

**The construct validity and fairness of the  
International Baccalaureate Diploma Programme  
Mathematics Internal Assessment:  
A case study of the assessment at United World  
College, China**

Jianbo Wang

A Research & Development Project

Submitted for the MSc Educational Assessment 2022

## DEPOSIT AND CONSULTATION OF THESIS

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

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

---

<b>Surname</b>	Wang
<b>First Name</b>	Jianbo
<b>Faculty Board</b>	Education
<b>Title of Dissertation</b>	The Construct Validity and Fairness of the International Baccalaureate Diploma Programme Mathematics Internal Assessment: A Case Study of the Assessment at United World College, China

### ***Declaration by the candidate as author of the dissertation***

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

## Abstract

This thesis outlines the findings of a mixed-methods study of the International Baccalaureate (IB) Diploma Programme (DP) Mathematics Internal Assessment (IA) carried out from students' and teachers' perspectives at an IB school in China. The present study examines the construct validity and fairness of the DP mathematics IA in the 2021–2022 academic year. The study took place in 2022, at the height of the COVID-19 pandemic in China, when IB examinations were cancelled in addition to school closures and other measures. The study of the validity and fairness of mathematics internal assessment in the context of a highly uncommon set of circumstances, as well as the lessons that may be learnt from these, is therefore given a new dimension.

The study used a mixed-methods approach, including two phases. Phase One included semi-structured interviews with five teachers and 13 students at UWC China. A survey of students' perceptions of assessment criteria with an evaluation of the validity of mathematics IA was also conducted in Phase One. Phase Two comprised quantitative analyses of Diploma Programme Year 2 students' mathematics internal assessment results and their final grades, as well as the relationship between IB DP mathematics performance and academic records of other domains.

The correlation coefficient was utilized to explain the degree to which the mathematics IA grades are associated with the mathematics final assessment grades awarded by the International Baccalaureate. The findings revealed a moderate correlation between IA scores and final mathematics grades. Furthermore, IB mathematics grades correlated significantly with physics ( $r = .556$ ) but weakly with

English ( $r = .27$ ), which indicates that students' IA grades were valid and fair. The interview with teachers and students suggested that clearer assessment criteria and new explanatory skills of the construct should be revisited.

These findings highlight some important implications for educators regarding the validity of IA grades, which may be used in the predicted grades or by international organizations for awarding the final marks. First, it is critical for both students and teachers to understand the subject and tiering of mathematics courses and comprehend the assessment objectives, which highlights a new requirement expected of learners rather than just at the subject level. Second, the construct of mathematics IA in requiring new skills has been revisited. In addition, it's recommended that there are five main aspects in order to support effective teaching and learning in empirical practice. Last but not least, it should be borne in mind that the study took place in China in 2022, when IB examinations were cancelled due to COVID-19. This research suggests that when conducting mathematics internal assessments throughout the epidemic, research and transferable abilities are highlighted for gains beyond construct validity comprehension.

## Table of Contents

Abstract .....	2
Chapter 1: Introduction .....	7
<i>1.1 The Construct of the DP Mathematics Course</i> .....	9
1.1.1 External Assessment .....	10
1.1.2 Internal Assessment .....	11
<i>1.2 Purpose of DP Mathematics Assessment</i> .....	13
<i>1.3 Research Aims</i> .....	17
Chapter 2: Literature Review .....	18
<i>2.1 Defining Internal Assessment</i> .....	19
2.1.1 The Nature of the Internal Assessment .....	19
2.1.2 Internal Assessment in Context of International Baccalaureate Diploma Programme .....	20
<i>2.2 The Validity of Mathematics Internal Assessment</i> .....	22
<i>2.3 The Reliability of Mathematics Internal Assessment</i> .....	26
<i>2.4 The Fairness of Mathematics Internal Assessment</i> .....	27
<i>2.5 Prior Empirical Studies on the Viewpoints of Students and Teachers</i> .....	28
<i>2.6 Research Questions</i> .....	29
Chapter 3: Methodology .....	31
<i>3.1 Research Aim</i> .....	31
<i>3.2 Research Design</i> .....	32
3.2.1 Overall Design .....	32
3.2.2 Quantitative Research Method .....	34
3.2.3 Qualitative Research Method .....	35
<i>3.3 Sampling and Participants</i> .....	36
3.3.1 Participants' Characteristics .....	37
<i>3.4 Instruments</i> .....	39
3.4.1 Survey .....	40
3.4.2 Semi-structured Interview .....	41
3.4.3 Mathematics Internal Assessment and Mock Examination Scores .....	42
<i>3.5 Pilot</i> .....	42
<i>3.6 Data Collection</i> .....	43
<i>3.7 Data Analysis</i> .....	44
<i>3.8 Ethical Considerations</i> .....	45
Chapter 4: Findings .....	47
<i>4.1 To what extent is the internal assessment valid and fair in the DP mathematics assessment? What are the relationships between IA scores and scores from the mathematics exam, the English writing assessment, or other IB subject exams such as science subject exams in 2022?</i> .....	47
4.1.1 The Relationship between Mathematics IA and Final Grades .....	48

4.1.2 IA Scores Sorted by Subjects and Levels. ....	48
4.1.3 Exploring the Fairness of IA. ....	50
4.2 <i>What are students' perspectives on the DP mathematics IA? How did the students perceive the validity and fairness of the DP mathematics internal assessment in 2022?</i> .....	52
4.2.1 Students' Perceptions of the Validity of Mathematics IA. ....	52
Students' understandings of the assessment criteria of IA .....	52
4.2.2 Students' Understandings of the Construct of IA. ....	55
Mathematics skills of IA .....	55
Research Skills of IA .....	57
Transferable Skills .....	58
Inquiry Skills of IA .....	60
4.2.3 Students' Understandings of the Fairness of the IA. ....	60
4.3 <i>How did the teachers understand the validity and fairness of the DP mathematics internal assessment?</i> .....	63
4.3.1 Underlying Skills. ....	63
4.3.2 Time Management. ....	64
4.3.3 Interaction between Teachers and Students. ....	65
4.4 <i>Summary of Findings</i> .....	66
Chapter 5: Discussion .....	68
5.1 <i>Construct Validity</i> .....	69
5.1.1 Subject and Tiering Differences. ....	70
5.1.2 Clearer Assessment Criteria. ....	72
5.1.3 New Explanatory Skills of the Construct. ....	74
Research Skills .....	75
Transferable Skills .....	76
Mathematics Literacy .....	77
5.2 <i>Fairness</i> .....	78
5.2.1 Engaging English Language Learners. ....	78
5.2.2 Authorship .....	80
5.3 <i>Implications</i> .....	81
5.4 <i>Limitations of this Research and Future Research Directions</i> .....	83
Chapter 6: Conclusion .....	86
References .....	88

## List of Tables and Figures

Figure 1: The Concept of DP Mathematics Assessments.....	13
Table 1: List of the Assessment Objectives from the IBDP Mathematics Guide.....	15
Table 2: Participant Characteristics by Gender at United World College, China.....	38
Table 3: Participant Characteristics by Level and Subject .....	39
Table 4: DP Year 2 Students' IA Scores by Level and Subject in 2022 .....	47
Table 5: Descriptive Statistics of the IB Subjects.....	49
Table 6: Bivariate Pearson Correlations among Mathematics Internal Assessment and Examinations (N = 135), English Language (N = 123), Physics (N = 62), Environmental Systems and Societies (N = 23), Biology (N = 26), Chemistry (N = 52), and Economics (N = 105).....	51
Table 7: English Proficiency of Students.....	54

## **Chapter 1: Introduction**

Construct validity and fairness play an important role in setting examination questions in real-world contexts (Ahmed & Pollitt, 2007; Messick, 1995; Wolming & Wikström, 2010). Construct validity examines whether the test is a good indicator of the construct, or the underlying skill being measured (Gipps, 2011, p. 49). Fairness refers to the need for tests to be thoughtfully developed, and the conditions for testing should be reasonable and equitable for all students (Camilli, 2006). Within the International Baccalaureate (IB) Diploma Programme (DP) assessment, the current IB qualification requires the student to be assessed in six subject groups, namely Studies in Language and Literature, Language Acquisition, Individuals and Society, Sciences, Mathematics, and the Arts. Students are required to choose three subjects at higher level (HL) and three at standard level (SL), as well as the mandatory core components of an Extended Essay, Theory of Knowledge and Creativity, Activity and Service (IBO, 2015a). The external assessment process of the IB DP is managed by the International Baccalaureate Organization (IBO) through written examination papers, while internal assessment (IA) coursework is marked by individual IB teachers from different IB schools and monitored anonymously by IB-employed teachers to ensure fair grading (IBO, 2019a; Mayer, 2010). Assessment in the context of the IB is therefore influenced by the school management, IB assessment objectives, and teachers' pedagogic behaviours and marking practices (IBO, 2015a, 2019a, 2019b; Zhang & McGrath, 2009). However, study on the internal assessment of the DP mathematics course is very limited. Therefore, it is of significant academic

interest to take one small step toward understanding the construct validity and fairness of the IB DP mathematics internal assessment.

Before reviewing the construct validity and fairness of the IB DP assessment, it is essential to consider the construct of the IB DP mathematics course, which has evolved significantly in recent years to include a number of subconstructs as assessment objectives, such as Knowledge and Understanding, Problem Solving, Communication and Interpretation, Technology, Reasoning, and Inquiry (IBO, 2019a, 2019b). The assessment objectives prioritize problem solving as central to learning mathematics and involve the acquisition of mathematical skills and concepts in a wide range of situations (IBO, 2019b). Notably, the specific purpose of the IA is to provide students with the opportunity to complete a piece of mathematical work over an extended period of time. The assessment criteria have been identified as presentation, mathematical communication, personal engagement, reflection, and use of mathematics (IBO, 2019a, 2019b). The "presentation" criterion evaluates the structure and coherence of the exploration. Students are expected to use level descriptors in their exploratory writing that highlight their positive achievements, such as coherent, well-organized and concise in English (IBO, 2019a). The "mathematical communication" criterion evaluates the extent to which the student has used the appropriate language (notation, symbols, and terminology), key terms, various types of mathematical representation, and a deductive method (IBO, 2019a, 2019b). The final DP mathematics examination and IA have several distinctions, which has resulted in a gap between teaching the summative assessments and the IA due to their different emphases. Consequently, the IB programme assessment is in the process of shifting paradigms, which need to be defined, from a curriculum-embedded assessment paradigm to a more comprehensive model of

educational assessment (Baccalaureate, 2020; Baird & Opposs, 2018; IBO, 2020c). The comprehensive model of educational assessment places a strong emphasis on mathematical thinking, presentation, communication, and problem solving in the actual world.

Aiming to investigate this gap between IA and final examinations, the first section of this chapter introduces the construct of the DP mathematics course, which includes the IA as writing contextualized mathematics. That might be a potential validity issue with the cognitive processes involving the extent to which IA assesses mathematical communication skills (IBO, 2019a). The second section of this chapter presents the purpose of the DP mathematics assessment in the context of internal and external assessments. The third section outlines the background of the research aims and rationale.

## **1.1 The Construct of the DP Mathematics Course**

The IB DP is intended for students between the ages of 16 and 19. It is currently available in more than 153 countries, and the majority of IB schools around the world utilize English as their primary language of instruction. The primary goals of assessments in the DP are to support academic objectives and advance student learning (IBO, 2018). IB DP mathematics coursework is assessed both externally and internally (IBO, 2015b, 2019a, 2019b).

### **1.1.1 External Assessment.**

External assessment mostly refers to a student's final IB written examination, which is often taken in their second year of the Diploma Programme, at the end of the DP course. The external assessment for DP mathematics courses is conducted using two or three examination papers (Papers 1, 2, and 3) with constructed-response items, up to standard level (SL) and higher level (HL) (Haladyna & Rodriguez, 2013; IBO, 2019a). While the relatively smaller November examination session mostly covers schools in the southern hemisphere, the May DP examination session primarily involves schools in the northern hemisphere.

The external assessment accounts for 80% of the student's final grade. A series of questions on functions, statistics, and probability make up the exam's substance (IBO, 2019a, 2019b). Its goal is to give pupils the opportunity to exhibit their knowledge and what they can do with it. However, a significant number of limitations of the final examinations for students have been noted in previous research. For example, the research of Sadler (1989) usefully reported that final examinations are essentially passive and do not normally have an impact on students' learning, as most students may stop learning after seeing their final examination scores. This finding was corroborated by the later research of Newton (2007), who found that summative assessment is used to support the purpose of summing up individual student achievement but does not highlight anything more significant about assisting learning. However, internal assessment provides promising alternative coursework that underpins students' learning by providing feedback to them during the process.

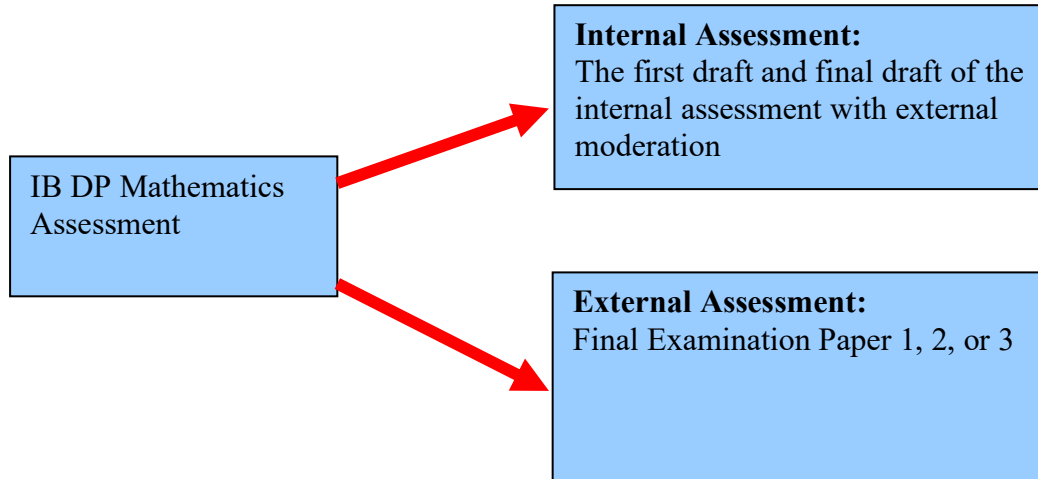
### **1.1.2 Internal Assessment.**

The IA coursework involves an exploration or investigation of mathematics and is an integral part of the mathematics course. Both SL and HL students must do it (IBO, 2019a, 2019b). The IA makes up 20% of a student's final IB mathematics grade. Its goal is to give students the freedom to explore their personal interests and demonstrate how their knowledge and skills are put to use. The opportunity for individual students to write about topics they are interested in is the first benefit of the IA (IBO, 2019c). As a result, the choice of topics for IA remains flexible while continuing to address a common set of competencies. Students are granted more control over their own learning in this way. Second, after the teacher's oral or written feedback is provided on how the work could be improved, having all student work internally assessed by a teacher is advantageous to students' learning (IBO, 2019a, 2019b). Wiliam et al. (2004) helpfully found that the IA can be thought of as a formative assessment (or assessment for learning), which results in measurable advantages including higher-quality learning. Finally, the IA enables educators to connect classroom learning to real-world applications of mathematics by putting students' learning in a local, cultural, or geographic context. Indeed, it has been reported that "[i]nternational schools, whose students often have a different cultural background from that in which the school is embedded, can use internally assessed work to develop a closer involvement in the local society or environment" (IBO, 2015a, 2015b). IA is driven by a rising desire to provide classroom teachers with more control over assessment (Allal, 2020; Clarke & Gipps, 2000; Wiliam et al., 2004; Yates & Johnston, 2018), as well as a growing awareness of the important impact of assessment on students' learning (Black & Wiliam, 1998).

As illustrated in Figure 1, the design of the DP mathematics internal assessment for the International Baccalaureate (IB) Diploma Programme (DP) provides chances to evaluate candidates internally and differently from final examinations (Black & Wiliam, 2018; Crooks et al., 1996; Elwood, 2006; Hopfenbeck, 2018; IBO, 2019c; Wiliam et al., 2004; Wright & Lee, 2014). Because IA is unique, a wide range of assessment tasks, like projects, fieldwork, and practical laboratory work, can be accomplished utilizing formative assessment. Internal assessment, which is administered in a classroom setting and is guided by a teacher to evaluate a student's learning, is noteworthy. The IA construct, however, has its own issues and difficulties. For example, the focus and evaluation environment of existing studies on IB DP mathematics coursework have been criticized for lacking coherence and fairness (Barrance, 2019; IBO, 2019a, 2019b). However, relatively few studies have examined IB DP mathematics assessment in greater detail and consequently analysed construct validity and fairness across a different cultural group.

