): Sage Publications. Available at: http://www.amazon.com/ [Accessed 27 December 2015].

Moss, P.A., 1998. The role of consequences in validity theory. *Educational Measurement: Issues and Practice.* 17 (2). 6-12.

Natriello, G., 1987. The Impact of Evaluation Processes on Students. *Educational Psychologist.* 22 (2). 155-175.

Pan, Y. C., 2014. Learner Washback Variability in Standardized Exit Tests. *Teaching English as a Second or Foreign Language,* 18(2) [online]. Available at: http://www.tesl-ej.org/wordpress/issues/volume18/ej70/ej70a2/ [Accessed 21 October 2015].

Perry, K., 2014. From O Levels to the GCSE: the landmarks. *The Telegraph* [online]. Available at: https://www.telegraph.co.uk/education/educationnews/10741337/From-O-Levels-to-the-GCSE-the-landmarks.html [Accessed 19 December 2018].

Peterson, A.D.C., 2003. *Schools Across Frontiers: The story of the International Baccalaureate and the United World Colleges.* 2nd ed. La Salle, IL: Open Court Publishing.

Pole, C. & Morrison, M., 2003. *Ethnography for Education.* Maidenhead: Open University Press.

Puttick, S., 2017. Performativity, guilty knowledge, and ethnographic intervention. *Ethnography and Education.* 12 (1). 49-63.

Rasooli, A., Zandi, H. & DeLuca, C., 2018. Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond. *Studies in Educational Evaluation.* 56. 164-181.

Reay, D. & Wiliam, D., 1999. 'I'll be a nothing': Structure, Agency and the Construction of Identity through Assessment. *British Educational Research Journal.* 25 (3). 343-354.

Reyes, V., 2018. Ethnographic toolkit: Strategic positionality and researchers' visible and invisible tools in field research. *Ethnography.*

Sahlberg, P., 2011. *Finnish lessons: What can the world learn from educational change in Finland?* New York, NY: Teachers College Press.

Salzinger, L. & Gowan, T., 2018. Macro-Analysis: Power in the Field. In: Jerolmack, C., and Khan, S., eds. *Approaches to Ethnography: Analysis and Representation in Participant Observation.* New York: Oxford University Press, 2018, pp. 61-94.

Sambell, K. & McDowell, L., 1998. The construction of the hidden curriculum: messages and meanings in the assessment of student learning. *Assessment & Evaluation in Higher Education.* 23 (4). 391-402.

Sambell, K., McDowell, L. & Brown, S., 1997. "But is it fair?": an exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation.* 23 (4). 349-371.

Scott, C., 2007. Stakeholder perceptions of test impact. *Assessment in Education: Principles, Policy & Practice.* 14 (1). 27-49.

Segers, M. & Tillema, H., 2011. How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation.* 37 (1). 49-54.

Shih, C., 2009. How tests change teaching: A model for reference. *English Teaching.* 8 (2). 188-206.

Silfver, E., Sjöberg, G. & Bagger, A., 2016. An 'appropriate' test taker: the everyday classroom during the national testing period in school year three in Sweden. *Ethnography and Education.* 11 (3). 237-252.

Stenlund, T., Eklöf, H. & Lyrén, P., 2017. Group differences in test-taking behaviour: an example from a high-stakes testing program. *Assessment in Education: Principles, Policy & Practice.* 24 (1). 4-20.

Struyven, K., Dochy, F. & Janssens, S., 2005. Students' Perceptions about Evaluation and Assessment in Higher Education: a Review. *Assessment & Evaluation in Higher Education.* 30 (4). 325-341.

Sultana, N. 2018. Investigating the Relationship between Washback and Curriculum Alignment: A Literature Review. *Canadian Journal for New Scholars in Education.* 9 (2). 151-158.

Taras, M., 2007. Machinations of assessment: Metaphors, myths and realities. *Pedagogy, Culture & Society.* 15 (1). 55-69.

Tolich, M., 2004. Internal confidentiality: When confidentiality assurances fail relational informants. *Qualitative Sociology.* 27 (1). 101-106.

Tsagari, D., 2009. *The Complexity of Test Washback: An Empirical Study*. Frankfurt: Peter Lang.