**Figure 1**

**The Concept of DP Mathematics Assessments**



**1.2 Purpose of DP Mathematics Assessment**

It is thought that the method of IB DP internal assessment is criterion related. It determines the outcomes of students' work based on their performance in respect to levels of achievement, not in comparison to that of other students. The focus is on communicating mathematically (including formulae, diagrams, graphs, and tables). Nevertheless, the IA, also known as “mathematical exploration”, relies significantly upon students' English writing skills and is considered a written assignment that involves researching a mathematical topic (IBO, 2019a). The IA uses criteria that may strongly emphasize high language demands (see Table 1) (IBO, 2019a). These excessive demands corroborate previous findings by Messick (1995): “skill in communicating mathematical ideas might well be considered irrelevant variance in assessing mathematical knowledge”. IBO and

Messick both emphasize that the evaluation of communicating mathematical concepts can be thought of as construct-irrelevant variance (IBO, 2020c; Messick, 1995). Additionally, it has to do with problems with test administration and fairness in aggregation (Baccalaureate, 2020; Crooks et al., 1996). Messick (1995) convincingly argues that construct irrelevance can also be described as "construct-irrelevant difficulty" (i.e. factors unrelated to the underlying construct that inappropriately make the task more difficult for some test takers, typically resulting in invalidly low scores for these people) and "construct-irrelevant easiness" (i.e. due to extraneous factors irrelevantly reducing the difficulty of the task, usually leading to invalidly high scores). When evaluating students' performance, these two types of construct-irrelevant variance could manifest themselves simultaneously.

Cureton (1950) described test validity as reflecting "how well the test meets the purpose for which it is used" (Cureton, 1951, p. 621). The assessment objectives of the IB DP mathematics are therefore outlined in Table 1 below, and are taken from the IB DP mathematics guide for the 2022 examination sessions (IBO, 2019a, 2019b).

**Table 1**

**List of the Assessment Objectives from the IBDP Mathematics Guide**

<b>Assessment Objectives</b>	<b>Problem solving</b> is central to learning mathematics and involves the acquisition of <b>mathematical skills</b> and <b>concepts</b> in a wide range of situations, including non-routine, open-ended, and real-world problems.
Knowledge and understanding	Recall, select, and use their knowledge of mathematical facts, concepts, and techniques in various familiar and unfamiliar contexts.
Problem solving	Recall, select, and use their knowledge of mathematical skills, results and models in both real and abstract contexts to solve problems.
Communication and interpretation	Transform common realistic contexts into mathematics; comment on the context; sketch or draw mathematical diagrams, graphs, or constructions on paper and using technology; record methods, solutions, and conclusions using standardized notation.
Technology	Use technology accurately, appropriately, and efficiently both to explore new ideas and to solve problems.
Reasoning	Construct mathematical arguments through precise statements, logical deduction and inference, and the manipulation of mathematical expressions.
Inquiry approaches	Investigate unfamiliar situations, both abstract and real-world, involving organizing and analysing information, making conjecture(s), drawing conclusions, and testing their validity.

Maul (2013) reconceptualized the underlying constructs used as composite variables in the curriculum-based assessments (Maul, 2013). IB DP mathematics assessment objectives, particularly for the mathematics internal assessment, comprise the correlations and interlinks of the variables or subconstructs, as shown in Table 1, while serving various mathematics purposes as composite variables (IBO, 2019a; Messick, 1984), which may

require fewer knowledge, understanding, and inquiry approaches and problem solving, and demand more communication and interpretation. These different and contradicting assessment objectives, to language learners, fail to include the fundamental construct of the IB DP mathematics courses and taint the scores with construct-irrelevant variance. In other words, IB DP mathematics IA may not significantly assess students' problem-solving skills but rather their English writing skills in the context of English language learners. Consequently, it seems highly questionable that the mathematics IA serves the stated purpose of DP mathematics assessment if a candidate's first language is not English (Obuon, 2019).

In addition, the test scores' interpretation or meaning can be inferred from the construct validity by integrating relevant criterion-related evidence (Messick, 1995). The IA test scores are marked and moderated using assessment criteria, the aspects of which might fail to include important facets of the focal construct. Meanwhile, IA might contain excess variance associated with other distinct constructs known as “construct-irrelevant variance” (Messick, 1995). The distinct constructs of IA therefore need to be further investigated in this study. This dissertation, in my opinion, contributes by offering a more thorough insight into the construct of DP mathematics IA as part of a wider practical framework for students from different cultures.

### **1.3 Research Aims**

This thesis examines the construct validity and fairness of the IB DP mathematics internal assessment from the perspectives of the students and teachers at an IB school, the United World College, in China. Emphasis was placed on the contribution to a more diverse culture. Its conclusions may help in recognizing IB assessment goals and in improving internal assessment implementation through policy, pedagogy, and other internal assessment-related changes in future. Following this introductory chapter, the literature review chapter outlines the context and relevant literature on this topic and discusses the research gaps. Chapter 3 presents the overall methodology of the thesis. Chapter 4 contains the findings of this study. Chapter 5 interprets these findings and outlines the implications and limitations of this research. Finally, the conclusion chapter highlights this study's contribution and key findings.

## **Chapter 2: Literature Review**

The purpose of this chapter is to situate this dissertation in a broader theoretical and conceptual framework in the field of the construct validity and fairness of the DP mathematics internal assessment. This chapter mainly reviews the previous literature that has quantified how the internal assessment affects the construct validity in the context of IB DP. There has not been much research done in this field of study. This is followed by a review of the validity and how the literature informs the IB DP mathematics internal assessment practice. This chapter aims to provide the theoretical and empirical foundations to guide the overall research design of this study.

The first section of this chapter defines the internal assessment in wider contexts and defines the framework used in IB DP. The second section of this chapter illustrates the limitation of previous studies that explored the validity, reliability, and fairness of the DP mathematics IA. The third section reviews the previous literature regarding perspectives of teachers and students on the internal assessment procedures. Following that, the effect of internal assessment is reviewed in terms of different influential factors. Reflections on, and the limitations of, the previous research are presented in this chapter's final paragraph of each part.

## **2.1 Defining Internal Assessment**

### **2.1.1 The Nature of the Internal Assessment.**

Previous research has discussed the nature of the IA on educational assessments from the varied curriculums implemented by different countries (Barrance, 2019; Brown, 1999; Luijten, 1988; Vitello & Williamson, 2017; Ziderman, 1984). The impact of the internal assessment has been mentioned in England (Vitello & Williamson, 2017), Dutch secondary school examinations (Luijten, 1988), and the Israeli Maturity Certificate (Ziderman, 1984). Notably, the main issues with internally assigned tasks are that: internal assessments might not be as valid and reliable as externally assigned tasks (Wolf, 2011); task completion may involve unethical behaviour, such as plagiarism and cheating; and students may receive useful assistance and advice from parents and teachers during a task. Barrance (2019) also made a note of an aspect of the assessment environment that allegedly has a substantial impact on students' performance regarding fairness whilst employing internal assessment for GCSE curriculums. However, it appears that there is a dearth of literature that discusses the internal assessment in the context of the International Baccalaureate Diploma Programme.

### **2.1.2 Internal Assessment in Context of International Baccalaureate Diploma Programme.**

Concerns have been voiced regarding the quality and manageability of Bunnell (2011) comprehensive evaluation of the International Baccalaureate Diploma Programme in this context. A major challenge for the future will be maintaining and managing a consistent and reliable assessment due to the current and anticipated growth of IB schools internationally (Bunnell, 2011). The reasons for the challenges are the recruitment of unqualified or undiligent examiners (Toze, 2008), the issues of "dualism" (Matthews, 2009), and a more diverse set of international schools. Due to "dualism" issues, the IB struggles to both satisfy its current customers of foreign schools and innovate in order to draw in new clients. In other words, the problem could be that the IB's organizational culture has undergone a lot of changes. It has evolved into a more corporate, business-oriented product distributor. As a result, the IB runs the risk of losing the loyalty of its current clients. However, the language component of the DP internal assessment has been overlooked in the IB DP assessments. Indeed, Maher et al. (2018) argued that to some extent mathematics assessment relies on language proficiency, while Barwell (2009) also remarked that mathematics is not a non-verbal topic because it depends heavily on language, both in oral and written forms (Barwell, 2009; Moschkovich, 2015). Similarly, Ní Ríordáin and O'Donoghue (2009) noted that "mathematics is not 'language free'" (p. 47). There was some evidence in the previous literature of a positive relationship between the ways that mathematical concepts were developed, and the language used to express these concepts (Ní Ríordáin & O'Donoghue, 2009; Vukovic & Lesaux, 2013). For example,

Barwell (2009) highlighted an example in a chapter written by Khisty and Chval (2002). It examined the writing practices of primarily Latino/a students in a mathematics classroom in the United States. The teacher in the classroom did not lower the expectations but kept a high level of mathematical language for those students who were developing English language proficiency. Feedback was used to advance students' current levels of mathematics writing forward. It was intriguing to see how this type of instruction can positively affect students' linguistic identities and their self-perceptions as practitioners of mathematics. A total of 199 maths lessons were audio-recorded, and 60 of those lessons were observed and carefully documented through field notes. The findings of the study provided evidence that writing plays a significant role in mediating students' learning of mathematics (Galbraith, 2009).

These findings suggested a learning process whereby teachers help students "give voice" to their problem-solving thoughts and actions. As a result, language supports the expression of mathematical concepts in a wide range of forms, including terminology. Language is important for learning and teaching mathematics in both oral and written forms, but it can also have a big impact on how valid mathematics assessments are. Students are expected to demonstrate their knowledge in standardized tests, which frequently include assessments buried in lengthy texts, such as complex linguistic puzzles (Maher et al., 2018).

## 2.2 The Validity of Mathematics Internal Assessment

The definition of "construct" is "a term used in psychology to label underlying skills or attributes" (Gipps, 2011). To ensure that the internal assessment (IA) examines the attribute it is supposed to, and that it is "valid", the construct being assessed is defined before the IA is developed. "Validity is about whether the form of assessment is fit for purpose and whether it accomplishes its purpose, including how the assessment grades are interpreted and used" (Gardner, 2012, p. 189). Messick (1995) claimed that "validity is an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other assessment" (p. 741).

As explained by Gipps (2011):

Construct validity itself relates to whether the test is an adequate measure of the construct, that is, the underlying (explanatory) skill being assessed. Important to the development of an assessment, then, is a clear and detailed definition of the construct; .... a full definition of reading as a construct would include reading aloud, reading comprehension, accuracy, and enjoyment of reading.

Messick (1995) stated that "construct validity comprises the evidence and rationales supporting the trustworthiness of score interpretation in terms of explanatory concepts that account for both test performance and score relationships with other variables". In IB DP mathematics internal assessment, construct validity is whether the IA is a sufficient measure of the extent to which the IA evaluates students' fundamental abilities to satisfy the assessment objectives of DP mathematics courses.

Threats to validity may come from construct-irrelevant variance or construct under-representation (Messick, 1995). The five criteria of presentation, mathematical communication, personal engagement, reflection, and use of mathematics are used to evaluate the students' internal assessments (IBO, 2019a, 2019b). However, in addition to the sub-objectives listed in Table 1, the IB guide (IBO, 2019a) designates problem-solving skill as central. The IA criteria's compatibility with the broader mathematics assessments is therefore in doubt. Newton (2007) reported three identified levels of "assessment purposes", namely the judgement level, the decision level, and the impact level (p. 150). The judgement level concerns the technical aim of an assessment. The purpose of the judgement level in IA is to derive a standards-referenced judgement, expressed as a score on a scale from 1 to 20. Each level descriptor is worth one or more marks. Each criterion is applied independently using a best-fit model. A best-fit approach means that compensation should be made when a piece of work matches different aspects of a criterion at different levels (IBO, 2019a). The mark awarded should be one that most fairly reflects the balance of achievement against the criterion. Teachers need to make a judgement on which grades shall be given based upon the level descriptors for each criterion until the teachers reach a descriptor that most appropriately describes the level of the work being assessed. In other words, it implies that the interpretation and content of the criterion could be evident as a construct-irrelevant variance. Newton (2016) also concluded there were five sources for conceptualizing validation evidence and analysis: Evidence based on (1) test content; (2) response processes; (3) internal structure; (4) relations to other variables; and (5) the validity and consequences of testing. The test content of the IA is selected by the students, and advice is provided by the subject teachers on their selected topics. The

selection of the topic generalizes the earlier notion of “content validity”. Meanwhile, the response process goes one step further by investigating the extent to which the proficiency specification faithfully represents responses. If IA is a written task that is confusingly wordy and students do not know how to write the understanding of mathematics in English, it might be a serious threat to validity that may involve other distinct constructs that affect responses in a manner irrelevant to the interpreted construct.

Similarly, Shuard and Rothery (1984) demonstrated the increased complexity of learning the language of mathematics as opposed to everyday English. When evaluating students' language of mathematics for the internal assessments, mathematics literacy frequently appears in the debates and discussions (Adams, 2009; Moschkovich, 2015; Roth et al., 2015). Students are expected to read in a mathematical context and correctly write mathematical derivations with understandings. Mathematics literacy is invoked in the sense of encompassing or organizing various processes. It is argued that "mathematics literacy is a composite variable, of which language is a dimension of the construct" (Letwinsky, 2017). Mevarech and Stern (1997) investigated the effects of sparse versus real contexts on the understanding of abstract mathematical concepts, and results of experiments showed that a sparse context facilitated understanding of abstract mathematical concepts more than real contexts. Similarly, Kotovsky et al. (1985) compared performance in the Tower of Hanoi problem-solving test in sparse versus enriched contexts and discovered that the latter was more challenging. Furthermore, Ahmed and Pollitt (2007) found that mathematics problems in context can be more difficult to solve than context-free problems in mathematics questions. This corroborates Messick's (1995) claim that "skill in

communicating mathematical ideas might well be considered irrelevant variance in assessing mathematical abilities" (p. 743). Therefore, real-world enriched context and abstract mathematics make up the validity of mathematics IA, and it is still unknown how challenging or complex this construct might be for a diverse culture group.

A further extension of language bias occurs when "construct-irrelevant variance is deliberately capitalized upon to produce desired social consequences, as in score adjustments for minority groups, and within-group norming sliding band procedures" (Traub & Rowley, 1991). There are a lot of non-native English speakers enrolled in IB courses. As a result, students might try to translate the texts from their original tongue into English while conducting maths inquiries, which presents issues concerning transadaptation while looking at mathematics assessment literacy (El Masri et al., 2016). In addition, teachers might provide feedback in the native tongue rather than English, which would make it harder for students to understand mathematics for English language learners (Lewkowicz & Leung, 2021). Consequently, the validity and fairness of the mathematics assessment would be compromised, and students would do worse than they should as a result of their language use. As a result, a number of studies have identified the existing validity issues in designing assessments (Black & Wiliam, 2018; Crooks et al., 1996; Pellegrino et al., 2016), approaches to learning underpinned by a monolingual bias that delegitimizes the experience of multilingual learners (Huckle, 2021), and fairness concerns about the authorship of students' work (Barrance, 2019).

## 2.3 The Reliability of Mathematics Internal Assessment

This section evaluates the reliability of internal assessment (Adams, 2005; Baird & Black, 2013). The reliability of student performance is related to its consistency (Gipps, 2011). When students perform consistently across various test settings, formats, and markers, an assessment is said to be reliable (William, 2009). Inter-rater reliability is the degree to which the marks awarded to the candidate by two or more examiners agree (Traub & Rowley, 1991).

Five criteria – presentation, mathematics communication, personal engagement, reflection, and use of mathematics – are taken into consideration while awarding marks in the DP mathematics IA (IBO, 2019a, 2019b). However, each teacher's interpretation of the criteria will be unique and subjective. Additionally, various IB teachers may operationalize their IA of a topic in different ways, thereby giving some aspects more weight than others. Different instructional interventions can be given to students whose mathematics IA results in marker errors. Additionally, rather than using traditional mathematics, the DP mathematics IA curriculum uses essay-style questions. Murphy (1982) asserted that questions marked in the form of essays might not be as accurately marked as questions marked in an organized, analytical manner. This is due to the subjective nature of the judgments made by the various markers and the complicated marking method (Gipps, 2011). The IA is affected as a result of teachers' inconsistencies. Due to this, the reliability and fairness of the DP mathematics IA are negatively impacted. However, Black and William (2012) asserted that domain-sampling errors can also put reliability in jeopardy.

The IA topic chosen for grading may have an impact on how well students do. The test's reliability is higher the more domains it includes in its sample (Johnson, 2013). When IA topics are limited to content that concentrates on a certain area of the mathematics curriculum rather than the full test syllabus, domain-sampling errors can also occur. The purpose of this study does not extend to evaluating the reliability of the mathematics IA. As a result, this study looks into cross-cultural fairness.

## **2.4 The Fairness of Mathematics Internal Assessment**

The fairness of IB DP Mathematics internal assessment can be threatened because of academic misconduct (Barrance, 2019; IBO, 2019c), generalizability (Lane et al., 1996), and cross-cultural language effects (El Masri et al., 2016). First, ensuring that the candidate is not engaged in academic misconduct, such as having someone else write their IA, is one of the key challenges during IA. The teacher has the challenging task of identifying which applicants are not presenting their own work as students compose their IAs after class. Second, the IA processes used by various IB institutions to generalize students' grades vary (IBO, 2015a, 2020b). The IB assigns students comparable ratings that must be combined depending on the historical performance of schools and other modifications (IBO, 2018; Lane et al., 1996). Therefore, a good student at a lower-ranking IB school may score better than another average student from a higher-ranking school, but the student from the underachieving school is put at a disadvantage when it comes to aggregation. It should be highlighted that for English language learners worldwide, these frameworks are primarily

built from instructors' and moderators' practices and viewpoints, with students' comments on the internal assessment being neglected.

## **2.5 Prior Empirical Studies on the Viewpoints of Students and Teachers**

From students' and teachers' perspectives, academic integrity and other crucial factors affecting student success have also been covered in previous studies on the International Baccalaureate Primary Years Programme (Bahn, 2022), Middle Years Programme (Storz & Hoffman, 2018), and Diploma Programme (Burbidge & Hamer, 2020; Culross & Tarver, 2011). As an illustration, Taylor and Porath (2006) stated that graduates spoke highly of the IB curriculum. The IB graduates believed that the extensive curriculum they were exposed to and the time management and critical thinking abilities they acquired were well worth the additional work needed to obtain an IB diploma. Additionally, students believed that their IB experience had prepared them adequately for university study. However, in addition to highlighting the IB DP's key components that are thought to affect student success, this thesis also provides a summary of how the research agenda developed, including how each study in the series led to new inquiries about the IB DP and how what started as a microstudy within a single school evolved into an analysis of the IB DP in a wider context (Culross & Tarver, 2011).

Walker et al. (2016) helpfully demonstrated strong measurement reliability and construct validity in addition to content validity from investigations and cross-validation with various

populations. Researchers think that the creation of the IB Learner Profile questionnaire (IBLPQ) will support educators interested in examining the connection between student outcomes and the Learner Profile in the Asia-Pacific region, which is home to an increasing number of IB schools. Researchers expect the IBLPQ to become a crucial first resource for a variety of stakeholders, which is significant given the current lack of systematically created and proven survey tools to measure the Learner Profile. Additionally, Burbidge and Hamer (2020) validly noted that a large portion of the existing research on academic integrity involves student questionnaires. In their study, survey responses from students, teachers, and school administrators were compared from institutions in 76 different nations. The surveys asked them about their knowledge about, attitudes toward, and perceptions of academic honesty as well as their experiences in school. According to the findings, academic integrity is mostly taught to students in the classroom. However, it appears that there is a dearth of material that discusses the validity and fairness of the internal assessment within the context of the International Baccalaureate Diploma Programme from students' and teachers' perspectives.

## **2.6 Research Questions**

The discussion above enables the development of a fictitious set of variables that may be significant to UWC China students and faculty, and which may have prompted the concerns that were raised with the validity and fairness issues of IB DP mathematics internal assessment. The following key research questions are, therefore, to be considered:

1. To what extent is the internal assessment valid and fair in the IB DP mathematics assessment? What are the relationships between IA scores and scores from the mathematics exam, the English writing assessment, and other IB subject exams such as the science subject exams in 2022?
2. What are the students' perspectives on the DP mathematics IA? How did the students perceive the validity and fairness of the DP mathematics internal assessment in 2022?
3. How did the teachers understand the validity and fairness of the IB DP Mathematics internal assessment?

The next chapter explains the methodology that was used to answer the research questions.

## **Chapter 3: Methodology**

### **3.1 Research Aim**

The study investigates the construct validity and fairness of IB DP mathematics internal assessment. To address the three research questions in this study, an explanatory mixed-methods research design was used (Hanson et al., 2005).

The quantitative approach was used in this procedure to examine the construct validity of the internal assessment in relation to the assessments from the mock exams and other DP courses. The fairness of the IB DP mathematics internal assessment was further examined using the qualitative technique from the perspectives of students and teachers. The qualitative data sought to provide different perspectives and answers to pertinent research questions, whereas quantitative data cannot elaborate (Creswell & Creswell, 2018). The aim of investigating these connections was to better understand the mathematics IA from the perspectives of UWC students and teachers, including their perceptions on how the mathematical IA is constructed, and the extent to which the mathematics IA measures the students' fundamental abilities in line with the IB assessment goals (IBO, 2019a). Additionally, the validity of the mathematics IA was investigated to determine whether it was fair for students who were learning English as a second language. Due to the COVID-19 pandemic in China, the DP May external assessments were cancelled in 2022. Instead, the validity of the DP mathematics courses was investigated using the results of the March mock examinations.

## **3.2 Research Design**

### **3.2.1 Overall Design.**

The research was conducted at an international school that offers the IB DP using a mixed-methods approach with an exploratory sequential design. Johnson and Onwuegbuzie (2004) helpfully found that a "mixed-methods research" study is one that combines quantitative and qualitative research methodologies, methods, approaches, concepts, and terminology. In terms of philosophy, mixed-methods research applies a pragmatic approach and philosophical system. Its method of investigation involves the use of induction, deduction, and abduction (De Waal, 2001). A study using mixed methods combines quantitative and qualitative research techniques (Gay et al., 2006). The provided study can be carried out using a variety of methods, but in order to choose the best approach, it is typically best to focus on only one kind of study. The value of mixing qualitative and quantitative research approaches in one study has been hotly contested since the 1960s (Creswell & Creswell, 2018; Curlette, 2006; Tashakkori & Teddlie, 1998; Thomas, 2003). Creswell and Creswell (2018) helpfully found that methods are strategies for data collection and analysis, such as a quantitative standardized instrument or a qualitative theme analysis of text data (Hanson et al., 2005).

Jackson (2017) suitably found that the advantage of using a mixed-methods study is that combining the two methods improves comprehension of the research findings. For instance, some of the remarks given by the students in their open-ended qualitative responses can be used to further clarify quantitative data. Using this method, a researcher can generalize from quantitative data and make analysed and enhanced generalizations further through

open-ended questions through qualitative responses. The technique also examined data showing discrepancies and even conflicts between multiple data sets (Lee & Greene, 2007).

In this thesis, the intention of the study's qualitative phase was to delve deeper into the drivers of students' and teachers' opinions on DP mathematics IA coursework, while in the quantitative phase there was an attempt to identify patterns in the student data connected to mathematics IA scores. Therefore, despite the fact that the qualitative phase of the study was given greater attention because of its more thorough analysis of the reasons for students' and teachers' opinions about mathematics IA, the quantitative phase of the study nevertheless responded to findings from the qualitative phase with comprehensive and practical foundations. Both quantitative and qualitative analysis can be used to answer the research questions. The COVID-19 pandemic was at its worst in April 2022, therefore the data for this study were gathered in stages. The qualitative data were first obtained and then analysed; the quantitative data were only gathered after the IB released the students' subject and component results in July. The results of the quantitative phase were used to enhance the qualitative phase. Both qualitative and quantitative approaches are combined to construct the discussion chapter for the thesis.

### **3.2.2 Quantitative Research Method.**

The philosophical stance of natural scientists is known as "positivism", which is characterized as a mode of inquiry based on realist ontology by Morcol (2001). Gill and Johnson (2010) point out that positivists view social reality as an element that exists outside of people and is wholly unrelated to people in their social networks. Researchers and academics who focus on maintaining an objective perspective by using logical justification and scientific methods frequently use the positivist approach (Hughes & Sharrock, 2016). Therefore, positivists contend that empirical data and observable facts should be employed in research to draw generalizations that are comparable to laws in order to maintain the precision and clarity of knowledge (Saunders et al., 2009).

Seale (1999) convincingly argues that when it comes to the analysis of particular subjects, positivists aim to use statistical measurements, value neutrality, and quantitative aspects in their research. Collis and Hussey (2009) helpfully corroborated Seale's findings in another study by proving that positivism is typically associated with quantitative research techniques when it comes to data analysis. Researchers that adhere to positivism frequently present testable questions and hypotheses. This method typically involves comparing variables or correlations to identify causal linkages. Burrell and Morgan (2016) validly proved that positivism focuses on identity and relationship patterns in order to anticipate the future and create scientific rules.

### **3.2.3 Qualitative Research Method.**

Contradictory theories pertaining to the philosophy of "positivism" can be found in the interpretivism school of thought. As a result, every philosophical viewpoint takes a completely distinct stance whilst doing investigations and research. Myers (2015) validly proved that social constructs are the only ways for research to have access to reality. The reality is also said to be socially dependent on things like common connotations, languages, and cultures. Similarly to this, Grix (2010) helpfully found that interpretivism-based researchers think that human interactions shape the world.

Guba and Lincoln (1994) assert that interpretivist scholars frequently hold to a relativist ontology. Denzin and Lincoln (2011) helpfully corroborated Guba's findings in another study by proving that positivism is constrained and subjective, therefore different people experience it in different ways. Additionally, when it comes to ideas, interpretivism alludes to subjectivist epistemological viewpoints. Levers (2013) convincingly argues that interpretations from specific persons should therefore be given a lot of weight. Likewise, interpretivist researchers believe in a diversity of truths that conflict with one reality, including universalism. As a result, Saunders et al. (2009) usefully claimed that researchers frequently object to generalizations like laws. Creswell (2003) applicably argues that interpretivists typically seek to develop new ideas based on information from qualitative research.

Finally, positivists commonly use statistical analysis and quantitative observations in their research. Some scholars point out that the empiricist viewpoint refers to human knowledge derived from considerable experience, which largely supports positivism referring to

Collins (2010). In contrast to positivists, interpretivist researchers focus primarily on the qualitative research approach, with a focus on both human experiences and cultural objects. Furthermore, positivists are committed to including their justifications in every step of the learning process.

The positivist concept serves as the foundation for this study, which will combine a qualitative and quantitative analytical method. Correlational analysis will be used by the researcher to evaluate a variety of hypotheses and provide answers to research questions.

### **3.3 Sampling and Participants**

Participants comprised 12th-grade IB students (54.7% female) and subject teachers from an IB school in China who took part in the study. In phase one, a total of 18 interviewees, consisting of 13 students and five teachers, participated in the interviews. Thirty-six students participated in the online survey (see Table 2). In phase two, 135 students signed the consent forms and agreed to share their test results, including IB subjects, and component scores were obtained and analysed for this research's quantitative analysis at United World College Changshu China. This institution was selected because the researcher was one of the teachers at UWC China. The fact that UWCs accept students from a variety of backgrounds and experiences was a major factor in my decision to select this school. In China, diversity includes variations in ethnic groups and nations, socio-economic backgrounds, and cultures. The majority of the pupils have travelled abroad to English-speaking nations. Changshu was where the school was situated. Although the

school offers needs-based scholarships to pupils from underprivileged homes, the tuition for two IB DP years cost between USD110,000 and USD120,000. The majority of the students chosen by national committees – more than 80% – receive full or partial scholarships.

However, 135 targeted DP Year 2 pupils could have completed the online survey that was originally scheduled. Due to the COVID-19 epidemic, China went into a prolonged state of "lockdown" in February 2022. From that time until May 2022, UWC Changshu China continuously practised closed-loop management. After receiving notification from the IB that all schools in East China would follow the non-examination protocol for May 2022, DP2 pupils were promptly instructed to vacate the premises in April 2022. Because of this, it was challenging to survey DP2 students online when they were not on campus. Due to this "early graduation", only 36 survey respondents (26% of the overall DP2 student body) responded to the online survey.

### **3.3.1 Participants' Characteristics.**

Among the participants, 99% were Chinese citizens due to the zero-COVID policy that was put in place in China (Normile, 2021), and international students were not given a visa to attend UWC Changshu China between 2021 and 2022. Within the sample, 99% of individuals identified as Chinese citizens, 2% as Chinese Americans, and one as an overseas student. Table 2 displays the participant characteristics broken down by gender.

**Table 2****Participant Characteristics by Gender at United World College, China**

Gender \ Year	Male	Female	Total
<b>DP Year 2</b>	61	74	135
<b>Survey participants</b>	16	20	36
<b>Interview participants</b>	8	10	18

The participant characteristics for the various levels and subjects of DP mathematics courses are summarized in Table 3. The mathematics classes these students were enrolled in had two modules: Analysis and Approaches (AA) and Application and Interpretation (AI). Students who select AA find it easy to manipulate algebraic formulas and enjoy spotting patterns and comprehending how to generalize these patterns mathematically. AI, in contrast, focuses more on applying mathematics to real-world situations and addressing real-world problems (IBO, 2019a, 2019b). Participants were split between the two groups, with 40% enrolling in the standard-level mathematics course and 60% in the higher-level course. Higher-level mathematics candidates possess more fundamental algebraic skills, enjoy working on problems for extended periods of time, and find delight in solving difficult problems in either pure mathematics or contexts that are relevant to their daily lives (David, 2007).

**Table 3**

**Participant Characteristics by Level and Subject**

Mathematics IA Coursework	Standard Level	Higher Level
Analysis and Approaches	30	64
Application and Interpretation	24	17

Teachers of mathematics were invited to take part in the semi-structured interviews. The maths teachers were colleagues of the researcher, but the difficulty was that teachers had heavy workloads at school and frequently put in far more time than was required to ensure the success of their students. Along with the demanding work hours at UWC, in-depth interviews were performed with five teachers who agreed to participate in the interviews. One limitation of the results may be that those five interviews cannot fully capture all the teachers' perspectives at UWCs. Being an insider, it was important to schedule and manage time for both teachers and students, as well as to get along well with colleagues. As a result, this thesis was written in appreciation of their vast contributions as my interviewees.

### **3.4 Instruments**

An online survey, semi-structured interviews, and statistical analysis of students' subject and component results in various IB subjects were the instruments used in the study. The online survey investigated the students' perspectives on assessment criteria with an evaluation of the validity of mathematics IA. To better answer the research questions, the interview explored how the mathematics IA has affected the validity and fairness of the

IBDP mathematics assessments. Students and teachers were questioned about how they perceived the mathematics IA and whether they thought it was consistent with mathematics exams and other IB disciplines like science under the IB DP coursework. Teachers were also questioned about obstacles they may have faced in promoting internal assessment more effectively in their lessons.

### **3.4.1 Survey.**

A survey provides the advantages of standardized and open responses from a large sample or population to a variety of topics (Cohen et al., 2018). An online self-assessment survey (see Appendix 3) was given to students in order to elicit their evaluations of their academic development and achievement in the mathematics IA exam in the last semester of DP Year 2 in 2022. The online survey includes 11 questions of a multiple-choice and Likert-scale type on a five-point scale and one open-ended question at the end of the study.

In the first section of the survey, students were asked to rate various aspects of their demographics and performance on a five-point scale, including their age and gender, English language proficiency, opinions on the IA's criteria, and their self-evaluated academic performance in accordance with the criteria. The second portion of the survey asked students to respond using a Likert scale and describe the extent to which the assessments for other subjects overlapped with the IA for mathematics in their academic work. The subsequent questions asked for students' perspectives on the extent to which the maths IA was graded fairly via an English written task to demonstrate mathematical skills

as English learners at UWC. Students could write about their views on whether the design of the IA is a good assessment to evaluate their mathematics ability and areas to improve. In total, 36 students responded to the online survey because of the COVID-19 pandemic and school restrictions.

### **3.4.2 Semi-structured Interview.**

The method used was a semi-structured interview since it gave participants the chance to explore conversational subjects and ask follow-up questions (Brinkmann & Kvale, 2015). The main purpose of the semi-structured interview was to investigate teachers' and students' in-depth comprehension and interpretation of the validity and fairness of the mathematics internal assessment. Some of the one-to-one interviews with students were conducted online via Zoom due to school closure.

The literature review in Chapter 2 served as the foundation for the development of the semi-structured interview protocols. All interview questions were developed to investigate the validity and fairness of the IA assessment, particularly how well the mathematics IA fits for assessment purposes and how much other IB subjects' skills overlapped with those of mathematics. Leading questions that might force an interviewee's answer in the interview questions were avoided.

### **3.4.3 Mathematics Internal Assessment and Mock Examination Scores.**

The data from the results of the mathematics internal assessment and mock examinations make up the quantitative portion of the research in this thesis. Scores from mock exams and internal assessments were gathered from academic departments at UWC China. To investigate the construct validity of the mathematics IA coursework, the scores for the IB science- and English-related courses were also gathered.

### **3.5 Pilot**

A comprehensive pilot test was carried out before the data collection phase. It is reported by Cohen et al. (2018) that pilot testing contributes to improving the questionnaire to ensure accuracy without instructions from the researcher. It assists in avoiding potential issues when conducting data analysis following the collection of the questionnaire. In this thesis, Internet surveys and interviews were piloted by IB graduates I taught previously. I contacted them personally via the online platform Zoom. Three participants, aged over 18, completed the pilot survey and interview, which took them about 30 minutes. Former IB students who had successfully completed the Diploma Programme were asked for their opinion on the interview questions for the prospective students. The interview questions were revised to incorporate more specific questions that better addressed the study issues based on feedback from three pilot participants. For instance, the pre-pilot version's first question on the student interview schedule was overly general: "How do you find the internal assessment course? Can you elaborate?" The question was changed to a more

focused one in the post-pilot version based on pilot participant comments: “What skills have you developed through the mathematics internal assessment?”