Vähäsantanen, K. & Saarinen, J., 2013. The power dance in the research interview: Manifesting power and powerlessness. *Qualitative Research.* 13 (5). 493-510.

Vernooij, E., 2017. Navigating multipositionality in 'insider' ethnography. *Medicine Anthropology Theory*. 4 (2). 34-49.

Wall, D., 2000. The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System.* 28 (4). 499-509.

Watanabe, Y., 2004. Methodology in Washback Studies. In: Cheng, L., Watanabe, Y., and Curtis, A., eds. *Washback in Language Testing: Research Contexts and Methods.* Mahwah, NJ: Lawrence Erlbaum Associates, 2004, pp. 19-36.

Watanabe, Y., 2008. The Importance of Motivation Theories for Understanding Washback to the Learner. *Bulletin of the Center for Educational Research and Practice, Faculty of Education and Human Sciences, Akita University* [online]. (s.l.): Akita University. Available at: http://air.lib.akita-u.ac.jp/dspace/bitstream/10295/650/4/jissen28a.pdf [Accessed 26 July 2015].

Wax, M.L., 1982. Research Reciprocity Rather than Informed Consent in Fieldwork. In: Sieber, J. E., ed. *The Ethics of Social Research: Fieldwork, Regulation, and* Publication. New York, NY: Springer, 1982, pp. 33-48.

Webb, N.L., 1997. Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education. *Research Monograph* 6. Washington, DC: National Institute for Science Education.

Wickins, E., 2014. *Principal Perspectives: Distributed Leadership in Hong Kong International Schools*. Ed.D. thesis, Bristol University.

Wickins, E. & Edwards, A., 2016. And a green tea frappuccino: Developing the local dimension of international schools. Manuscript submitted for publication.

Wiles, R., 2012. *What are Qualitative Research Ethics?* London: Bloomsbury Academic.

Wiliam, D., 1994. *Towards a Philosophy for Educational Assessment*. Paper presented to the annual conference of the British Educational Research Association, Oxford.

Wiliam, D. & Black, P., 1996. Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal.* 22 (5). 537-548.

Wischmann, A. & Riepe, V., 2019. Resistant agency in German lessons in primary education in Germany. *Ethnography and Education.* 14 (2). 136-152.

Wolcott, H. F., 2008. *Ethnography: A Way of Seeing.* 2nd ed. [Kindle version]. (s.l.): AltaMira Press. Available at: http://www.amazon.co.uk/ . [Accessed 21 July 2016].

Wyatt-Smith, C. & Klenowski, V., 2013. Explicit, latent and meta-criteria: Types of criteria at play in professional judgement practice. *Assessment in Education: Principles, Policy & Practice.* 20 (1). 35-52.

Yeo, R. & Dopson, S., 2018. Getting lost to be found: The insider-outsider paradoxes in relational ethnography. *Qualitative Research in Organizations and Management.* 13 (4). 333-355.

Yerkes, R.M. & Dodson, J.D., 1908. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology.* 18 (5). 459-482.

Zhan, Y. & Andrews, S., 2014. Washback effects from a high-stakes examination on out-of-class English learning: insights from possible self theories. *Assessment in Education: Principles, Policy & Practice.* 21 (1). 71-89.

Zhao, Y. 2014, *Who's Afraid of the Big Bad Dragon? Why China Has the Best (and Worst) Education System in the World.* San Fracisco, CA: Jossey-Bass.

Research Consent Form - Student Participant

Dear Students and Parents,

You may remember me writing previously to let you know about a research project I am conducting as part of my doctoral studies with Nottingham Trent University in the UK. As part of my earlier message I mentioned that, during the course of the research, I may invite some students to take part in a short interview about their experiences of assessment in their studies.

You are receiving this message because you are / your child is one of the people I wish to invite to participate in such an interview.

The interview will last for around 20-30 minutes and will be scheduled at a time of convenience to the participant, probably a break or lunchtime. The interview will be based on a series of questions about the student's experience of assessment. To aid the conduct of the interview I may send some questions in advance to the participant so that they can think about potential responses, but it does not matter if they do not get a chance to review the questions beforehand.

Participants are also asked to write a short 'diary entry' in advance of the interview, based on a series of prompts which I will provide. This should not be an extensive piece of writing and a few sentences or bullet points will be fine.