Interviews questions were revised according to pilot participants’ opinions and feedback. One interview protocol was developed specifically for subject teachers, and the other one was modified for Diploma Programme final-year students. Teachers were questioned in interviews about their views on instruction and the degree to which the mathematics internal assessment construct is applicable to the science assessment and the English assessment. The ethical clearance in Appendix 4 was attached and received from the University of Oxford (CUREC Approval).

### **3.6 Data Collection**

Data collection was carried out at UWC Changshu China in the period 2021–2022. Clearance was granted by Oxford University's Central University Research Ethics Committee (CUREC). Thereafter, the principal of UWC China granted his approval for the data collection at the end of February 2022. The parents of the students were contacted through emails to obtain their permission with consent forms and informational pamphlets. The consent forms were completed online by the end of March 2022. In the meantime, a major proportion of student interviews were performed using the online platform Zoom. Each interview took around 30 minutes and was audio-recorded. Immediately after the interviews had been conducted, they were transcribed. I conducted more than five in-person interviews with subject teachers at UWC China in May 2022. Based on the developing

themes and significant findings for the teacher and student interviews, I analysed the interviews to answer the research questions. Additionally, the recruitment process involved sending out emails to 135 students after a presentation was delivered regarding the validity and fairness of mathematics IA in a cohort meeting. The online survey was sent out by the researcher. It would take approximately 15 minutes to complete. Meanwhile, all data pertaining to the March IB mock examinations, including Papers 1, 2, and 3, were gathered. So was the information regarding their internal assessment scores. The study's results were analysed and discussed for the first draft of the dissertation by the end of July 2022 in relation to earlier literature.

### **3.7 Data Analysis**

Both quantitative and qualitative methods were used in the study. First, the qualitative analysis involved conducting interviews to gather viewpoints from students and teachers. Data were categorized and organized using the thematic framework in accordance with important themes, concepts, and emerging categories (Ritchie et al., 2014). The themes were chosen in accordance with the study questions as well as the labelling, sorting, and synthesizing processes, which deepened understanding of the available evidence. As a result, each research has a unique thematic framework made up of a number of major topics that are further separated into associated components and aspects as seen from the perspectives of the interviewees. Second, to ascertain the construct validity of the internal mathematics assessment, quantitative data were analysed using bivariate correlations, and analysis of variance (ANOVA) interpretation of student records from IB Diploma

Programme Year 2 students in the 2021–2022 academic year ( $n = 135$ ). The information was made up of responses from the online survey, and that will be added to the results of the students' test scores. The participants' replies were tabulated, their frequencies were calculated, and their statistical significance was evaluated using the Statistical Package for the Social Sciences (SPSS) computer program. For use with various statistical techniques, such as the examination of bivariate correlations and variance, data were converted to the SPSS. The extent to which the construct of the mathematics IA has been applicable to mathematics examinations, English examinations, and science examinations was determined by the Pearson correlations between the internal assessments and mock examinations produced by the SPSS software.

### **3.8 Ethical Considerations**

Cohen et al. (2014) valuably found that ethical issues are widely recognized in research and may be found "in the expanding relevant literature and regulatory codes created by multiple organizations" (p. 75). Similarly, Kumar (2011) helpfully contends that most professions are bound by a code of conduct that governs how research must be conducted. The research participants are present in both qualitative and quantitative research. It is necessary to look at the ethical concerns in relation to them. This study follows the British Economic and Social Research Council's (ESRC) guidelines for ethical research.

A research proposal was submitted to the CUREC of the University of Oxford, and formal application approval for this research to proceed is attached in Appendix 4. I have gained

approval from the principal of UWC Changshu China. A total of 135 participants agreed to sign the consent forms and the parents of participants who had consented to participate received the consent forms and information sheets through emails (see Appendices 5 and 6). The information sheet covered the study's purposes and rationale, research questions, data collection techniques, intended uses for the data, and confidentiality policies. Only people who received and signed the consent forms indicated their willingness to take part, and they were allowed to do so. Participants were kept anonymous throughout the research procedure, and only the researcher and supervisor had access to the raw data obtained. The researcher emphasized before the data collection that participants could withdraw and have their data excluded whenever they felt anxious, uncomfortable, or embarrassed and no longer wanted to participate.

## Chapter 4: Findings

### 4.1 To what extent is the internal assessment valid and fair in the DP mathematics assessment? What are the relationships between IA scores and scores from the mathematics exam, the English writing assessment, or other IB subject exams such as science subject exams in 2022?

Statistical analysis was performed to investigate students' performance in the DP mathematics internal assessment. The sample included 54 standard-level students and 81 higher-level students from two different subjects: mathematics Analysis and Approaches (AA) and mathematics Applications and Interpretation (AI) (see Table 4). The overall total score for mathematics IA is 20 points. Mathematics IA scores ranged from 7 to 17 for students at the standard level and from 3 to 18 for students at the higher level.

**Table 4**  
**DP Year 2 Students' IA Scores by Level and Subject in 2022**

<b>Mathematics IA Coursework</b>	<b><i>n</i></b>	<b><i>M</i></b>	<b><i>SD</i></b>
Analysis and Approaches Standard Level	30	11.67	2.454
Analysis and Approaches Higher Level	64	11.66	2.993
Application and Interpretation Standard Level	24	11.29	2.758
Application and Interpretation Higher Level	17	9.88	2.446

*Note.* N = 135. Students' total scores in the internal assessment from a maximum available mark of 20.

#### **4.1.1 The Relationship between Mathematics IA and Final Grades.**

The correlation coefficient between mathematics final grades (scaled overall IB grades) and mathematics IA scores was 0.539, which denoted a moderate to high level of validity. The strong association indicates that students' overall performance in the mathematics IA can account for 29.1% of the variance. In other words, mathematics IA results approximately predicted the IB DP mathematics assessment in light of the cancellation of the final exams.

#### **4.1.2 IA Scores Sorted by Subjects and Levels.**

There were no notable disparities between the groups, with the exception of the subjects of the IA scores, according to the two-way analysis of variance (ANOVA). The analysis was performed on internal assessment scores using two subjects and two levels of mathematics courses (higher level and standard level) (AA, AI). At the .05 level of significance, there was no main effect of mathematics course levels on IA scores,  $F(1, 135) = 1.750, p = .188$ , indicating that IA mean scores for higher-level mathematics ( $M = 11.28, SD = 2.963$ ) and standard-level mathematics ( $M = 11.50, SD = 2.575$ ) did not differ significantly. There was a main effect of subjects (AA versus AI) on IA scores,  $F(1, 135) = 1.974, p = .121$ . The post hoc analysis revealed that AA students achieved significantly higher main IA scores ( $M = 10.71, SD = 2.695$ ) than AI students ( $M = 10.7, SD = 2.695$ ).

In contrast, rather than subjects, the IB final mathematics grades, referred to as “scaled overall mean scores”, showed a significant difference between the levels. At the .05 level of significance, there was no main effect of mathematics course subjects on IB mathematics

final scores,  $F(1, 135) = 2.028, p = .157$ , indicating that overall mean scores for the AA subject ( $M = 51.30, SD = 18.5$ ) and the AI subject ( $M = 46.63, SD = 2.72$ ) did not differ significantly. There was a main effect of mathematics course levels on overall scores,  $F(1, 135) = 15.209, p = .001$ . The post hoc analysis revealed that higher-level students achieved significantly better overall scores ( $M = 55.37, SD = 2.23$ ) than standard-level students ( $M = 42.56, SD = 2.33$ ).

In the online survey, one of the students indicated that because AAHL did not lend itself to developing mathematical abilities thoroughly in real-world contexts, it was challenging to stimulate their interest in investigating abstract and sophisticated mathematics.

**Table 5**  
**Descriptive Statistics of the IB Subjects**

<b>Variable</b>	<b><i>n</i></b>	<b><i>M</i></b>	<b><i>SD</i></b>
Maths Internal Assessment <sup>a</sup>	135	11.37	2.807
Maths Paper 1	134	67.16	22.513
Maths Paper 2	134	59.35	24.625
Maths Paper 3	81	23.57	11.161
Maths Mock Examination	134	60.53	19.28
IB Maths Scaled Total	133	50.34	17.73
Physics <sup>b</sup>	62	55.58	13.2
Environmental System Society <sup>b</sup>	23	52.30	11.66
English	123	67.82	11.42
Biology <sup>b</sup>	26	64.62	16.45
Chemistry <sup>b</sup>	52	65.44	11.33
Economics	105	58.37	13.68

<sup>a</sup> The criteria marks in internal assessment

<sup>b</sup> Students' total scores in science taken in 2022

<sup>c</sup> Students' total scores in English papers  $p < .01$

### 4.1.3 Exploring the Fairness of IA.

The total marks for IB mathematics final grades did not include final examinations because the May IB exams were cancelled. Instead, the final IB grade was composed of the grade predicted by the subject teachers and students' IA score. Pearson correlations were performed to examine the relationship between mathematics IA scores and scores from the mathematics examinations. The results demonstrated that there was a range of  $r$  values between 0.29 and 0.381 for the correlations between the mathematics IA score and the four variables (mock examinations Papers 1, 2, and 3, and the total scores), as shown in Table 6. Lower moderate correlations were shown by these variables. Despite this, this study further explored the relationships between mathematics and other domains to gain a better understanding of the construct tested by mathematics assessments. It is clear that the other IB courses, including physics, environmental systems and societies, and biology ( $r = 0.5656$ ,  $r = .500$ , and  $r = 0.5435$ ), had a close relationship with mathematics IA. Compared to English ( $r = 0.27$ ) and chemistry ( $r = 0.23$ ), the correlation coefficients were slightly higher for the subjects of physics, environmental systems and societies, and biology. This research also explored the relationships between mathematics IA and English achievements to understand whether there are fair mathematics performances among students with different English proficiency levels.

**Table 6**

**Bivariate Pearson Correlations among Mathematics Internal Assessment and Examinations (N = 135), English Language (N = 123), Physics (N = 62), Environmental Systems and Societies (N = 23), Biology (N = 26), Chemistry (N = 52), and Economics (N = 105)**

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
1. Maths Internal Assessment												
2. Paper 1	.290											
3. Paper 2	.309	<b>.845</b>										
4. Paper 3	.333	<b>.753</b>	<b>.753</b>									
5. Mock Exam	.387	<b>.818</b>	<b>.781</b>	<b>.850</b>								
6. Maths IB Scaled Total	<b>.539</b>	<b>.594</b>	<b>.561</b>	<b>.749</b>	<b>.857</b>							
7. Physics	<b>.556</b>	<b>.676</b>	<b>.613</b>	<b>.549</b>	<b>.635</b>	<b>.614</b>						
8. Environmental Systems and Societies	<b>.500</b>	<b>.507</b>	.471	.147	<b>.538</b>	.487	-1					
9. English	.270	.364	.391	.174	.461	.428	.308	.403				
10. Biology	<b>.535</b>	<b>.520</b>	<b>.647</b>	<b>.720</b>	.403	.372	-	-	.293			
11. Chemistry	.234	<b>.565</b>	<b>.576</b>	<b>.635</b>	.497	.400	<b>.535</b>	<b>1.000</b>	.316	<b>.795</b>		
12. Economics	.471	<b>.570</b>	<b>.554</b>	.375	<b>.530</b>	.480	<b>.707</b>	<b>.702</b>	<b>.524</b>	<b>.779</b>	<b>.546</b>	

## **4.2 What are students' perspectives on the DP mathematics IA? How did the students perceive the validity and fairness of the DP mathematics internal assessment in 2022?**

To answer the second research question, the qualitative findings from the perceptions of 13 students were obtained from the interview analysis. A survey of 36 students' perceptions of assessment criteria was also conducted. The students, however, may have had a different understanding of the validity and fairness of mathematics IA. The students who participated in the interviews and online survey in this study offered a variety of viewpoints on the validity of the mathematics IA. Themes from the interviews were categorized based on the perspectives of the students, and it was shown that research and mathematical ability were essential for getting higher IA scores. In addition to being mentioned as secondary to IA attainment, inquiry and transferable skills were also highlighted. From students' perspectives, the following aspects of IA have an impact on the validity and fairness of IB DP mathematics assessments:

### **4.2.1 Students' Perceptions of the Validity of Mathematics IA.**

#### *Students' understandings of the assessment criteria of IA*

The students' understandings of assessment criteria of mathematics IA were investigated by the online survey. The online survey also explored 36 students' opinions on the issues of the assessment criteria, and how students perceive the validity of IA. These students had

varying levels of English language proficiency, as shown in Table 7: beginner (n = 0), intermediate (n = 13), advanced (n = 16), and proficient (n = 7) (see Appendix 3).

With regard to students' understanding of the assessment criteria of IA, 83.3% of students believed that they understood the level descriptors from the criteria of the mathematics IA, 16.7% of students had concerns that the level descriptors in the assessment criteria were vaguely worded, and 19.4% of students found it difficult to understand some of the criteria, such as the personal engagement criterion. One student mentioned having difficulty understanding the level descriptors in the assessment criteria, particularly the terms expressing the band gap between descriptors.

Differences between marks in the level descriptors are evident in the criteria. If focusing on the same criterion with different achievement levels, I can understand what is needed to reach higher scores through a comparison. However, most of the time the descriptors written in the criteria, such as "partially", "good", "thorough", "significant", and "outstanding", are ambiguous and it is difficult for a second-language speaker to understand these different words. In this case, I am unsure what should be included to reach "outstanding" but not "significant". To better support rubric understanding, rubrics should be broken down with further explanations.

Another student was concerned that the adjectives in the specific achievement levels in the assessment criteria, such as "meaningful", "substantial", and "outstanding", are not descriptive enough to follow. When it came to the "use of mathematics" and "mathematical communication" criteria, 30.6% of students said that they had trouble providing the evidence that was required, whereas students found the mathematical terminology notations were entirely new to them and needed further improvements.

**Table 7**

**English Proficiency of Students**

<b>Level of English Language Proficiency</b>	<b>Total Responses</b>	<b>% of IB Students</b>
Beginner	0	0
Intermediate	13	36.1%
Advanced	16	44.4%
Proficient	7	19.4%
Total	36	100%

As regards the “presentation” and “mathematical communication” criteria, six students voiced dissatisfaction with assessment criteria that prioritized English writing skills over mathematical skills. The students were asked if they had a reason for their concerns, with the majority of students (37.5%) reporting language barriers, two students (6%) reporting a “personal engagement” criterion, and the remaining students (27.3%) giving neutral responses. It is plausible that participants’ mathematical competence was at least partially moderated by their English language proficiency in students’ perspectives.

Sixteen students (44.4%) showed wide disagreement regarding the validity of the mathematics IA. Ten students reported additional factors that affected the validity of the mathematics IA adversely, including English writing experience, unfairness to students whose second language is English, limited marks designed for the “use of mathematics” criterion, lots of opportunities for cheating and unfair advantages, authentication of the work, and tiering of the mathematics subjects (AA or AI). From students’ perspectives, there are several reasons why students may not believe that the IA is a good assessment for evaluating their mathematics abilities. First, students had concerns that only six out of 20 marks in the rubric are intended to evaluate students' competence in mathematics. Second, language proficiency may limit pupils' mathematical abilities, particularly for those who perform well in maths exams. Finally, AA HL students are expected to do more theoretical

and abstract mathematics than students in other subjects, including statistics, a subject area in which students regard AA HL for IA topics as uninteresting. However, 22 students (61.1%) reported that IA is a good assessment that examines their conceptual understanding, technology use, problem-solving, exploration, experimentation, and application abilities as it emphasizes the construct of mathematics IA as a deeper understanding of the selected topic, revealing shortcomings in learning mathematics. Students' understanding of the assessment criteria of IA varied. In addition, one student who participated in the interviews in this study said that language learners should be given preference when it comes to assessment criteria of communication and presentation. However, there is a discrepancy between how students comprehend the assessment criteria and how the students perceive the construct of IA.

#### **4.2.2 Students' Understandings of the Construct of IA.**

##### *Mathematics skills of IA*

During the interviews, students were asked to what extent the mathematics IA assessed their skills. The student interviewees had various opinions on how significant mathematics abilities were to achieving IA. The majority defined mathematical skills as comprehensive mathematical reasoning, an in-depth understanding of mathematical concepts and principles, the ability to apply mathematical theory to real-life scenarios, and the exploration of new mathematical terminologies and notations.

I gained the ability to assess mathematical concepts and their consequences for mathematical derivations after passing the mathematics IA. I spend the majority of my time doing computations, organizing data, thinking, using logic, and reading

academic articles. IA is unique and difficult as it requires a deep understanding of mathematics. (Student 6, UWC China)

Due to their diverse purposes, the diverse test content, and the skills being evaluated, the interviewees tended to distinguish internal assessment from external assessment as a separate construct. As two assessments examine different facets of one's mathematics abilities, there seemed to be some disparities between internal and external assessments. While examinations concentrate on the applications of abstract and contextual mathematics, IA tests mathematical presentation and communication in a particular topic. As one respondent stated:

IA did not assist me in getting ready for my exams because the examination does not evaluate my ability to conduct research, write, or format. Furthermore, IA only asks me to use mathematics in relation to a very small fraction of the overall syllabus, therefore there are not many mathematics topics involved. Instead of focusing on learning mathematics, IA examines how we communicate mathematically to explain our ideas while also researching new areas of mathematics. (Student 7, UWC China)

Other students claimed that the mathematics IA assessed a particular area, such as calculus or polynomial function, which students researched and showed a stronger understanding of the particular domain of mathematics.

In my IA, I used algebra, polynomial functions, and graphs to come to a conclusion. This technique is also examined in external assessments, and my IA familiarized me with the technique for solving algebraic questions and helped me to prepare for the examination on the topic. (Student 2, UWC China)

Overall, the responses from interviewees on the value of mathematical abilities showed a definite endorsement of the internal assessment, with students having greatly improved their mathematical abilities through completing coursework for the internal assessment. Most participants in this study believed, however, that the internal assessment and examinations placed different emphasis on the various uses for one's mathematical abilities.

### *Research Skills of IA*

Seven out of the 13 students said that IA had assessed their research skills throughout the IA coursework. According to these students, research skills include the capacity to solve problems in both abstract and practical situations, and the ability to independently explore oneself, and uncover pertinent and trustworthy information. In addition, studying mathematical modelling to address a real-world problem was seen as crucial from the viewpoint of these students. Data collection, processing, analysis, and interpretation were also seen as important research skills that could be developed through IA work:

I have an intermediate level of research and analysis developed from my previous studies and an understanding of IB standard-level mathematics content. I found research skills to be important to complete the maths IA. For the assessment, a student was asked to study a topic that's beyond what's taught in the classroom. Therefore, being able to find relevant and reliable information is crucial for completing the internal assessment. (Student 2, UWC China)

In addition, one student defined research skills as the planning and execution of scientific techniques in experimental design. This process entails applying theory to real-world situations and drawing fresh conclusions from a mathematical model.

I think I am learning the procedure of how to design an experiment because my IA mainly focused on an investigation to model the speed of a machine...I had to create the process, calculate the water volume, and plot the machine into a diagram. I analysed the data to understand the speed. So, one of the skills I learned was developing experiment design using mathematics. (Student 4, UWC China)

One participant explained how it was challenging to demonstrate evidence in personal engagement. Personal engagement called for utilizing various strategies to solve problems in various unique circumstances.

It challenged me and helped significantly with data analysis for an application because it gave me an entirely new avenue to truly analyse the problem independently and the situation at source, saying how this person would have seen it as opposed to how it was. (Student 11, UWC)

In conclusion, students' experiences with the internal assessment were reflective of the many different research skills. The participants' experiences highlighted the significance of data interpretation and analysis, novel problem solving in practical situations, and the pursuit of pertinent and trustworthy information.

### *Transferable Skills*

In this study, when asked how the mathematics IA compared to other IB subject assessments like English, physics, biology, economics, chemistry, and environmental systems and societies, the student interviewees believed that what IA assessed overlapped with the ESS assessment. Seven out of the 13 participants claimed that the ESS evaluated

the data analysis abilities acquired during the maths IA. For example, seven students spoke of ESS assessing the descriptive statistics in mathematics. Similarly, the IA also requires students to analyse data for mathematical modelling in descriptive statistics.

Environmental systems and societies (ESS) is the only science subject I chose. I used a lot of mathematics in ESS – using the chi-square test to see data and a t-test of the two samples in analysis. ESS requires a lot of data collection as well as presenting and analysing data. I think this overlaps with the maths IA in terms of skills. (Student 6, UWC China)

More interviewees reported that the mathematics IA prepared them for science-related subjects – particularly for environmental systems and societies (ESS) and physics rather than for history and English, where knowledge was, according to them, more about problem solving than contextual writing. Many students emphasized that environmental systems and societies (ESS) was one of the scientific courses comparable to the mathematics IA. This is because there is a lot of mathematics in ESS, and students used statistical t-tests and chi-square tests to analyse data. Data gathering, interpretation, and analysis are essential in statistics. Maths IA and ESS frequently overlap in the field of inferential statistics. These findings were in line with the online survey, with 55.6% of students responding that science-related subjects had more significant overlaps with the mathematics IA.

### *Inquiry Skills of IA*

The interviewees in this study stated that the IA significantly reflects the ability to ask one's own questions and provide an answer to the inquiry in the personal engagement criterion. One interviewee, for example, stressed that seeking support from the teacher could help her effectively understand the IA requirements:

I have to prepare questions and tools to draw mathematical graphs throughout the process of IA. I have to ask teachers and my peers for their advice and help. (Student 5, UWC China)

Students also believed that strong inquiry skills may help them prioritise their inquiries for the mathematical investigation in critical thinking and synthesise the data with insightful responses. Students must take the initiative to connect with teachers about the exploration topic, recognise any obstacles they may be facing, and seek additional clarification with support in order to ask a meaningful question.

#### **4.2.3 Students' Understandings of the Fairness of the IA.**

Students had mixed opinions on how writing abilities affect their IA scores when asked about the impact of English writing skills on the fairness of mathematics IA. English writing abilities were mentioned by three interviewees as being important for finishing their IA successfully. The remaining three interviewees did not consider proficiency in English writing to be necessary:

Yeah, undoubtedly, for report writing skills, I think we learnt a lot of different techniques of citation and how we can apply these skills in formal writing – learning the structure of formal writing. For example, understanding how to write a table of contents and how to use fonts and margins that adhere to the guidelines of the maths IA are all part of mathematics communication. (Student 11, UWC China)

My English skills do not seem to have grown or improved. Compared to other internal assessments I took, including history and English, my writing abilities were limited in this mathematics IA. Language is not a key component of the mathematics IA. (Student 10, UWC China)

The majority of interviewees had positive opinions about the mathematics IA, indicating that written examinations were not the same as other methods of evaluating one's mathematical abilities. Therefore, although two interviewees complained that the internal assessment process was unfair to English language learners, other interviewees seemed to be less concerned about this. The interviewees gave their explanations for why the majority of university applications were written in English. For their university preparations, writing skills were useful:

I am aware that mathematics written in English is somewhat biased toward Chinese learners. However, I'd add that some people think Chinese people have their own unique way of thinking when Chinese students study mathematics. For students from different countries or cultures, I'm not sure how studying mathematics will differ. In English, I have not fallen into these issues. I believe it to be fair based on my own experience. Likewise, I believe it can be a little harsh for students who excel in mathematics but struggle to communicate in written English. However, in other courses, like TOK, students are permitted to write in Chinese. (Student 4, UWC)

Some students thought that completing the writing portion of their internal assessment was enjoyable as it enabled pupils to select their own subject. However, others believed it to be a tortuous process that was not comparable to final examinations as it requires writing long and tedious reports and students found it challenging to submit these reports before their deadlines. In addition, one student who responded to the online survey expressed concern about cheating in the mathematics IA and suggested that he may have his parent do it without the IB examiner knowing. It is unfair for pupils who cannot get outside help.

### **4.3 How did the teachers understand the validity and fairness of the DP mathematics internal assessment?**

#### **4.3.1 Underlying Skills.**

Five teacher interviewees agreed that the mathematics IA tested students' abilities to apply theory to practice. When questioned about the particular skills that students gained in the various IA coursework, a maths teacher at UWC stated that his teaching concentrated on helping students to identify "data", extracting and analysing crucial references from data. Other maths teachers at UWC described how the Diploma Programme encouraged students to analyse sources and potential applications within these sources, and study how students could compare different theories and use them to explore diverse real-life situations. For example, one mathematics teacher interviewed in this study explained the disparities between a theoretical framework and real practice. Teachers get to understand the difference between requiring pupils to learn theoretical concepts and expecting them to strive toward a mathematical project while unforeseen events occur in real-world practice.

Similarly, one maths teacher explained the flexibility of applying mathematical theory to an application. He claimed that a student needs to be flexible to adjust the methodology throughout the IA process. This kind of mathematics, regarding the method, or the adjusting process of mathematics application, needs to be considered in the actual situation. He perceived IA as different mathematical approaches to investigating real-world problems. Meanwhile, another teacher interviewee mentioned the importance of reflection in mathematics exploration. He indicated that students need to demonstrate critical reflection

by considering what would come next, discussing the implications of results, discussing the strengths and weaknesses of approaches, and considering different perspectives. The reflection requires English writing proficiency, which is expected to meet the aforementioned requirements. Meanwhile, this teacher stated that students from all backgrounds could communicate in mathematics since it is a universal language. Therefore, students' ability to complete IA mathematical assignments would not be impacted by their English proficiency. However, other teachers believed that the mathematics IA evaluated communicating and presenting mathematical ideas in addition to working out formulas and mathematical models. Teachers thought that the quality of English comprehension and writing was what determined how well a mathematics presentation was evaluated. To some extent, the construct validity of mathematical IA is therefore considerably influenced by English writing skills.

#### **4.3.2 Time Management.**

In addition to the underlying abilities, one teacher emphasized the significance of time management skills, which students may be assessed on across all IB subjects. He asserted that in order to finish the IA tasks, students must manage their time effectively.

I've seen some students struggle and not learn how to manage time...I asked for the first and final drafts of their internal assessments at a certain time, and this was something that students were able to do at the beginning, but maybe 40% of students did not know how to manage their time. Time management skills are essential. It is about making a plan to produce one's own timetable. (Teacher, UWC China)

Teachers also believed that students with good time management abilities may organize their tasks for the mathematics IA in a systematic manner and apply those skills to other IB subjects. In order to transfer their skills, students must take the initiative to engage with teachers about the IA planning and identify connections between the mathematics IA and science related IB subjects.

### **4.3.3 Interaction between Teachers and Students.**

One of the teachers who participated in the interviews said that the interactions between teachers and pupils have encountered the problems of supervising pupils with regard to the fairness of the mathematics IA. The interactions helped pupils understand by responding to their inquiry with teachers' assistance. The students make contact with teachers more frequently, receive feedback that is of better quality, and ultimately earn IA grades that are higher than those of their peers. While students are making progress in their exploration of mathematics, teachers might not be aware of the fairness in interactions with them:

Culturally, I realize that Chinese students lack active communication with teachers, and I've missed some feedback from students regarding their understanding of the mathematics IA. Help and support should have been requested from the student to the teacher, but some students are passive. Some may not be passive, but students do not ask questions.

In conclusion, while acknowledging the difficulties of the internal assessment in providing feedback on students' learning, the teacher interviewee gave a clear sense of the negative impact that inadequate consideration of the interactions between students and teachers may have on teachers' predicted results and, potentially, on students' mathematics attainment.

The degree to which feedback is given to individuals also contributed to the fairness problems.

#### **4.4 Summary of Findings**

The results of the quantitative analysis reveal the relationship between IA scores and final exam grades, as well as the relationship between IBDP mathematics performance and the academic records of other domains in other IB subjects. Statistical differences in IA scores were affected by the different levels of the mathematics courses and this remained true even after a number of pertinent associations were examined using the propensity scores of the IA results and the grades from the mock exams. It should be noted that the analysis also indicated that for the use of tiering in DP mathematics courses demonstrated by two distinct subjects (AA, AI), scores were different between the groups of the Diploma Programme. The qualitative data echo the findings from the quantitative data, showing the construct validity and fairness from students' and teachers' perspectives. Students' understandings of assessment criteria, as well as the constructs of IA and fairness, provide encouraging evidence that participation in the DP mathematics IA may improve research and transferable skills as measured from students' and teachers' perspectives.

Students who participated in the one-to-one interviews and online surveys in this study said that the mathematics IA improved their implementation in IB learning profiles and gave them a deeper understanding and knowledge of mathematics. Students also said that the instruction in mathematical exploration improved their academic writing for their

upcoming university studies. The IA's transferable skills were also shown to be better suited for science courses such as physics, according to the quantitative analysis in this study. In line with the findings from the qualitative data, the majority of students thought that the environmental systems and societies course overlapped with mathematics IA in terms of skill sets. However, students and teachers were more critical of how this course was graded, particularly the use of what students felt were unclear assessment criteria.

## **Chapter 5: Discussion**

This research set out to explore the construct validity and fairness of the IB mathematics internal assessment by using a mixed-methods research approach. Both the online survey and semi-structured interviews showed that most students and subject teachers in the sample valued the mathematics IA and considered it to be valid and fair. However, a few interviewees believed that the validity and fairness of the mathematics IA were challenged because the IB guide's description of the assessment purposes was ambiguous. This finding replicates the emphasis of Newton (2007) on the importance of clarifying assessment purposes at different levels and not categorizing assessment purposes misleadingly. The technical goal of an assessment is addressed and distinguished at the judgement level whereas teachers translate the observation of performance into a particular type of assessment judgement. Teacher assessments serve as the basis for assessment judgement in the IB DP mathematics IA. The outcome of an assessment can be based on the teacher's knowledge, skills, and understanding, which can have an impact on the assessment judgement made regarding the mathematics IA.

This chapter also discusses the key findings of this research to further answer the research questions in this study. This discussion focuses on analysing the construct validity of the mathematics IA in terms of subject and level, exploring the new skills of the construct, and investigating clearer assessment criteria. Possible fairness issues of language and authorship are also explained from students' and teachers' perspectives. This chapter provides a comprehensive summary of both the qualitative and quantitative analysis and

explains their relationships to identify the potential impact of mathematics IA. The findings will first be discussed by looking at the contribution of this study to develop an understanding of the impact of the DP mathematics IA upon its construct validity and fairness. Following that, the significance and implications of the findings are presented, which is followed by an outline of the limitations of this study and potential future research directions.

## **5.1 Construct Validity**

The findings of the quantitative analysis revealed a moderate correlation between the mathematics IA scores and final assessment grades awarded by the International Baccalaureate. These findings are consistent with the previous literature on the validity of mathematical assessment reviewed in Chapter 2. However, the findings helpfully provide more nuanced indications that the final assessment grades during the COVID-19 pandemic were aggregated based on teachers' predicted grades and teacher internal assessment marks. Oğuzhan Ö et al. (2022) validly proved that the correlation coefficient was used to indicate how closely the final assessment scores given by the International Baccalaureate matched the predicted grades given by teachers. The results of this quantitative study similarly found that there were low-to-moderate connections (ranging from .42 to .73 at standard level and from .17 to .73 in high-level courses). Likewise, the correlation coefficient between the maths IA marks and final assessment grades was 0.539, which was similar to the correlation coefficient reported by Oğuzhan Ö et al. (2022), falling in the low-to-moderate connection range. The analysis also found that there was a statistically moderate relationship between

mathematics IA scores and the mock examination, including its component results (Paper 1, Paper 2, and Paper 3).

The qualitative findings of this study suggest that students believed that the DP mathematics IA requires mathematical communication skills that are significantly different from examination skills in mathematics. Viro and Joutsenlahti (2020) suitably found that pupils' grades in project work correlated positively with their overall grades in mathematics in their study. The results of this study align with the previous findings suggesting that mathematical thinking can be developed and associated with cognitive constructs, as well as the construct openness to project experience, both of which were also associated with IB internal and external assessments. This implies that students who tend to perform better in the mathematics IA are more likely to perform well in terms of their overall grades, but the combined quantitative and qualitative findings of this study offer encouraging preliminary evidence that the internal assessment and external assessment (examinations) are intertwined to assess mathematical skills and students' research abilities.

### **5.1.1 Subject and Tiering Differences.**

In this study, the IA mean scores for different student groups differed statistically, especially for two DP mathematics subjects: Analysis and Approaches (AA) and Application and Interpretation (AI). To be more specific, students who learned AA achieved higher mean scores than those who learned AI. One possible reason for this could be that AA concentrates on the domain of mathematics such as algebra, arithmetic, and

calculus. Students who learned AA mathematics typically performed well in a deeper understanding of the above topics. Likewise, the achievements of AI students with respect to their IA scores were impacted by the fact that AI focuses more on the interpretations of data analysis, statistics, and probability. Only individuals with a higher degree of language proficiency can demonstrate their contextualized mathematical performance because the interpretations of mathematics demand more written communication skills. Vukovic and Lesaux (2013) usefully proved that language proficiency predicts improvements in the many domains of mathematical cognition in data analysis, probability, and geometry rather than in arithmetic or algebra. In other words, language proficiency and mathematical aptitude can be seen as two different variables evaluating pupils' mathematical IA abilities.