The interview will be audio-taped to allow for transcription – a transcript of each participant's interview will be given to them some time after the interview has taken place, so that they can comment on any issues they see as relevant.

Ethical Statement

The purpose of this statement is to assure you of good ethical practice during the conduct of the research.

-      All participants have the right to decline to participate and this right extends up to 14 December 2017, meaning that even if you withdraw consent after data has been collected it will be discarded and not published
-      All participants' information will remain confidential and pseudonyms will be used in writing up the research
-      All data will be stored securely
-      Participants may contact me at any time via james.smith022015@my.ntu.ac.uk to ask for further information about the research

If, after having reviewed the above, you are comfortable to participate / for your child to participate in this short interview, I would be grateful if you would please fill in the form below to indicate this. Please note that for consent to be given it is necessary for both the student and their parent to fill in a version of the form. Once you have filled in the form you will see a confirmation message acknowledging the submission.

I am very grateful for the opportunity to conduct this interview and I know that your / your child's reflections on their experience of learning will be a valuable addition to the research. If you have any questions please don't hesitate to get in touch with me.

Many thanks,

Mr James Smith

*Required

1. **I have read and understood the details of the research activities proposed in this message and I wish to take up the invitation to participate / I am happy for my child to take up the invitation to participate.** *
   Mark only one oval.

   ◯ Agree

   ◯ Disagree

2. **I am...** *

Mark only one oval.

      ⬭   The student participating

      ⬭   A parent of the student participating

3. **My name is...** *

_____

4. **(If you are a parent) My child's name is...**

_____

Powered by

Google Forms

Appendix B - Interview Schedule (Phase 1) – Ways in which students construe and interpret assessment

Aiming to address aspects of the following research questions:

#1	#2
What meanings and constructions do the students attach to the assessments in which they participate?	What aspects of the students' learning environment act as mediating influences on the formation of washback effects?

Non-indented questions will be asked of all participants and will be sent in advance for participants to consider.

Indented questions are possible follow-up questions and will be used selectively alongside other questions that are generated during the interview

In advance of the interview, participants will be asked to create or bring an image or object that represents their thoughts and feelings about assessment – this will lead into Question 1.

Participant briefing:

- Thank participant for the opportunity to interview
- Explain that the interview is being conducted to help understand the participant's thoughts and feelings about assessment
- Point out that the interview is being recorded and will be transcribed
- State that the participant will later receive a copy of the transcript to see if they wish to add anything
- Explain that the interview will be based on the questions they have received in advance, with possible follow-up questions for clarification and extension
- Remind that everything said will be anonymised and data held securely, and confidential to the researcher and members of the University involved with the EdD assessment process (+ RtW)
- Ask if the participant is comfortable and ready to start, or if they have any questions before the interview begins

---

1. Tell me about the image / object you have brought and how it represents your thoughts and feelings about assessment.

  - What do you understand 'assessment' to mean?
  - Are there any other ways in which you would define assessment?
  - What thoughts and feelings do you associate with assessment?

2. Tell me about a past experience of assessment that is memorable to you.

  - Why is this experience memorable to you?
  - How do you feel this assessment experience has influenced you?
  - How do you feel about this experience now?

3. Tell me about how you are assessed in your MYP English course this year.

  - Are there any other ways in which you are assessed?
  - Is this similar or different to how you are assessed in your other courses?
  - Is this similar or different to your past experiences with assessment?

4. What do you see the purpose of assessment in your MYP English course as being?

5. Take a moment to think about how you are assessed in your MYP English course. What is going through your mind?

  - Is this similar or different to your feelings about other assessment experiences?

6. How will your assessment in English this year make a difference to you going forward?

  - To what extent might your assessment in English affect you positively?
  - To what extent might your assessment in English affect you negatively?
  - Are there any potential ways in which assessment might make a difference to you but in a neutral way, not positive or negative?

Offer thanks for participating and remind participant that they will receive a copy of the subsequent transcript for further comment.

Appendix C - Interview Schedule (Phase 2) – the influence of assessment washback on the student experience

Aiming to address aspects of the following research questions:

#1	#3
What meanings and constructions do the students attach to the assessments in which they participate?	In what ways does washback influence the student experience?

Non-indented questions will be asked of all participants and will be sent in advance for participants to consider.

Indented questions are possible follow-up questions and will be used selectively alongside other questions that are generated during the interview

In advance of the interview, participants will be asked to write a short 'diary' in response to the following prompts: 1. What assessment is happening in your English lessons at the moment? 2. What emotion(s) do you associate with the assessment that is happening in English at the moment? Why? 3. How is the assessment that is happening in English at the moment influencing your experience as a student?

Participant briefing:

- Thank participant for the opportunity to interview
- Explain that the interview is being conducted to help understand the participant's thoughts and feelings about assessment
- Point out that the interview is being recorded and will be transcribed
- State that the participant will later receive a copy of the transcript to see if they wish to add anything
- Explain that the interview will be based on the questions they have received in advance, with possible follow-up questions for clarification and extension
- Remind that everything said will be anonymised and data held securely, and confidential to the researcher and members of the University involved with the EdD assessment process (+ RtW)
- Ask if the participant is comfortable and ready to start, or if they have any questions before the interview begins

---

1. Tell me about the assessment you mentioned in your diary entry.

    - What do you understand 'assessment' to mean?
    - Are there any other ways in which you would define assessment?

2. What do you see the purpose of assessment in your MYP English course as being?

3. Tell me about the emotion(s) you mentioned in your diary entry.

    - To what extent do you feel the emotion is specific to this point in time?
    - To what extent do you feel the emotion is specific to your studies in English, versus other subjects?
    - To what extent do you think other people in your class feel the same?

4. Describe what it is like to be a student in your English class at the moment.

    - Which aspects of this experience do you feel are influenced by Assessment? How so?
    - Tell me about your comments in your diary about the way assessment is influencing your experience
    - Why do you think assessment has this influence?

Offer thanks for participating and remind participant that they will receive a copy of the subsequent transcript for further comment.

Appendix D – Table of Participants

Name	Gender	Nationality	Role	Research Phase
Amy	F	Chinese (Hong Kong)	Student	2
Ana	F	Brazilian	Student	1
Celine	F	Brazilian / Dutch	Student	2
Chloe	F	Chinese (Hong Kong)	Student	1
Elise	F	British	Student	2
Emma	F	American	Teacher	1
Gladwin	M	American / Chinese (Hong Kong)	Student	2
Luca	M	Swiss	Student	1
Parth	M	Indian	Student	1
Ronnie	M	Indian	Student	2
Susan	F	Chinese (Hong Kong)	Student	1
Tom	M	Australian	Teacher	2

N.B. the students listed above are those who participated in interviews.

Appendix E – Typographical tools used in interview transcription

Tool	Example	Usage
Ellipsis	...	The interviewee pauses
Hyphen	-	The interviewee switches from one line of thought to another
Single quotation mark	'	The interviewee seems to be quoting from another person, either real or hypothetical
Italics	*emphasis*	The interviewee stresses or emphasizes a word or phrase in a way that seems to suggest importance
Square brackets	[laughter]	The interviewee gives a non-verbal cue which seems materially to affect how their words should be interpreted

Appendix F – Map of codes and themes

Emotion

- Child → **Stress and anxiety**
  - Child → **Falling behind vs. getting ahead**
- Child → **Confidence**
- Child → **Strengths**
- Child → **Positive emotions**

Behaviour

- Child → **Spilling over into rest of life**
- Child → **Hyper-awareness of assessment issues among students**
- Child → **Preparation and assessment lead-up strategies**
- Child → **Presenting thoughts in the open**

Motivation and driving purpose

- Child → **Who is assessment for**
- Child → **Attribution of cause for success or failure**
- Child → **Motivation**
- Child → **Stakes and significance**
- Child → **Working for the teacher (as distinct from working for the outcome itself)**
- Child → **Quid pro quo**
- Child → **Peers**
- Child → **Expectations**
- Child → **Self-improvement as a purpose of assessment**
- Child → **Preparation for life**
- Child → **Working harder**
- Child → **Possibility of future regret**
- Child → **The future**
- Child → **Stress as a positive driver**
- Child → **Link between enthusiasm and success**
- Child → **Knowing your strengths and weaknesses**
- Child → **Your better self**
- Child → **Parents**
- Child → **Proof**
- Child → **The IBDP**
- Child → **Difficulty and level of challenge**