Another interesting finding in this study is that tiering differences are the main factor behind the statistical differences between final assessment grades awarded by the International Baccalaureate. The analysis revealed that higher-level students achieved significantly better overall scores than standard-level students. This may also be because students who initially select a higher-level subject may have better underlining mathematical and literacy competencies among all groups. The final assessment grades for the IB are determined by predicted grades and IA scores. This study found that higher-level teachers frequently tended to predict grades that are higher than those of standard-level pupils. Generally, subject and tiering interact with the other characteristics considered in this study such as teachers' predicted grades, affecting the construct validity of the DP mathematics courses. This finding is in line with Oğuzhan Ö et al.'s (2022) detailed discussion on the relationship between predicted grades and final assessment grades in the IB DP. But it is beyond the scope of this study. Consequently, AA students performed better in IA scores, and higher-level students did well in their final grades. In other words,

the AAHL mathematics course appears to benefit from being graded at both IA and final assessment grades.

However, AAHL students in this study reported that they did not easily adapt to the application of mathematical skills in real-world contexts and that it was consequently challenging to engage in explorations of abstract mathematics. This finding is somewhat surprising as it makes sense that students expect to strengthen their mathematical thinking skills both in abstract mathematics and mathematics applications as a result of the purposes of IA coursework. Considering the mathematics in Section 1.2, where the assessment objective focuses on students' ability to "recall, select, and use their knowledge of mathematical skills, results, and models in both abstract and real-world contexts to solve problems" (IBO, 2019b, p. 23), it seems that it is important to understand students' perspectives on assessment criteria in real-life contexts and the construct of mathematics IA coursework.

### **5.1.2 Clearer Assessment Criteria.**

The teachers and students in this study said that they occasionally struggled with perplexing assessment criteria. The assessment practices revealed by the findings of this study indicate that the DP assessment objectives prioritize problem solving within the assessment framework. This is particularly clear in the "use of mathematics" criterion as well as the "personal engagement" criterion. This is crucial as DP mathematics assessments, especially high-stakes exams, must make students feel valued and rewarded for developing higher-order thinking abilities such as mathematical reasoning (Shepard, 2000).

According to the interviews, both teachers and students felt that the official IB assessment standards were too general and not subject-specific enough. In regard to what was expected of students, there was some uncertainty reported due to the perceived vagueness of the criteria. Although teachers claimed that they had recently received updated assessment instructions, they were still confused about how to interpret some of the assessment criteria. The “personal engagement” criterion assesses the extent to which students engage with the topic independently, or creatively, presenting mathematical ideas in their own way, exploring the topic from different perspectives. This could be challenging for Chinese students as rote learning is emphasized in pre-university education due to the influence of Confucianism (Li & Cutting, 2011). In addition, students reported concerns about the personal engagement criterion because it could be marked subjectively by examiners. The lack of clarity surrounding the concepts of personal engagement leaves students wondering how it is assessed. These concerns about the assessment criteria of the teachers and students interviewed in this study further support this finding. In the marking process, it was emphasized how important it was to adhere to assessment criteria and prevent teachers from serving these as learning obstacles for students. Without teachers' and students' complete comprehension of the reasons behind the principal examiner's marking practice, marks were modified up or down for pupils. Because of this, IB recommendations in publications for teachers and students may be able to help decrease these moderation errors by precisely explaining and conceptualizing the insights of the assessment criteria. One interviewee in this study said that by recognizing this, IB educational organizations could demonstrate a better understanding of marking practice in detail to students and teachers.

### **5.1.3 New Explanatory Skills of the Construct.**

It is noteworthy that the student interviewees in this study believed that the mathematics IA requires an in-depth understanding of mathematics, mathematics in real-life applications (Graven et al., 2009), mathematical communication (Stylianides & Stylianides, 2008), and reflection (Taylor & Porath, 2006). Based on the finding of a similar study by Schoenfeld (1992), a plausible explanation for this is that thinking mathematically involves problem solving, metacognition, and sensemaking. Problem solving focuses on using techniques such as proof, arithmetic calculation, statistical analysis, and mathematical modelling, with minimal attention to having students justify their use of these techniques in a deeper procedure. In addition to the previous findings in the literature, the analysis in this study determined that existing assessment purposes were insufficient and misleading for students' comprehension of the underlying skills assessed in the International Baccalaureate Diploma Programme mathematics courses. The existing assessment purposes of the IA from the IB guide were originally taught by teachers. In addition to assessment purposes, “mathematical presentation” and “communication” criteria were seen by the UWC students as important criteria for evaluating mathematical ability. However, it was discovered through the surveys and interviews in this study that IA assessed additional new skills as a construct that were not mentioned in the IB guide, without which students' initial understanding of the construct of IA would be inadequate. As a result, the qualitative research in this study highlighted a number of potential constructs that the mathematics IA could use to influence the validity and fairness of the DP mathematics

assessment. These constructs were organized into three themes: (1) research skills; (2) transferable skills; and (3) mathematics literacy.

### ***Research Skills***

The results of this study showed that the students regarded the IA as a necessary component of their mathematics learning process and that using research skills during this process could improve their independent and critical thinking ability. Teachers also believed that developing students' research skills during the IA process was beneficial for students. These findings show that research skills are perceived as the ability to solve problems in both abstract and real-world contexts, especially the mathematical modelling of real-world problems. This finding implies that the importance of research skills is recognized by both teachers and students, which is in line with the recommendations from the meta-analysis of Dickson et al. (2018), which stated that opportunities for debate and exposure to real-world situations and examples should be used to teach research skills. However, the teachers in this study also suggested that there is room for developing flexible methodologies in research skills throughout the IA process. It is believed that incorporating some debate and real-world examples in class could help students develop their research skills.

The results of this study build on the existing evidence from the IB DP mathematics guide (IBO, 2019a), suggesting that there are clear benefits from developing research skills in the IA from students' and teachers' perspectives. The findings of this study also suggest that

some of the benefits in research and critical thinking skills developed are attributable to students' IB learning profiles, such as inquirers, communicators, and reflective and international mindedness (Walker et al., 2016). Specifically, stakeholders widely noted that IB programmes develop research and creative skills in line with the “whole person” educational philosophy of the IB (Dickson et al., 2018). There remains a significant advantage of IB learning profiles on inquiry-based assessments over and above these assessment objectives (IBO, 2019c).

### ***Transferable Skills***

The findings of this study showed that UWC China students thought that the mathematics IA was related to the assessments of other IB topics. While putting transferable skills to work in the mathematics IA, this could also help with other IB subjects' assessment abilities. The assessments for environmental systems and societies and physics have been deemed the most pertinent when compared to mathematics IA. These results corroborated earlier research by Boaler (1993), who helpfully reported on a research project that considered the transfer of students' mathematical understanding across different task contexts. The research suggested that students were more able to transfer their learning across contexts if they had learned in an environment characterized by the integration of process and content. The integration of process, content, context, and culture was considered in relation to real-world transfer and the effectiveness of different learning environments. Teachers did note that there is a need for students to improve their time management abilities. It is

thought that including interdisciplinary subjects in the classroom and collaboration between the same school's teachers can assist students in gaining transferable abilities.

### ***Mathematics Literacy***

The qualitative analysis in this study demonstrated that students had mixed opinions about the validity of using English writing abilities to evaluate students' mathematics IA. Their English writing skills in real-life contexts during the IA process may be determined by their mathematics literacy. Teachers also believed that developing students' mathematical communication skills during the IA process was beneficial. The results showed that both IB teachers and students recognize, to some extent, the value of mathematics literacy. Graven et al. (2009) validly considered the ways in which the mathematical literacy assessment taxonomy provides spaces for the problem solving and reasoning identified as critical to mathematical literacy competence. Furthermore, this could result in greater motivation for developing in-depth mathematical thinking among students as well as a better understanding of the mathematical principles in the IB context. This is particularly noteworthy given that mathematical literacy provides learners with an awareness and understanding of the role of mathematics that are ultimately valued in the modern world (Graven et al., 2009). Although there is no conclusive proof that connecting mathematical literacy to these higher-order thinking skills is helpful in fostering problem-solving abilities, previous research indicated that students are more engaged with content when they understand its worth and purpose (Yeager et al., 2012). The purpose of mathematics IA shall be engaging English language learners in addition to the existing assessment

objectives. In order to guarantee the fairness of mathematics IA, it is suggested that teachers could provide more support materials and IA samples in their lessons.

## **5.2 Fairness**

### **5.2.1 Engaging English Language Learners.**

The findings demonstrated that some students believed language to be an essential part of their mathematical IA process, while engaging English language learners during the process could improve their mathematical communication and presentation. Teachers also believed that additional language support for students taking the mathematics internal assessment with pedagogic discourse was necessary. In the DP mathematics IA, students in China who do not speak English as their first language have educational difficulties when it comes to criteria such as mathematical communication and presentation. The COVID-19 pandemic's impact on the external assessment due to the cancellation of examinations has directly shifted the requirements for the mathematical assessment further in the direction of a report-writing task. This finding implies that the importance of the language is recognized by both IB teachers and students, which is in line with Messick's (1995) study. Messick (1995) suitably argued that English writing skills might be considered as construct-irrelevant variance, especially for English language learners. When the IA's construct-irrelevant variance is purposefully used to create desired social outcomes, as in score adjustments for minority groups and within-group norming sliding band techniques, this leads to further elaboration of the construct validity of mathematics

(Cascio et al., 1995). It has been debated whether or not language effects are construct relevant. Therefore, the language construct of mathematics may be seen as a component of a "composite" construct instead (Maul, 2013). Khisty and Chval (2002) suitably discussed the role and nature of pedagogic discourse, particularly in the learning of racially, ethnically, and linguistically diverse students. It was argued that teachers who use rich mathematical language play an important role in students' learning. Barrance (2019) helpfully corroborated Khisty and Chval's (2002) findings in another study by proving that the assessment environment affected students' performance in the internal assessment for a multi-cohort study. In addition, Richardson and Wilkinson (2005) compellingly found that students do not perform as well in assessments not because of their poor academic background, but because of cultural and linguistic barriers. For the DP mathematics course, this perspective on the language construct is problematic. Therefore, language bias and its implications might provide a thread to the fairness of mathematics IA in evaluating problem-solving abilities.

However, the interviewees in this study provided a useful counterargument to this claim by pointing out that IB assessments were not necessary to satisfy the needs of non-native language learners because not all students taking an IB assessment intend to progress to English-speaking universities. Students further argued that since tests for other IB subjects are also administered in English, the construct of mathematics IA is fair for language learners. Supporting evidence for the fairness of assessments came from the qualitative analyses in this study, which showed that students' perceptions of the mathematics IA were well balanced. This indicates that most students perceived the IA as fair. Conversely, if

assessments are not fair to students who are good at mathematics because of language barriers, the students may choose not to apply to take the qualifications in the first place. The needs of language learners are interrelated, therefore if there is significant failure in the fairness of assessment, it is likely that the IBO will not be able to find appropriate candidates internationally.

### **5.2.2 Authorship.**

The results of the online survey showed that students believed the IA process might have enabled academic fraud. Authorship could teach them how to practise academic integrity and stay away from plagiarism throughout the process. Teachers also believed that developing students' sense of authorship during the IA was beneficial. However, one pupil in the online survey stated that cheating in the IB internal assessment was simple. Teachers and IB examiners may not be aware of students having their parents complete the maths IA. This is unfair to pupils who cannot get external assistance. This finding implies that the authenticity of student work whilst writing IA is recognised by both teachers and students, which is in line with Tattersall's (1994) study. Tattersall (1994) validly concerns about the authenticity of students' work were overlooked. Despite this researcher's useful findings, Bishop et al. (1999) and Ipsos's (2006) study were most applicable to the issues of authorship because they convincingly found popular views among students and teachers that there was too much cheating in the coursework and that students could submit work authored by others. As a result, teachers suggested that there is room for improvement in

authenticity. It is believed that incorporating some interviews and academic honesty software in class can help students understand the importance of authorship.

### **5.3 Implications**

IB learner profiles, such as developing inquiry and research abilities and becoming an autonomous inquirer, have received a lot of attention from IB education. Both teachers and students who were interviewed for this study expressed their appreciation for these principles as a crucial competence for the DP coursework. Students also expressed how challenging it had initially been to adopt an inquiry approach whilst preparing for the IA. The concepts of time management skills for the IA could be introduced at the start of the DP curriculum with both generic and subject-specific examples. This conclusion is based on both the teacher and student interviews carried out in this study, which discovered that students are engaged in the DP mathematics IA courses and generally love learning new things. The findings indicated that Diploma Programme students would gain from writing internal assessments in mathematics and would apply new assessment criteria that students could comprehend compared to existing purposes. It would be beneficial if teachers had more support in fostering their students' ability in terms of research and transferable skills in their areas of interest through the IA process with other IB subjects.

First, at a broad level, the findings of this study have extensive policy ramifications for STEM education in the International Baccalaureate Diploma Programme, especially for science-related courses. The development of research and creative skills is not just restricted to students' university study but can also be developed through their study of subjects such as mathematics and physics once IB students have selected the IB DP. Policymakers in China should be aware that research skills can be found in all sorts of IB DP IA and extended essays.

Second, a student's success in science-related assessments could benefit from the transferable skills of the mathematics IA. According to the students' comments in the interviews in this study, studying statistics in DP mathematics makes it easier to study subjects such as biology and economics. The students in this study believed that they completed ESS with higher overall grades, completed their physics IA more quickly, and were more likely to succeed in science-related courses because of the skills developed from mathematics IA. As a result, it is essential that teachers from interdisciplinary departments work together at the same school. This is in line with the suggestions made by Lynch et al. (2019) in their meta-analysis, who found a substantial correlation between student outcomes and in-school collaboration (teachers taking part in PD alongside their colleagues) in their final model.

Third, linguistic fairness may not eventually present a challenge for language learners' mathematics IA attainment from students' and teachers' perspectives, while mathematics literacy during the IA process could use alternative terminology or vocabulary to define and capture these constructs, which may make it easier for students to generalize and apply

mathematical thinking abilities to real-world problems (Graven et al., 2009). Teachers also believed that providing feedback during the IA process was beneficial. In the online survey, students stated that oral and written feedback were integrated into their learning. Some students were able to act on this feedback and appreciated it. These qualitative findings support earlier research that found that students' learning depends on timely and high-quality feedback from their teachers (Steen-Utheim & Hopfenbeck, 2019), and that students are motivated to act upon teachers' feedback (Shute, 2008).

#### **5.4 Limitations of this Research and Future Research Directions**

The intended research design of the present study initially called for 136 targeted DP2 students to complete the online survey. Only 18 respondents took part in the interviews and 36 respondents participated in the online survey. This was because of the COVID-19 pandemic, which caused China to enter a prolonged "lockdown" in February 2022. From February to May 2022, UWC Changshu China continuously implemented closed-loop management, and as a result, it was disclosed that the IB exams would not take place in April 2022. DP2 students were instructed to leave the school when the IB notified them that all schools in East China would follow the non-examination practice. Due to this "early graduation", only 36 survey respondents – or 26% of all DP2 students – replied to the online survey.

The interviews, however, represented 13 DP2 students I taught at UWC China. The interviewees in this study are a reasonable representation of UWC students in China, offering a helpful starting point for understanding the construct validity and fairness of mathematics IA within the small sample size. The students I taught in DP Year 2 were all taking DP standard-level mathematics, which may have skewed the sample results, as higher-level students might have different viewpoints.

The generalizability of the findings of this study is constrained by the fact that data were only gathered at one IB institution. Therefore, the research of other IB schools at comparable institutions and institutions with high schools that serve mainstream public schools should be investigated in future quantitative study. In addition, as this study was cross-sectional and was conducted during the COVID-19 pandemic, I was unable to correlate the opinions of specific students with their mathematics achievements. Future studies should overcome this shortcoming by combining survey data that gauge each student's opinions on construct validity with other psychological and linguistic behavioural factors that the literature associates with mathematical achievement. The direct and indirect correlations between these factors could be understood at the individual level more fully by integrating all components in various instruments and data sets. Of course, it's still possible that the underlying abilities of the 13 student interviewees were overlooked by the survey and the student interviews. By recruiting more students to take the survey in future studies, these problems will be addressed.

Finally, after IB moderations, a sizable fraction of students' IA scores changes. These grade differences may be due to teachers' bias in assigning grades or examiners' moderation being based on varying perspectives and interpretations of the assessment criteria. Additional examiner moderation may give English language learners more credit if teacher assessment is thought to be more accurate than examiners' moderation (Johnson, 2013). For instance, language learners could take the IA coursework in Chinese or get extra marks for the maths presentation assessment criterion. In doing so, the mathematics IA will be more representative of the entire population. If this is the case, further study is required to understand how teachers' assessments of students' IA grades are moderated by examiners.

## Chapter 6: Conclusion

This study explored the construct validity and fairness of mathematics internal assessment for the delivery of DP mathematics courses, using a mixed method from students' and teachers' perspectives. The results showed that those tasks with revamping assessments for IB qualifications assess new explanatory skills of the construct with considerations of the subject, tiering, and the assessment criteria. Engaging English language learners and authorship in fairness has been discussed. Pedagogy suggestions were given regarding the issues of the construct validity and fairness of internal assessments at the United World Colleges. It is hoped that this study has taken one small step toward enhancing the understanding about this issue. It is recommended that some guidance shall be provided for students, teachers, researchers, policymakers, examiners, and stakeholders, notably:

- i) Subject and tiering. To ensure construct validity, different subjects and levels of DP mathematics assessment will have different aggregations in the final exam results.
- ii) New explanatory skills of the construct. The IB guide on the policy should mentioned and emphasized research skills and transferable skills.
- iii) Clearer assessment criteria. More clarity in the personal engagement and more elaborations in the different achievement descriptors of the same criterion.

iv) Engaging English language learners. In order to distinguish between language learners and native speakers when evaluating students' mathematical skills, internal assessment grades for English language learners use different marking criteria.

v) Feedback: more interactions between teachers and students during conversations based on instruction and assessment in the classroom.

Future mathematics internal assessment developers may also need to take into account the possible effects of large national disruption to assessment arrangements, such as those brought about by COVID-19.

In addition to the direct significance for students, teachers, and examiners, it is envisaged that this greater understanding may have broader applicability. Additionally, in a different cultural setting, it has been emphasized that in order to successfully implement internal assessment, instructors and stakeholders must connect effectively with the many levels of assessment aims and established criteria inside the United World Colleges.

## References

- Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation, 31*(2–3), 162–172. doi:10.1016/j.stueduc.2005.05.008
- Adams, A. (2009). Teachers' use of literacy strategies in secondary mathematics classes. In K. Guilfoyle (Ed.), ProQuest Dissertations Publishing.
- Ahmed, A., & Pollitt, A. (2007). Improving the quality of contextualized questions: An experimental investigation of focus. *Assessment in Education: Principles, Policy & Practice, 14*(2), 201–232. doi:10.1080/09695940701478909
- Allal, L. (2020). Assessment and the co-regulation of learning in the classroom. *Assessment in Education: Principles, Policy & Practice, 27*(4), 332–349. doi:10.1080/0969594X.2019.1609411
- Baccalaureate, T. I. (2020). The assessment and awarding model for the Diploma Programme May 2020 session.
- Bahn, C. M. (2022). *Teachers' perspectives implementing inquiry-based learning in the international Baccalaureate primary years program*. ProQuest Dissertations Publishing,
- Baird, J.-A., & Black, P. (2013). The reliability of public examinations. *Research Papers in Education, 28*(1), 1–4. doi:10.1080/02671522.2012.754232
- Baird, J.-A., & Opposs, D. (2018). The Standard Setting Project: Assessment paradigms. In J.-A. Baird, T. Isaacs, D. Opposs, & L. Gray (Eds.), *Examination standards: How measures and meanings differ around the world* (pp. 2–25). London: UCL IOE Press.
- Barrance, R. (2019). The fairness of internal assessment in the GCSE: The value of students' accounts. *Assessment in Education: Principles, Policy & Practice, 26*(5), 563–583. doi:10.1080/0969594X.2019.1619514
- Barwell, R. (2009). *Multilingualism in mathematics classrooms : Global perspectives*. Blue Ridge Summit, PA: Multilingual Matters.
- Bishop, K., Bullock, K., Martin, S., & Thompson, J. (1999). Users' perceptions of the GCSE. *Educational Research, 41*(1), 35–49.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. doi:10.1080/0969595980050102

- Black, P., & Wiliam, D. (2012). *Assessment and learning* (2nd ed.). London  
Thousand Oaks, CA: SAGE.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575.  
doi:10.1080/0969594X.2018.1441807
- Boaler, J. (1993). Encouraging the transfer of ‘school’ mathematics to the ‘real world’ through the integration of process and content, context and culture. *Educational Studies in Mathematics*, 25(4), 341–373. doi:10.1007/BF01273906
- Brinkmann, S., & Kvale, S. (2015). *InterViews : Learning the craft of qualitative research interviewing* (3rd ed.). Los Angeles: SAGE.
- Brown, R. (1999). Mathematics around the world: A look at mathematics in the International Baccalaureate. *I8(3)*, 98–105. doi:10.1093/teamat/18.3.98-a
- Bunnell, T. (2011). The growth of the International Baccalaureate Diploma Program: Concerns about the consistency and reliability of the assessments. *The Educational forum (West Lafayette, Ind.)*, 75(2), 174–187.  
doi:10.1080/00131725.2011.552700
- Burbidge, T., & Hamer, R. (2020). Academic honesty in the International Baccalaureate Diploma Programme: Student, teacher, and school perspectives. *Journal of International Students*, 10(2), 265–285. doi:10.32674/jis.v10i2.667
- Burrell, G., & Morgan, G. (2016). *Sociological paradigms and organisational analysis : elements of the sociology of corporate life*. Routledge.
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221–256). Westport, Ct: Praeger Publishers.
- Cascio, W. F., Outtz, J., Zedeck, S., & Goldstein, I. L. (1995). Statistical implications of six methods of test score use in personnel selection. *Human Performance*, 8(3), 133–164. doi:10.1207/s15327043hup0803\_2
- Clarke, S., & Gipps, C. (2000). The role of teachers in teacher assessment in England 1996–1998. *Evaluation & Research in Education*, 14(1), 38–52.  
doi:10.1080/09500790008666960
- Cohen, I. G., Amarasingham, R., Shah, A., Xie, B., & Lo, B. (2014). The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Affairs*, 33(7), 1139–1147. doi:10.1377/hlthaff.2014.0048
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Collins, H. (2010). *Creative research : The theory and practice of research for the creative industries*. Lausanne : La Vergne, TN: AVA Academia  
Distributed in the USA & Canada by Ingram Publisher Services.

- Collis, J., & Hussey, R. (2009). *Business research : A practical guide for undergraduate & postgraduate students* (3rd ed.). Palgrave Macmillan.
- Creswell, J. W. (2003). *Research design : Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Sage.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design : Qualitative, quantitative & mixed methods approaches* (5th edition, International student ed.). SAGE.
- Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice*, 3(3), 265–286. doi:10.1080/0969594960030302
- Culross, R., & Tarver, E. (2011). A summary of research on the International Baccalaureate Diploma Programme: Perspectives of students, teachers, and university admissions offices in the USA. *Journal of Research in International Education*, 10(3), 231–243. doi:10.1177/1475240911422139
- Cureton, E. E. (1950). Validity, reliability, and baloney. *Educational and Psychological Measurement*, 10(1), 94–96. doi:10.1177/001316445001000107
- Cureton, E. E. (1951). Validity. In *Educational measurement*. (pp. 621–694). American Council on Education.
- Curlette, W. L. (2006). A framework for research studies: Mixed methods through combining Bayesian statistics and qualitative research in individual psychology. *Journal of Individual Psychology*, 62(3), 338-349. Retrieved from <https://search.ebscohost.com/login.aspx?direct=true&db=aqh&AN=23596616&site=ehost-live>
- David, K. (2007). International Baccalaureate Diploma Programme: Mathematics SL, A Review. *Nonpartisan Education Review*, 3(4), 1–8.
- De Waal, C. (2001). *On Peirce* (Vol. 38): Wadsworth.
- Denzin, N. K., & Lincoln, Y. S. (2011). *The Sage handbook of qualitative research*. Sage.
- Dickson, A., Perry, L. B., & Ledger, S. (2018). Impacts of International Baccalaureate programmes on teaching and learning: A review of the literature. *Journal of Research in International Education*, 17(3), 240–261. doi:10.1177/1475240918815801
- El Masri, Y. H., Baird, J.-A., & Art, G. (2016). Language effects in international testing: The case of PISA 2006 science items. *Assessment in Education: Principles, Policy & Practice*, 23(4), 427–455. doi:10.1080/0969594X.2016.1218323
- Elwood, J. (2006). Formative assessment: Possibilities, boundaries and limitations. *Assessment in Education: Principles, Policy & Practice*, 13(2), 215–232. doi:10.1080/09695940600708653

- Galbraith, D. (2009). Cognitive models of writing. *German as a foreign language GFL*, 2–3, 7–22.
- Gardner, J. (2012). *Assessment and learning* (2nd ed.). SAGE.
- Gay, L. R., Mills, G. E., & Airasian, P. W. (2006). *Educational research : Competencies for analysis and applications* (8th ed.). Pearson Merrill Prentice Hall.
- Gill, J., & Johnson, P. (2010). *Research methods for managers* (4th ed.). SAGE Publications Ltd.
- Gipps, C. (2011). *Beyond testing (classic edition) : Towards a theory of educational assessment*. Taylor & Francis Group.
- Graven, M., Nalube, P., Venkat, H., & Lampen, E. (2009). Critiquing the mathematical literacy assessment taxonomy : Where is the reasoning and the problem solving? *Pythagoras (Pretoria, South Africa)*, 2009(70), 43–56. doi:10.4102/pythagoras.v0i70.38
- Grix, J. (2010). *The foundations of research* (2nd ed.). Palgrave Macmillan.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, 2(163–194), 105.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
- Hanson, W. E., Creswell, J. W., Plano Clark, V. L., Petska, K. S., & Creswell, J. D. (2005). Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology*, 52(2), 224–235. doi:10.1037/0022-0167.52.2.224
- Hopfenbeck, T. N. (2018). Classroom assessment, pedagogy and learning: Twenty years after Black and Wiliam 1998. *Assessment in Education: Principles, Policy & Practice*, 25(6), 545–550. doi:10.1080/0969594X.2018.1553695
- Huckle, J. (2021). Multilingualism and the International Baccalaureate Diploma Programme (IBDP): An analysis of language learning in the IBDP in light of the ‘multilingual turn’. *Journal of Research in International Education*, 20(3), 263–280. doi:10.1177/14752409211059267
- Hughes, J. A., & Sharrock, W. W. (2016). *The philosophy of social research* (3rd ed.). Routledge.
- IBO. (2015a). Diploma programme: From principle into practice (for use from August 2015). . *The Hague, Netherlands: International Baccalaureate Organization*.

- IBO. (2015b). IB Mathematics: Comparability study: Curriculum & assessment comparison. Retrieved from [https://resources.ibo.org/data/g\\_0\\_ibres\\_sur-ir\\_1601\\_1b\\_e.pdf](https://resources.ibo.org/data/g_0_ibres_sur-ir_1601_1b_e.pdf). [https://resources.ibo.org/data/g\\_0\\_ibres\\_sur-ir\\_1601\\_1b\\_e.pdf](https://resources.ibo.org/data/g_0_ibres_sur-ir_1601_1b_e.pdf)
- IBO. (2018). Assessment principles and practices: Quality assessments in a digital age. *International Baccalaureate Organization*.
- IBO. (2019a). Mathematics: analysis and approaches guide. *International Baccalaureate Organization*.
- IBO. (2019b). Mathematics: applications and interpretation guide. *International Baccalaureate Organization*.
- IBO. (2019c). Assessment principles and practices—Quality assessments in a digital age. *International Baccalaureate Organization*.
- IBO. (2020). IB Assessment for May 2020 examination session during COVID-19 pandemic. Retrieved from <https://www.ibo.org/globalassets/news-assets/coronavirus/assessment-model-letter-may-2020-en.pdf>
- Ipsos, M. (2006). Teachers' views on GCSE coursework: Research study conducted for the QCA. Retrieved from the Gov. uk website: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/605733/0506\\_QCA\\_teachers-views-coursework.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/605733/0506_QCA_teachers-views-coursework.pdf)
- Jackson, E. A. (2017). Media review: Practical research methods: A user-friendly guide to mastering research techniques and projects. *Journal of Mixed Methods Research, 11*(3), 417–418. doi:10.1177/1558689816668103
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14–26. doi:10.3102/0013189X033007014
- Johnson, S. (2013). On the reliability of high-stakes teacher assessment. *Research Papers in Education, 28*(1), 91–105. doi:10.1080/02671522.2012.754229
- Khisty, L. L., & Chval, K. B. (2002). Pedagogic discourse and equity in mathematics : When teachers' talk matters. *Mathematics Education Research Journal, 14*(3), 154–168. doi:10.1007/BF03217360
- Kumar, R., Mehta, S., & Kalra, R. (2011). Knowledge of staff nurses regarding legal and ethical responsibilities in the field of psychiatric nursing. *Nursing & Midwifery Research Journal, 7*(1), 1-11.
- Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*(2), 248–294.

- Lane, S., Liu, M., Ankenmann, R. D., & Stone, C. A. (1996). Generalizability and validity of a mathematics performance assessment. *Journal of Educational Measurement*, 33(1), 71–92. doi:10.1111/j.1745-3984.1996.tb00480.x
- Lee, Y.-J., & Greene, J. (2007). The predictive validity of an ESL placement test: A mixed methods approach. *Journal of Mixed Methods Research*, 1(4), 366–389. doi:10.1177/1558689807306148
- Letwinsky, K. M. (2017). Examining the relationship between secondary mathematics teachers' self-efficacy, attitudes, and use of technology to support communication and mathematics literacy. *International Journal of Research in Education and Science*, 3(1), 56.
- Levers, M.-J. D. (2013). Philosophical paradigms, grounded theory, and perspectives on emergence. *SAGE Open*, 3(4), 2158244013517243. doi:10.1177/2158244013517243
- Lewkowicz, J., & Leung, C. (2021). Classroom-based assessment. *Language Teaching*, 54(1), 47–57. doi:10.1017/S0261444820000506
- Li, X., & Cutting, J. (2011). Rote learning in Chinese culture: Reflecting active Confucian-based memory strategies. In *Researching Chinese learners* (pp. 21-42). Palgrave Macmillan, London.
- Luijten, A. J. M. (1988). Internal versus external assessment in the Dutch examinations at 16+ and 18. *Educational Psychology (Dorchester-on-Thames)*, 8(4), 237–246. doi:10.1080/0144341880080404
- Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM Instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, 41(3), 260–293. doi:10.3102/0162373719849044
- Maher, C. A., Sigley, R., Sullivan, P., & Wilkinson, L. C. (2018). An international perspective on knowledge in teaching mathematics. *The Journal of Mathematical Behavior*, 51, 71–79. doi:10.1016/j.jmathb.2018.05.002
- Matthews, M. (2009). Challenging the IB: A personal view, post-Seville. *The International Educator*, 24(2), 19–20.
- Maul, A. (2013). Method effects and the meaning of measurement. *Frontiers in Psychology*, 4(169). doi:10.3389/fpsyg.2013.00169
- Mayer, A. P. (2010). Factors influencing the implementation of an International Baccalaureate Diploma Program in a diverse urban high school. *Journal of Advanced Academics*, 22(1), 78–105. doi:10.1177/1932202X1002200104

- Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, 21(3), 215–237. doi:10.1111/j.1745-3984.1984.tb01030.x
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741. Retrieved from <https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&AN=00000487-199509000-00001&D=ovft&PDF=y>
- Mevarech, Z. R., & Stern, E. (1997). Interaction between knowledge and contexts on understanding abstract mathematical concepts. *Journal of Experimental Child Psychology*, 65(1), 68–95.
- Morcol, G. (2001). Positivist beliefs among policy professionals: An empirical investigation. *Policy Sciences*, 34(3/4), 381–401. doi:10.1023/A:1012749120909
- Moschkovich, J. N. (2015). Academic literacy in mathematics for English Learners. *The Journal of Mathematical Behavior*, 40, 43–62. doi:10.1016/j.jmathb.2015.01.005
- Murphy, R. J. L. (1982). A further report of investigations into the reliability of marking of GCE examinations *British Journal of Educational Psychology*, 52(1), 58–63. doi:10.1111/j.2044-8279.1982.tb02503.x
- Myers, M. (2015). Ten years of qualitative research in organizations and management: Some reflections. *Qualitative Research in Organizations and Management*, 10(4), 337–339. doi:10.1108/QROM-07-2015-1307
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170. doi:10.1080/09695940701478321
- Newton, P. E. (2016). Macro- and micro-validation: Beyond the 'five sources' framework for classifying validation evidence and analysis. *Practical Assessment, Research & Evaluation*, 21. doi:10.7275/f75k-1y75
- Ní Ríordáin, M., & O'Donoghue, J. (2009). The relationship between performance on mathematical word problems and language proficiency for students learning through the medium of Irish. *Educational Studies in Mathematics*, 71(1), 43–64. doi:10.1007/s10649-008-9158-9
- Normile, D. (2021). 'Zero COVID' is getting harder – but China is sticking with it. *Science*, 374(6570), 924–924. doi:10.1126/science.acx9657
- Obuon, O. P. (2019). *Assessing mathematics or language proficiency? Relationship between English language proficiency and mathematics achievement among English learners*. ProQuest Dissertations Publishing,

- Oğuzhan Ö, Z., Akın, M., & Dinçer, A. (2022). The relationship between predicted grades and final assessment grades in the International Baccalaureate Diploma Program. *Hayef: Journal of Education*, 19(2), 155–163.  
doi:10.5152/hayef.2022.22011
- Pellegrino, J. W., Dibello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist*, 51(1), 59–81.  
doi:10.1080/00461520.2016.1145550
- Richardson, S., & Wilkinson, M. (2005). A compilation of survey data and focus group findings: Texas secondary mathematics teachers' views of needed professional development to support mathematics instruction for English language learners. *Austin, TX: Texas State University System Mathematics for English Language Learners Initiative*.
- Ritchie, J., Lewis, J., McNaughton Nicholls, C., & Ormston, R. (2014). *Qualitative research practice : A guide for social science students and researchers* (2nd ed.). SAGE Publications Ltd.
- Roth, W.-M., Ercikan, K., Simon, M., & Fola, R. (2015). The assessment of mathematical literacy of linguistic minority students: Results of a multi-method investigation. *The Journal of Mathematical Behavior*, 40, 88–105.  
doi:10.1016/j.jmathb.2015.01.004
- Sadler, D. (1989). Formative assessment and the design of instructional systems. *An International Journal of Learning and Cognition*, 18(2), 119–144.  
doi:10.1007/BF00117714
- Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson Education.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). Macmillan  
Maxwell Macmillan Canada  
Maxwell Macmillan International.
- Seale, C. (1999). *The quality of qualitative research*. Sage Publications.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. doi:10.3102/0013189X029007004
- Shuard, H., & Rothery, A. (1984). *Children reading mathematics*. John Murray.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. doi:10.3102/0034654307313795

- Steen-Utheim, A., & Hopfenbeck, T. N. (2019). To do or not to do with feedback. A study of undergraduate students' engagement and use of feedback within a portfolio assessment design. *Assessment and Evaluation in Higher Education*, 44(1), 80–96. doi:10.1080/02602938.2018.1476669
- Storz, M., & Hoffman, A. (2018). Becoming an International Baccalaureate Middle Years Program: Perspectives of teachers, students, and administrators. *Journal of Advanced Academics*, 29(3), 216–248. doi:10.1177/1932202X18770171
- Stylianides, A. J., & Stylianides, G. J. (2008). Studying the classroom implementation of tasks: High-level mathematical tasks embedded in ‘real-life’ contexts. *Teaching and Teacher Education*, 24(4), 859–875. doi:<https://doi.org/10.1016/j.tate.2007.11.015>
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology : Combining qualitative and quantitative approaches*. Sage.
- Tattersall, K. (1994). The role and functions of public examinations. *Assessment in Education: Principles, Policy & Practice*, 1(3), 293–304. doi:10.1080/0969594940010305
- Taylor, M. L., & Porath, M. (2006). Reflections on the International Baccalaureate Program: Graduates’ perspectives. *Journal of Secondary Gifted Education*, 17(3), 149–158. doi:10.4219/jsge-2006-398
- Thomas, R. M. (2003). *Blending qualitative & quantitative research methods in theses and dissertations*. Corwin Press.
- Toze, D. (2008). Concerns about IB exam results undermining confidence in program. *International Educator*, 23(2), 6.
- Traub, R. E., & Rowley, G. L. (1991). An NCME Instructional module on. Understanding reliability. *Educational Measurement: Issues and Practice*, 10(1), 37–45. doi:10.1111/j.1745-3992.1991.tb00183.x
- Viro, E., & Joutsenlahti, J. (2020). Learning mathematics by project work in secondary school. *LUMAT*, 8(1), 107–132. doi:10.31129/LUMAT.8.1.1372
- Vitello, S., & Williamson, J. (2017). Internal versus external assessment in vocational qualifications: A commentary on the government’s reforms in England. *London Review of Education*, 15(3), 536–548. doi:10.18546/LRE.15.3.14
- Vukovic, R. K., & Lesaux, N. K. (2013). The language of mathematics: Investigating the ways language counts for children’s mathematical development. *Journal of Experimental Child Psychology*, 115(2), 227–244. doi:10.1016/j.jecp.2013.02.002
- Walker, A., Lee, M., & Bryant, D. A. (2016). Development and validation of the International Baccalaureate Learner Profile Questionnaire (IBLPQ). *Educational Psychology*, 36(10), 1845–1867. doi:10.1080/01443410.2015.1045837

- Wiliam, D. (2009). *Assessment for learning : Why, what and how?* University of London, Institute of Education.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, *11*(1), 49–65. doi:10.1080/0969594042000208994
- Wolf, A. (2011). *Review of vocational education : The Wolf report*. London: TSO.
- Wolming, S., & Wikström, C. (2010). The concept of validity in theory and practice. *Assessment in Education : Principles, Policy & Practice*, *17*(2), 117–132. doi:10.1080/09695941003693856
- Wright, E., & Lee, M. (2014). Developing skills for youth in the 21st century: The role of elite International Baccalaureate Diploma Programme schools in China. *International Review of Education / Internationale Zeitschrift für Erziehungswissenschaft / Revue Internationale de l'Education*, *60*(2), 199–216. Retrieved from <http://www.jstor.org/stable/24636723>
- Yates, A., & Johnston, M. (2018). The impact of school-based assessment for qualifications on teachers' conceptions of assessment. *Assessment in Education : Principles, Policy & Practice*, *25*(6), 638–654. doi:10.1080/0969594X.2017.1295020
- Yeager, D. S., Bundick, M. J., & Johnson, R. (2012). The role of future work goal motives in adolescent identity development: A longitudinal mixed-methods investigation. *Contemporary Educational Psychology*, *37*(3), 206–217. doi:10.1016/j.cedpsych.2012.01.004
- Zhang, Y., & McGrath, I. (2009). Teacher–student relationships in an International Baccalaureate school in China. *Journal of Research in International Education*, *8*(2), 164–190. doi:10.1177/1475240909105203
- Ziderman, A. (1984). External examination versus internal assessment: The Israeli maturity certificate. *Comparative Education Review*, *28*(3), 477–484. doi:10.1086/446451

## Appendices

### Appendix 1: Subject teachers

Thank you so much for accepting to take part in the interview. To better understand how internal assessment is taught and evaluated in the Diploma Program, this will be highly beneficial for my research. The interview will last about 30 minutes, and I'm interested in learning about your experience preparing for the Diploma Programme Mathematics Internal Assessment as a subject teacher. All replies and personal data will be treated in strict confidence and discarded as soon as it is no longer required for research. There will not be any reports with your name or the name of your school. You are able to skip a question or leave the interview at any time. To help with data collecting, I would like to audio-record the interview. Do you mind if I audio-record our conversation?

1. What skills do you teach students through the process of the mathematics internal assessment? some examples of skills such as Mathematical skills, language skills or research skills.
2. How does the Internal assessment help you plan and prepare to teach the external examinations? (Paper 1, Paper 2, or Paper 3).
3. To what extent does the Math IA assess students' mathematics skills or English language skills, or other skills in your opinion?
4. The mathematics assessment objectives are Problem solving, Knowledge and understanding, communication and interpretation, technology, reasoning, and Inquiry approaches. Do you think the internal assessment ultimately meets the IB guide's assessment objectives?

5. To what extent was the internal mathematics assessment fair and valid for English language learners?
6. Thinking back on our interview conversation, is there anything else you would like to add that you believe is of importance?

## Appendix 2: Interview schedule for Diploma Programme students

Thank you so much for accepting to take part in the interview. This will be really beneficial for my research in helping me better grasp what the Mathematics Internal Assessment construct is and how it has been assessed in the Diploma Programme. We'll talk for around 20 minutes, and I'm interested in learning about your experience as Diploma Programme students making preparations for the Mathematics Internal Assessment. All replies and individual information will be treated in strict confidence and discarded as soon as it is no longer required for research. There will not be any reports with your name or the name of your school. You are able to skip a question or leave the interview at any time. Do you mind if I audio-record our conversation?

1. What skills have you developed through the Mathematics Internal assessment?

Some examples of skills such as Mathematical skills, Language skills and research skills.

2. What do you assume about the skills and knowledge you possess before writing the Internal Assessment of Mathematics? Examples of skills such as problem-solving skills, writing skills and investigation skills.

3. How did the Internal assessment help you in planning and preparing for your external examinations? (Paper 1, Paper 2, or Paper 3)

4. How do you relate the existing knowledge of mathematics IA with other subjects, such as Physics, Biology, English, TOK, and Extended Essay in your IBDP course? Give an example.

5. To what extent does the Math IA assess your mathematics skills, English writing skills, or the skills of other science-related subjects?

6. To what extent was the mathematics internal assessment fair and valid for English language learners?

7. The mathematics assessment objectives are Problem solving, Knowledge and understanding, communication and interpretation, technology, reasoning, and Inquiry

approaches. Whilst writing towards the Math IA, do you think the written task meets the assessment objectives in the IB Guide?

8. Thinking back on our interview conversation, is there anything else you would like to add that you believe is of importance?

### Appendix 3: Online Survey

This study aims to explain students' perspectives on the International Baccalaureate Diploma Programme (IBDP) Mathematics Internal Assessment. It evaluates whether the results of an internal assessment can indicate students' mathematical ability. This study also investigates the extent to which the internal assessment affects the assessment purposes of the IBDP Mathematics assessment among English language learners aged 16 to 19 years old, studying at an IB school named the United World College Changshu, China. The potential benefits could provide suggestions for assessment development and implementation to improve the validity and fairness of the mathematics internal assessment.

To take part in the study, I am looking for participants who are at least 16 years old. You would receive an invitation to complete a brief online survey. You would need to take 10 to 15 minutes to it. You will be requested to provide views or feedback on the internal assessment coursework for maths following the online survey.

If you are interested and would like more information, please contact XX at the Department of Education, 15 Norham Gardens OX2 6PY Oxford, via email.

There is no obligation to take part, and you are free to withdraw at any time you see fit. Thank you!

Note: It is a semi-structured online survey, and questions might be added at the moment.

#### ONLINE SURVEY

1. I understand the purpose of the study, and I am willing to participate in this online survey.  
Yes/No
2. Age:
3. Gender: Female/ Male
4. What do you think of your English Language Proficiency?  
  
1 Beginner 2 Intermediate 3 Advanced 4 Proficient
5. Please review your final mathematics IA and the full rubric for the Maths IA at the following link from the IB Guide (Read Page 84 to 88): Do you understand the criteria? If not, why? \_\_\_\_\_
6. Please self-evaluate yourself and share below: \_\_\_\_\_

- |                           |      |
|---------------------------|------|
| Presentation              | 1- 4 |
| Mathematics Communication | 1- 4 |
| Personal Engagement       | 1- 3 |
| Reflection                | 1- 3 |
| Use of Mathematics        | 1- 6 |
7. Which criteria do you think you have done well? Which criteria do you think there are issues with? Why? \_\_\_\_\_
8. Which assessment in the other IB courses requires skills that overlap with the skills needed for the IB Maths Internal assessment?
0. None
  1. Mathematics external assessment
  2. English assessment
  3. Physics assessment
  4. Environmental Systems and Societies assessment
  5. Chemistry assessment
  6. Biology assessment
9. Please rate the relevance of the following subjects' assessments with respect to Math IA. (0 = no relevance at all; 5 = highly relevant)
- |   |     |
|---|-----|
| 0. None   |     |
| 1. Mathematics external assessment                | 0-5 |
| 2. English assessment                             | 0-5 |
| 3. Physics assessment                             | 0-5 |
| 4. Environmental Systems and Societies assessment | 0-5 |
| 5. Chemistry assessment                           | 0-5 |
| 6. Biology assessment                             | 0-5 |
10. Considering how many students are English language learners, it is fair that the IB Mathematics course bases the Internal Assessment grade entirely on a written task that requires students to communicate effectively through English writing.  
1 Strongly disagree 2 Disagree 3 Neutral 4 Agree 5 Strongly Agree
11. In your view, whether the design of the Internal Assessment is a good assessment to evaluate your mathematics ability? Why? (50 words) \_\_\_\_\_

## Appendix 4: CUREC Approval

**Subject:** CUREC Ref: CIA-22-083  
**Date:** Tuesday, February 8, 2022 at 23:38:03 China Standard Time  
**From:** Hamish Chalmers  
**To:** [REDACTED]  
**CC:** [REDACTED] Student CUREC  
**Attachments:** attachment.png

Dear [REDACTED]

Title: The construct validity and fairness of International Baccalaureate Diploma Programme Mathematics Assessment in inquired-based approaches assessment – a case study around assessment in United World College, China (To be changed)

Ref: CIA-22-083

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

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

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse

<http://www.nspcc.org.uk/globalassets/documents/information-service/factsheet-child-abuse-reporting-requirements-professionals.pdf>

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to [research.office@education.ox.ac.uk](mailto:research.office@education.ox.ac.uk) for consideration.

Good luck with your research study.

Best wishes

Hamish Chalmers  
Member of the DREC

## Appendix 5: Information for participants

In partnership with researchers at the University of Oxford, United World College China has agreed to take part in a study investigating How does internal assessment affect the construct validity across different cultural groups? For non-native speakers of the English language. Is Internal Assessment fair for Chinese students? We would like to invite your child to be part of this study. We very much hope you would like your child to take part, but before you decide, it is important that you understand why the study is being done and what it will involve.

### **What are we trying to find out?**

The research aims to study how Internal Assessment affect the construct validity of IB DP Mathematics across different culture groups. It could help to IB students in the internal assessment for teaching and learning and find out the issues during the study.

### **Why has my child been invited to take part?**

We are inviting your child to take part because they are a young person, aged between 16 and 19 years, attending United World College Changshu China.

We are inviting 135 young people to take part.

### **Does my child have to take part?**

No. You can ask questions about the study before deciding whether or not to allow your child to participate. If you do agree to participation, you may withdraw your child and their data from the study at any time, without giving a reason and without penalty, by advising the researchers of this decision.

### **What will happen if my child takes part?**

You are invited to attend a study examining the validity and fairness of Internal Assessment (IA) for IB DP Mathematics coursework. The intervention will involve a 30-minutes self-assessment activity after the students submit their final draft of the IA, where students will evaluate their own work against five different criteria or assessment objectives. And have verified their feedback against the teachers' feedback for the first drafts. All feedback will be written, and online survey will be conducted.

Your final evaluation of the scores will be compared with the teacher assessment and subsequently compared with IB component scores at beginning of July.

### **What are the advantages / disadvantages of taking part?**

There will be no direct or personal benefit to you from taking part in this research.

### **What happens to the data provided? <sup>1</sup>**

The information you or your child provide during the study is the **research data**. Any research data from which you or your child can be identified such as your name, your date of birth, is known as **personal data**.

**Personal / sensitive data** will be destroyed when the research file with coded pseudonyms is created.

**Other research data** will be stored for at least 3 years after publication or public release of the work of the research.

The researcher and his supervisor will have access to the research data. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.

I would like your permission to use de-identified data in future studies, and to share data with other researchers (e.g. in online databases). All personal information that could identify you will be removed or changed before information is shared with other researchers or results are made public.

### **Will the research be published?**

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

The research will be written up as a student's thesis. On successful submission of the thesis, it may be deposited both in print and online in the University archives to facilitate its use in future research. If so, the thesis will be openly accessible.

### **What will happen to any samples taken from my child?**

---

We will use your child's samples to write up as a student's thesis.

### **Who is conducting this research?**

The research project is organised by a graduate student at the Oxford University, who is a master student. The research is funded by the student's saving. This study has been reviewed by Dr. XX, and received ethics clearance through, the University of Oxford's Central University Research Ethics Committee, CUREC reference number CIA-22-083.

### **What if there is a problem?**

If you have a concern about any aspect of this study, please contact the student, and we will do our best to answer your query. I/we will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Chair, **Medical Sciences Inter-Divisional Research Ethics Committee**; Email: [ethics@medsci.ox.ac.uk](mailto:ethics@medsci.ox.ac.uk); Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

**OR**

Chair, **Social Sciences & Humanities Inter-Divisional Research Ethics Committee**; Email: [ethics@socsci.ox.ac.uk](mailto:ethics@socsci.ox.ac.uk); Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

[For applications reviewed by the **Oxford Tropical Research Ethics Committee (OxTREC)**, please insert the contact details for the local ethics committee that has reviewed your study]

### **Data Protection**

The University of Oxford is the data controller with respect to your child's personal data and, as such, will determine how your child's personal data is used in the study.

The University will process your child's personal data for the purpose of the research outlined above. Research is a task that we perform in the public interest.

Further information about your rights with respect to your child's personal data is available from <https://compliance.web.ox.ac.uk/individual-rights>.

### **What should I do next?**

Please fill in the enclosed form and return it to your child's subject teacher if you would like your child to take part in this study. Please remember that you may withdraw your child at any time, without penalty and without giving a reason, by notifying the researcher.

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

Department of Education  
University of Oxford, 15 Norham Gardens

